Mapping the Concentration of Ambient Air Pollutants for Central Massachusetts

Jerome Chia-Rung Yang, Hong Xia, Suzanne Edmunds, Lydia Meintel-Wade

Abstract

To understand how air pollution affects human health, mapping the concentration of ambient air pollutants is indispensable. In this study, we collected the air pollutant data of particulate matter (PM_{10} and $PM_{2.5}$) at 60 monitoring locations in central Massachusetts. We employed the "location-allocation" approach to optimize the spatial configuration of these monitoring stations. To characterize temporal variation, each location was measured twice, one during peak hours and one during non-peak. A DustScan Scout Model 3020 Aerosol Monitor and a Garmin GPS tracker were used to obtain measurements and geographic locations. After collecting data from the field, kriging methods were used to generate concentration maps. We successfully created concentration maps for $PM_{2.5}$ and PM_{10}, while suffering from the issues of insufficient sensor sensitivity, suboptimal allocation of monitoring locations due to previous data availability, and the neglect of small-scale variation. Although the highest concentrations in the results never exceeded the NAAQS, further studies on the actual effect of pollution on human health are needed.

Introduction

Most cities in Massachusetts have been going through the process of post-industrialization in the past century. Numerous abandoned industrial buildings were left when manufacturers moved out, while the descendants of minority laborers, who are relatively vulnerable to environmental hazards, still live in this area. It is vital to understand how residents in this area are affected by environmental pollution. Thus, we proposed a project "Exploring the Association between Low Birth Weight and Exposure to Air Pollution in Massachusetts" as a case study on the association between LBW and six pollutants (CO, lead, NO_2, PM_{10}, $PM_{2.5}$, O_3 and SO_2) on EPA's National Ambient Air Quality Standards (NAAQS) in Massachusetts (Table 1).

Table 1 Six Pollutants Listed on the National Ambient Air Quality Standards (NAAQS)

Pollutant	Primary/ Secondary	Averaging Time	Level	Form
Carbon Monoxide	primary	8-hour	9 ppm	Not to be exceeded more than once per year
		1-hour	35 ppm	
Lead	primary and secondary	Rolling 3 month average	0.15 μg/m^3 (1)	Not to be exceeded

Nitrogen Dioxide		primary	1-hour	100 ppb	98th percentile, averaged over 3 years
		primary and secondary	Annual	53 ppb (2)	Annual Mean
Ozone		primary and secondary	8-hour	0.075 ppm (3)	Annual fourth-highest daily maximum 8-hr concentration, averaged over 3 years
Particle Pollution	$PM_{2.5}$	primary and secondary	Annual	15 $\mu g/m^3$	annual mean, averaged over 3 years
			24-hour	35 $\mu g/m^3$	98th percentile, averaged over 3 years
	PM_{10}	primary and secondary	24-hour	150 $\mu g/m^3$	Not to be exceeded more than once per year on average over 3 years
Sulfur Dioxide		primary	1-hour	75 ppb (4)	99th percentile of 1-hour daily maximum concentrations, averaged over 3 years
		secondary	3-hour	0.5 ppm	Not to be exceeded more than once per year

(1) Final rule signed October 15, 2008. The 1978 lead standard (1.5 $\mu g/m^3$ as a quarterly average) remains in effect until one year after an area is designated for the 2008 standard, except that in areas designated nonattainment for the 1978, the 1978 standard remains in effect until implementation plans to attain or maintain the 2008 standard are approved.
(2) The official level of the annual NO_2 standard is 0.053 ppm, equal to 53 ppb, which is shown here for the purpose of clearer comparison to the 1-hour standard.
(3) Final rule signed March 12, 2008. The 1997 ozone standard (0.08 ppm, annual fourth-highest daily maximum 8-hour concentration, averaged over 3 years) and related implementation rules remain in place. In 1997, EPA revoked the 1-hour ozone standard (0.12 ppm, not to be exceeded more than once per year) in all areas, although some areas have continued obligations under that standard ("anti-backsliding"). The 1-hour ozone standard is attained when the expected number of days per calendar year with maximum hourly average concentrations above 0.12 ppm is less than or equal to 1.
(4) Final rule signed June 2, 2010. The 1971 annual and 24-hour SO_2 standards were revoked in that same rulemaking. However, these standards remain in effect until one year after an area is designated for the 2010 standard, except in areas designated nonattainment for the 1971 standards, where the 1971 standards remain in effect until implementation plans to attain or maintain the 2010 standard are approved.

This project has been going for one year. In the past year, we completed both descriptive and multivariate statistic analysis of the LBW dataset from the Massachusetts Department of Public Health (MassDPH). This dataset recorded 623,844 births from 2000 to 2007 with variables including infant's birth weight, sex and plurality as well as the mother's age, marital status, race, education, gestational age and diseases. We also applied a statistic, geographically weighted logistic regression (GWLR), to characterize the spatial heterogeneity of the association between LBW and air pollutant concentration. Those results were summarized in our paper "Exploring the Association between Low Birth Weight and Exposure to Lead Using Geographically Weighted Logistic Regression".

The biggest problem in the previous project is the lack of air pollutant concentration monitoring data in Massachusetts. In previous studies, two approaches have been employed to obtain the air pollutant concentration or exposure data within an area: modeling the dispersion from pollutant sources, or interpolating a smooth surface based on point observations. Two results of the former provided by the EPA are the National Air Toxics Assessment (NATA) and Risk-Screening Environmental Indicators (RSEI) model. Among six chemicals, only Diesel Particulate Matter (DPM) is available in the NATA, and ozone and lead are available in the RSEI. Their concentration values are estimated based on modeling instead of direct observations. The latter approach generates a smooth surface of air pollution concentration based on observations. However, currently there are only 28 unevenly located monitoring sites in Massachusetts (Figure 1). This number is insufficient for generation of a concentration surface characterizing the spatial heterogeneity within the state.

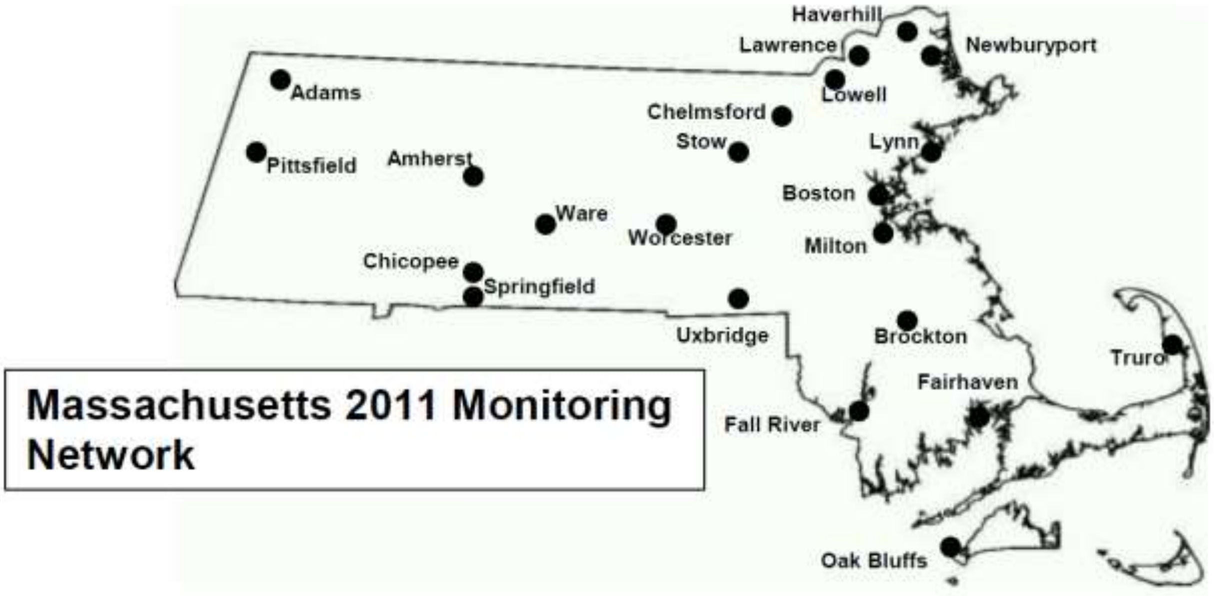


Figure 1 Configuration of the Massachusetts 2011 Monitoring Network

Therefore, we realized that collecting air pollution concentration data from the field is needed. We propose this project to obtain observations using a mobile monitoring station, which is composed of chemical monitoring equipment and a GPS unit installed on a car. These observations will be employed to interpolate an air pollutant concentration surface for each chemical. Although the ultimate goal is to measure six ambient air pollutants listed on the National Ambient Air Quality Standards (NAAQS) for the entire Commonwealth of Massachusetts, in this project we only focused on central Massachusetts. Only $PM_{2.5}$ and PM_{10} were completely measured and analyzed, while the experience of this study would enable us to explore the greater area of Massachusetts, as well as include other pollutants in the experiment.

Objectives

The primary objective of this project is to obtain the concentration observations of particulate matters ($PM_{2.5}$ and PM_{10}). As mentioned above, EPA's air pollution models (NATA and RSEI) only estimate the concentrations of diesel particulate matter (DPM), lead and ozone. Most of the EPA air monitoring sites provides data for all of the six NAAQS chemicals including $PM_{2.5}$ and PM_{10}, whereas the number of existing monitoring stations is not enough to interpolate a concentration surface, by which the concentration at any location in the state could be estimated. Only by collecting data from the field could we obtain more observations to generate the surface. Those observations were obtained at the locations and times we choose, which were systematically distributed within the state. A concentration surface was produced using interpolation tools in GIS for each chemical we explored. Using this surface with geocoded addresses of individuals, we were able to identify the amount of chemical concentration that each mother and baby exposed to.

The secondary goal of this project is to suggest candidate locations for additional ambient air pollution monitoring stations. When distributing our monitoring sites, we identified the locations that have a higher demand for an additional monitoring site. The demand was based on the spatial variation and dependency of concentrations across the study area. If any new temporary or permanent monitoring station is to be added to this area in the future, the locations we identified could be considered as the priority.

Methodology

Experiment Design

We measured pollutant concentrations at 60 locations within fourteen cities and towns (Auburn, Boylston, Clinton, Fitchberg, Grafton, Holden, Lancaster, Leominster, Lunenburg, Millbury, Shrewsbury, Sterling, West Boylston, Worcester; refer to Figure 2) around Worcester. Each location was measured twice. The configuration was decided by the spatial variation of concentrations from previously modeled data. Two times were chosen based on the temporal trend from previous monitoring data. The concentrations of $PM_{2.5}$ and PM_{10} in the air were measured in the experiment. The instruments that were used, as well as the approach to decide locations and times of measurements, are specified below.

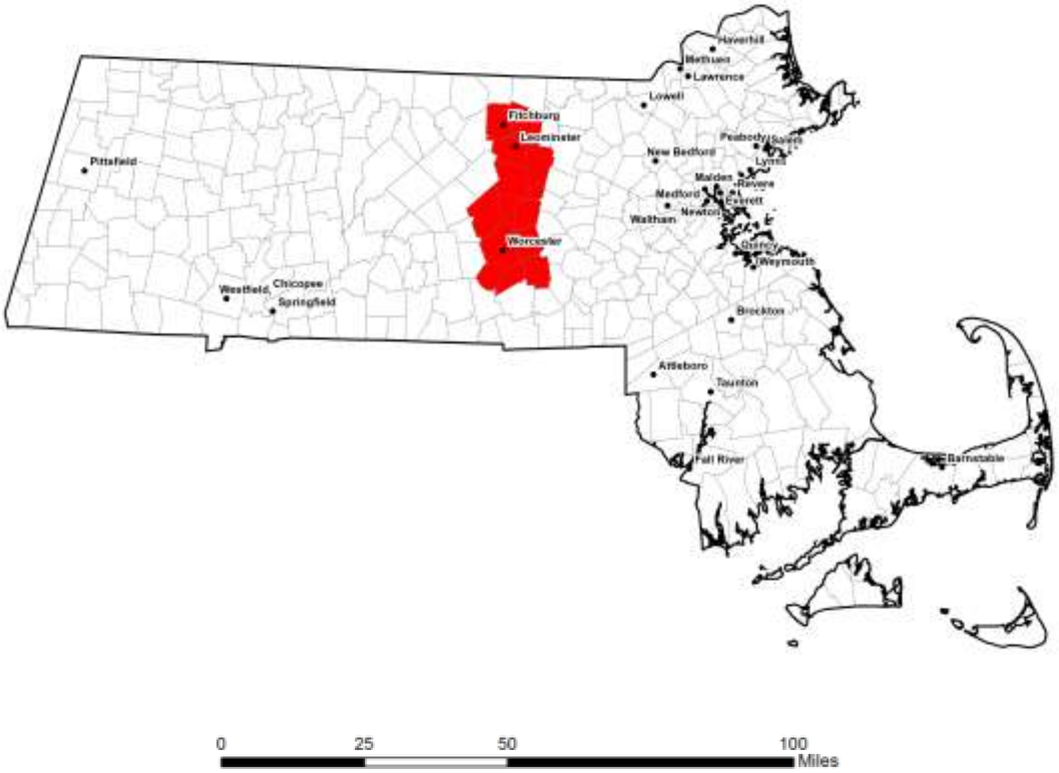


Figure 2 Location of 14 cities and towns included in the study area

1. Instrument

A DustScan Scout Model 3020 Aerosol Monitor (Figure 3) was employed to measure concentrations. Measurement ranges and correspondent errors are shown in Table 2. Since there was only one DustScan 3020 available to us, $PM_{2.5}$ and PM_{10} were measured by the same sensor with two different filters. Each type of PM was measured for one minute at one location at a time.

Figure 3 DustScan Scout Model 3020 Aerosel Monitor

Table 2 Measured pollutants and correspondent instruments

Pollutant	Range	Error	Instrument
$PM_{2.5}$, PM_{10}	0-0.5 mg/m³ 0-1 mg/m³ 0-10 mg/m³ 0-99.999 mg/m³	0.002 mg/m³ 0.002 mg/m³ 0.010 mg/m³ 0.100 mg/m³	DustScan Scout Model 3020 Aerosol Monitor

A Garmin GPS tracker was used to record the location for each measurement. This GPS tracker is able to constantly capture our real-time locations. By relating the time-stamps of these locations to the time-stamps on DustScan, we could easily get accurate coordinates for each measurement. This GPS tracker was also connected to a laptop. Google Earth was on the laptop with our real-time tracklogs, satellite imagery and marks of monitoring locations. This set-up enabled us to visualize our location on Google Earth during field trips (Figure 4).

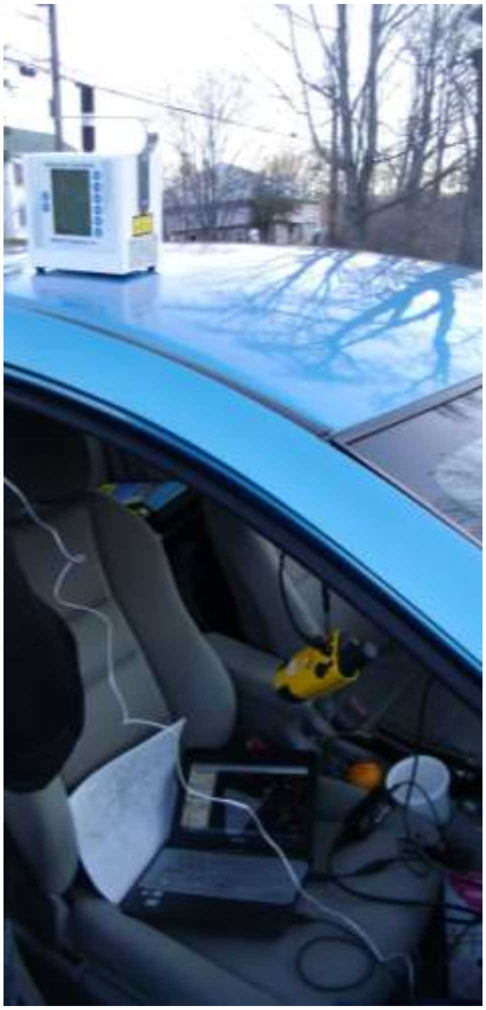


Figure 4 The sensor, laptop and GPS tracker were used together to provide spatial and pollution data at the same time

2. Spatial Optimization

Previous studies used the concept of "resource allocation" to decide where to put monitoring sites. Based on this concept, spatial variation is regarded as the demand for measurement. There is a higher demand at locations where the spatial variation is greater, and vice versa. In this approach, the process of site selection is generally composed of two steps: demand map generation and site allocation.

In the first step, researchers create a demand surface based on previous data. They calculated the spatial variation around each specific location. Greater variations indicate a greater demand for monitoring site. Kanaroglou et al. (2005) and Su et al. (2007) calculated total semivariance within a specific distance from each location and used this value as the "demand" at that location. Kumer et al. (2009) claimed that Kanaroglou et al. failed to consider spatial autocorrelation. They suggested the exclusion of variation within the range from each location (i.e. the range on semivariogram) can successfully avoid redundant site selection due to spatial autocorrelation.

In this study, we do not have a raster-based monitoring map available. We got the modeling data from NATA 2005, which were based on census tracts (Figure 5). Unlike three studies mentioned above, we cannot deal with this discrete (vector-based) data with the way they did on raster maps. We treated each census tract as an individual object and explore the variation among these objects in study area. In total there are 90 tracts. We adopted the approach that Kumer et al. (2009) used, which includes variation of each tract with its surrounding tracts while excluding variation within its range of spatial autocorrelation. For tract i with concentration z_i , the autocorrelation-adjusted spatial variance (σ^2) is:

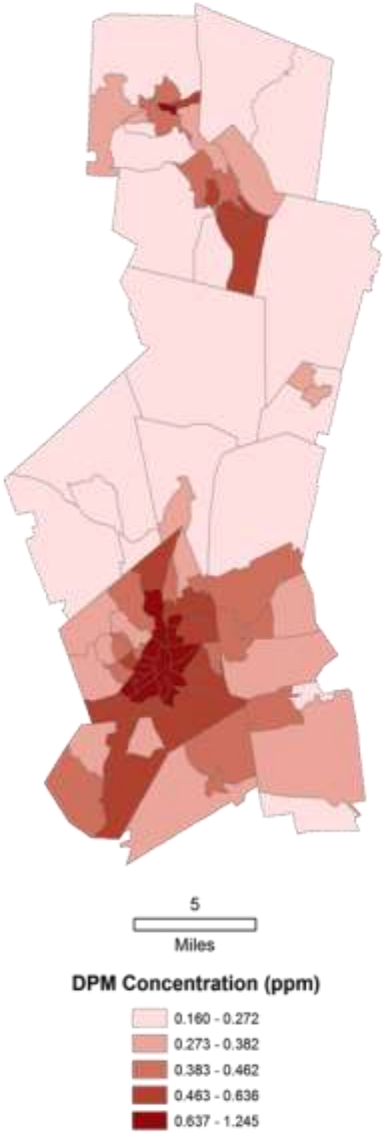
$$\boldsymbol{\sigma^2}_{\boldsymbol{i}} = \frac{\mathbf{1}}{\sum_{j=1}^{k} \forall_{ij}} \sum_{j=1}^{k} \left(z_i - z_j\right)^2 \forall_{ij} \qquad \text{...... } \textbf{Equation 1}$$

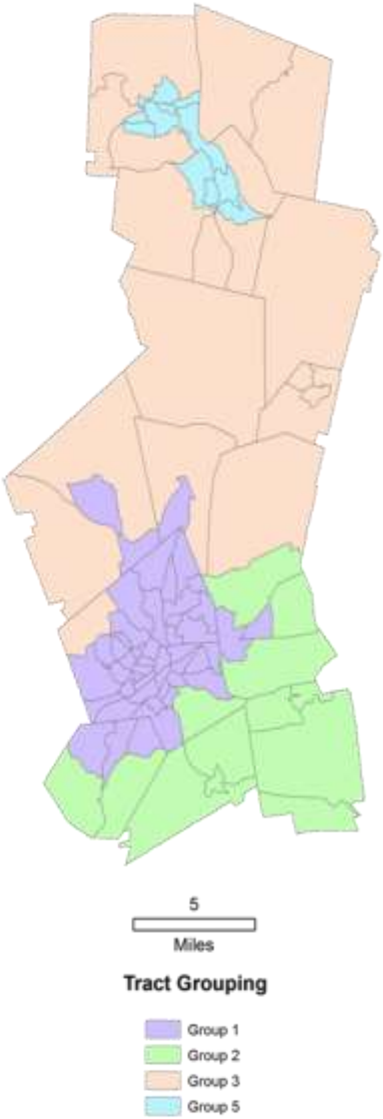
where j is a neighboring tract of i with concentration z_j; k is the number of neighbors around i; $i \neq j$; d_{ij} is the distance between I and j; h_i is the bandwidth for i determined by semivariance; $\forall_{ij}=0$ when $d_{ij}<h_i$, $\forall_{ij}=1$ otherwise.

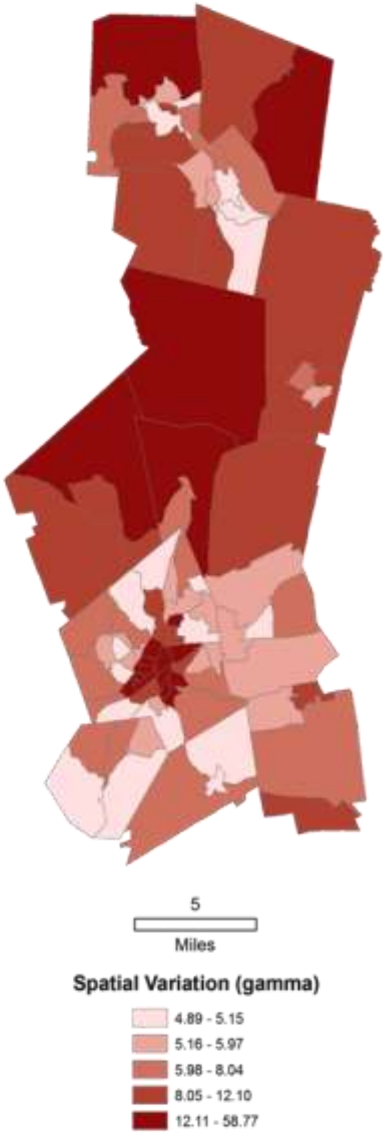
This "autocorrelation-adjusted" variance is regarded as the demand for monitoring sites of each census tract. Considering spatial heterogeneity within our study area, 90 tracts were divided into four regions based on their areas and concentration levels. (Figure 6) A bandwidth (h_i) was calculated for each region and applied to all tracts within that region.

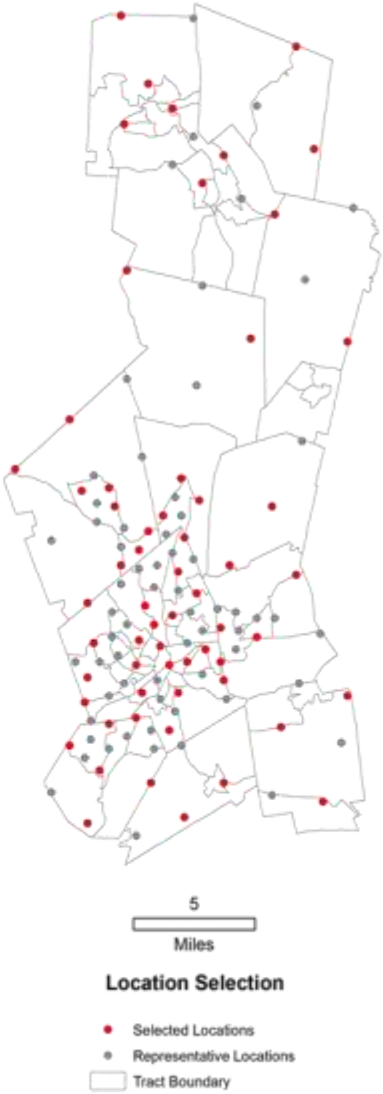
In the second step, we would like to allocate monitoring sites on this demand surface. Every potential location in study area was treated as a "candidate location" and chosen according to specific criteria. Kanaroglou et al. (2005) and Su et al. (2007) used the location-allocation tool in Network Analyst with the "maximum attendance" option to allocate monitoring site. This tool allowed them to ensure a given number of points are situated at locations where they together can capture maximum amount of variation. Kumer et al. (2009) used a program they composed to allocate a monitoring site to the locations that can characterize most variation within their study area. After the allocation of one site, the buffering area within its autocorrelation range was excluded from candidate locations. The next site was then chosen from the remaining locations. The same process until their targeted ratio of variation was attained within these monitoring sites.

In our study, we adopted the approach of Kumer et al. (2009). Each tract was divided into numerous 200*200m cells. Each cell inherits the demand value from the tract and was regarded as an individual candidate location (Figure 7). Those cells inaccessible from road were excluded from the candidate set. With the constrain of minimum distance between any two points, which equals the bandwidth (h_i), we used this process until no more location can be chosen from the study area. 124 representative locations were obtained (Figure 8). Then we manually selected 60 locations based on the following conditions: within the Worcester and Fitchberg urban areas, we avoided redundant selection of sites that cover the same or any nearby communities; within the remaining area, locations were chosen if they are closer to town centers while a complete coverage over the study area can be maintained (Figure 9).









From left to right:

Figure 5 Diesel particulate matter concentrations from NATA 2005 by census tracts
Figure 6 Four groups of tract used to determine adaptive bandwidth
Figure 7 Autocorrelation-adjusted spatial variation (gamma) by census tracts
Figure 8 Representative locations determined by a computer program
Figure 9 Manually selected 60 monitoring locations

Temporal Optimization

The goal of temporal optimization is to distribute two monitoring times optimally to capture the most temporal variation. Each location was only visited twice, one during peak hours and the other during non-peak. Since previous PM data were incomplete at the state monitoring station in Worcester, based on previous daily carbon monoxide monitoring data in April, we identified peak/non-peak hours as the period when the concentration is above/below daily average (Figure 10). To prevent an overlap between peak and non-peak, a buffer period was defined as the period when the concentration is less than 0.5 standard deviations above/below the daily average. No measurement was taken during the buffer period. Thus, peak hours are from 5am to 8am and non-peak hours are from 9am to 1:30pm.

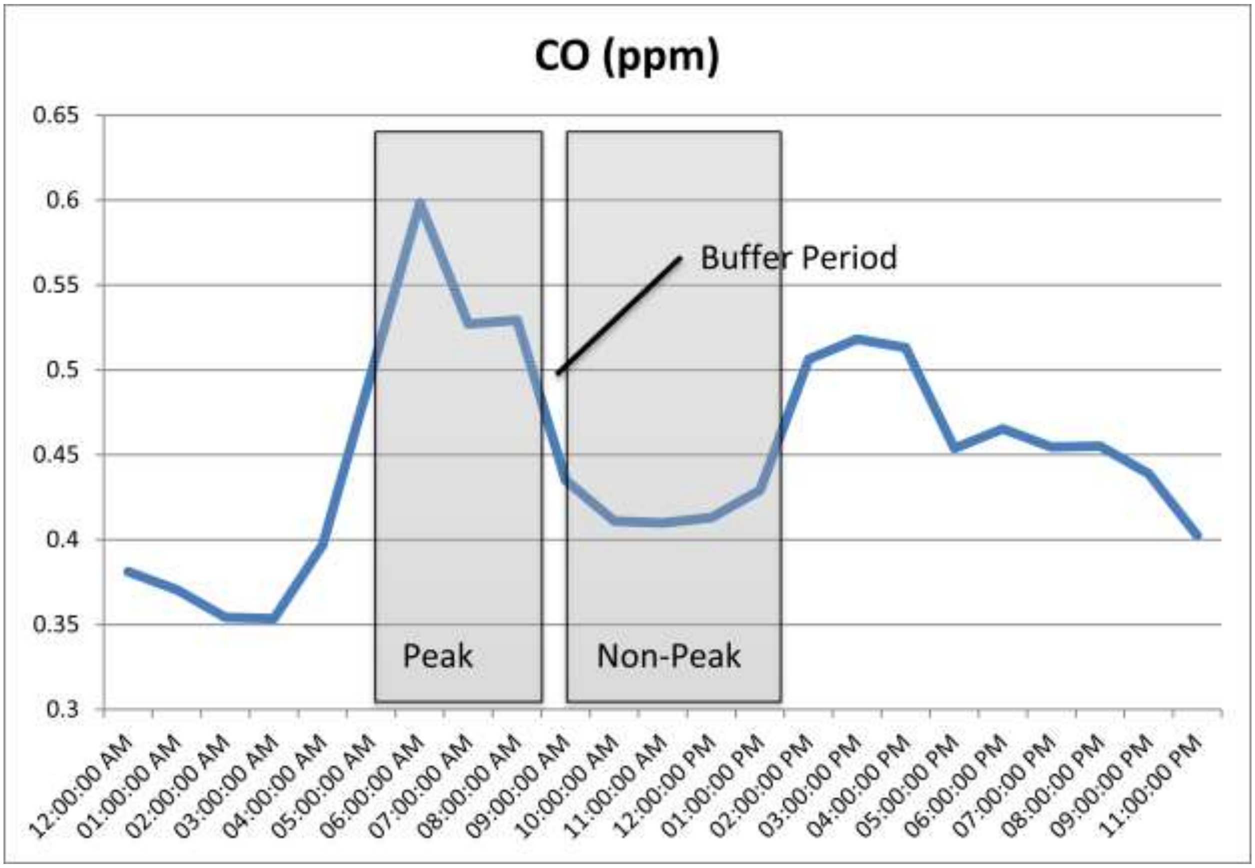


Figure 10 The temporal trend of CO concentration at Worcester in April

Map Production

We used the Geostatistical Analyst in ESRI ArcGIS 10.0 to create concentration maps. Four datasets were used to create four maps: $PM_{2.5}$ during peak hours, $PM_{2.5}$ during non-peak hours, PM_{10} during peak hours, and PM_{10} during non-peak hours. Each input dataset contains all point-based measurements we obtained (60 for non-peak and 59 for non-peak). We interpolated a concentration surface layer from these point-based observations.

In the Geostatistical Analyst, we first examined the normality of each dataset using a histogram and a normal Q-Q plot. The histogram and normal Q-Q plot also helped us indentify outliers. After this examination, only one measurement of PM_{10} during non-peak was excluded since it is quite deviant from other measurements.

Before creating kriging maps, the tool is able to examine if there is a global trend in study area. If so, it is recommended to get rid of the trend before examining semivariogram to explore spatial dependency. We chose global trend removal models that can minimize prediction errors in our results. Afterward, semivariogram was employed to decide which regression model would be used to describe the relationship between distance and semivariance. In addition, we got to decide if an anisotropic model characterizes this relationship better than an isotropic model (i.e. which yields smaller prediction errors). If an anisotropic was chosen, regression models with different parameters were applied to different directions. We found that anisotropic models are more suitable for the maps during non-peak hours.

After the model and parameters were determined, the tool also automatically offered a cross-validation report. This report indicated prediction errors in our models. These models, parameters and the results of cross-validation were shown in Table 3 and Appendix 2.

Table 3 Model Summary and Cross-Validation Results

Data		**PM2.5 (Peak)**	**PM10 (Peak)**	**PM2.5 (Non-Peak)**	**PM10 (Non-Peak)**
Sample Size		59	59	60	59
Model Summary (unit of input data: mg/m^3)					
Trend Removal	Type	Local Polynomial Interpolation	Local Polynomial Interpolation	Local Polynomial Interpolation	Local Polynomial Interpolation
	Power	2	2	2	3
Kriging Model	Type	Gaussian	Stable	Exponential	Spherical
	Nugget	$6.74 \cdot 10^{-7}$	$1.36 \cdot 10^{-6}$	$3.07 \cdot 10^{-7}$	$4.58 \cdot 10^{-7}$
	Parameter	N/A	0.2	N/A	N/A
	Range	3228	56640	14000	10000
	Partial Sill	$1.46 \cdot 10^{-6}$	0	$7.80 \cdot 10^{-7}$	$5.18 \cdot 10^{-7}$
	Function	0.346x+0.001	0.353x+0.002	-677.144x+1.326	0.379x+0.001
Anisotropy	Minor Range	N/A	N/A	4685	3386
	Direction	N/A	N/A	95°	107°
Results of Cross Validation – Prediction Error (unit of input data: mg/m^3)					
Mean		$3.45 \cdot 10^{-6}$	$2.19 \cdot 10^{-5}$	$-5.74 \cdot 10^{-6}$	$-6.23 \cdot 10^{-6}$
Root-Mean-Square		0.001253	0.001208	0.000957	0.000912
Mean Standardized		0.005398	0.017915	-0.001899	-0.002566
RMS Standardized		0.865121	1.006222	0.988293	0.946582
Ave. Std. Error		0.001455	0.001204	0.000976	0.000976

Results

Figure 11 demonstrates four concentration maps we made by spatial interpolation. Several high concentrations of $PM_{2.5}$ were found during peak hours. These hot spots are located at the Main South area in Worcester, I-90 near Grafton, and the center of West Boylston. Concentrations at the former two exceeded 5 µg/m³. As for PM_{10}, no single hot spot was observed, while the concentration increased gradually from Worcester to the southern boundary of Auburn. On the southern boundary of Auburn, concentrations higher than 5 µg/m³ were observed. During non-peak hours, both $PM_{2.5}$ and PM_{10} formed two strips of high concentration, one stretched from Holden to West Boylston and another from Worcester to Grafton. The patterns of them are fairly similar. Monitoring locations within these two strips include several locations near I-190, as well as a location near the intersection at Holden and another near West Boylston. Maximum concentrations in these two strips are slightly above 4 µg/m³.

These maps also enabled us to compare the concentration of any location to the NAAQS in Table 1. The NAAQS specifies the following exposure as hazardous: annual average 15 µg/m³ or 24-hour average 35 µg/m³ of $PM_{2.5}$, as well as 24-hour average 150 µg/m³ of PM_{10}. Based on our measurements and analyses, no observation was comparable to the EPA standard. Even the maximum concentrations we observed during peak hours did not exceed these standards.

Through the process of model optimization, the cross-validation results of kriging yield average standardized prediction errors around or less than 0.001 mg/m³ (1 µg/m³) for these four kriging maps (Table 3). This value is even smaller than the measurement errors of the DustScan. Numerical distributions of these errors are very likely to be Gaussian (Appendix 2). Looking at the spatial distribution of prediction errors (Figure 12), we found that $PM_{2.5}$ of non-peak hours, PM_{10} of peak and non-peak hours yield relatively small amount of errors, which are less than 0.001 mg/m³ (1 µg/m³). Apparently, $PM_{2.5}$ of peak hours has greater errors in most of the study area. Only areas near monitoring sites yield smaller prediction errors. Since the spatial patterns of non-peak hours were more spread out, anisotropic models are fairly suitable to map non-peak concentrations (Table 3).

The sources of error in this study include: (1) Process errors associated with experiment design, including short sampling period at one location, limited number of monitoring sites and times, and fluctuation in temperature. Because we were not able to include all the 127 representative locations into our experiment, some spatial variations were consequently missed by our experiment. The same issue also exists in the process of capturing temporal variation. The effects of temperature on measurements were also neglected in this study. (2) Measurement errors associated with the sensitivity and accuracy of the DustScan and Garmin GPS. The accuracy of DustScan is 0.002 mg/m³ (2 µg/m³) in our experiment. In other words, the measurement errors could be as high as 100% of the measurements when the measured value is 0.002 mg/m³ or less. (3) Analysis errors that came from the kriging method we chose. Based on the cross-validation results, they are around or less than 0.001 mg/m³ (1 µg/m³).

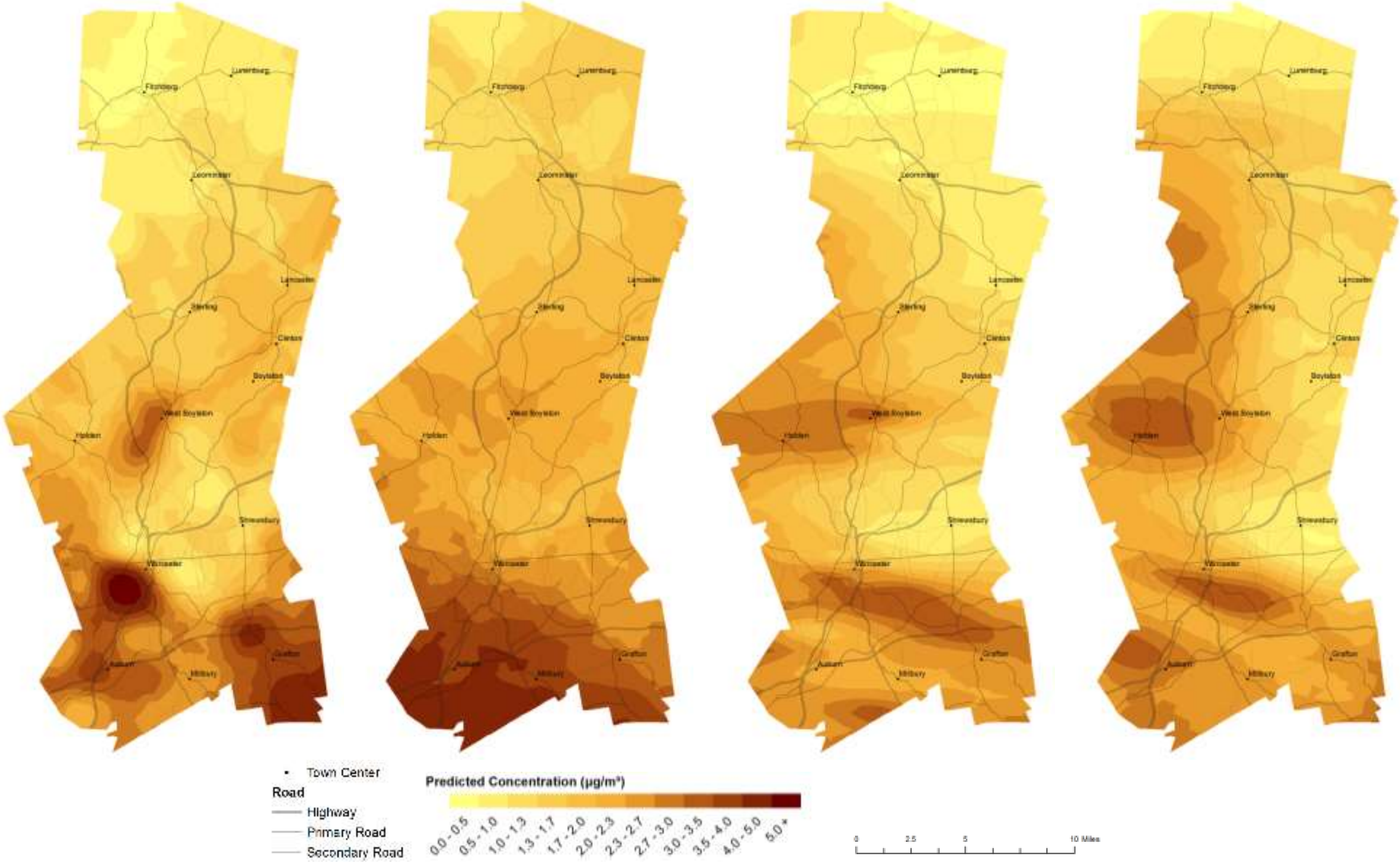


Figure 11 Results of Interpolation for $PM_{2.5}$ (Peak), PM_{10} (Peak), $PM_{2.5}$ (Non-Peak) and PM_{10} (Non-Peak) (from left to right)

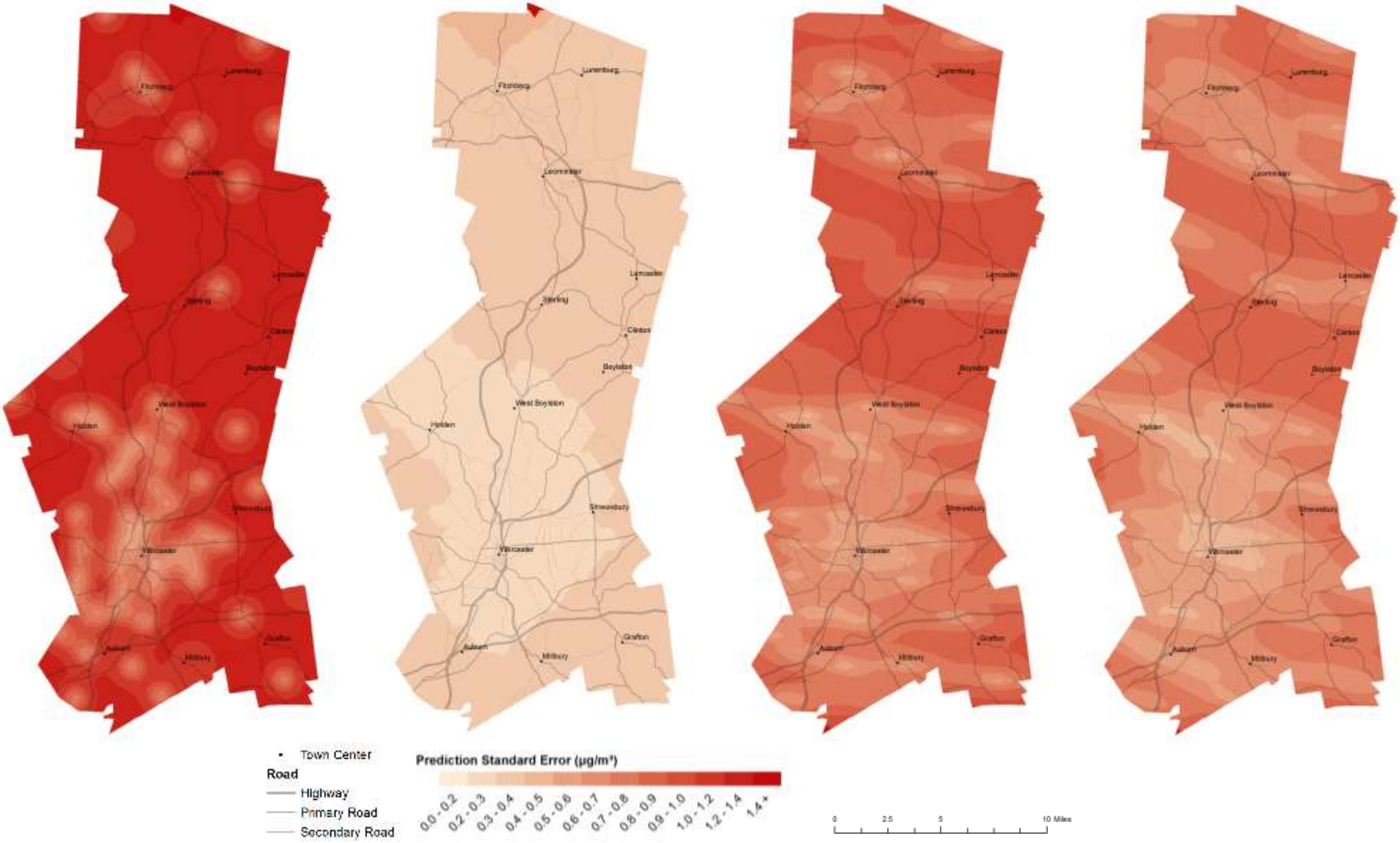


Figure 12 Spatial Distribution of Prediction Errors of $PM_{2.5}$ (Peak), PM_{10} (Peak), $PM_{2.5}$ (Non-Peak) and PM_{10} (Non-Peak) (from left to right)

Discussion

In this study, we successfully obtained measurements at 60 locations in our study area. In contrast to the state ambient air pollution monitoring network, which contains only one station in downtown Worcester near I-190 and Union Station, our monitoring network contains locations in various environments. This network is able to better characterize the heterogeneity within our study area. In addition, 60 is a statistically large sample site. Such a sample size enables us to examine data with more statistical approaches.

However, this study also suffered from several issues. First and the most problematic, the sensitivity of our sensor is insufficient to provide us with more accurate measurements. Although we did obtain non-zero measurements and were able to see the difference between low and high values, the measurement errors could be as high as 100% of measurements when monitored values are very low. Unfortunately, such low values of measurement appeared very frequently in our dataset. Since the general concentrations of PMs in the outdoor environment in our study area are pretty close to the measurement errors of the sensor, a sensor with higher sensitivity is needed to measure the concentrations with a higher accuracy.

Second, although we adopted the method that Kumar et al. (2009) used to allocate monitoring locations while excluding the effects of spatial dependency, there is still a room for optimization. When Kanaroglou et al. (2005), Su et al. (2007) and Kumer et al. (2009) employed the location-allocation approach to attain similar goals, their input previously monitored data were raster files. In this study, however, since no raster data were available to us. NATA 2005 only estimated concentrations based on census tracts and block groups. Restricted by data availability, we could only calculate demand values tract by tract. Because of this, all candidate locations (i.e. those 200*200m cells mentioned in the Spatial Optimization section) in the same census tract would always have the same demand value. When the monitoring sites were being allocated, they were actually evenly distributed within a census tract since the demands are equal everywhere. Although we successfully allocated monitoring stations to those tracts with higher demands, aside from reinforcing a minimum distance between any two locations, we did not control the spatial configuration within one single tract. This also resulted in an inappropriate allocation of monitoring locations at the edge of some large tracts to the northern and western boarders, which might not be the most representative locations in those tracts.

Third, by allocating 60 monitoring locations in an area of 877 square kilometers (i.e. one location per 14.6 square kilometers on average), we were not able to characterize small-scale spatial variations, such as the variation among a community or the variation among various distances from major highways. At the scale of this study, the measurements we obtained could be representative for the overall condition within a census tract, while not representative for any single block or residential community within that tract. When interpreting the results, it is very important to be aware of the scale we chose. These small-scale variations might also have brought exaggerations of high concentration values into our data. For example, several points we chose are very close to highway intersections. The measurements we got at these locations might be much higher than the general condition in nearby areas. Since these small-scale variations were ignored, the results of this project do not imply that the PM concentrations are not hazardous in our study area, even if none of our predicted values exceeds the standard.

To make improvements on air monitoring projects in the future, we suggested that there should be a more sensitive sensor available. Many gas sensors at lower prices were designed for indoor exposure. The outdoor concentration is often too low for them to measure accurately. In addition, there should be a better way to allocate monitoring stations. Ideally a raster-based previously monitored data can be employed to create the initial surface. Under some circumstances, there is point-based monitoring data at some locations although no raster data available. If the number of these point-based measurements can be regarded as a large sample, land-use regression (LUR) can be used to create the raster and thus generate the initial surface. If there is no way to obtain or create a raster based on previously monitored data, as the case of this study, a new approach to allocating resources based on a vector-based initial surface would be needed. From another perspective of location-allocation, regardless using raster or vector data, in this study the spatial optimization and temporal optimization were done separately. In fact, it is possible to integrate them as a spatiotemporal process. Instead of a “demand surface”, we can create a “demand cube” in a three dimensional space composed of two spatial axes and one temporal axis. Monitoring location can be allocated in this cube in the way that the most spatiotemporal variation within this cube can be characterized. Furthermore, the small-scale variation near any monitoring site should be explored. Knowing the small-scale variation would contribute to the understanding of the actual concentration that individuals are exposed to.

Conclusion

We clearly identified certain areas with higher concentrations, such as areas near major highways and intersections. Certain hot spots stood out more during peak hours. The kriging models we employed also indicated that anisotropic models are more suitable for non-peak hours. Although the highest concentrations in the results never exceeded the NAAQS, the actual human exposure might be quite different from our results since our network was not able to characterize small-scale and indoor exposure. Further studies on the effect of pollution on human health are needed to verify the applicability of the NAAQS.

References

1. Environmental Protection Agency (EPA). (2011). *National Ambient Air Quality Standards (NNAQS).* www.epa.gov/air/criteria.html [Retrieved on 4/27/2012]

2. Environmental Protection Agency (EPA). (2012). *Particulate Matter (PM).* www.epa.gov/pm/index.html [Retrieved on 4/27/2012]

3. Kanaroglou, P.S. and Jerrett, M. and Morrison, J. and Beckerman, B. and Arain, M.A. and Gilbert, N.L. and Brook, J.R. 2005. Establishing an air pollution monitoring network for intra-urban

population exposure assessment: a location-allocation approach. *Atmospheric Environment* 39(13): 2399-2409.

4. Kumar, N. 2009. An optimal spatial sampling design for intra-urban population exposure assessment. *Atmospheric Environment* 43(5): 1153-1155.

5. Su, J.G. and Larson, T. and Baribeau, A.M. and Brauer, M. and Rensing, M. and Buzzelli, M. and others. 2007. Spatial modeling for air pollution monitoring network design: example of residential woodsmoke. *Journal of the Air & Waste Management Association* 57(8): 893-900.

6. Thermo Electronics Corporation. (2005). *Features Sheet: DustScan Model 3020 Aerosol Monitor.*

Acknowledgements

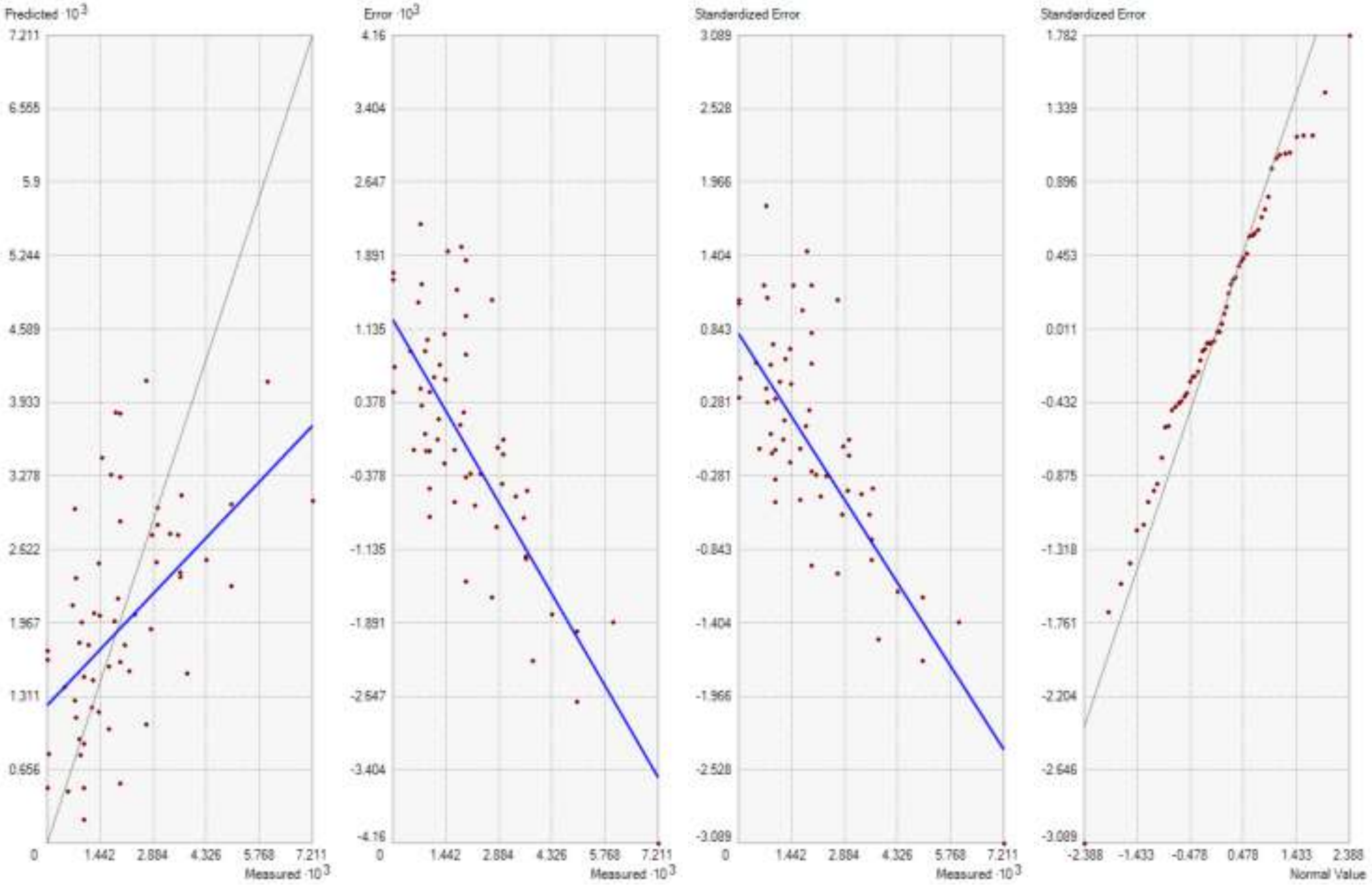
Prof. Abigail Mechtenberg and Prof. Yelena Ogneva-Himmelberger, for their guidance and support of this project; George Perkins Marsh Institute, for funding the project; John V. Carvalho III and Brandon Faillace at Apollo Safety Inc., for their technical supports.

Appendix 1: The Master Table of Monitoring Results

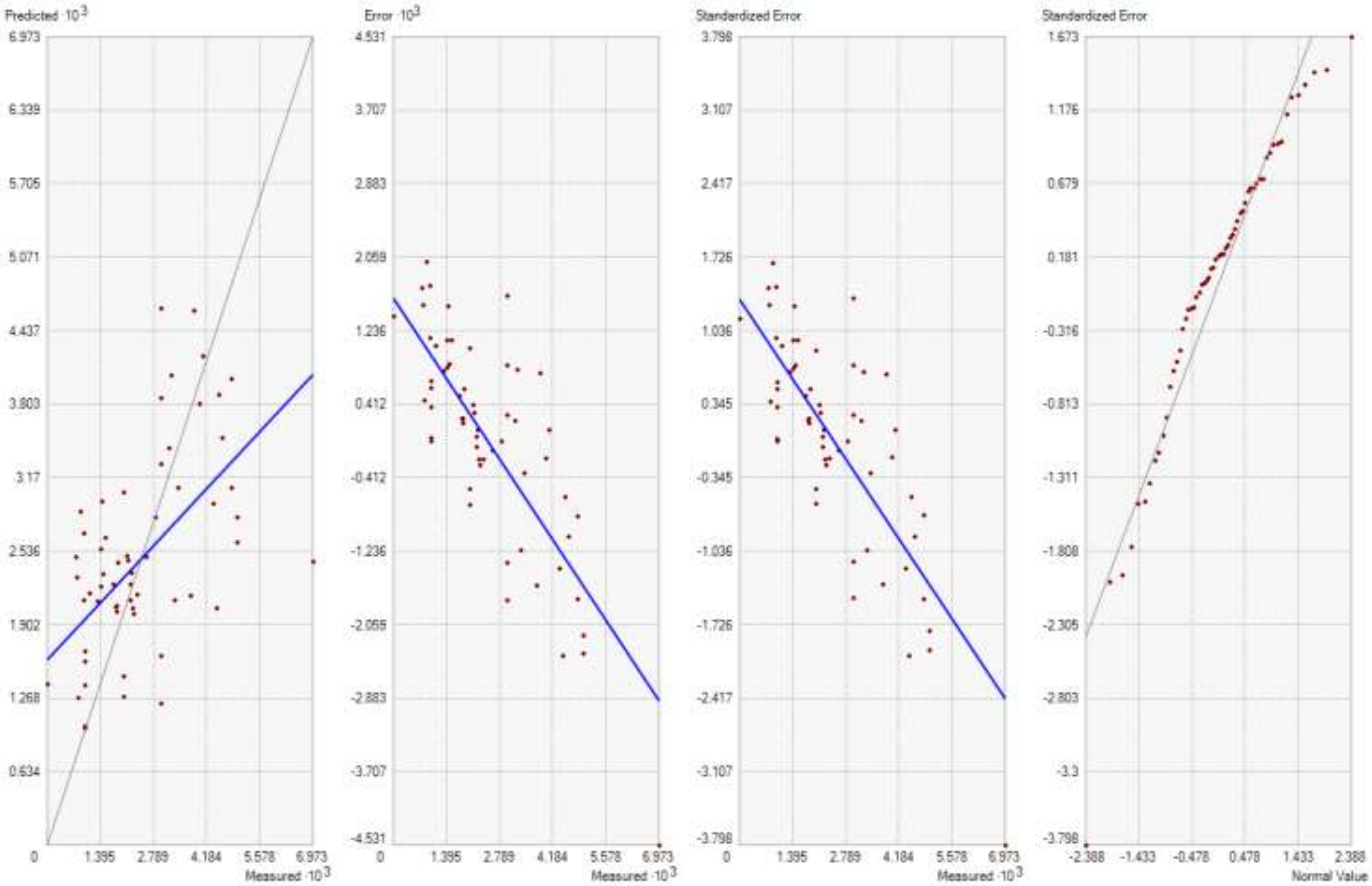
ID	Peak				Non-Peak			
	Longitude	Latitude	PM2.5	PM10	Longitude	Latitude	PM2.5	PM10
1	-71.792332	42.260089	0.002970	0.006973	-71.793775	42.260087	0.002000	0.009000
2	-71.820768	42.260602	0.007211	0.004377	-71.820830	42.260581	0.004000	0.004000
3	-71.787130	42.244589	0.003561	0.004857	-71.787130	42.244589	0.003571	0.004606
4	-71.801720	42.271745	0.001915	0.002594	-71.801808	42.271623	0.002000	0.001000
5	-71.694962	42.618956	0.001000	0.001000	-71.695101	42.619085	0.000806	0.000750
6	-71.677418	42.559862	0.000000	0.001000	-71.677891	42.559865	0.000880	0.001400
7	-71.875955	42.403013	0.002222	0.001741	-71.875955	42.403013	0.002326	0.001704
8	-71.919808	42.374853	0.001417	0.001820	-71.919808	42.374853	0.002154	0.002175
9	-71.828455	42.488312	0.000000	0.003000	-71.828315	42.488331	0.002800	0.003767
10	-71.729374	42.449893	0.002000	0.003000	-71.729712	42.450191	0.000931	0.001189
11	-71.792663	42.288951	0.001220	0.001854	-71.792196	42.288905	0.000111	0.000857
12	-71.769032	42.356024	0.000775	0.003343	-71.769024	42.356026	0.001307	0.001864
13	-71.710681	42.522457	0.001000	0.002000	-71.710825	42.522604	0.000815	0.000762
14	-71.652866	42.447065	-	-	-71.652866	42.447065	0.000714	0.001289
15	-71.834613	42.637914	0.001000	0.001000	-71.834552	42.637658	0.000970	0.000654
16	-71.808591	42.597745	0.000000	0.001000	-71.808384	42.597788	0.000577	0.000615
17	-71.859279	42.294451	0.002814	0.002097	-71.859312	42.294607	0.002000	0.002000
18	-71.815891	42.244540	0.006000	0.003000	-71.815891	42.244540	0.003000	0.001000
19	-71.712374	42.352837	0.002688	0.001806	-71.712516	42.353622	0.000909	0.001220
20	-71.746394	42.318126	0.000486	0.001821	-71.746545	42.318140	0.001000	0.000947
21	-71.668781	42.182869	0.005000	0.005000	-71.668781	42.182869	0.002000	0.002000
22	-71.832528	42.574127	0.000059	0.000818	-71.832735	42.574165	0.000030	0.001462
23	-71.814283	42.294043	0.000757	0.000781	-71.814233	42.294173	0.001750	0.001944
24	-71.806023	42.284299	0.000909	0.001417	-71.806886	42.284246	0.002622	0.003083
25	-71.793572	42.582628	0.000568	0.000000	-71.793670	42.582904	0.000000	0.000833
26	-71.844320	42.362524	0.001290	0.002222	-71.842861	42.362942	0.002795	0.003585
27	-71.863699	42.361478	0.001677	0.002366	-71.862924	42.361147	0.003328	0.003589
28	-71.837768	42.353409	0.002125	0.002205	-71.836628	42.351151	0.003043	0.002886
29	-71.812172	42.338824	0.003000	0.003769	-71.812159	42.338379	0.002704	0.002738
30	-71.818713	42.326644	0.002397	0.002277	-71.818713	42.326644	0.001800	0.002257
31	-71.832119	42.317781	0.001840	0.001320	-71.832119	42.317781	0.001231	0.001750
32	-71.779436	42.262034	0.000704	0.001421	-71.777665	42.264640	0.003000	0.002000
33	-71.752945	42.555934	0.002000	0.002000	-71.752768	42.555821	0.001375	0.002694
34	-71.649795	42.243919	0.003000	0.002000	-71.649795	42.243919	0.002000	0.002000
35	-71.706789	42.225258	0.005000	0.005000	-71.706789	42.225258	0.004000	0.002000
36	-71.808323	42.191041	0.003618	0.004093	-71.808298	42.191339	0.002000	0.002000
37	-71.752025	42.192998	0.002000	0.003000	-71.752025	42.192998	0.002000	0.003000
38	-71.781149	42.171039	0.002000	0.003000	-71.781149	42.171039	0.004000	0.002000
39	-71.862137	42.240194	0.003343	0.003194	-71.862031	42.240056	0.003000	0.002000

40	-71.753502	42.263490	0.000778	0.001520	-71.753647	42.263649	0.002000	0.003000
41	-71.750086	42.250210	0.000949	0.001449	-71.750069	42.250459	0.004000	0.004000
42	-71.786564	42.370241	0.003794	0.000767	-71.785253	42.368643	0.004140	0.003308
43	-71.800134	42.347170	0.003610	0.004447	-71.800134	42.347170	0.003127	0.003309
44	-71.780786	42.334473	0.001426	0.002182	-71.780911	42.334565	0.000933	0.001975
45	-71.788337	42.314604	0.001250	0.000957	-71.787619	42.314802	0.001083	0.000800
46	-71.855256	42.272736	0.002841	0.002833	-71.855098	42.272934	0.002000	0.002000
47	-71.857580	42.253093	0.001860	0.003450	-71.857809	42.253197	0.002000	0.002000
48	-71.818031	42.275810	0.000767	0.000875	-71.818031	42.275810	0.001000	0.002000
49	-71.843320	42.226249	0.003639	0.004520	-71.843118	42.225999	0.001000	0.003000
50	-71.874685	42.213589	0.001750	0.003256	-71.874685	42.213589	0.003600	0.003957
51	-71.850124	42.199926	0.004333	0.004857	-71.849988	42.199878	0.002000	0.003000
52	-71.821075	42.230195	0.002700	0.004595	-71.820956	42.230133	0.002000	0.002000
53	-71.693581	42.313251	0.000886	0.002250	-71.693723	42.313340	0.000927	0.000960
54	-71.769557	42.539739	0.001000	0.001000	-71.769758	42.540266	0.000000	0.001182
55	-71.773451	42.302543	0.001415	0.001111	-71.772800	42.301943	0.000935	0.001171
56	-71.743883	42.282457	0.001667	0.000957	-71.753587	42.282469	0.000947	0.000862
57	-71.764872	42.271028	0.000880	0.002143	-71.764621	42.271299	0.001000	0.002000
58	-71.860373	42.168772	0.001500	0.003848	-71.860412	42.168878	0.002000	0.003000
59	-71.724192	42.276575	0.001133	0.001480	-71.724090	42.276672	0.000756	0.000333
60	-71.795902	42.221784	0.002000	0.004000	-71.795902	42.221784	0.002000	0.002000

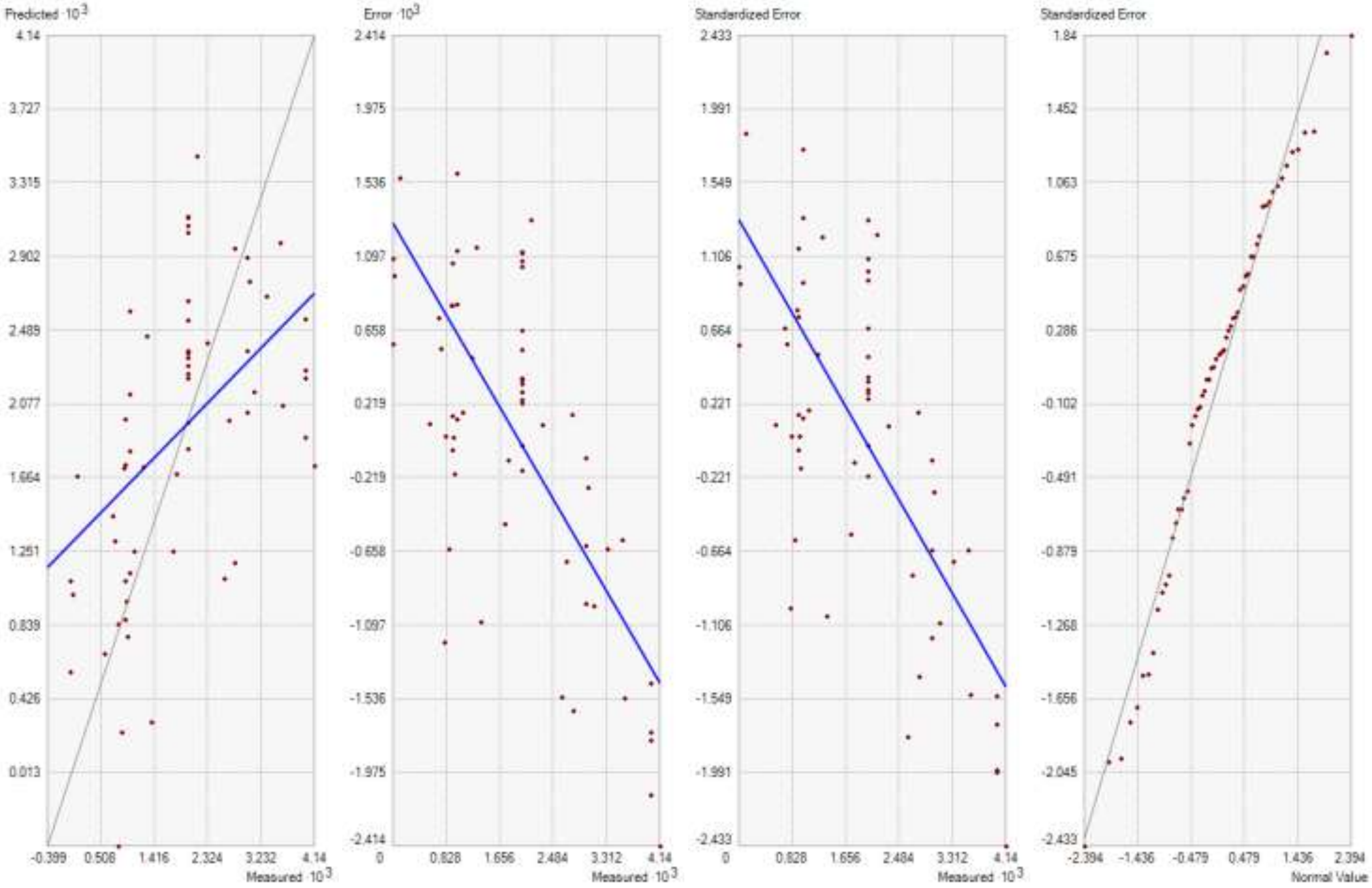
Appendix 2: Results of Cross-Validation (charts)



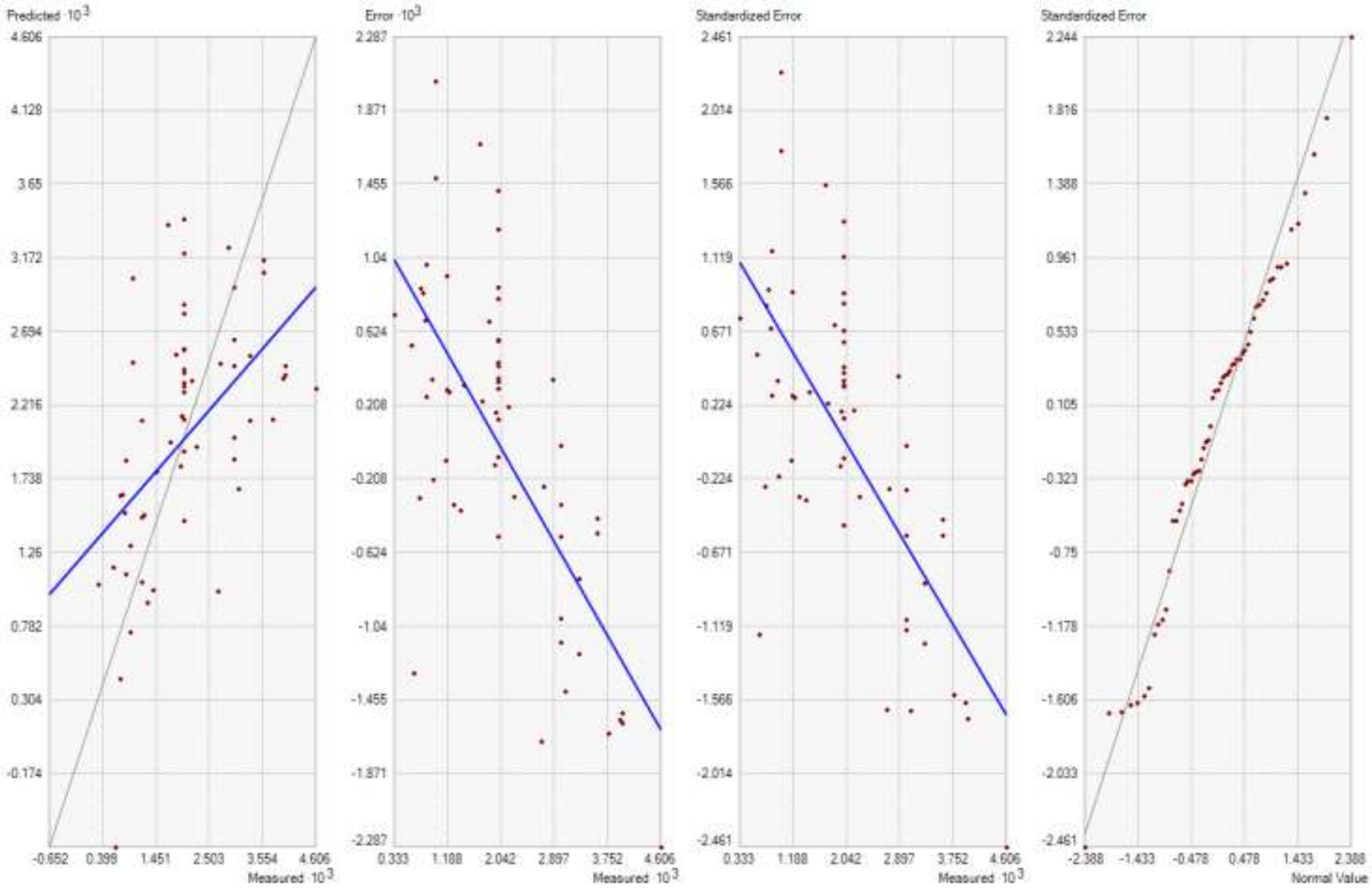
$PM_{2.5}$, Peak


Predicted ·10^3
6.973
6.339
5.705
5.071
4.437
3.803
3.17
2.536
1.902
1.268
0.634
0
1.395
2.789
4.184
5.578
6.973
Measured ·10^3
Error ·10^3
4.531
3.707
2.883
2.059
1.236
0.412
-0.412
-1.236
-2.059
-2.883
-3.707
-4.531
0
1.395
2.789
4.184
5.578
6.973
Measured ·10^3
Standardized Error
3.798
3.107
2.417
1.726
1.036
0.345
-0.345
-1.036
-1.726
-2.417
-3.107
-3.798
0
1.395
2.789
4.184
5.578
6.973
Measured ·10^3
Standardized Error
1.673
1.176
0.679
0.181
-0.316
-0.813
-1.311
-1.808
-2.305
-2.803
-3.3
-3.798
-2.388
-1.433
-0.478
0.478
1.433
2.388
Normal Value

PM_{10}, Peak



$PM_{2.5}$, Non-Peak



PM_{10}, Non-Peak