Supplementary Information for

Land Use Models for Elemental Components of Particulate Matter in an Urban Environment: A Comparison of Regression and Random Forest Models

Cole Brokamp, Roman Jandarov, MB Rao, Grace LeMasters, Patrick Ryan

Land Use Predictors

Transportation. Roadway data (both distance and total length) were extracted from 2002 TIGER/Line shapefiles [1] for each of the 5 classes of roads and railroads. Class 1 roads are primary highways with limited access; class 2 roads are primary roads without limited access; class 3 roads are secondary and connecting roads; class 4 roads are local, neighborhood, and rural roads; and class 5 roads are vehicular trails [2]. Annual average daily truck count on both interstates and highways were obtained from the Ohio Department of Transportation based on data from 2002 - 2005. Bus routes were obtained from local transit authorities and intersections were identified as locations where class 3 or 4 roads intersected one another.

Physical Features. An elevation raster supplied by the US Geological Survey was used to identify the elevation and the mean elevation within a buffer radius. The standard deviation of elevation as well as the fraction of elevation points more than 20 meters uphill (or downhill) were separately calculated as a measure of the elevation gradient.

Community Characteristics. Population totals for each census block were retrieved from the 2000 US Census [3]. For varying buffer radii, the population count was defined as the sum of the total population of all census blocks for which the census block centroid was contained within the buffer radius. The population density was the population count divided by the total area of the census blocks which were included in the population count.

Greenspace. Greenspace was estimated using satellite-derived normalized difference vegetation index (NDVI) images. A raster image of the Cincinnati area was obtained from the United States Forest Service and the average NDVI within varying buffer radii of each sampling site was extracted. NDVI ranges from -1 to 1 and a higher value represents more surrounding greenspace. Briefly, a cloud-free composite image with a resolution of roughly 100 by 100 feet for all of the Cincinnati area was created based on individual images collected in June of 2000 that differed by no more than 15 calendar days. Imagery digital numbers were converted to top of atmosphere reflectance (ToAR) using the standard Landsat calibration process. ToAR was then converted to surface reflectance by using the 6S atmospheric correction procedure as described previously [4].

Land Cover. The 2001 National Landcover Database from the United States Geological Survey was used to extract the percentage of each land class within varying buffer radii from each location. The raster file classifies 30 by 30 meters grids of land into 15 different land use classes: open water, developed open, developed low, developed medium, developed high, barren, deciduous forest, evergreen forest, mixed forest, shrub, grassland, pasture, crops, woody wetlands, herbaceous wetlands.

NEI Point Sources. The 2011 National Emissions Inventory (NEI) is a national compilation of emissions sources collected from state and local agencies as well as information from the Environmental Protection Agency (EPA) emissions programs including the Toxics Release Inventory (TRI). Point source sites and total emissions were obtained from the NEI for PM2.5, PM10, and the available modeled elements (Ni, Pb, and Mn). Land use models extracted from the NEI data included the distance to the nearest point source, total number of point sources, total point source emissions, average point source emissions, and point source emissions weighted by inverse distance to the source.

Land Use Regression Final Model Coefficients

Note that since the elemental concentrations were log transformed prior to modeling the coefficient estimates and standard errors are on the log scale.

Table 1: Regression coefficients from Al LUR Model

Term	Estimate	Standard Error	p-value
Intercept	3.23	0.07	< 0.001
developed.high_1200	1.77	0.31	< 0.001
lines.length_bus_100	0.00	0.00	0.003

Table 2: Regression coefficients from Cu LUR Model

Term	Estimate	Standard Error	p-value
Intercept	0.60	0.07	< 0.001
developed.high_1000	2.34	0.26	< 0.001
shrub_1500	-74.20	24.29	0.006
interstate.truck_800	0.00	0.00	0.039

Table 3: Regression coefficients from Fe LUR Model

Term	Estimate	Standard Error	p-value
Intercept	4.09	0.04	< 0.001
developed.high_1000	2.63	0.20	< 0.001
interstate.truck_800	0.00	0.00	0.012

Table 4: Regression coefficients from K LUR Model

Term	Estimate	Standard Error	p-value
Intercept	3.68	0.14	< 0.001
distance.to.roads_2	0.00	0.00	0.019
population.density_1750	139.29	44.94	0.006
lines.length_bus_150	0.00	0.00	0.048

Table 5: Regression coefficients from Mn LUR Model

Term	Estimate	Standard Error	p-value
Intercept	0.60	0.06	< 0.001
developed.high_1000	2.26	0.35	< 0.001
lines.length_railroads_1000	0.00	0.00	0.025

Table 6: Regression coefficients from Ni LUR Model

Term	Estimate	Standard Error	p-value
Intercept	-0.55	0.15	0.001
barren_1100	1371.28	361.11	0.001

Table 7: Regression coefficients from Pb LUR Model

Term	Estimate	Standard Error	p-value
Intercept	0.64	0.10	< 0.001
lines.length_bus_900	0.00	0.00	< 0.001
population.density_500	140.35	44.39	0.005

Table 8: Regression coefficients from S LUR Model

Term	Estimate	Standard Error	p-value
Intercept	7.41	0.06	< 0.001
highway.truck_350	-0.00	0.00	0.004

Table 9: Regression coefficients from Si LUR Model

Term	Estimate	Standard Error	p-value
Intercept	4.28	0.05	< 0.001
developed.high_1100	1.78	0.22	< 0.001
lines.length_bus_100	0.00	0.00	< 0.001

Table 10: Regression coefficients from V LUR Model

Term	Estimate	Standard Error	p-value
Intercept	-1.19	0.12	< 0.001
developed.high_1500	2.00	0.50	0.001
mixed.forest_1100	-417.16	211.71	0.063

Table 11: Regression coefficients from Zn LUR Model

Term	Estimate	Standard Error	p-value
Intercept	2.38	0.10	< 0.001
lines.length_bus_850	0.00	0.00	< 0.001
distance.to.roads_3	0.00	0.00	0.475

Table 12: Regression coefficients from TRAP LUR Model

Term	Estimate	Standard Error	p-value
Intercept	-1.26	0.08	< 0.001
developed.high_1000	1.14	0.30	0.001
interstate.truck_800	0.00	0.00	0.062
lines.length_roads1_1000	0.00	0.00	0.063
elevation.uphill_1000	0.28	0.17	0.117
shrub_1500	-12.63	25.55	0.627

Table 13: Regression coefficients from PM2.5 LUR Model

Term	Estimate	Standard Error	p-value
Intercept	2.66	0.06	< 0.001
lines.length_bus_350	0.00	0.00	< 0.001
lines.length_railroads_150	0.00	0.00	0.085
woody.wetlands_1300	-301.30	153.03	0.064
distance.to.roads_2	0.00	0.00	0.128

References

[1] 2002 tiger/line files [machine-readable data files]. ua 2003. us department of commerce, geography division, us census bureau.

[2] 2002 tiger/line files technical documentation. ua 2003. us department of commerce, geography division, us census bureau.

[3] 2000 census of population and housing. ua 2004. us census bureau.

[4] Eric F Vermote, Didier Tanré, Jean Luc Deuzé, Maurice Herman, and J-J Morcette. Second simulation of the satellite signal in the solar spectrum, 6s: An overview. *Geoscience and Remote Sensing, IEEE Transactions on*, 35(3):675–686, 1997.