Land use regression models for exposure assessment of non-tailpipe PM constituents and sources

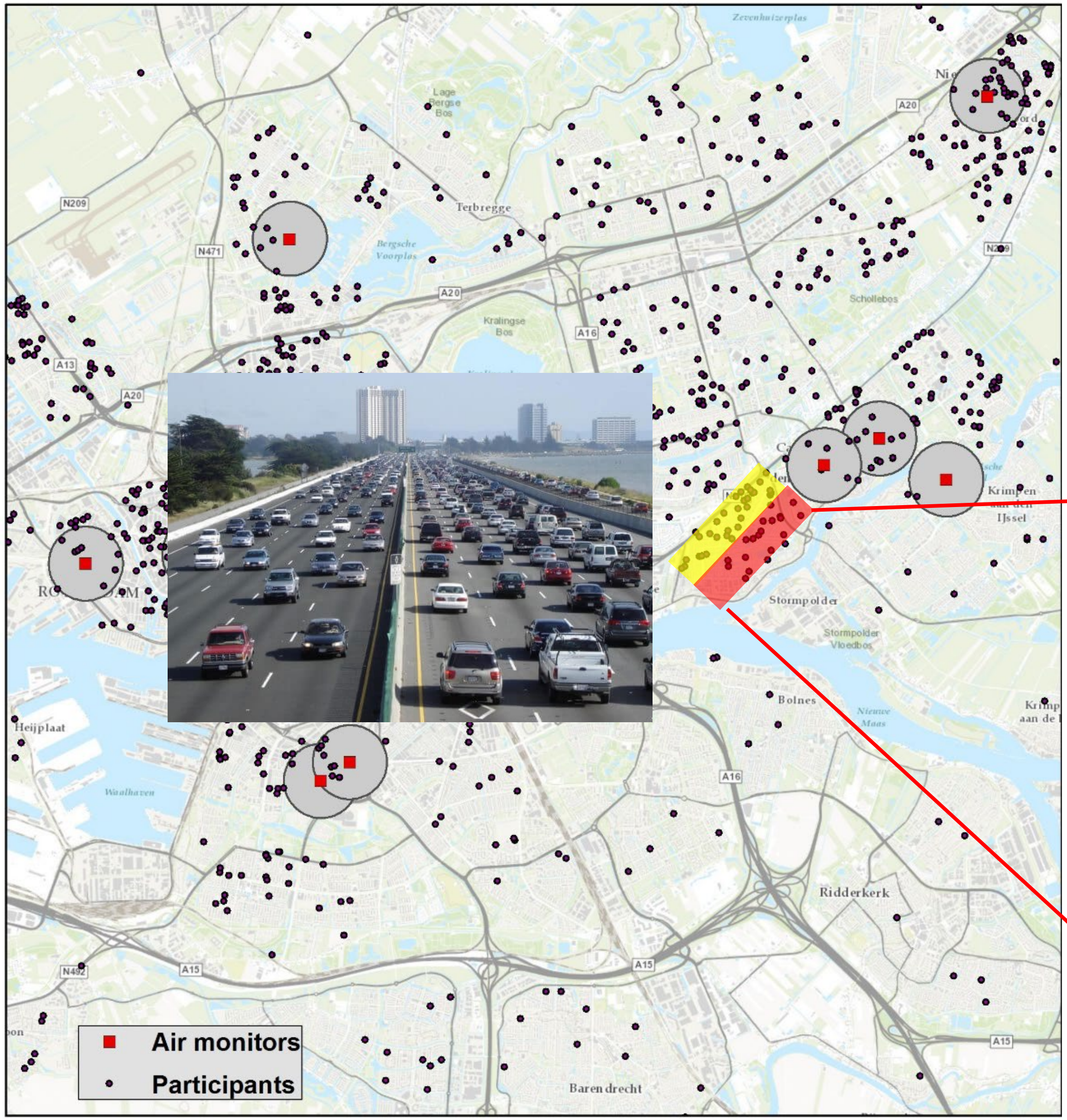
Meng Wang PhD
Department of Epidemiology and Environmental Health
University at Buffalo, SUNY
2020-11-13

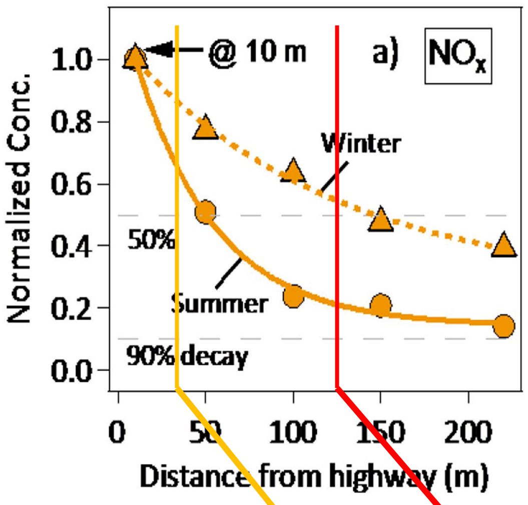
HEI Non-Tailpipe Workshop

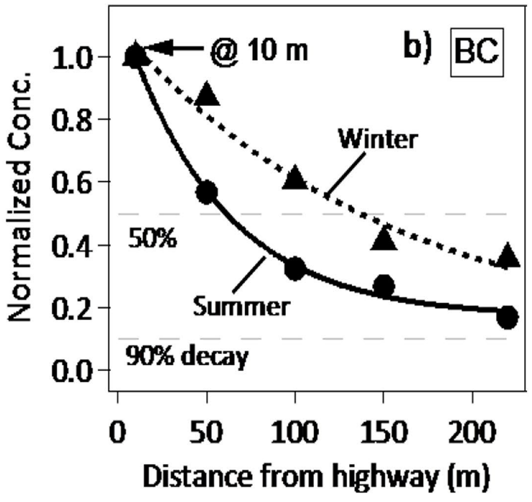
Outline

- What is land use regression model?
- Land use regression model for non-tailpipe PM constituents
- Land use regression model and source apportionment for PM mixtures

Challenges of assessing exposure to traffic air pollution







Saha et al. 2018 Atmo Environ

Land Use Regression Modeling

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_n X_n + c$$

Air pollution measurement

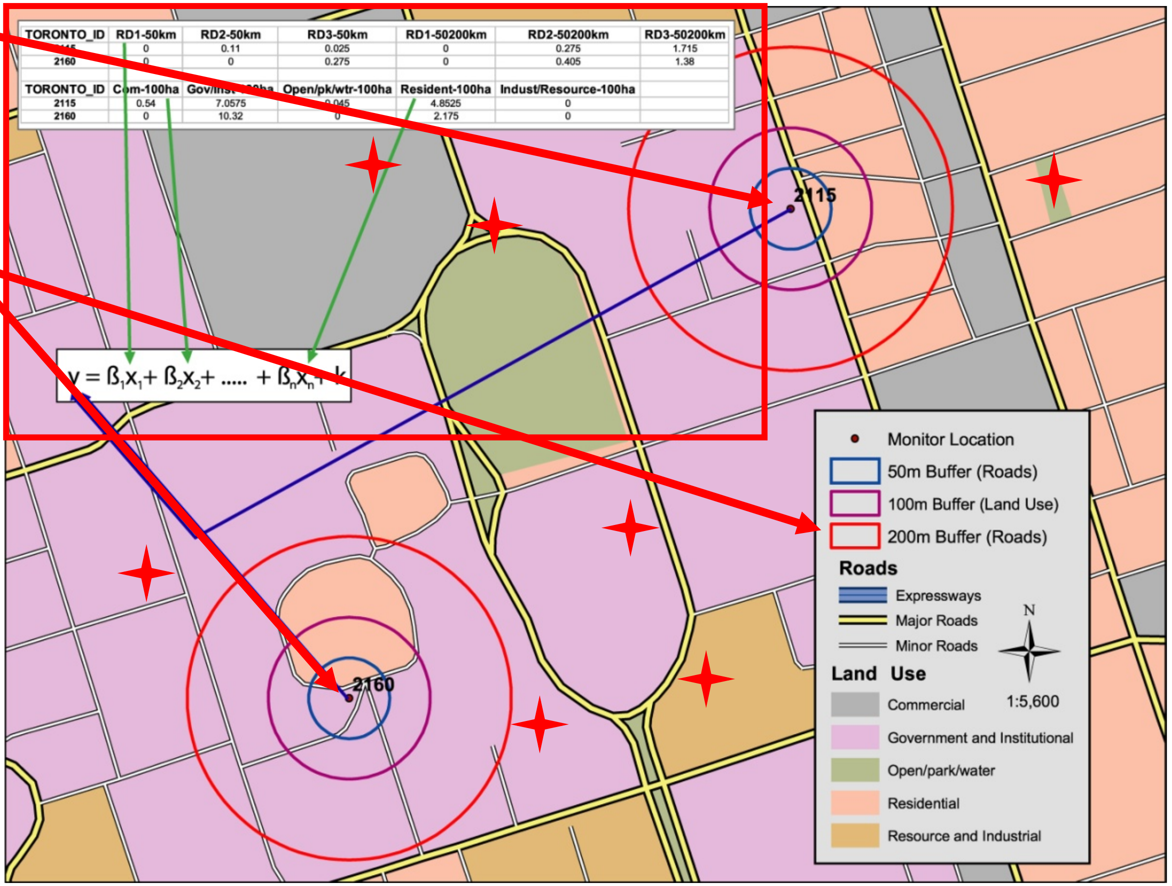
Collect GIS and other variables

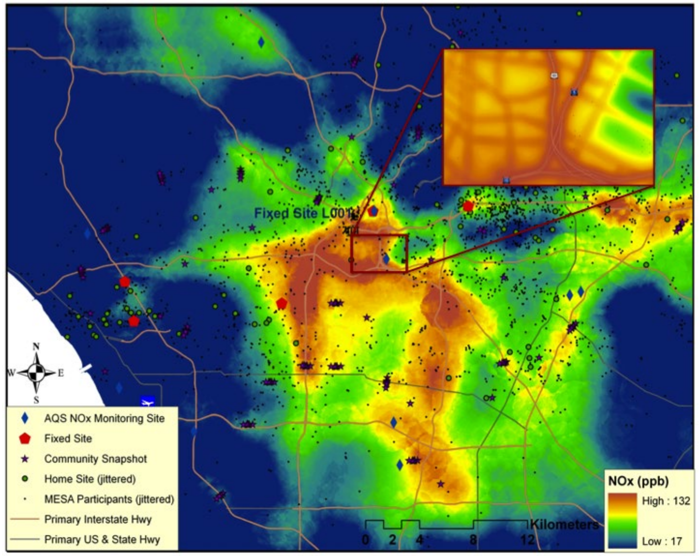
Statistical modeling and evaluation

- Linear
- Non-linear
- Machine learning

Prediction

- Geocoded homes
- Spatial mapping





Jerrett et al. 2007

Outline

- What is land use regression model?
- Land use regression model for non-tailpipe PM constituents
- Land use regression model and source apportionment for PM mixtures

Non-tailpipe PM components

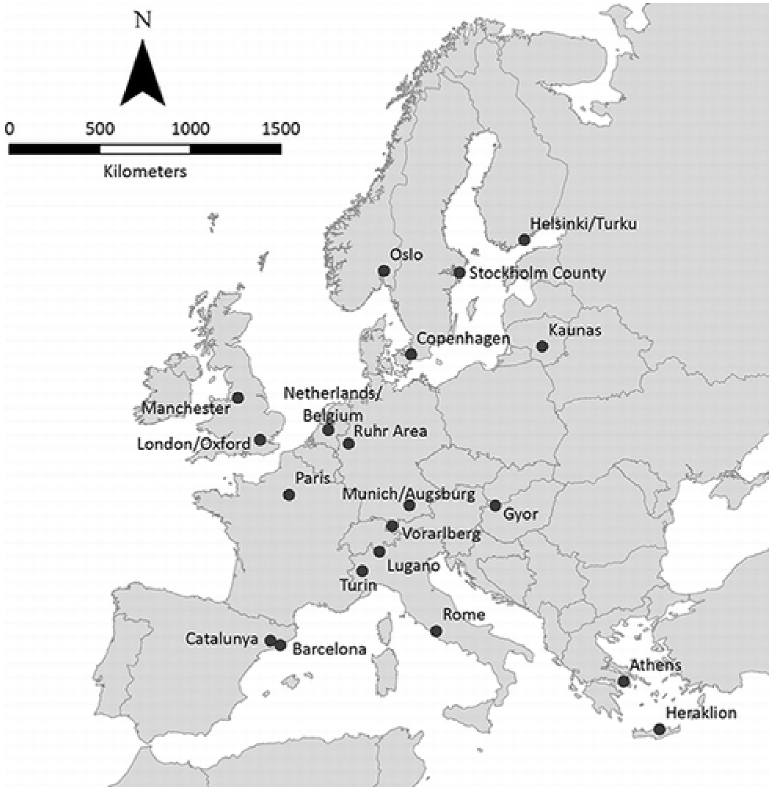
- Non-tailpipe PM is becoming more important but detailed information is relatively scarce
- Non-tailpipe emissions are typically characterized by trace metals in PM (e.g. Cu, Fe, Zn).
- LUR model for trace metals in PM has been studied in cities of Europe, U.S., and Canada.

Key tracers used for non-exhaust PM.

Reference	Brake wear	Tyre wear	Re-suspension
Adachi and Tainosho (2004)	Fe, Ba, Cu, Sb, Zr	Zn	–
Schauer et al. (2006)	Fe, Cu, Ba		–
Grieshop et al. (2006)	Cu, Sb, Ba and Ga		
Wahlin et al. (2006)	Cr, Fe, Cu, Zn, Zr, Mo, Sn, Sb, Ba and Pb	Al, Si, K, Ca, Ti, Mn, Fe, Zn and Sr (together with road dust)	
Tanner et al. (2008)	Cu, Cd	Zn	
Canepari et al. (2008)	Ba, Fe, Sb, Sr	–	–
Harrison (2009)	Ba, Cu	–	Al
Dongarra et al. (2009)	Cu, Mo, Sb	–	
Fabretti et al. (2009)	Cu, Zn, Sb, Sn (vehicular abrasion)	–	Rb, Sr, Mn, Fe, As
Keuken et al. (2010)	Cu	Zn	–
Bukowiecki et al. (2010)	Fe, Cu, Zn, Zr, Mo, Sn, Sb and Ba		
Pey et al. (2010)	–	–	Fe, Ca, Sb, Sn, Cu, Zn
Perez et al. (2010)	Sb, Cu, Ni, Sn (wear of brake, tire and other parts)		Fe
Amato et al. (2011a)	Fe, Cu, Zn, Cr, Sn, Sb	OC, S, Zn	Al, Ca, Fe, V
Apeagyei et al. (2011)	Fe, Ti, Cu, Ba	Zn, Ca, W, K, Fe, Ti, Cr, Mo	–
Duong and Lee (2011)	Ni, Cu	Zn	
Ondráček et al. (2011)	Cu, Ba, Fe, Zn	–	
Song and Gao (2011)	Sb, Cu, Fe, Pb	Zn, Co	
Sahu et al. (2011)	Zn (brake and tyre wear)	–	
Peltier et al. (2011)	–	–	Al, Si, Ti, Fe
Harrison et al. (2012b)	Ba, Cu, Fe, Sb	Zn	Si, Al

Pant et al. 2013 Atoms Environ

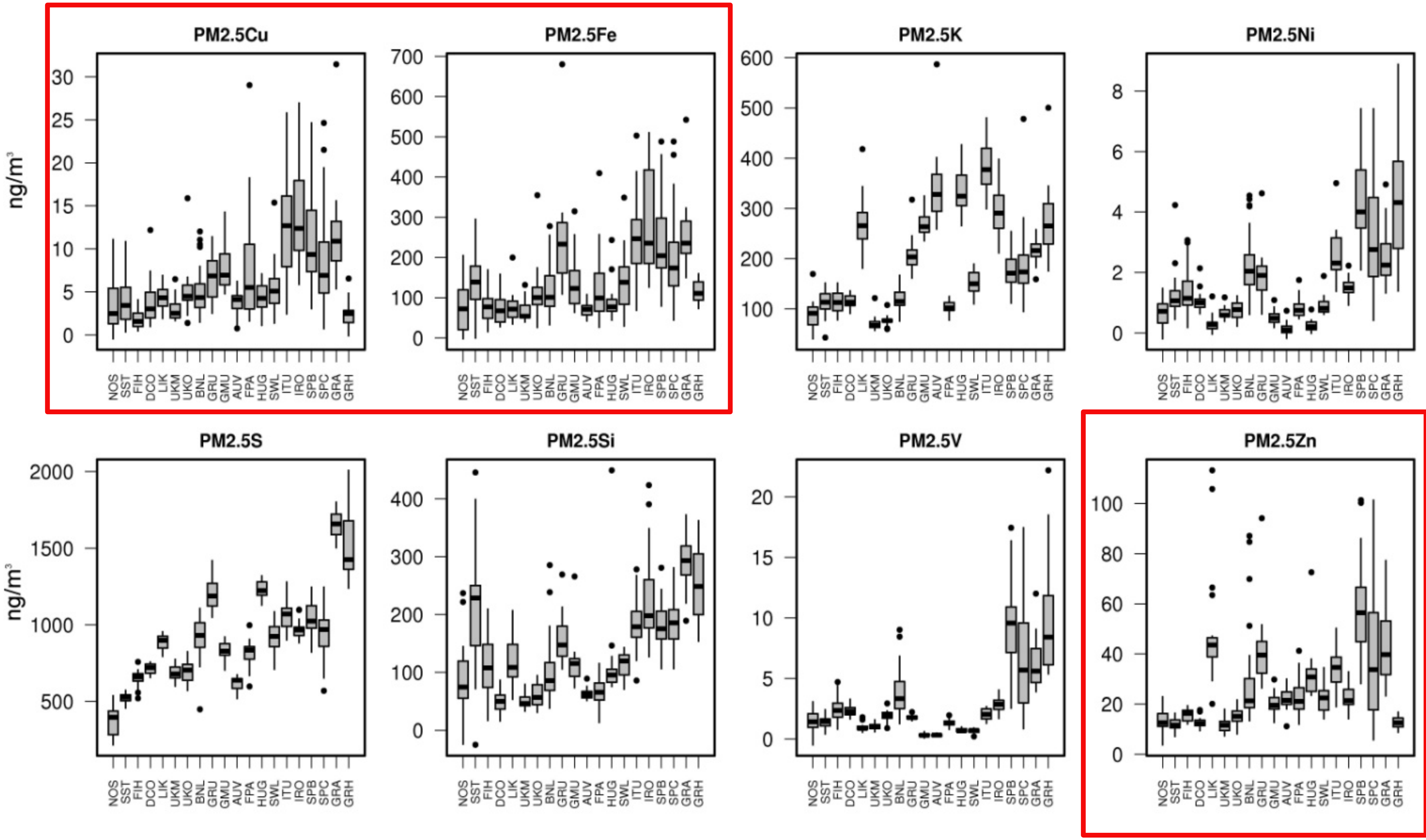
$PM_{2.5}$ Metal Elements in 20 European Cities



European Study of Cohorts for Air Pollution Effects (ESCAPE) (www.escapeproject.eu)

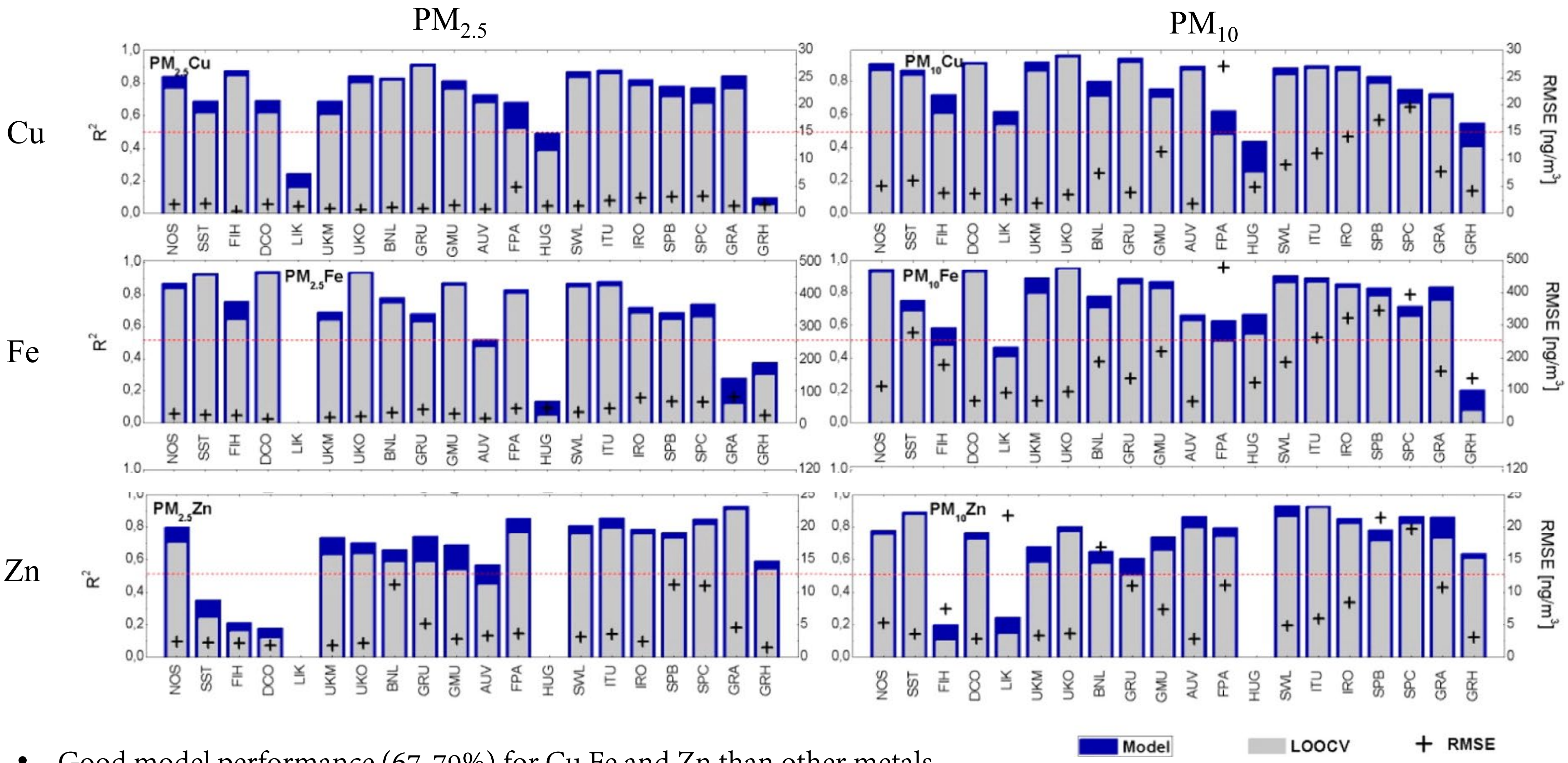
- 20-40 sites per city
- 3 seasons per site

Annual mean concentrations of trace metals in $PM_{2.5}$ across the cities



de Hoogh et al. 2013 ES&T

Performance of LUR Models for metal elements in $PM_{2.5}$ and PM_{10}

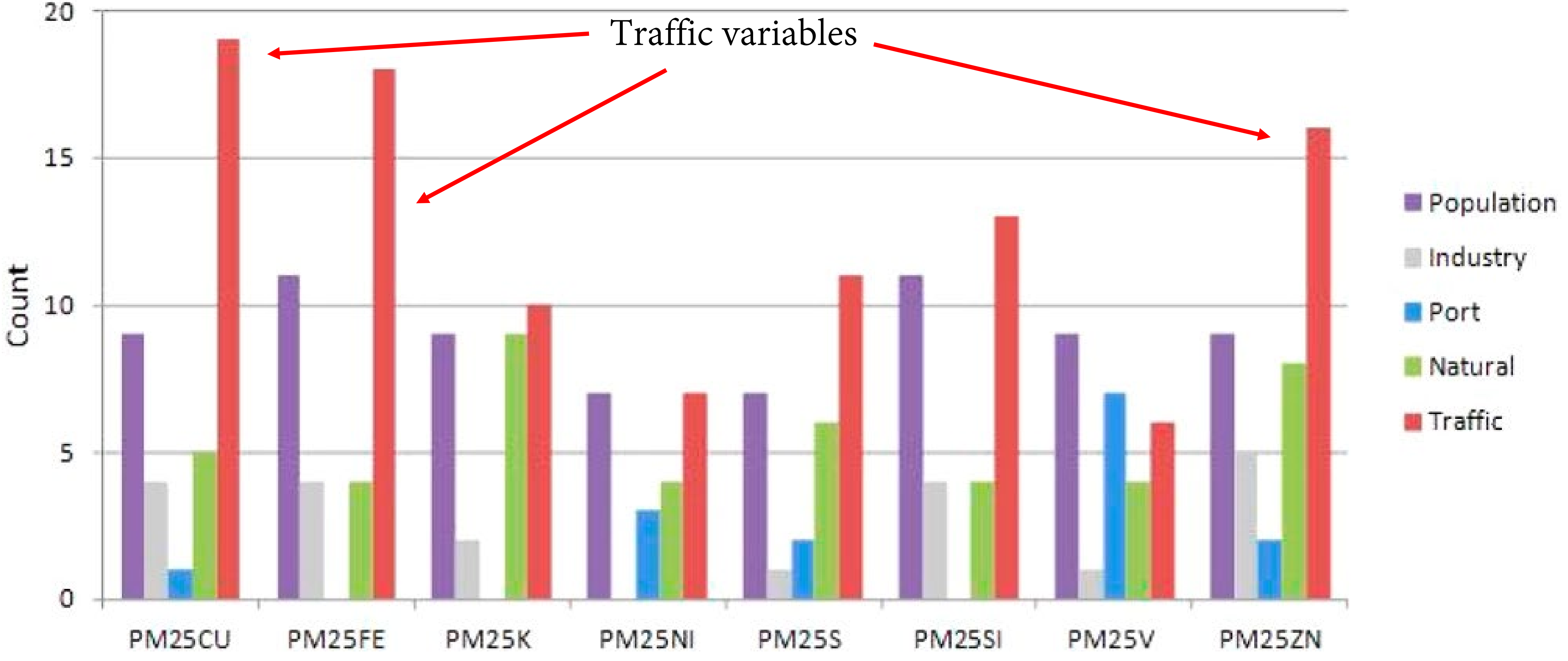


- Good model performance (67-79%) for Cu Fe and Zn than other metals
- Better model in PM_{10} than in $PM_{2.5}$

- Model: model fitting R^2
- LOOCV: leave-one-out-cross-validation R^2
- RMSE: root mean square error

de Hoogh et al. 2013 ES&T

Predictor variables for metal elements in $PM_{2.5}$ and PM_{10}



Traffic variables were the dominant predictors in Cu, Fe, and Zn, reflecting non-tailpipe emissions.

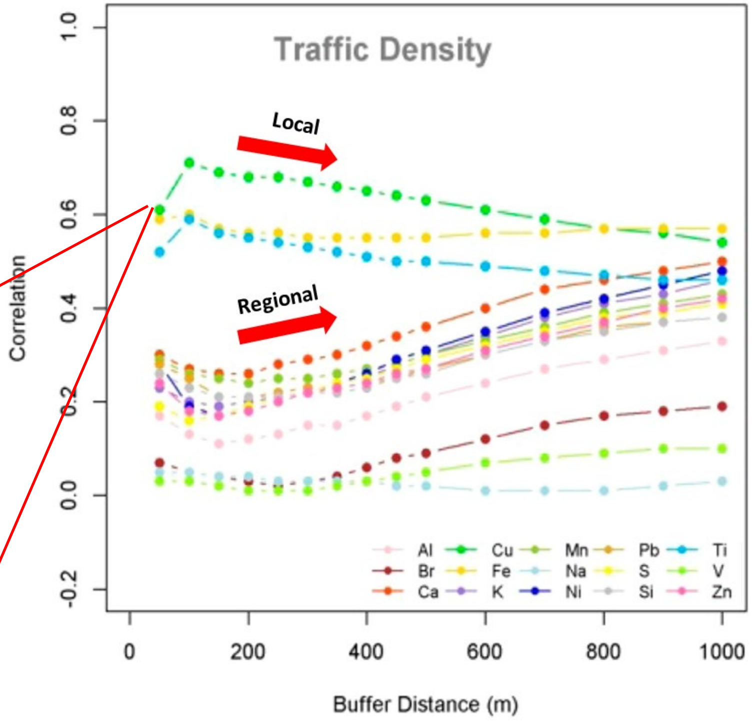
de Hoogh et al. 2013 ES&T

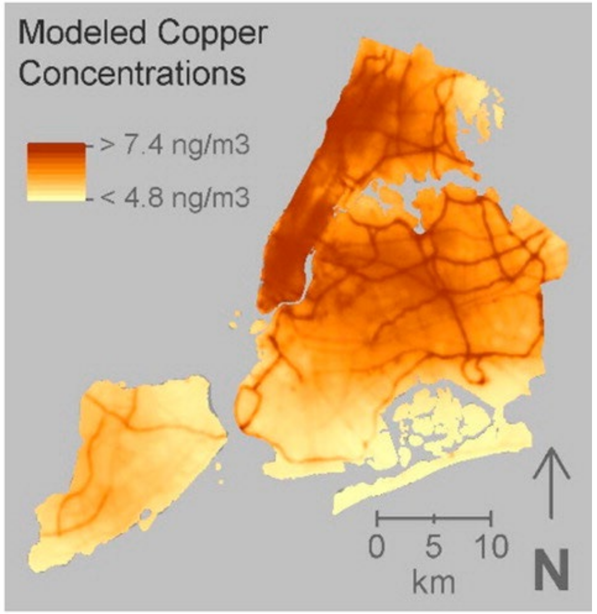
$PM_{2.5}$ Metal Elements in New York City

Table 1. R^2 Values for Temporal Adjustment Model and LUR Models Developed Using the Data From 150 Sites during the First Year of Sampling[a]

		spatial land-use regression (N = 150)		
	temporal trends (N = 703)	emission indicators only	with spatial autocorrelation	leave-one-out-cross-validation
Al	0.43	0.45	0.48	0.50
Br	0.62	0.43	n/a	0.24
Ca	0.74	0.75	0.82	0.85
Cu	0.69	0.58	0.64	0.68
Fe	0.86	0.63	n/a	0.60
K	0.68	0.55	0.61	0.64
Mn	0.55	0.50	n/a	0.36
Na	0.53	n/a	n/a	n/a
Ni	0.69	0.72	0.84	0.85
Pb	0.35	0.24	0.33	0.38
S	0.94	0.52	0.63	0.67
Si	0.65	0.64	0.67	0.69
Ti	0.60	0.51	n/a	0.49
V	0.68	0.67	0.80	0.82
Zn	0.73	0.54	0.77	0.80
$PM_{2.5}$	0.93	0.74	0.80	0.83

[a]December, 2008–November, 2009.



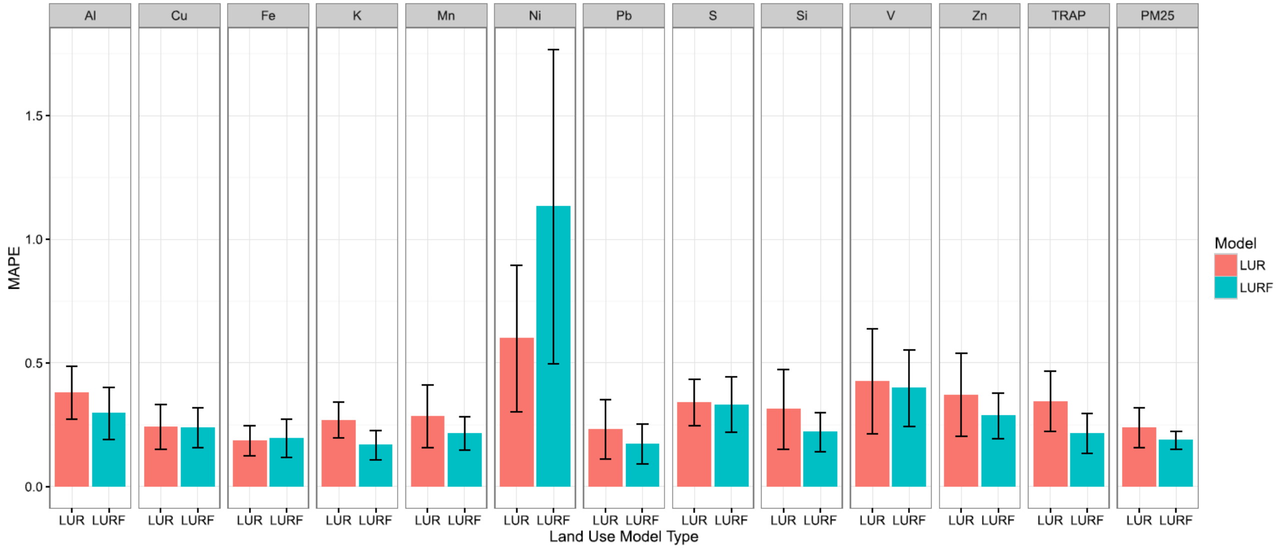


Ito et al. 2016 ES&T

Country & continent	Cities	N sites	Non-tailpipe metals	CV R^2	Predictors ranks 1-3	Correlation (R^2) with tailpipe emissions
Europe ($PM_{2.5}$ PM_{10}) De Hoogh, 2013, ES&T	20 cities summary	20-40 sites, 1-year	Cu Fe Zn	0.73 0.70 0.66	Traffic (1st) Population (2nd) Industrial (3rd) Traffic (1st) Population (2nd) Natural (3rd) Traffic (1st) Population (2nd) Natural (3rd)	0.55-0.60 (NO2/NOx, BC) 0.59-0.72 (NO2/NOx, BC) 0.29-0.66 (NO2/NOx, BC)
U.S. ($PM_{2.5}$) Ito, 2016, ES&T	New York City NY	150, 3-year	Cu Fe Ti	0.68 0.60 0.49	Traffic (1st) Res oil boiler (2nd) Industrial (3rd) Traffic (1st) Res oil boiler (2nd) Cooking (3rd) Traffic (1st) Industrial (2nd) Res oil boiler (3rd)	NA
U.S. ($PM_{2.5}$) Brokamp, 2016, AE	Cincinnati OH	24, 1-year	Cu Fe Zn	0.92* 0.94* 0.85*	Land cover (1st) Shrub (2nd) truck count (3rd) Land cover (1st) truck count (2nd) Bus routes (1st) Distance to A3 road (2nd)	0.55 (EC) 0.58 (EC) 0.56 (EC)
U.S. ($PM_{2.5-10}$) Zhang, 2014, EHP	Chicago IL St. Paul MN Winston-Salem NC	33, 1-year 34, 1-year 35, 1-year	Cu Zn Cu Zn Cu Zn	0.65 0.73 0.86 0.40 0.41 0.36	Traffic (1st) CALINE (2nd) Land cover (3rd) Traffic (1st) Industrial (2nd) Traffic (1st) Residential (2nd) Land cover (3rd) Traffic (1st) Commercial(2nd) Land cover (3rd) Traffic (1st) Commercial(2nd) Land cover (3rd) Traffic (1st) Residential (2nd) Land cover (3rd)	NA
U.S. ($PM_{2.5}$) Tunno, 2016, STOTEN	Pittsburgh, PA	36, 1-year	Fe (summer) Fe (winter)	0.63 0.04	Bus density (1st) Parking garage (2nd) Reference site	0.11 (BC) 0.23 (BC)
Canada ($PM_{1.0}$) Zhang, 2015, AE	Calgary Alberta	25, 1-year	Cu (summer) Cu (winter) Fe Fe Ti Ti Zn Zn	0.74 0.82 0.86 0.74 0.82 0.75 0.73 0.77	Commercial (1st) Industrial (2nd) Industrial (1st) Traffic (2nd) Residential (3rd) Industrial (1st) Commercial (2nd) Traffic (3rd) Industrial (1st) Residential (2nd) Commercial (3rd) Industrial (1st) Commercial (2nd) Industrial (1st) Traffic (2nd) Industrial (1st) Commercial (2nd) Residential (3rd) Industrial (1st) Traffic (2nd) Residential (3rd)	NA

Challenges for non-tailpipe trace metal models

- Over-fitting due to small number of monitoring sites
- Relatively short-term measurements (mostly 2-3 seasons in 1 year)
- Traffic-related predictor variables cannot disentangle tailpipe and non-tailpipe PM species
  - Accounting for traffic congestion, such as traffic speed and speed limit.
  - Incorporate dispersion model predictions.
- Trace metals (e.g. Cu, Fe, Zn) may not exclusively represent non-tailpipe emissions
- Modeling approach was mostly based on linear regression approach
  - Variable selection vs. machine learning



Brokamp, 2016, Atmo Environ

Outline

- What is land use regression model?
- Land use regression model for non-tailpipe PM constituents
- Land use regression model and source apportionment for PM mixtures

Land Use Regression Model for PM Mixtures

PM species:
- Metals
- PAHs
- EC/OC

→

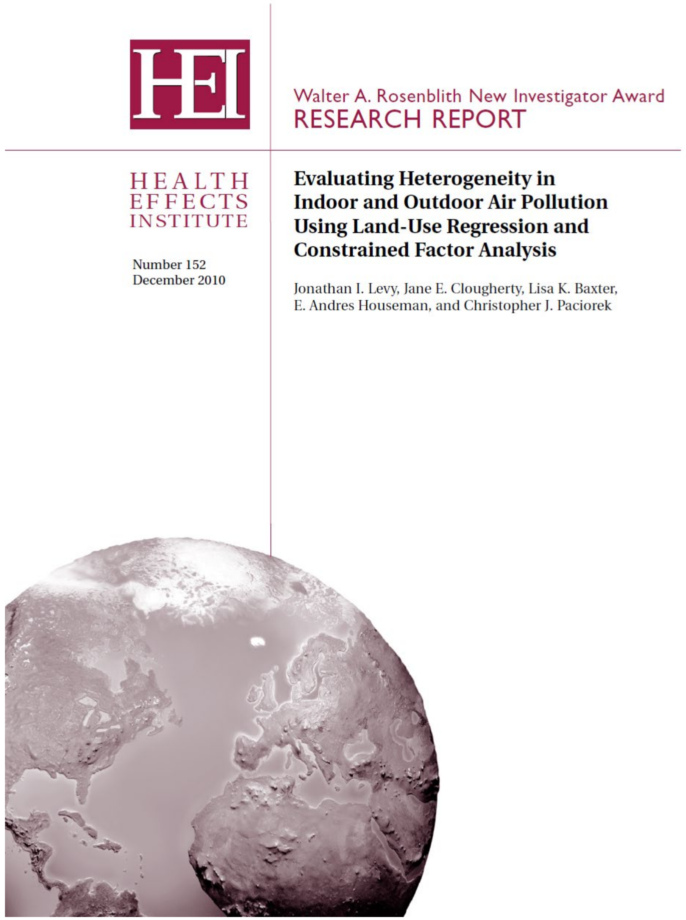
Source apportionment:
- Positive Matrix Factorization
- Principal component analysis
- K-means clustering

→

LUR model for spatially-derived factor scores as indicators of emission sources

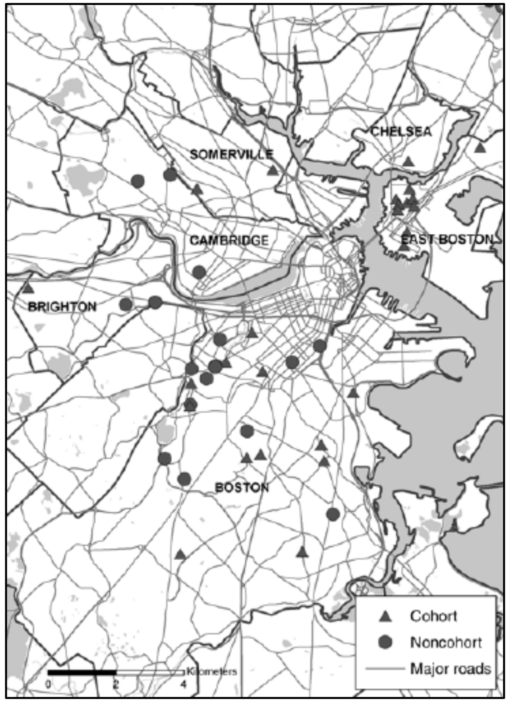
→

Predicting spatial distribution of each factor score for exposure assessment


HEI
Walter A. Rosenblith New Investigator Award
RESEARCH REPORT
HEALTH EFFECTS INSTITUTE
Number 152
December 2010
Evaluating Heterogeneity in Indoor and Outdoor Air Pollution Using Land-Use Regression and Constrained Factor Analysis
Jonathan I. Levy, Jane E. Clougherty, Lisa K. Baxter, E. Andres Houseman, and Christopher J. Paciorek

Air Pollutants (N=38 sites, 2 seasons)

PM2.5 mass, EC, NO_2, 23 $PM_{2.5}$ constituents by ICP-MS and XRF



Boston, MA

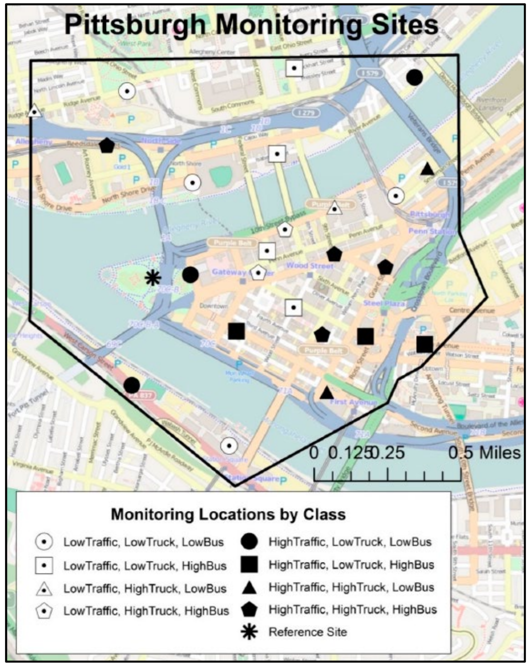
Table 11. Multivariate LUR Model Results for Outdoor Concentration Factor Scores, Adjusted for Repeated Measures by Household[a]

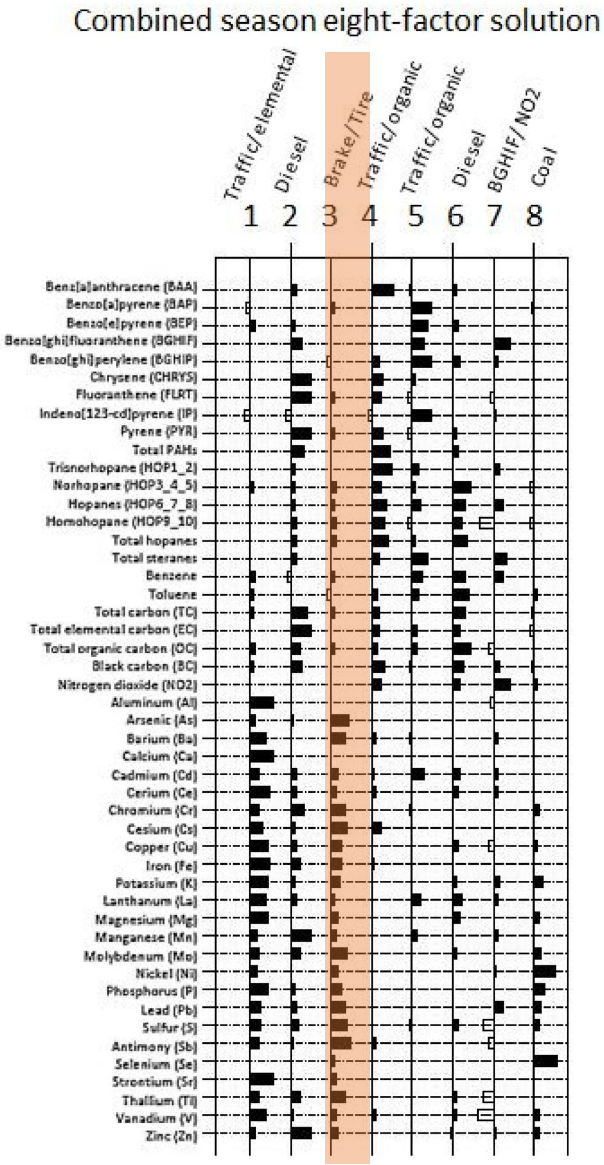
Factor: Interpretation / Covariate Type	Variable(s)	Univariate Coefficient	Univariate (P Value)	Seq $R^{2\,b}$
Factor 1: Long-Range Transport				
Central site	$PM_{2.5}$ (μg/m³)	0.129	(< 0.0001)	0.63
Season	Cooling season*	0.527	(< 0.0001)	**0.69** ←
Traffic terms	—			
Local sources / population	—			
Modified traffic and local sources	—			
Factor 2: Brake Wear and Local Traffic				
Central site	—			
Season	—			
Traffic terms	Roadway length within 100 m weighted by wind direction (m)	0.00369	(0.16)	0.02
Local sources / population	High density residential land use*	−0.362	(0.006)	0.11
Modified traffic and local sources	Roadway length within 100 m weighted by wind direction × cooling season (m)	0.00329	(0.04)	**0.16** ←
Factor 3: Diesel Exhaust				
Central site	$PM_{2.5}$ (μg/m³)	0.050	(0.001)	0.08
Season	Cooling season*	−0.268	(0.06)	0.13
Traffic terms	Diesel fraction on nearest major road (%)	6.07	(0.03)	0.26
Local sources / population	—			
Modified traffic and local sources	Diesel fraction on nearest major road × floor of building (%)	−1.72	(0.06)	**0.32** ←
Factor 4: Fuel Oil Combustion				
Central site	NO_2 (ppb)	59.7	(0.002)	0.14
Season	—			
Traffic terms	—			
Local sources / population	Population density within 200 m (pers/km²)	5.3×10^{-5}	(0.0003)	0.31
Modified traffic and local sources	Population density within 200 m × percent hours of wind < 2 m/sec (pers/km²)	2.1×10^{-4}	(0.02)	**0.41** ←
Factor 5: Road Dust and Resuspension				
Central site	—			
Season	Cooling season*	−0.144	(0.008)	0.08
Traffic terms	—			
Local sources / population	High density residential land use*	0.240	(0.002)	0.16
Modified traffic and local sources	High density residential land use × monitor obstructed from nearest road*	−0.145	(0.08)	**0.20** ←

Land Use Regression Model for PM Mixtures

Air Pollutants (N=36 sites, 2 seasons)

PM2.5 mass, BC, EC/OC, NO_2, 25 $PM_{2.5}$ constituents, PAHs, VOCs





Combined winter and summer factor score LUR (land use regression) results (n = 72).

Factor (% variance)	Proposed Sources	Final LUR Modeling Covariates (R^2)	Covariates Most Strongly Correlated with Factor Scores (r)
1 (36%)	Traffic-related elemental	Signaled intersections, 125 m (R^2 = 0.07)	Bus stop use, 100 m (r = 0.34) Signaled intersections, 125 m (r = 0.27) Commercial land use, 25 m (r = 0.25) Bus density, 25 m (r = 0.25)
2 (16%)	Diesel (chrysene, fluoranthene, Mn, pyrene, total carbon, total EC, Zn)	Bus density, 200 m Commercial land use, 25 m (R^2 = 0.37)	Signaled intersections, 200 m (r = 0.57) Bus stop use, 200 m (r = 0.56) Bus density, 200 m (r = 0.55) Truck density, 200 m (r = 0.51)
3 (11%)	Brake/ tire wear (Cu, Mo, Sb)	Primary and secondary roadways, 50 m (R^2 = 0.10)	Primary and secondary roadways, 50 m (r = 0.23)
4 (7%)	Traffic-related organic (benz[a]anthracene, hopanes, total PAHs)	Traffic density, 200 m (R^2 = 0.06)	Traffic density, 200 m (r = 0.25) Primary and secondary roadways, 200 m (r = 0.25)
5 (5%)	Traffic-related organic (benzo[a]pyrene, benzo[e]pyrene, benzo[ghi]perylene, indeno[123-cd]pyrene, and total steranes)	Primary roadways, 125 m $PM_{2.5}$ emissions, (R^2 = 0.19)	Primary roadways, 125 m (r = 0.33) Road complexity, 150 m (r = 0.30) $PM_{2.5}$ emissions (r = 0.30)
6 (4%)	Diesel (norhopane, toluene, total hopanes, total OC)	Bus density, 200 m (R^2 = 0.10)	Bus stop use, 175 m (r = 0.35) Bus density, 200 m (r = 0.32) Truck density, 200 m (r = 0.30)
7 (3%)	Benzo[ghi]fluoranthene, NO_2	Temperature Bus density, 200 m (R^2 = 0.73)	Temperature (r = −0.81) Wind speed (r = 0.56) Bus density, 200 m (r = 0.23)
8 (3%)	Coal (Ni, Se)	Commercial land use, 25 m Building density, 25 m Railroads, 200 m (R^2 = 0.11)	Commercial land use, 25 m (r = 0.40) Building density, 50 m (r = 0.33) Railroads, 200 m (r = 0.29)

- Model performed much better for exhausted sources than non-exhausted sources (i.e. brake/tire wear)

Tunno, 2018, Int. J. Environ. Res. Public Health

Summary

- Land use regression model has moderately good prediction ability for non-tailpipe PM metal elements, such as Cu, Fe, and Zn
- Land use regression model has limited power to predict source-apportioned non-exhaust sources.
- More advanced statistical approaches and more specific traffic-related predictor variables are needed for non-tailpipe PM and source predictions.

Thank You !

Welcome for Questions

mwang54@buffalo.edu