APPENDIX AVAILABLE ON THE HEI WEB SITE

Research Report 178

National Particle Component Toxicity (NPACT) Initiative Report on Cardiovascular Effects

Sverre Vedal et al.

Section 1: NPACT Epidemiologic Study of Components of Fine Particulate Matter and Cardiovascular Disease in the MESA and WHI-OS Cohorts

Appendix D. The NO_2 Model

Note: Appendices that are available only on the Web have been assigned letter identifiers that differ from the lettering in the original Investigators' Report. HEI has not changed the content of these documents, only their identifiers.

Appendix D was originally Appendix C

Correspondence may be addressed to Dr. Sverre Vedal, University of Washington, Department of Environmental and Occupational Health Sciences, Box 354695, 4225 Roosevelt Way NE, Suite 100, Seattle, WA 98105-6099; email: *svedal@uw.edu*.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR–83234701 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. For the research funded under the National Particle Component Toxicity initiative, HEI received additional funds from the American Forest & Paper Association, American Iron and Steel Institute, American Petroleum Institute, ExxonMobil, and Public Service Electric and Gas. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

This document was reviewed by the HEI NPACT Review Panel but did not undergo the HEI scientific editing and production process.

APPENDIX C: The NO2 Model

- **AQS Data for NO_x and NO_2 Models**
- **Building and Validating the Gaseous pollutant Spatio-Temporal Models**
- **Results**

AQS Data for NO_x and NO_2 Models

The EPA Air Quality System (AQS) repository contains hourly monitoring data for NO_x and NO_2. MESA Air's exposure database converts these into daily averages, using the following criteria:

1. The recorded hourly zeros (which are rounded down by AQS from actual nonzero observations) are converted to 0.5 ppb.
2. To calculate an average, daily data at a given monitor must include at least 18 hourly values, including 4 each during the morning and afternoon diurnal lows and highs.

From these daily averages, 2-iweekly averages were calculated for time intervals compatible with the local MESA Air monitoring schedule, subject to the following limitations:

1. Some time intervals for certain monitors, deemed unreliable after statistical quality analysis and communication with agencies, were excluded.
2. Two-week averages were only calculated if at least 9 daily averages were available (analogous to the MESA Air's monitoring system data quality standard).

The AQS dataset for the model originally included all available data from the 17 state planes containing or adjoining to MESA metropolitan areas, from 1999 through 2009. This was later restricted to the final, smaller modeling regions around the cities of interest, based on geographic relevance (e.g., for the Chicago model only data from the Chicago suburb of Gary, IN were used, while data further afield in Indiana, such as Indianapolis, were excluded). Additionally, monitors of the following types were excluded:

1. Purely seasonal (summer-only) monitors were excluded, as they seemed to suffer from chronic quality problems at season startup, and their time series is also incompatible with our time-trend modeling technology.
2. Monitors with less than one year of data (i.e., less than 26 two-weekly averages) were excluded as well, because the advantage from including them was outweighed by the risk to quality and by potential model-compatibility issues.

The table below presents that data quantities – number of sites and number of biweekly averages - from both sources (AQS and MESA-Air) used in each metropolitan-region model.

NO_x

Sites	Baltimore	Chicago	LA	NY	St. Paul	Winston-Salem	Total
AQS	13	8	27	17	5	3	73
Fixed	5	7	7	3	4	4	30
snapshot	104	129	252	157	107	121	870
Home	87	113	120	119	129	117	685
Total	209	257	406	296	245	245	1658
Observations							
AQS	2371	1472	5423	2723	810	569	13368
Fixed	387	448	599	246	345	371	2396
snapshot	306	302	611	409	285	308	2221
Home	173	255	217	244	270	270	1429
Total	3237	2477	6850	3622	1710	1518	19414

NO_2

Sites	Baltimore	Chicago	LA	NY	St. Paul	Winston-Salem	Total
AQS	12	8	27	24	5	3	79
Fixed	5	7	7	3	4	4	30
snapshot	104	129	252	157	107	121	870
Home	87	112	120	119	129	117	684
Total	208	256	406	303	245	245	1663
Observations							
AQS	3126	1556	5423	4273	884	762	16024
Fixed	387	448	599	245	345	370	2394
snapshot	306	302	611	409	285	307	2220
Home	173	255	217	244	270	269	1428
Total	3992	2561	6850	5171	1784	1708	22066

Building and Validating the Gaseous pollutant Spatio-Temporal Models

A. Modeling Framework

Gaseous pollutant models were developed for NO_x and NO_2 to allow individual-level gaseous pollutant variables to be incorporated into our health effects models. Because of the greater density and duration of gaseous pollutant monitoring, modeling of the gaseous co-pollutants allowed a richer model than the model developed for PM components. The model can be described as

$$Y(s,t) \sim N(\mu(s,t), \Sigma_\varepsilon(\varphi_\varepsilon)), \quad (1)$$

indicating that the array of exposures Y across space s and time t indices is normally distributed with a spatiotemporal mean field μ and a geostatistical (kriging) error field.

This in itself is a fairly standard specification. The innovation lies in modeling the mean field as

$$\mu(s,t) = [\mathbf{X}_0(s)\alpha_0 + K_0(s \mid \varphi_0)] + \sum_{i=1}^{T} F_i(t)[\mathbf{X}_i(s)\alpha_i + K_i(s \mid \varphi_i)], \qquad (2)$$

with the *X*'s indicating matrices of traffic and land-use covariates, the *F*'s indicating smooth trends which are functions of time only, the *K*'s indicating spatial kriging fields without a random-error component, and α, φ indicating data-estimable parameters characterizing the magnitude and spatial extent of the various model components. This model was first described by Sampson et al. (2011), modeling $PM_{2.5}$ exposure for the MESA-Air cohort over the period 2000-2006, based on concepts from Fuentes et al. (2006). Szpiro et al. (2010), in our group, introduced simultaneous maximum-likelihood estimation (MLE) for the α and φ parameters in (1) and (2), and demonstrated its application by modeling NO_x exposure using data from 18 metropolitan LA regulatory (AQS) sites over a period of 2.5 years. The time trends *F* are estimated separately before the model is fitted, using data from regional long-term time series and singular value decomposition (SVD) as detailed in Fuentes et al. (2006). Specifying the model matrices **X** in (2) is challenging. The MESA Air solution as implemented for gaseous pollutants is described below.

Lindström et al. in our group further developed the model by adding covariates *Z(s,t)* varying in both time and space to the mean function (2), scaled by data-estimable magnitude parameters γ. This expansion of the model enables the incorporation of spatio-temporally varying effects such as local weather, or of output of a deterministic dispersion model, under the same exposure-prediction framework. An optimization code for the entire model, with or without $Z(s,t)\gamma$, has been developed, allowing investigators to rapidly produce estimates and predictions for large datasets and long time spans, whereas existing tools require weeks to produce such estimates. The modeling tools and their relative speed also enable a high level of exploration, selection, tuning and diagnosing. The code has been written in R and is available in the CRAN repository as the package "SpatioTemporal".

B. Variable Selection for Metropolitan-Region Models

The model (1)-(2) was applied to 2-week NO_x and NO_2 monitoring data. A separate model was developed for each of the six MESA cities, over the 11 years 1999-2009. NO_x concentrations were log-transformed. The model was fit without a time-varying covariate (the deterministic dispersion covariate from Caline). However, the long-term means of Caline3QHCR predictions at each location, for distance buffers from 1500 to 9000 meters, were used as variables in the spatial "land-use" matrices **X**.

Most regions had a single time-trend function *F*, except Los Angeles where two were needed due to strong seasonal and directional variations in wind patterns. Model selection for the **X** matrices was performed mostly using the MESA Air community Ogawa snapshot campaigns, which were specifically designed to gauge the near-road effect (Mercer et al., 2011). Initial variable selection via LASSO was followed by all-subset cross-validation, performed several times on different re-shuffling configurations

of CV groups. The most robust model across all CV iterations was chosen. Models were then subsequently tuned on the full spatio-temporal dataset,using cross-validation results for MESA home-outdoor monitoring sites and long-term AQS and MESA Air fixed sites, diagnostic analysis for artifacts or lack of fit, and comparison and alignment between MESA cities for scientific plausibility.

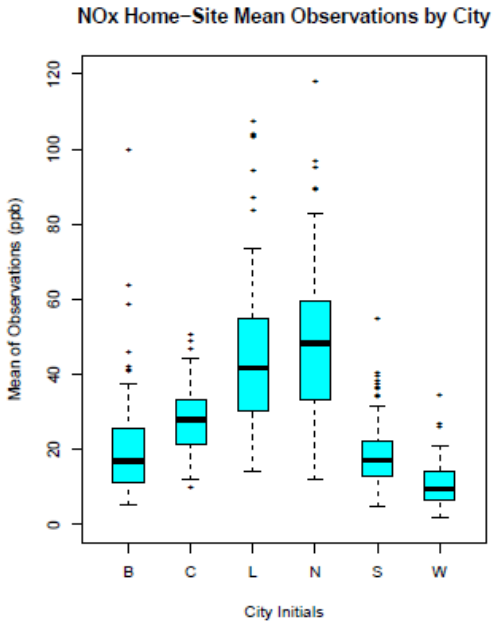
For NO_x the complete model-selection process was carried out. For NO_2 the final NO_x models were used as a starting point and selection proceeded directly to the tuning stage. The final models were identical to NO_x except for one variable added in New York and one removed in Chicago. The long-term baseline matrix $\mathbf{X}_0$ in the final models had between 6 and 9 variables; matrices for the time-trend amplitudes, $\mathbf{X}_1$, $\mathbf{X}_2$ required only subsets of 2-3 variables from $\mathbf{X}_0$. A full list of selected variables appears in the table below. Those variables present in both $\mathbf{X}_0$ and $\mathbf{X}_{1,2}$ are indicated in blue.

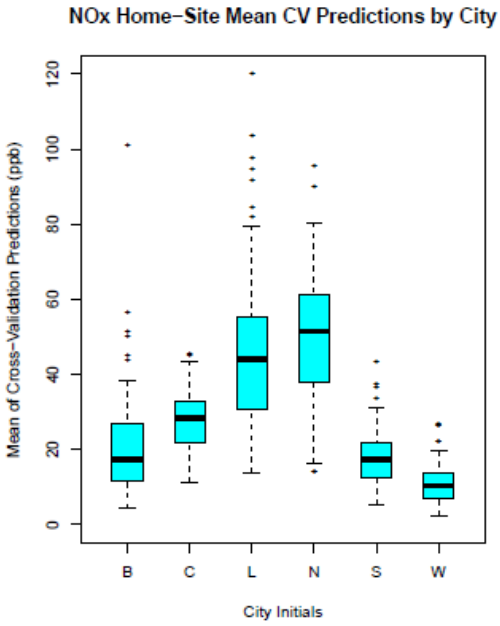
City	Log distance to point/line source	Road Buffers	Population Buffers	Regional Plume Variables	High-emission Land Use (area source)	Clean Air Source / Other Land Use
Baltimore	(major) Road	CalineMean6k (log)	1k (log)	m.to.downtown (i.e., city hall)	Hi-intensity3k Transportation1k	
Chicago	Road A1 Maj. Airport	CalineMean7.5k (NOx only) A1_15k (regional!)	0.5k(log)	Nox_15k	Hi-intensity1k	Bays1k (Lake Michigan)
LA + Riverside	Road A1	CalineMean9k (log)	5k		Hi-intensity3k	m.to.coast Open3k
NY +Rockland	Road (NO2 only: maj. Airport)	CalineMean9k (log)	2k(log)	m.to.downtown (log)	Commercial1k	Residential3k Stream0.75k
St. Paul	A1	CalineMean7.5k (log) A23_150m	1.5k		Industrial3k	Open1.5k
Winston-Salem	Road A1		1.5k	m.to.downtown (log)	Hi-intensity1.5k	Residential3k Residential: downtown interaction

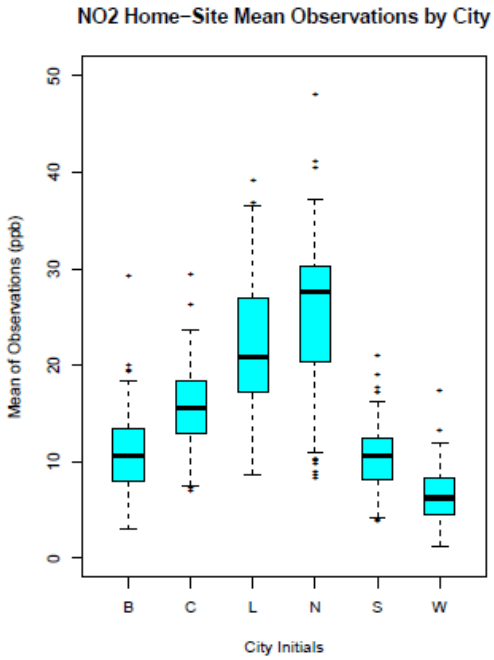
As can be seen, population density in buffers ranging from 500 to 5000m was found to be useful in all six regions. The long-term Caline3QHCR mean for buffers of 6000-9000m was also selected in all five regions for which it was available. Additionally, at least one land-use variable representing high-emission area sources with radii of 500-3000m was selected in each city, but the exact nature of this variable differed between cities. Distance to major road (log-transformed and truncated at 10m) was selected in all but one city, and in 4 of 5 cases also made it into the trend-amplitude matrices; in the remaining two cities it was “represented” by the distance to A1 roads only (similarly transformed, and truncated at 20m and less). Conversely, variables representing sum-of-roads in buffer were rarely selected for the final model.

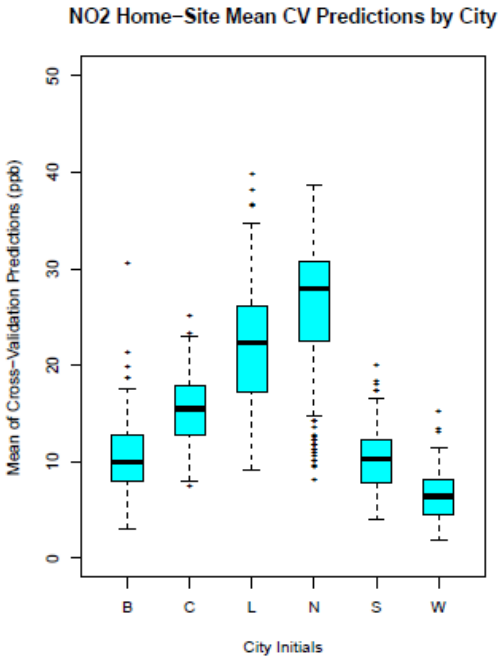
Results

The plots below compare observed home-site means for all sites with multiple observations (left) and the corresponding means of predictions (right) in each city, for NOx and NO2. As can be seen, the models reproduce fairly accurately the relative ordering and variability of exposures between and within MESA cities.



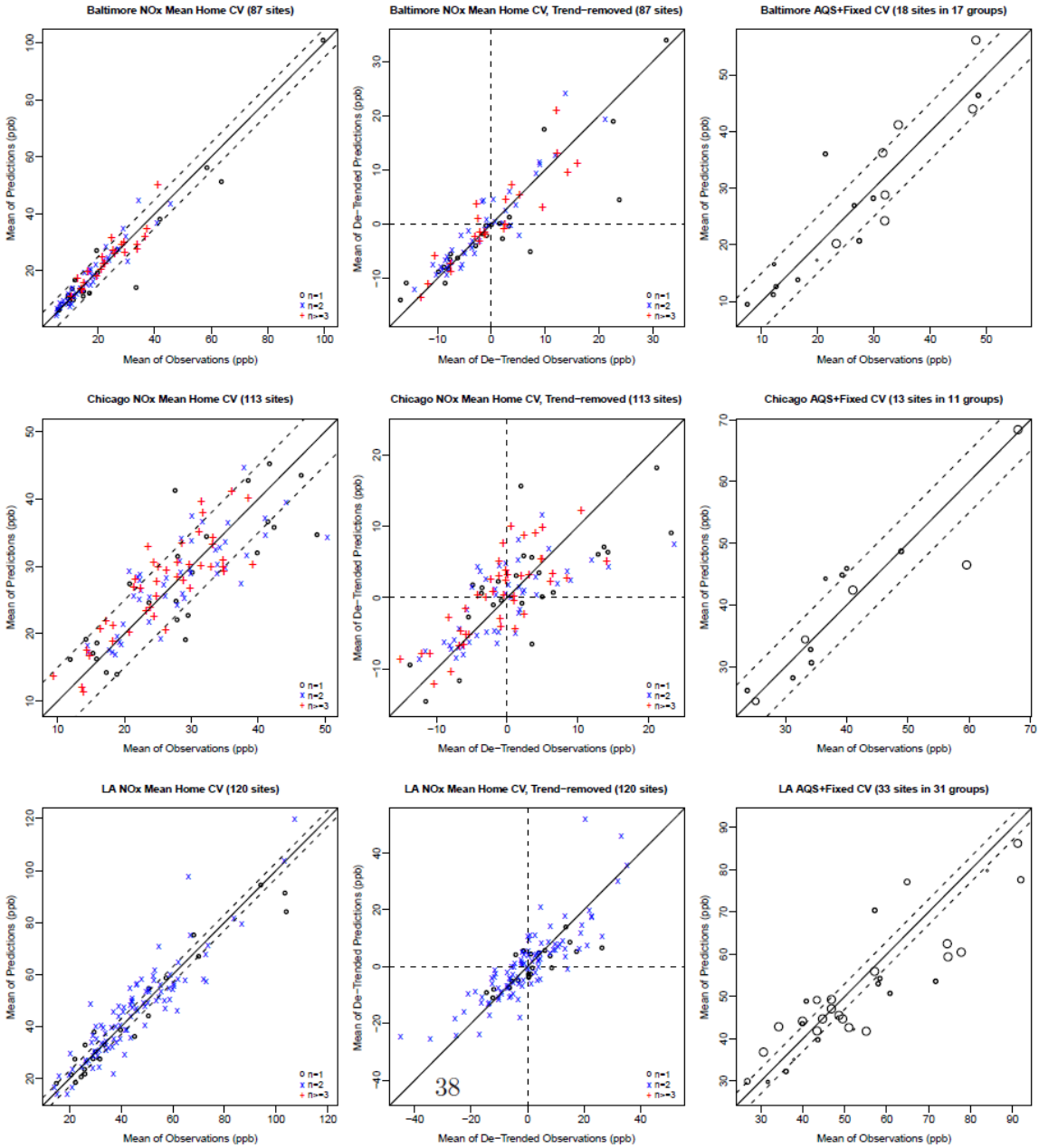


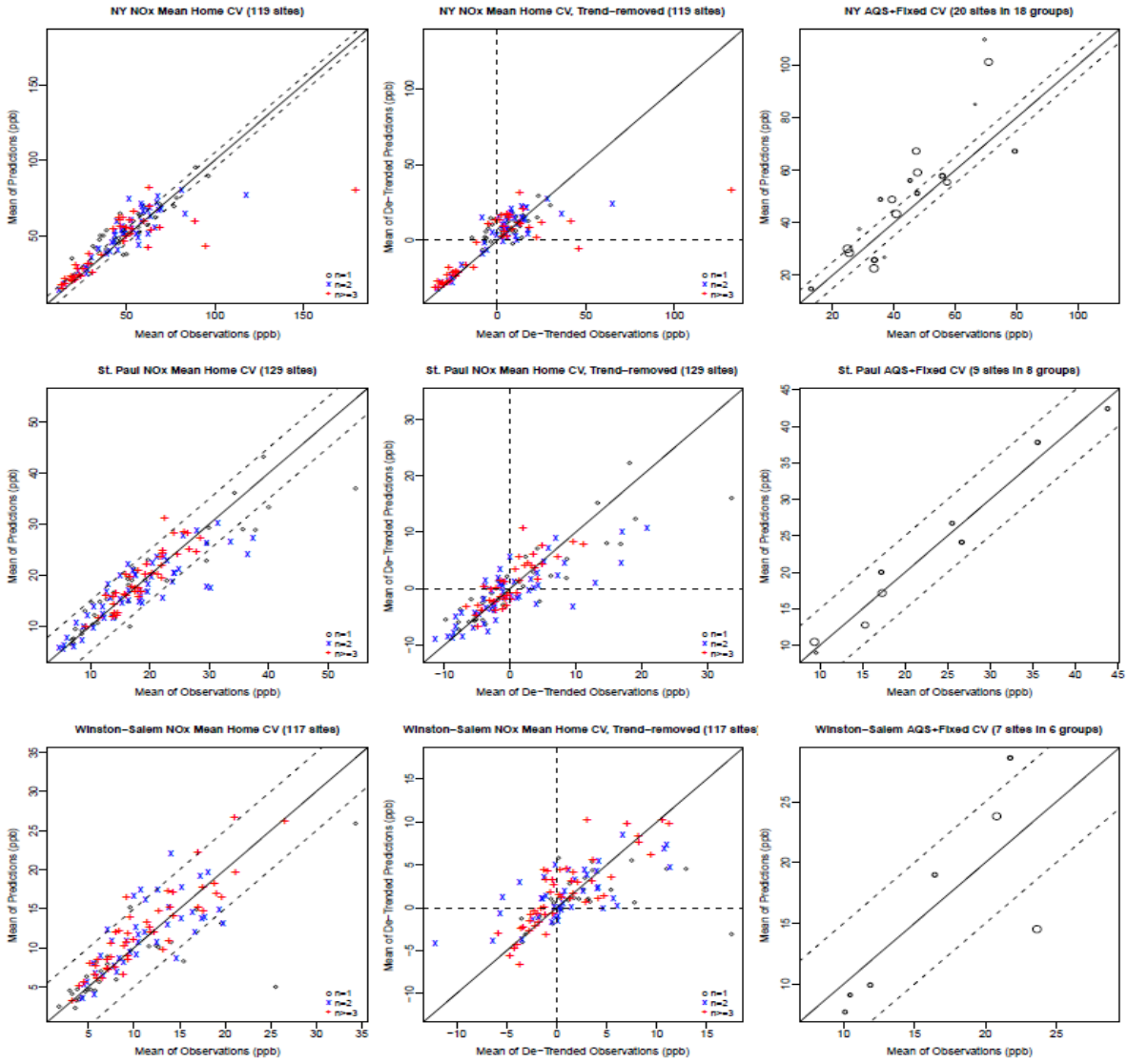




Cross-validation plots, shown here for NOx only, reveal further detail. For each city (in alphabetical order), the left frame plots home-site means vs. their 10-fold CV predictions; the center frame plots the same data, after the seasonal trend was removed, to reveal the extent of spatial determination; and the right frame plots the means of long-term sites (AQS and MESA-Air fixed) vs. the leave-one-site-out CV prediction means.

Predictions in Baltimore were especially accurate for home sites, while in St. Paul they were almost perfect for fixed-site means. In Winston-Salem, the absolute prediction error was lowest (absolute errors of +/- 5ppb for NOx are indicated by dashed lines), but the relative determination was not as good due to the much smaller variability in actual exposures.





For the larger cities, Chicago and LA, predictions were fairly good overall, even though the error magnitude was larger due to the larger scale of variability. In New York, the lesser accuracy can be attributed to known deficiencies in the available set of variables. Residual-oil burning for heating is estimated to produce as much NOx as road traffic in northern Manhattan; a geographical layer quantifying such combustion has been obtained after model completion, and will be incorporated into future models. Additionally, land-use classification for New York is almost uniform across all of Manhattan; no variable was available to account for vertical "street canyon" effects. Summary CV statistics for home-sites – R^2, spatial-only R^2, RMSE, relative error and bias - are shown in the table below, using both classical and robust (quantile and rank-based) statistics.

Classical:

City	R^2	Detrended R^2	RMSE(ppb)	Rel.Err(%)	Bias(ppb)
Baltimore	0.93	0.82	3.9	13.0%	0.1
Chicago	0.67	0.55	4.9	14.7%	0.2
LA-Riverside	0.87	0.70	7.3	12.6%	0.5
NY-Rockland	0.64	0.59	13.9	15.4%	0.7
St. Paul	0.79	0.70	3.9	14.3%	-0.5
Winston-Salem	0.64	0.42	3.5	21.6%	0.1

Robust:

City	Rank R^2	Detrend Rank R^2	APE_{68}(ppb)	Rel.Err(%)	Bias(ppb)
Baltimore	0.91	0.84	2.7	9.3%	0.2
Chicago	0.72	0.58	4.8	12.6%	0.6
LA-Riverside	0.86	0.74	6.4	10.5%	0.3
NY-Rockland	0.73	0.55	8.1	11.2%	2.2
St. Paul	0.81	0.70	3.0	13.0%	-0.1
Winston-Salem	0.71	0.47	2.5	16.4%	0.0

The corresponding statistics for NO_2 were:

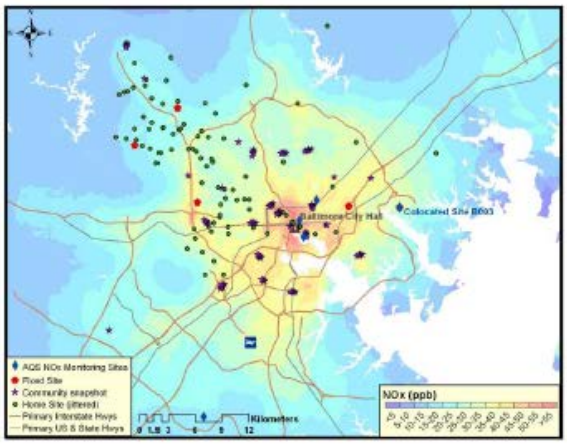
Classical:

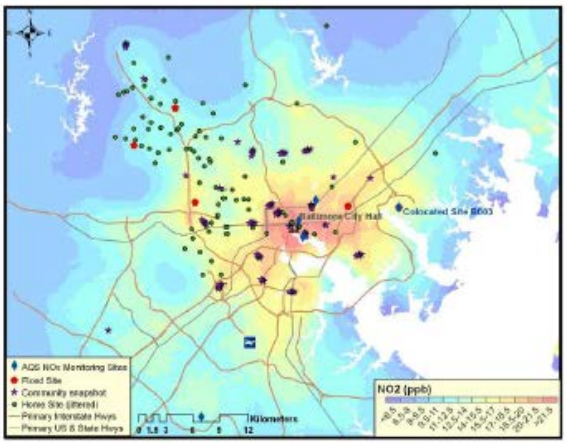
City	R^2	Detrended R^2	RMSE(ppb)	Rel.Err(%)	Bias(ppb)
Baltimore	0.89	0.84	1.6	11.3%	-0.1
Chicago	0.74	0.71	2.1	10.0%	0.0
LA-Riverside	0.81	0.68	2.9	9.6%	0.3
NY-Rockland	0.80	0.80	3.7	10.2%	0.1
St. Paul	0.85	0.70	1.3	9.6%	-0.2
Winston-Salem	0.76	0.68	1.3	17.7%	-0.0

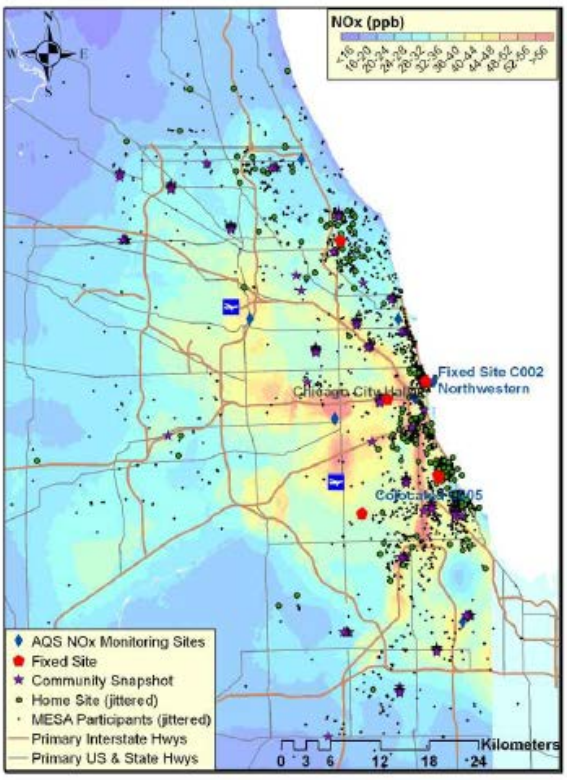
Robust:

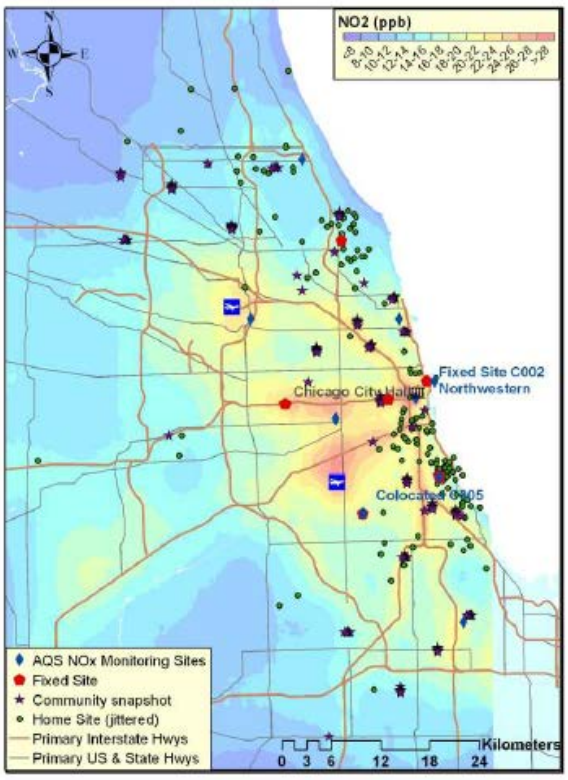
City	Rank R^2	Detrend Rank R^2	APE_{68}(ppb)	Rel.Err(%)	Bias(ppb)
Baltimore	0.87	0.81	1.5	9.4%	-0.3
Chicago	0.76	0.73	1.9	7.0%	0.1
LA-Riverside	0.80	0.74	2.1	7.0%	0.5
NY-Rockland	0.61	0.54	2.5	6.5%	0.2
St. Paul	0.81	0.69	1.1	7.2%	-0.0
Winston-Salem	0.76	0.65	1.2	12.9%	-0.1

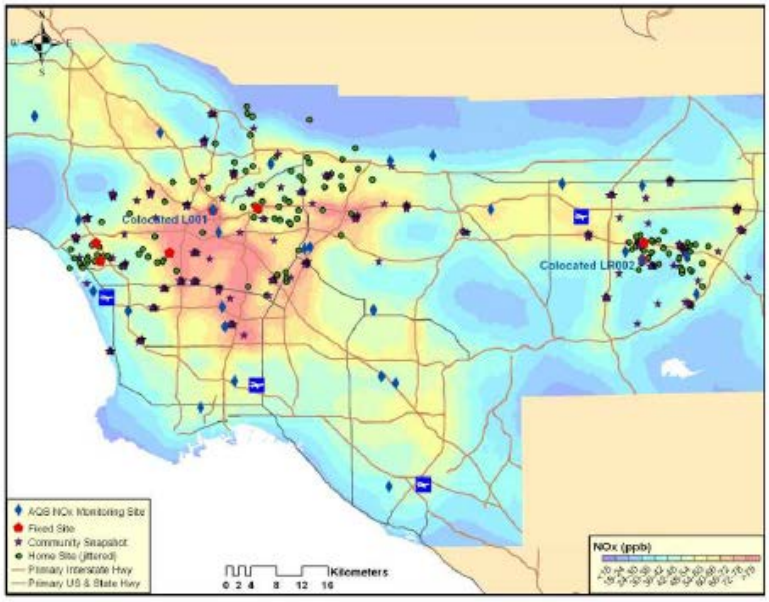
Finally, prediction maps for NO_x and NO_2 are presented below for each city.

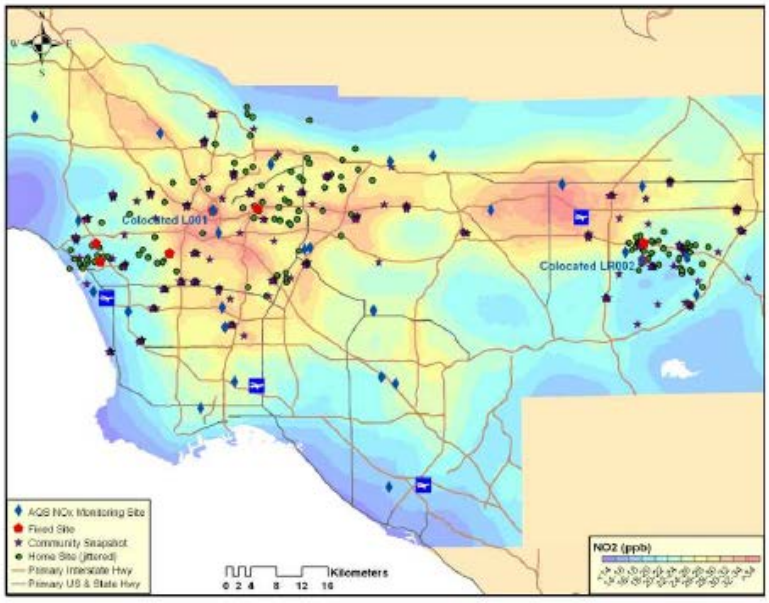

Baltimore City Hall
Colocated Site B003
AQS NOx Monitoring Sites
Fixed Site
Community snapshot
Home Site (jittered)
Primary Interstate Hwys
Primary US & State Hwys
Kilometers
NOx (ppb)

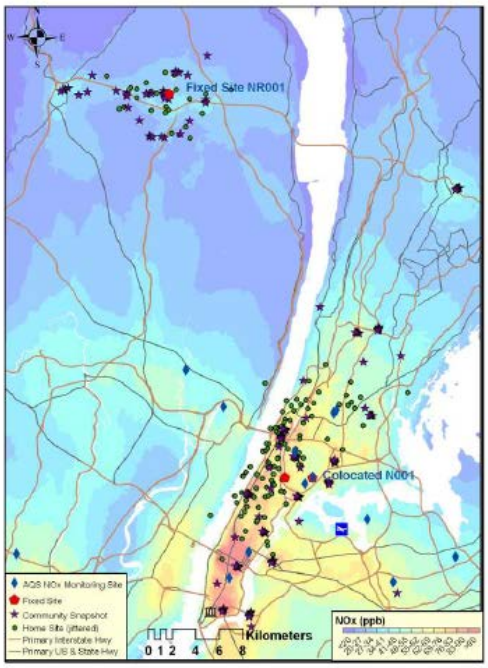

Baltimore City Hall
Colocated Site B003
AQS NOx Monitoring Sites
Fixed Site
Community snapshot
Home Site (jittered)
Primary Interstate Hwys
Primary US & State Hwys
Kilometers
NO2 (ppb)

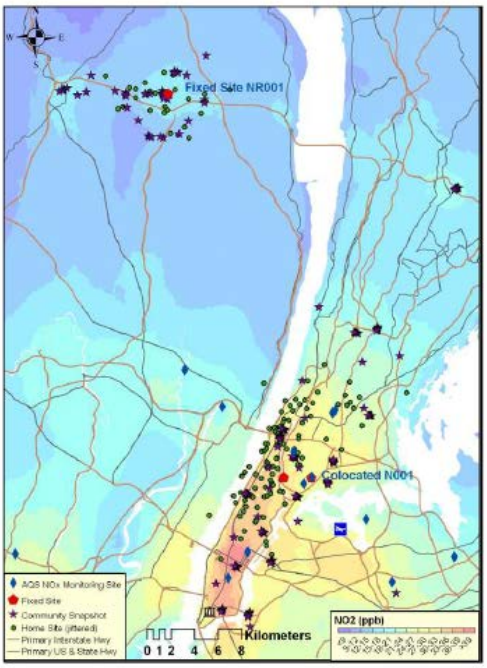

NOx (ppb)
Fixed Site C002
Northwestern
AQS NOx Monitoring Sites
Fixed Site
Community Snapshot
Home Site (jittered)
MESA Participants (jittered)
Primary Interstate Hwys
Primary US & State Hwys
0 3 6 12 18 24
Kilometers

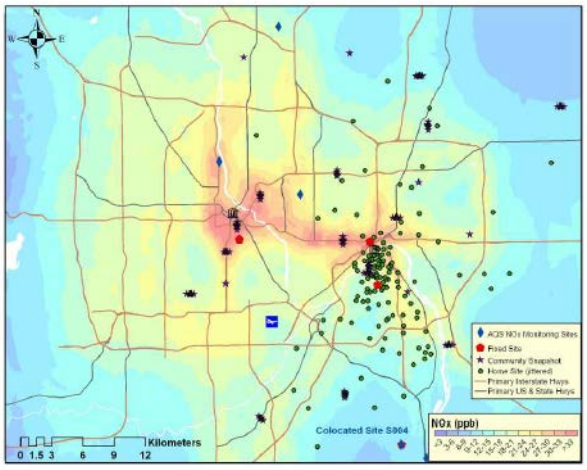

NO2 (ppb)
Fixed Site C002
Northwestern
AQS NOx Monitoring Sites
Fixed Site
Community snapshot
Home Site (jittered)
Primary Interstate Hwys
Primary US & State Hwys
0 3 6 12 18 24
Kilometers

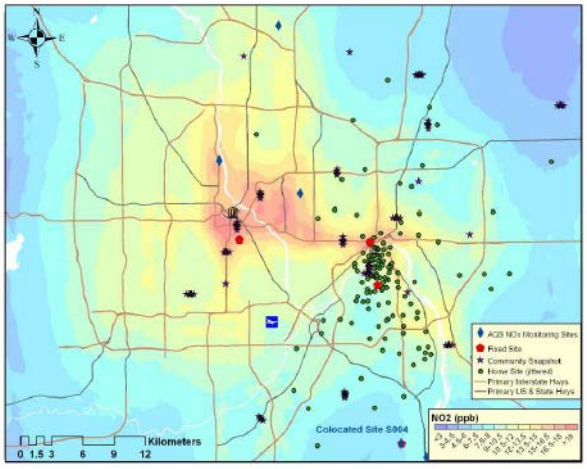

Colocated L001
Colocated LR002
AQS NOx Monitoring Site
Fixed Site
Community Snapshot
Home Site (jittered)
Primary Interstate Hwy
Primary US & State Hwy
0 2 4 8 12 16
Kilometers
NOx (ppb)

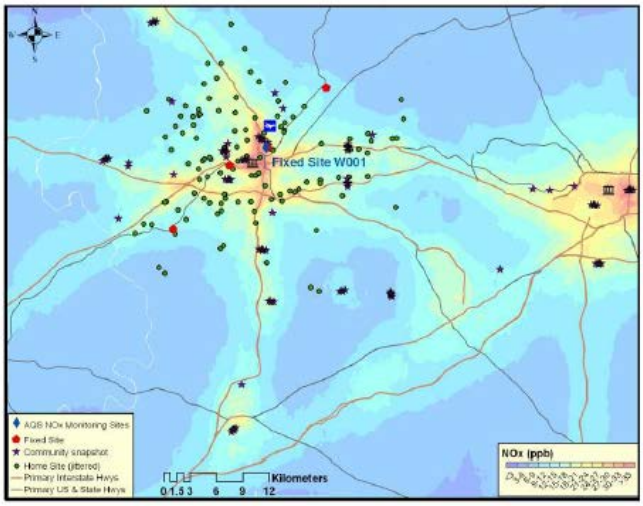

Colocated L001
Colocated LR002
AQS NOx Monitoring Site
Fixed Site
Community Snapshot
Home Site (jittered)
Primary Interstate Hwy
Primary US & State Hwy
0 2 4 8 12 16
Kilometers
NO2 (ppb)

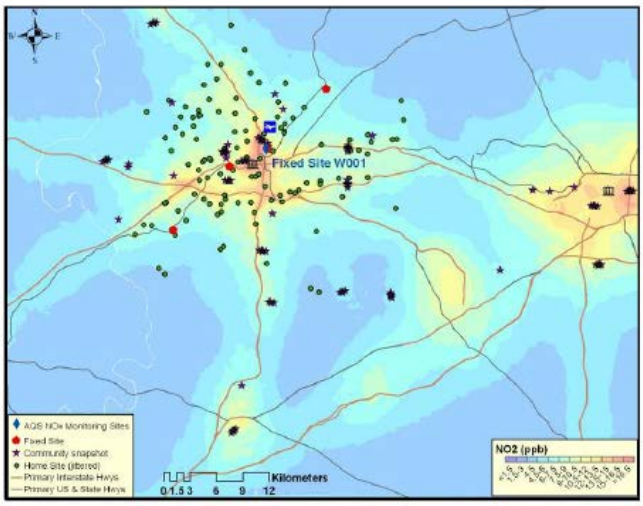

Fixed Site NR001
Colocated N001
AQS NOx Monitoring Site
Fixed Site
Community Snapshot
Home Site (jittered)
Primary Interstate Hwy
Primary US & State Hwy
0 1 2 4 6 8
Kilometers
NOx (ppb)


Fixed Site NR001
Colocated N001
AQS NOx Monitoring Site
Fixed Site
Community Snapshot
Home Site (jittered)
Primary Interstate Hwy
Primary US & State Hwy
0 1 2 4 6 8
Kilometers
NO2 (ppb)


AQS NOx Monitoring Sites
Fixed Site
Community Snapshot
Home Site (jittered)
Primary Interstate Hwys
Primary US & State Hwys
NOx (ppb)
Kilometers
0 1.5 3 6 9 12
Colocated Site S004


AQS NOx Monitoring Sites
Fixed Site
Community Snapshot
Home Site (jittered)
Primary Interstate Hwys
Primary US & State Hwys
NO2 (ppb)
Kilometers
0 1.5 3 6 9 12
Colocated Site S004


Fixed Site W001
AQS NOx Monitoring Sites
Fixed Site
Community snapshot
Home Site (jittered)
Primary Interstate Hwys
Primary US & State Hwys
0 1.5 3 6 9 12
Kilometers
NOx (ppb)


Fixed Site W001
AQS NOx Monitoring Sites
Fixed Site
Community snapshot
Home Site (jittered)
Primary Interstate Hwys
Primary US & State Hwys
0 1.5 3 6 9 12
Kilometers
NO2 (ppb)