Development of Satellite-Based Emission Inventories and Indoor Exposure Prediction Models for PM2.5

Development of Satellite-based Emission Inventories and Indoor Exposure Prediction Models for $PM_{2.5}$

Chia-Hsi Tang

A Dissertation Submitted to the Faculty of

The Harvard T.H. Chan School of Public Health

in Partial Fulfillment of the Requirements

for the Degree of *Doctor of Science*

in the Department of Environmental Health

Harvard University

Boston, Massachusetts.

Novemember, 2016

Dissertation Advisor: Dr. *Petros Koutrakis* *Chia-Hsi Tang*

Development of Satellite-based Emission Inventories and Indoor Exposure Prediction Models for $PM_{2.5}$

ABSTRACT

Epidemiological studies have documented significant relationships between outdoor particle exposure and adverse health effects whilst indoor residential $PM_{2.5}$ exposure is found to be the most influential on total $PM_{2.5}$ exposure as people spend more than 80% of their time indoors. Accurate exposure assessments for ambient and indoor environments are therefore of equal importance. In this dissertation, we aim to develop methodologies that enhance our ability to quantify fine particulate matter ($PM_{2.5}$) exposure in both macro and micro perspective.

With advanced remote sensing technologies becoming more prevalent and less expensive, there is great potential in employing satellite date to analyze and illustrate ambient air quality in real-time over large geographic areas. In Chapter 1, we introduced a top down approach to construct $PM_{2.5}$ emission inventory through the integration of mass balance and satellite retrieved daily Aerosol Optical Depth (AOD) at 1km x 1km resolution. The satellite-based inventory provides spatially- and temporally-resolved emission estimates as opposed to the conventional source-oriented inventory that has time lag issues with limited spatial variability to the extent of its data source (usually ground monitoring network). Subsequently in Chapter 2, we quantified the temporal and spatial trends of $PM_{2.5}$ emission in the North East U.S. using the satellite-based emission inventory. Satellite-based emission trends are in agreement with that of the source-oriented inventories released by the US EPA, showing major reductions achieved in urban areas

as well as along important traffic corridors. The technique of this part of the study can be applied to nation-wide or global remote sensing data for estimating $PM_{2.5}$ emissions and hence improving the quantification of fine particles effects on climate, air quality and human health.

While big data may be the game changer for resolving ambient air quality problems, we still face the challenge of data scarcity in microenvironments. In Chapter 3, prediction models that utilize few available samples to assess indoor $PM_{2.5}$ exposure were developed. We estimated infiltration rate of ambient particles penetrating indoors using sulfur as the tracer element at 95 residences in the Greater Boston Area. Mixed effects model was employed in order to predict infiltration for individual residences. We then estimate indoor levels of $PM_{2.5}$ and its black carbon (BC) content using outdoor measurements and the estimated infiltration factor. We cross validated the aforementioned models to evaluate their predictive power specifically at dates without indoor information. Cross validation results of the infiltration model (R^2=0.89) indicates that mixed effects captured infiltration rates for individual households adequately. We also found strong predictability when sulfur infiltration surrogate and outdoor measurements of $PM_{2.5}$ and BC were used in predicting indoor exposure levels (R^2= 0.79 [$PM_{2.5}$], 0.76 [BC]).

Altogether, the methodologies introduced in this dissertation may serve as frameworks to (1) quantify and illustrate ambient emission of $PM_{2.5}$ or other pollutants in a macro perspective and (2) determine the relationships between outdoor and indoor air quality and to predict indoor air pollution which are critical information for developing solutions of micro-level air quality problems.

TABLE OF CONTENTS

FIGURES AND CAPTIONS

TABLES AND CAPTIONS

ACKONWLEDGEMENTS

This dissertation has been the most challenging test of my own commitment and perseverance and I have been so fortunate to receive help, support and encouragement from so many people along the way.

First and foremost, I would like to thank my advisor, Dr. Petros Koutrakis, for his mentorship, patience and encouragement throughout the development of this work. I would also like to express my deepest gratitude to the rest of my research committee members for providing invaluable inputs and guidance. I would like to thank Dr. Brent Coull for sharing his expertise in statistical methods and Dr. Joel Schwartz for his insightful analytical suggestions. Furthermore, I would like to acknowledge Dr. Rose Goldman, Dr. Alexandra Chudnovsky, and Dr. Joy Lawrence for discussing research problems and sharing their knowledge with me. In your own ways, you each have added to this project and helped shaped me as a student as well as a researcher.

Thank you to my graduate school colleagues at Harvard. I am especially thankful to Wan Chen Lee, Alice Smyth, Qian Di, Souzana Achilleos, Shahir Masri, Annelise Mesler, and Cheng Peng.

Last but not least, I would like to thank my parents, Claudia and LiHui, my little sister Alice, and my family and friends at Taiwan for their unconditional faith and support over the years in which this project came in to fruition.

Finally, thank you to my loving husband, Wayne who always stands by me.

Jessie Chia-Hsi Tang

Boston, Massachusetts

INTRODUCTION

Air pollution has occupied the public attention in the U.S throughout the 20th century. This awareness drives the U.S. to reduce its air pollution to only a fraction of past levels. On the flip side, recent research reveals that particulate matter with aerodynamic diameters less than 2.5 μm, or $PM_{2.5}$, can still pose adverse effects(Pope et al. 2002, Dominici 2003, Zanobetti and Schwartz 2006, Peng et al. 2009, Levy et al. 2012) on human health even concentrations of $PM_{2.5}$ has reached historical low(Shi et al. 2016). With a fairly clean atmosphere in the U.S., further improving air quality becomes even more challenging as unnecessary collateral damage could be easily made due to obscured or sometimes exaggerated exposure levels of $PM_{2.5}$. Better understanding the origin, distribution and infiltration to microenvironments of $PM_{2.5}$ is therefore indispensable prior to implementing particle control strategies.

Sources of $PM_{2.5}$ are commonly categorized into primary emissions and secondary formations. Primary $PM_{2.5}$ is directly released into the atmosphere from natural and anthropogenic activities while secondary $PM_{2.5}$ is formed within the atmosphere from precursor gases such as Sulfur dioxide (SO_2), ammonia (NH_3), nitrogen oxides (NO_x), and volatile organic compounds (VOCs) via gas-phase photochemical reactions or through liquid phase reactions in clouds and fog droplets. Currently, the U.S. Environmental Protection Agency (EPA) monitor primary sources by estimating unit emission rate (or emission factor) and frequency of activity for each known sources (EPA 2005a, 2008, 2011). Modeling emission factors often rely on ground-level networks for data collection and thus most emission estimates have time lags with limited temporal and spatial information. While a spectrum of primary $PM_{2.5}$ sources are documented in the EPA's emission inventory, this source-oriented method is not directly

applicable to assess secondary particle formations which constitute a substantial portion of the total $PM_{2.5}$ in the atmosphere. At present, sources of the precursor gases that contribute to or aid secondary particles formation are voluntarily reported to the EPA. The formation processes for inorganic particles is fairly well understood so far but much is to be learned regarding the organic portion of the process. In essence, quantifying secondary formation of $PM_{2.5}$ based on precursor gases is still difficult and the high uncertainties of these estimation undermines our efforts to understand trends of $PM_{2.5}$(Fine, Sioutas, and Solomon 2008). Ultimately, this source-oriented approach becomes very expensive and unrealistic as new sources are rapidly emerging while research on known sources has yet to be completed. In order to enhance the emission control efficiency, cost effective ways to quantify particle emissions over broad areas with enough temporal and spatial details are needed.

An alternative source of data that could help us address the aforementioned issue has emerged along with the major advancement in detecting air pollution from outer space (Streets et al. 2013). Satellite data has the potential for quantifying emissions near real time with high spatial resolution global coverage. Furthermore, it can be applied to measure area sources (e.g. biogenic sources, biomass burning, etc.) which are difficult to assess using traditional source-oriented methods and uncertain emissions from neighboring countries (e.g. Mexican and Canadian emissions) where ground network data may be unavailable. At present, inverse modeling is the most common technique to constraint emissions using satellite data. However, the forefront limitation of this method is that it is computationally unattractive. Consequently, the emission inventories generated from inverse models are still restricted to coarse spatial resolution with little temporal information. For example, Dubovik et al., predicted global aerosol emissions at 2 º x 2.5º using two weeks of aerosol data simulated by the Goddard Chemistry

Aerosol Radiation and Transport (GOCART) model (Dubovik et al. 2008) and more recently, Huneeus et al., estimates monthly averaged aerosol emission fluxes at 1º x 1º in 2002 using Aerosol Optical Depth (AOD) from the Moderate Resolution Imaging Spectroradiometer (MODIS)(Huneeus, Chevallier, and Boucher 2012). In the aforementioned studies, actual measurements of emissions with comparable spatial and temporal scale as the satellite-derived emissions are still limited and therefore uncertainties of these models are often indirectly determined. Furthermore, satellite measurements (e.g. AOD) represent the vertical loading of aerosols while ground-based observations reflect particles near ground and this discrepancy could pose another challenge to evaluate the uncertainties of satellite-derived emissions.

In Chapter 1 we introduced a new emission inventory method, **P**articles **E**mission **I**nventory using **R**emote **S**ensing (PEIRS), predicting total $PM_{2.5}$ emissions through integration of mass balance and high resolution 1km x 1km AOD data. AOD is one of the most robust measures for fine aerosols, also referred to as aerosol optical thickness (AOT). The relationship between AOD and $PM_{2.5}$ has been well studied and calibrations were performed using various statistical techniques including simple linear regression(Tai, Mickley, and Jacob 2010), neural networks(Hooyberghs et al. 2005) and hybrid spatial and temporal models(Kloog et al. 2011) to predict ambient $PM_{2.5}$ concentrations with high accuracies. Building upon the sophisticated calibration techniques, we first calibrated AOD into ground-level $PM_{2.5}$ concentrations. This calibration allows us to predict near-ground particle emissions using a mass balance equation and further evaluate emission estimates with monitor observed or estimated emission from the National Emission Inventory released by the U.S. EPA. The trajectory of air mass movement was captured using wind direction obtained from the North American Regional Reanalysis database and subsequently included in the emission model as surrogate of upwind particle

contributor. Local emissions including primary and secondary particles within each 1km x 1km grid cell are estimated across the North East U.S. over the period of 2002 to 2013.

In Chapter 2, as an implication of the inventory method proposed in Chapter 1, we applied the PEIRS approach to build and analyze temporal and spatial trends of triennial averaged emission inventories across Northeast U.S. during 2002 to 2013. Temporal emission trends were estimated using a linear quantile regression model in order to examine the variation of trends at different quantiles of the emission distribution. Source-oriented trends were quantified using land use regression models. To further determine the seasonality of emission trends, the aforementioned trend analyses were also conducted separately for warm and cold season specific satellite-based emission inventories.

While ample amount of resources and great advancement of modeling techniques have become readily available for ambient air quality research, just as we demonstrated in the first two chapters, indoor air quality studies are still restricted to very few field samples and questionnaire data as surrogates of indoor sources or penetrations of ambient air pollutants. As people spend the majority of their time indoor, it is equally or even more critical to adequately assess the exposure levels in the micro-environment as of the macro-environment. More importantly, in view of the clear association between health effects and ambient $PM_{2.5}$ concentrations demonstrated by many studies worldwide in the past decade, the correlation between the components of micro-environmental exposure due to outdoor particles have also drawn the interests of epidemiologists. Many studies have focused on building prediction models for indoor particles using various data such as household characteristics, land use parameters, tracer elements. However, reliable predictions from previous studies were rather limited due to the

scarcity of indoor samples, variation of infiltration rate by the size distribution and composition of fine particles, and mediation of infiltration rates by human behaviors.

To address these issue, Chapter 3 presented an analysis to model infiltration factor using sulfur as its tracer for each individual households in the Greater Boston Area. Cross validation is used to evaluate the stability and predictability of the model and the appropriateness of the use of I/O S ratio as a surrogate of infiltration. More importantly, the cross validation was designed to examine the predictability when indoor data are not available which is crucial for epidemiological studies. We then incorporated the estimated I/O S ratio in predicting indoor $PM_{2.5}$ and BC concentrations as a secondary evaluation. The relationship between the predicted sulfur ratio and human activities that modifies the infiltration rates such as use of AC, purifier, and open windows were also examined in this study.

CHAPTER 1

DEVELOPING PARTICLE EMISSION INVENTORIES USING REMOTE SENSING (PEIRS)

Chia Hsi Tang [a]

Brent Coull [b]

Joel Schwartz [a]

Alexei I. Lyapustin [c]

Qian Di [a]

Petros Koutrakis [a]

[a] Department of Environmental Health, Harvard T.H Chan School of Public Health

[b] Department of Biostatistics, Harvard T.H Chan School of Public Health

[c] National Aeronautics and Space Administration Goddard Space Flight Center

ABSTRACT

Information regarding the magnitude and distribution of $PM_{2.5}$ emissions is crucial in establishing effective PM regulations and assessing the associated risk to human health and the ecosystem. At present, emission data is obtained from measured or estimated emission factors of various source types. Collecting such information for every known source is costly and time consuming. For this reason, emission inventories are reported periodically and unknown or smaller sources are often omitted or aggregated at large spatial scale. To address these limitations, we have developed and evaluated a novel method that uses remote sensing data to construct spatially-resolved emission inventories for $PM_{2.5}$. This approach enables us to account for all sources within a fixed area, which renders source classification unnecessary. We applied this method to predict emissions in the northeast United States during the period of 2002-2013 using high- resolution 1 km x 1km Aerosol Optical Depth (AOD). Emission estimates moderately agreed with the EPA National Emission Inventory (R^2=0.66~0.71, CV = 17.7~20%). Predicted emissions are found to correlate with land use parameters suggesting that our method can capture emissions from land use-related sources. In addition, we distinguished small-scale intra-urban variation in emissions reflecting distribution of metropolitan sources. In essence, this study demonstrates the great potential of remote sensing data to predict particle source emissions cost-effectively.

INTRODUCTION

Fine particulate matter ($PM_{2.5}$) is a major public health burden associated with a range of adverse health effects (Schwartz and Dockery 1992, Schwartz 2001, Pope et al. 2002, Schwartz, Laden, and Zanobetti 2002, Ren, Williams, and Tong 2006, Zanobetti and Schwartz 2006, Turner et al.

2011, Lepeule et al. 2012). As such, identifying particle sources and quantifying their emissions is of paramount importance to the development of air quality standards and the enforcement of emission reduction policies. The US Environmental Protection Agency (EPA) is responsible for developing a nationwide particle emission inventory, the National Emission Inventory (NEI), which is the most comprehensive database for criteria air pollutants (CAP) and hazardous air pollutants (HAP) emissions. The NEI is updated every three years using data collected from state, local and tribal air agencies as well as EPA emission trading programs. Collecting information on area and mobile emission sources can be challenging especially when there are ample amount of small sources widely dispersed over a large area. For instance, smaller stationary sources such as wood furnaces and stoves are not defined as major point sources in the NEI and thus are not rigorously regulated. However, these smaller sources may represent a large fraction of the total emissions as bigger industrial sources decrease in intensity due to strict regulations and improved technology. For PM, the current NEI aggregates mobile sources and nonindustrial sources at county level. Information for such broad geographic areas, in combination with measurement error and modeling uncertainties may limit effectiveness to implement emission regulations. Coarse spatial resolution also limits the utility of the NEI data to assess human health risks or develop air pollution models (EPA 2008).

Satellite data is increasingly important for air pollution exposure assessment because of scarce and ad hoc spatial-temporal coverage of the federal monitoring network. Overtime, satellite data has been incorporated in a variety of air quality applications, including tracking long-term pollution transport, identifying exceptional events such as wild fires, fireworks and dust storms, and estimating ground level pollution concentrations (Duncan et al. 2014). Satellite-based

sensors have provided information on important air pollutants, such as nitrogen dioxide (NO_2), sulfur dioxide (SO_2), volatile organic compounds (VOCs) and fine particulate matter ($PM_{2.5}$) (Martin 2008). Specifically, a growing number of studies have successfully employed AOD data to characterize properties and patterns of $PM_{2.5}$ (Gupta and Christopher 2009, Liu, Paciorek, and Koutrakis 2009, Kloog et al. 2011, Lee et al. 2012, Hu et al. 2014, Kloog et al. 2014).

AOD is a dimensionless measure of the attenuation of light due to the presence of aerosols that prevent light transmission via absorption or scattering. This fundamental property of AOD, with proper correction for absorption and scattering of gases in the atmosphere, makes it a suitable surrogate of the aerosol loading in the atmosphere (Hoff and Christopher 2009). Nevertheless, vigilant calibration of the AOD data is required due to several physical differences among AOD and ground-level $PM_{2.5}$. A critical concern is that AOD represents the total amount of aerosols in the entire atmospheric column while measured $PM_{2.5}$ only reflects particles at ground level. The relationship between AOD and $PM_{2.5}$ is therefore highly dependent on the vertical distribution of aerosols. Moreover, like all satellite measurements, AOD readings are snapshots of the aerosol distribution at that exact moment, while filter-based $PM_{2.5}$ samples are collected over a 24-hour period. This temporal mismatch also influences the association between AOD and ground level $PM_{2.5}$. To address these concerns, multiple techniques including neural network, generalized additive models and hybrid models have been used to generate AOD-derived $PM_{2.5}$ concentrations (Gupta and Christopher 2009, Lee et al. 2012, Hu et al. 2014, Kloog et al. 2014). Recent advancements in various predictive statistical models have enabled scientists to assess daily $PM_{2.5}$ exposures with continuous spatial coverage, which are crucial for legislation

development and health effects studies. Nonetheless, satellite data has yet to become a prime resource in predicting particle emissions.

In this chapter, we introduce a method using satellite AOD data to predict emission inventories for $PM_{2.5}$. As a demonstration of the proposed method, we constructed spatially resolved $PM_{2.5}$ emission inventories in the Northeast U.S. using 1 km x 1 km daily AOD retrievals during the period of 2002-2013. We derive the emission model based on the concept of one compartment model, which is often used to estimate emissions of indoor sources (Spengler, Samet, and McCarthy 2001). Our approach has the potential to generate comprehensive emission inventories cost- effectively as oppose to the existing ones.

METHODS

Data and Materials

$PM_{2.5}$ ground monitoring concentration data. The study domain is the Northeast U.S. that includes the states of Connecticut (CT), Massachusetts (MA), Maine (ME), New Hampshire (NH), New York (NY), Rhode Island (RI) and Vermont (VT). We obtained daily averaged $PM_{2.5}$ measurements, mostly from integrated filter samples, measured at 124 monitoring sites among the U.S. Environmental Protection Agency (EPA) Compliance Network, Air Quality System(AQS) and the Interagency Monitoring of Protected Visual Environments (IMPROVE) database during the period of 2002 to 2013. Daily averaged monitoring $PM_{2.5}$ is used to calibrate AOD data into AOD-derived-$PM_{2.5}$ concentrations (Figure 1).

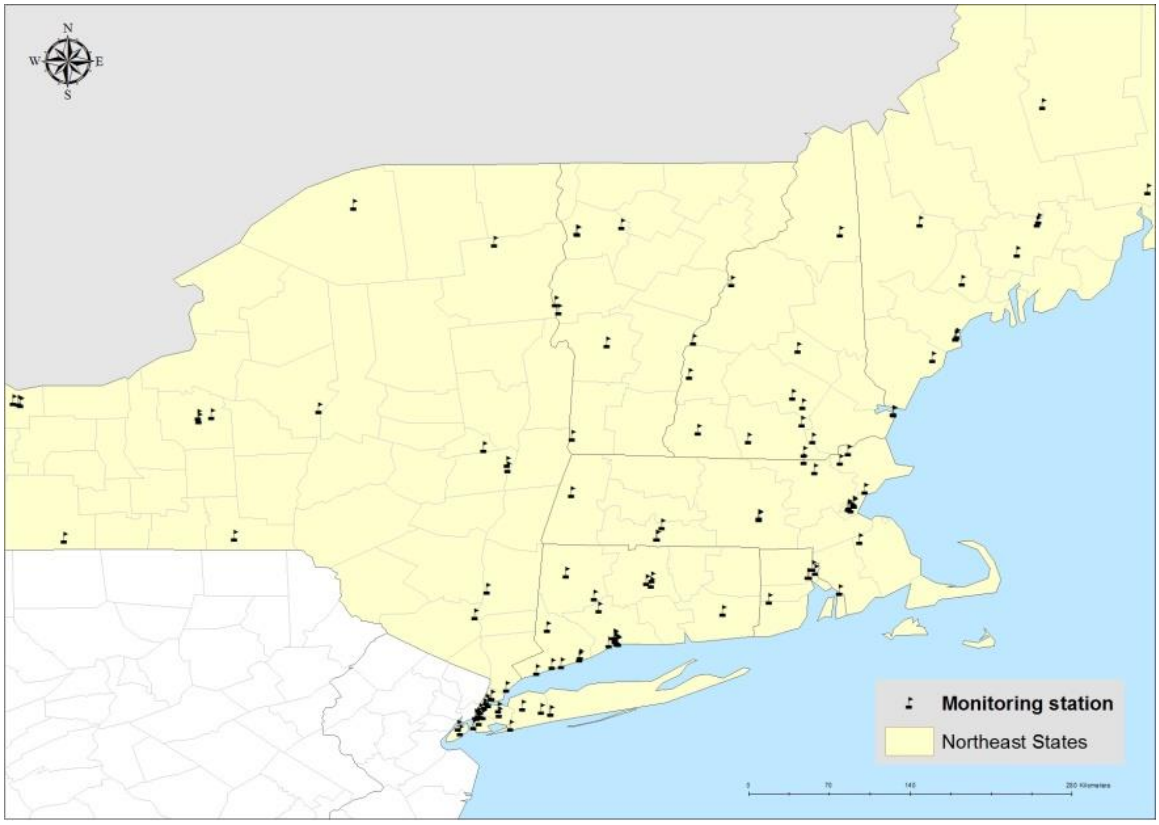


Figure 1. Study area and EPA monitoring network.

Satellite Aerosol Optical Depth. The Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on board the Earth Observing System (EOS) Aqua satellite provides numerous aerosol measures including AOD product reflecting fine particle loading. In 2011, an advanced algorithm ,the Multi-angle Implementation of Atmospheric Correction (MAIAC), was presented (Lyapustin et al. 2011) providing a set of AOD product with much finer resolution (1 km x 1km) compared to the standard MYD04 product at 10km x 10km resolution. A study evaluating the MAIAC AOD product concluded that it is more robust under partly cloudy conditions with fewer non retrieval days and pixels than the standard product. For this reason, MAIAC AOD also pertains improved ability in capturing spatial patterns of particle loading (Chudnovsky et al. 2013). Furthermore, calibration of the MAIAC AOD product was shown to be successful for the

New England Area (Kloog et al. 2014). Here we took advantage of the MAIAC AOD Aqua product to predict spatially resolved (1 km x1 km) emission inventories.

Meteorological data. Meteorological parameters such as vertical (VWND) and horizontal wind speed (UWND), relative humidity (RH), boundary layer height (PBL), snow coverage, precipitation (PRCP) and temperature (TEMP) for the period from 2002 to 2013 were obtained from the National Oceanic and Atmospheric Administration (NOAA) North America Regional Reanalysis (NARR) database. This dataset assimilates multiple sources of measurements and optimizes estimation of meteorological fields as described by (Kalnay et al. 1996). The reanalysis dataset provides surface level meteorological variables at a spatial resolution of 32 km x 32 km and temporal resolution of 1 day. The PBL is used to estimate the columnar volume of air on a given day, and other meteorological variables are used to calibrate the AOD/$PM_{2.5}$ relationship. All NARR daily meteorological variables were linearly interpolated to 1 km×1 km resolution using a Matlab package, scatteredInterpolant. More detail on the package algorithm can be found in the MathWorks online documentation (http://www.mathworks.com/help/matlab/ref/scatteredinterpolant-class.html).

In building the emission prediction model, wind speed is used to estimate the residence time of air mass inside a volume of 1km x 1km x PBL km, while wind direction is the key factor to identify the location of the upwind adjacent grid cell. After interpolating both wind field parameters into 1km x 1km resolution as described above, we calculated wind speed (WS) as the square root of sum of u^2 and v^2 and wind direction (WD) as the vector sum of UWND and VWND. We assumed that the daily wind direction and wind speed was constant within at all altitudes at the surface level defined by the NOAA land surface model (Mesinger et al. 2006).

Land use variables. Traffic-related variables including major roads (A1-A3) density and other roads (A4) density were gathered from the StreetMap USA database. Roads were classified using the Feature Class Code (A1-A4) from the US Census Bureau,
Topologically Integrated Geographic Encoding and Referencing (TIGER) system. Annual averaged traffic count for major roads was obtained from the Highway Performance Monitoring System (HMPS) database. We used a Kernel density algorithm to calculate grid (1km x 1km) averages of the major traffic density parameters using ESRI® ArcMap software. Land cover data for the entire Northeast U.S. were obtained from the 2011 collection of the National Land Cover Database (NLCD). With more detailed classification of land use, land cover data for Massachusetts were also gathered from the Massachusetts Department of Environmental Protection (Mass DEP). Elevation raster data was obtained from the ESRI® database. All land use parameters except elevation were used only for emission validation and are excluded from both AOD/$PM_{2.5}$ calibration and the emission model.

NEI emission data. Point, nonpoint, and mobile emissions were obtained from the 2008 and 2011 U.S. EPA emission inventories. According to EPA's definition, point emission contains larger industrial sources while nonpoint refers to smaller stationary sources that are inventoried at county level. Such sources (or sectors) include residential wood combustion, field burning, and consumer solvent use etc. Mobile sources pertains mostly transportation emissions such as road traffic, locomotive, aircraft and commercial marine vessels. Detailed sector descriptions can be found in the 2008 NEI report. For this study, we aggregated the nonpoint and mobile EPA NEI emission data to evaluate the predicted particle emissions. The locations of NEI point emissions were intersected with the corresponding 1km x 1km AOD grid cell. A variable indicating the

presence of NEI point sources was created and was included in the land use regression validation models.

Statistical Analysis

The PEIRS approach encompasses three analytical stages: First, we calibrated AOD to obtain AOD-derived $PM_{2.5}$ predictions for all 1km x 1km grids in the study domain. Second, we fit an emission model for each grid cell to predict emissions released within the grid. Third, we deployed land use regression models and compared county-level predicted and NEI-reported emissions to evaluate our method.

AOD/$PM_{2.5}$ Calibration. Let us consider an air mass enclosed in a box with a base of 1km x 1km (Pixel Area) and a height equal to that of the boundary layer (PBL). We can assume that AOD is proportional to the particle mass inside this box based on the previously established relationship between AOD and fine particles(Hoff and Christopher 2009). Subsequently, we can express the particle mass as the product of the average particle concentration ($C_{PM2.5}$) and the box volume:

$$AOD \propto Mass_{PM2.5} = C_{PM2.5} \times Volume = C_{PM2.5} \times PBL \times Pixel\ Area \quad (1)$$

Since the pixel area remains constant (1 km^2) and most particles are usually below the boundary layer, we can translate the relationship in eq 1. into the basis of a calibration model as follows:

$$C_{PM2.5} = \beta_0 + \beta_1 \frac{AOD}{PBL} \quad (2)$$

Where β_0 is the intercept and β_1 is the slope or conversion factor of the simple calibration model. Because PBL, relative humidity and particle composition vary daily, we performed daily calibrations using a mixed-effects regression model. Including random slopes (α_{1i} , i: day) and intercepts (α_{0i}) by day enables the model to capture day-to-day variability in the AOD/$PM_{2.5}$ relationship (Lee et al. 2012, Kloog et al. 2014). Relative humidity (RH), wind speed (WS), snow coverage, precipitation, temperature (Temp) and elevation are also included as covariates to adjust for site-specific characteristics. We also added interaction terms between AOD and all meteorological parameters to further control their effects on particle extinction efficiency. However, we did not include any source related land use here because AOD-derived-$PM_{2.5}$ concentrations are used to estimate emissions. For this reason, source related parameters should be excluded from the calibration process. The final calibration model is as follows:

$$C_{PM2.5} = \alpha_{0i} + \beta_0 + (\alpha_{1i} + \beta_1)\frac{AOD}{PBL} + \beta_2 WS + \beta_3 RH + \beta_4 Temp + \beta_5 SnowCoverage + \beta_6 Elevation + \beta_7 Precipitation + \beta_8 Temp \times \frac{AOD}{PBL} + \beta_9 SnowCoverage \times \frac{AOD}{PBL} + \beta_{10} RH \times \frac{AOD}{PBL} + \beta_{11} WS \times \frac{AOD}{PBL} + \beta_{12} Precipitation \times \frac{AOD}{PBL} \quad (3)$$

Subsequently, a 10-fold cross validation is conducted to evaluate the calibration model (eq. 3).To perform cross validation, data from the monitoring stations are randomly separated into a 10% held-out set and a 90% training set. The calibration is a supervised linear regression model fitted based on the 90% training set and thereafter to predict the 10% held-out data. The same procedure is repeated 10 times until all data are predicted once. We then compare the calibrated prediction to the observed $PM_{2.5}$ concentration and examine the variability explained by the prediction (R^2) and bias (slope and intercept). The difference in R^2 between the 10-fold

calibration and training (where all data is used to fit the model) should be less than 10% otherwise the model is likely over fitted. After careful evaluation, we obtained AOD-derived-$PM_{2.5}$ concentrations for all 1km x 1km cells in the study domain.

Emission model. We considered each box as a single compartment and modeled $PM_{2.5}$ concentration (C) using the mass balance concept as illustrated in Figure 2:

$$\frac{dC(t)}{dt} = \sum Sources - \sum Sinks \quad (4)$$

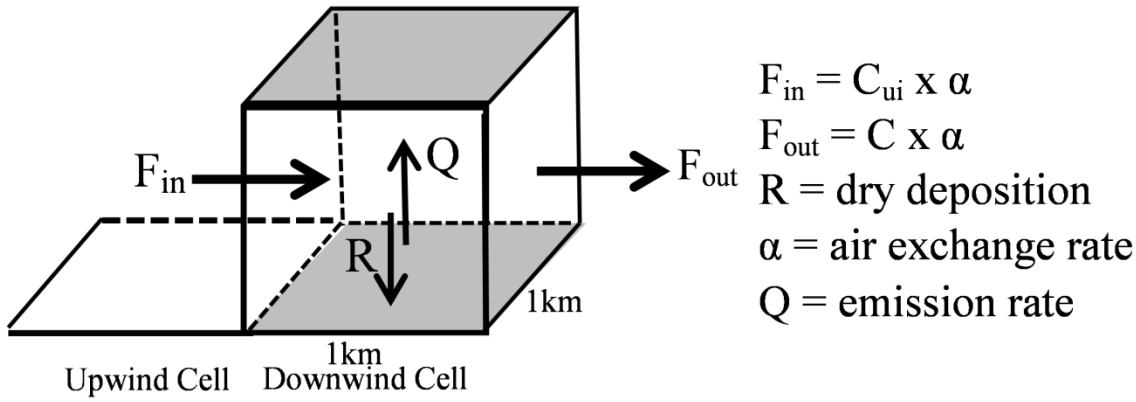


Figure 2. Box model dynamics.

Within a box, particles are transported from upwind cell (s) ($C_{u,}$ $PM_{2.5}$ concentration in the upwind cell) or released by sources located inside the box (local emissions). This includes primary emissions from local sources as well as formation of secondary particles from precursor gases emitted from local or distant sources. On the other hand, particle losses (sinks) within an atmospheric column are due to dry deposition (d), wet deposition (w) and air exchange (α) transport to the downwind cell. Wet deposition is not accounted for in this model because we omitted AOD retrievals during days with rain or clouds. As previously stated, we used PBL to

estimate the volume of each 1 km x 1 km cell, and thus the predicted local emission Q is expressed in tons/km^2/year. Moving forward we will discuss particle transport and model derivation in 2-dimensional space.

$$\frac{dC(t)}{dt} = \alpha C_u + \frac{Q}{PBL} - (\alpha + d)C(t) \quad (5)$$

Since dry deposition is usually considerably slower than the air exchange rate, we can simplify the mass balance equation by neglecting dry deposition:

$$\frac{dC(t)}{dt} = \alpha C_u + \frac{Q}{PBL} - \alpha C(t) \quad (6)$$

Assuming equilibrium, we can solve the differential eq 6. and obtain the following solution:

$$C = C_u + \frac{Q}{\alpha \times PBL} + C_1 e^{-\alpha t} \quad (7)$$

In addition, if we assume that the transported pollution C_u is independent of local emissions Q, the integration constant C_1 is equal to zero, and we can simplify eq 7. as follows:

$$C = C_u + \frac{Q}{\alpha \times PBL} \quad (8)$$

Particle transport primarily occurs between two neighboring cells as shown in Figure 2. However, emissions may travel further depending on the wind speed and elevation of the source. Given the fine resolution (1km x 1km cells) of this study, transported emission is likely travelling from further than one upwind cell. In fact, $PM_{2.5}$ concentrations of upwind cells within 3km distance strongly correlate to the concentration of the corresponding downwind cell. For

this reason, 3 upwind cells are included in the emission model to account for all transported particles. The algorithm used to locate upwind cell(s) depends on wind direction in the downwind cell as illustrated in Figure 3.

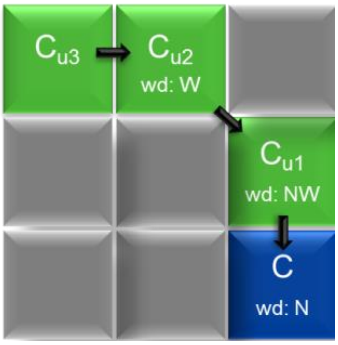


Figure 3. Example of upwind cells identification.
wd: wind direction, N: North, NW: Northwest, W: west.

In addition to the within-cell emission, secondary particles are formed from gaseous emissions originated from sources situated within and outside the downwind cell. In order to control for the secondary particles formed outside the downwind cell and were not captured by the upwind cell concentrations, we include temperature(ºK) in our model as a surrogate of the various weather parameters associated with particle formation (Vehkamaki and Riipinen 2012):

$$C = \sum_{i=1}^{3}(C_{ui} \times Temperature) + \frac{Q}{\alpha \times PBL} \quad (9)$$

We fitted the above model (eq.9) to predict emissions, Q, for all 1 km x 1 km cells within the study domain.

Land use regression. Since emissions are often closely related to land use parameters, we fitted regression (LUR) models to examine relationships between model predictions (Q) and source types, which is an indirect approach to evaluate PERIRS. The land cover data from MASS DEP provides a more detailed classification than that of the NEI. Therefore, we fitted a land use model specifically for Massachusetts to examine potential relationships between model predictions and different land cover types. In addition, we also included point emission estimates from the NEI in the LUR model to account for industrial or larger point sources that may not be included in the land use database. Toward this end, we included an indicator variable for point emission in the LUR models to determine whether higher Q values are associated with major point sources inside the 1km x 1km grid cell.

County level evaluation model. US EPA reports PM emission from all sources at the county level, and for this reason, we averaged our 12-year-averaged-emission predictions by county and compared them to the county NEI emission densities as a secondary validation (eq. 10). We performed two county-level evaluations using NEI 2008 and 2011, respectively. In addition, we conducted state-specific evaluations to examine differences among model predictions and NEI emissions. Although a better evaluation is to compare annual PEIRS emissions to the NEI in corresponding years, we currently do not predict annual emissions with the PEIRS model. The reason being annual emission predictions could be biased due to imbalanced data since missing in AOD occurs seasonally.

$$NEI\ Emission = \beta_0 + \beta_1 Q_{county\ averaged} + \epsilon \quad (10)$$

RESULTS AND DISCUSSION

AOD/$PM_{2.5}$ Calibration Performance

Table 1 depicts the performance of AOD/$PM_{2.5}$ calibration. The model training R square (0.85) is moderately high suggesting that the calibration model was well fitted. The cross-validation analysis demonstrated that model predictive accuracy is high for temporal variability (R^2=0.80) and moderate for spatial variability (R^2=0.66). Among AOD-$PM_{2.5}$ calibration models, that of Kloog et al., has so far the strongest predictive power (temporal R^2 = 0.87, spatial R^2 = 0.87) (Kloog et al., 2014). While the temporal predictability is similar between our calibration and that reported by of Kloog et al., our spatial predictability is much lower. A possible reason for this difference is that Kloog et al., used a hybrid approach that includes not only meteorological covariates and elevation, but also many other land use variables, which are crucial PM_{25} predictors enhancing spatial predictability. As previously stated, the major goal of our model is to predict emissions based on remote sensing data, and it would not be appropriate to include source-related covariates such as land use terms.

Table 1. 10 fold cross validation

R^2 (dimensionless) RMSE ($\mu g/m^3$)	**Temporal**		**Spatial**	
	R^2	**RMSE**	**R^2**	**RMSE**
Training Performance	0.85	2.56	0.75	0.90
10 fold Cross Validation	0.80	2.85	0.66	1.06

Predicted emissions

Figure 4 depicts emissions predicted over the period of 2002 to 2013 in the northeast US. We observed emission hotspots with more than 35 tons/year/km^2 in most highly populated areas such

as Boston, New York City, and Long island. In addition, we observed transportation emissions along major highways. Finally, we predicted low emission levels for rural areas (<20 tons/year/km^2).

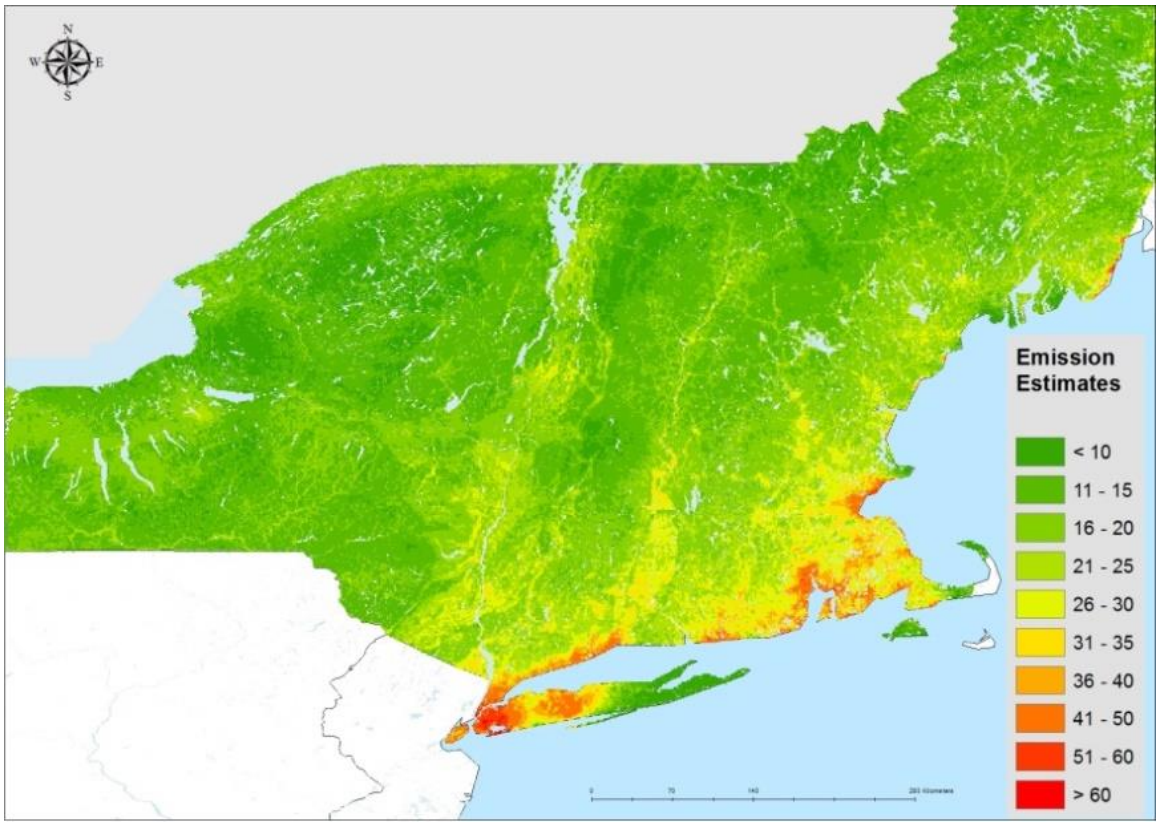


Figure 4. Estimated emission in north east US.

Figure 5 shows the estimated emissions, land cover and population distributions in the Greater Boston, New York, and Providence areas. In general, we found similar spatial patterns among highly populated areas such as the Greater Boston and New York, reflecting typical urban activities. PEIRS predicted high emissions levels in developed areas (Figure 5d and 5e, color in red) in east Boston, New York City, and Long Island areas with a mixture sources including residential heating, transportation, industrial, and commercial activities. More importantly, we were able to capture intra-urban variation. For example, we predicted lower emissions for the

east Cambridge where the Massachusetts Institute of Technology (MIT) Campus is located. As shown in figure 5a, emissions are noticeably lower (25~30 tons/year/km^2) as opposed to surrounding areas such as Somerville and west Cambridge. The campus area is less populated compared to neighboring cities and the school has been actively promoting green building and energy sustainability programs over the past decade, which both attribute to lower emission. Furthermore, we predicted high emissions at densely populated and commercially active areas in northern Brookline and lower emissions at parks, large green spaces, and residential areas in the southwestern part of the town (Figure 5a). Our results in New York (Figure 5b) show comparable intra-urban spatial pattern of $PM_{2.5}$ to those reported by a previous study (Clougherty et al. 2013). Clougherty et al., found that large combustion boilers for residential heating, mostly burning oil, are concentrated in midtown and downtown Manhattan as well as in some neighborhoods of Brooklyn and Queens. These areas are all highly developed with similar population density and land use. Even surrounded by high intensity emission sources, we are able to identify intense oil-burning emissions in some of the grid cells that exceeded 60 tons/year/km^2 in these neighborhoods.

Figure 5. (a) Estimated emission in Greater Boston.

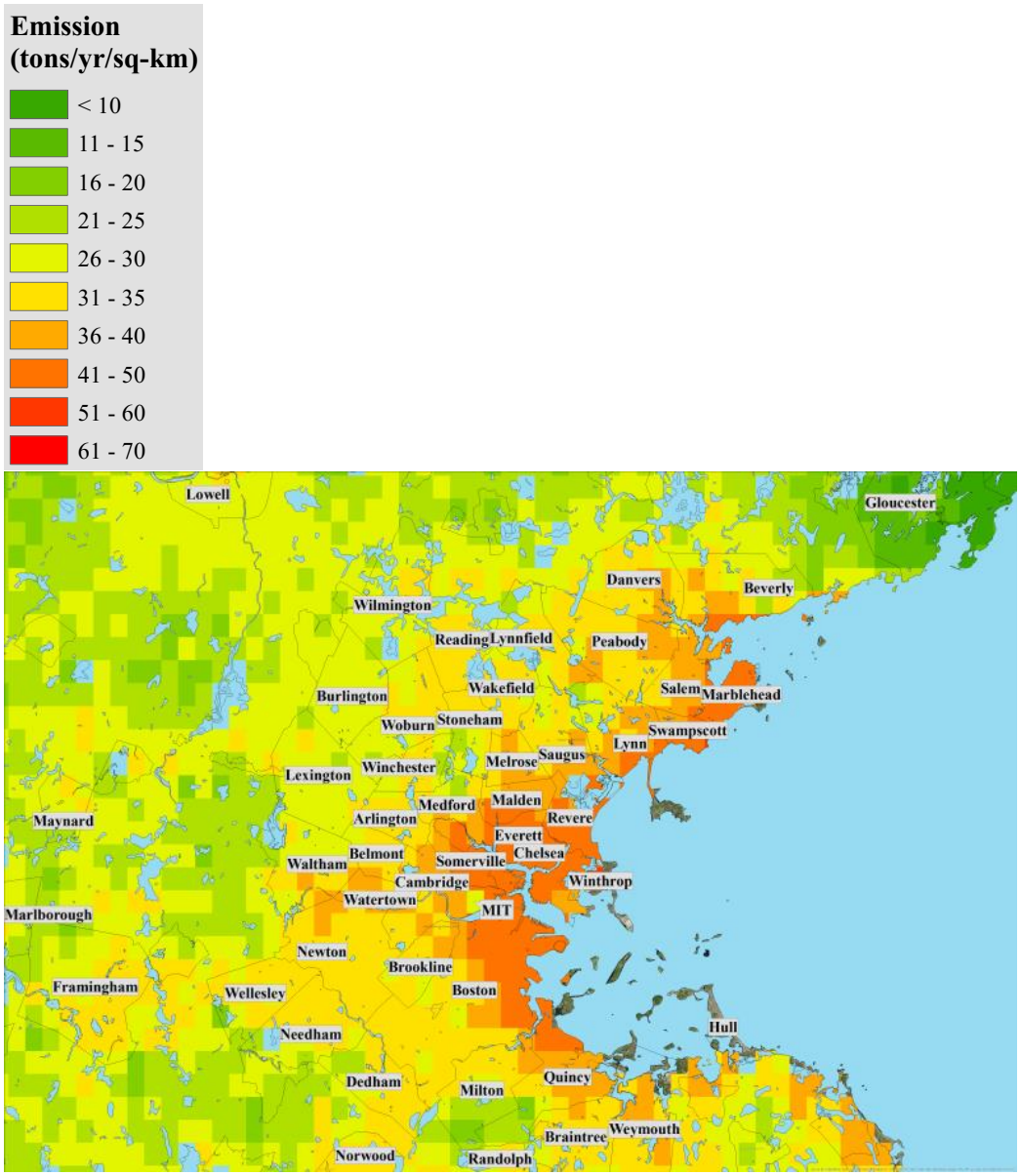


Figure 5. (b) Land cover type from NLCD 2011 database in Greater Boston.

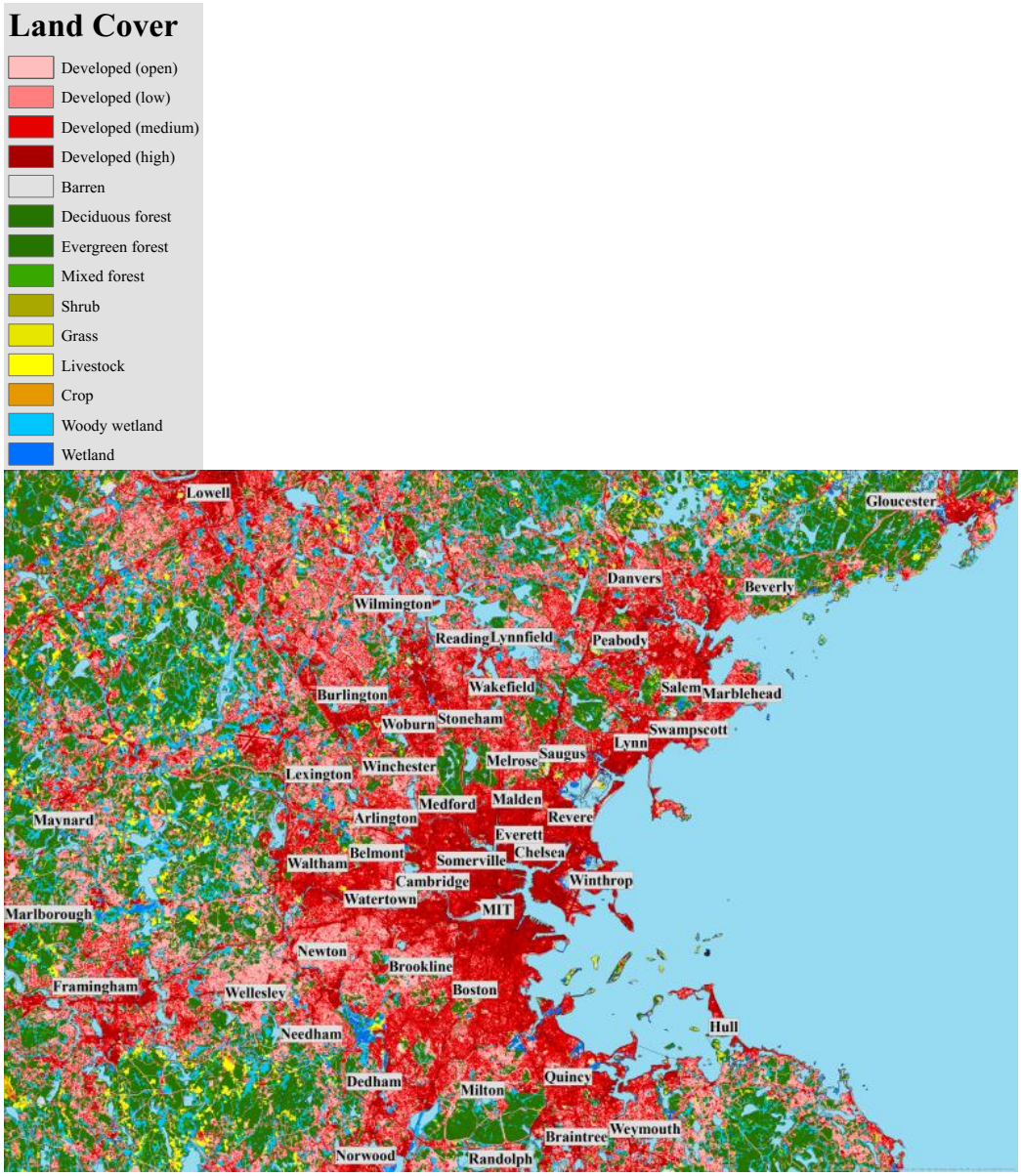


Figure 5. (c) Population of year 2000 in Greater Boston.

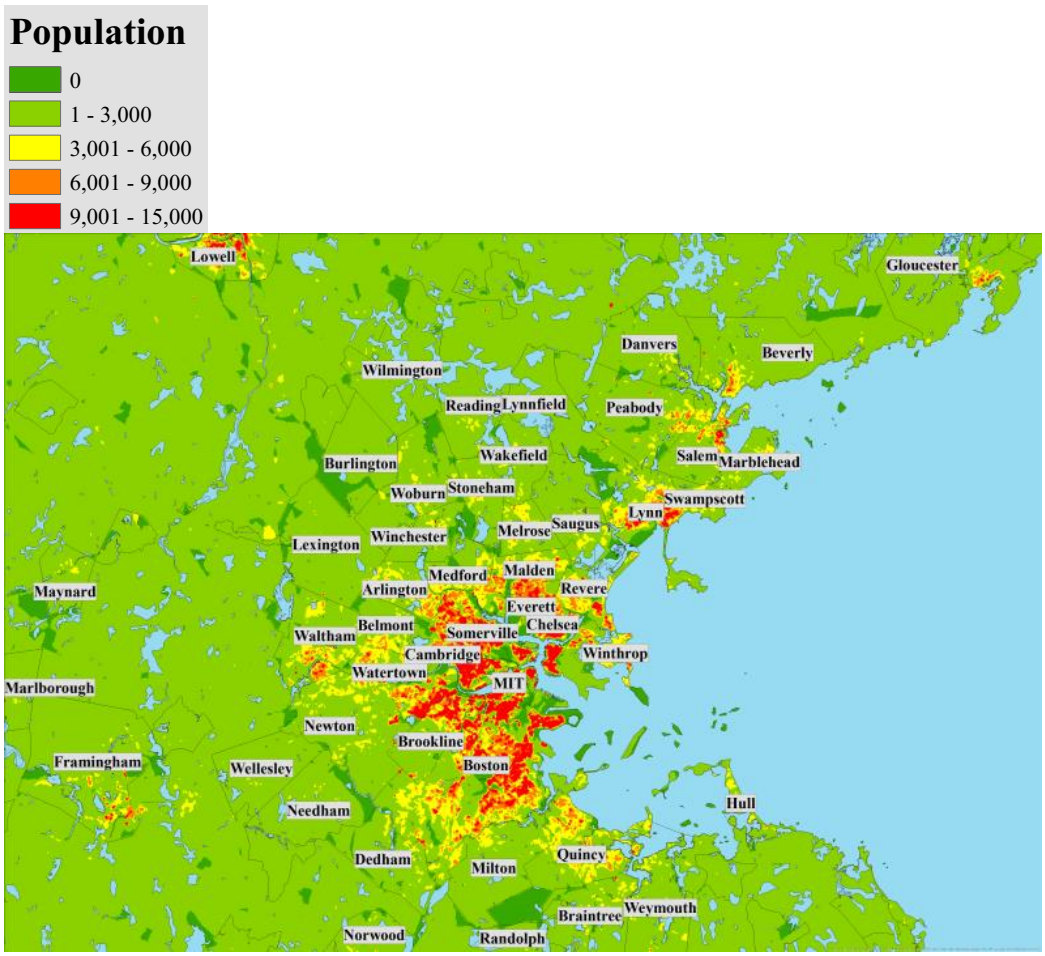


Figure 5. (d) Estimated emission in New York.

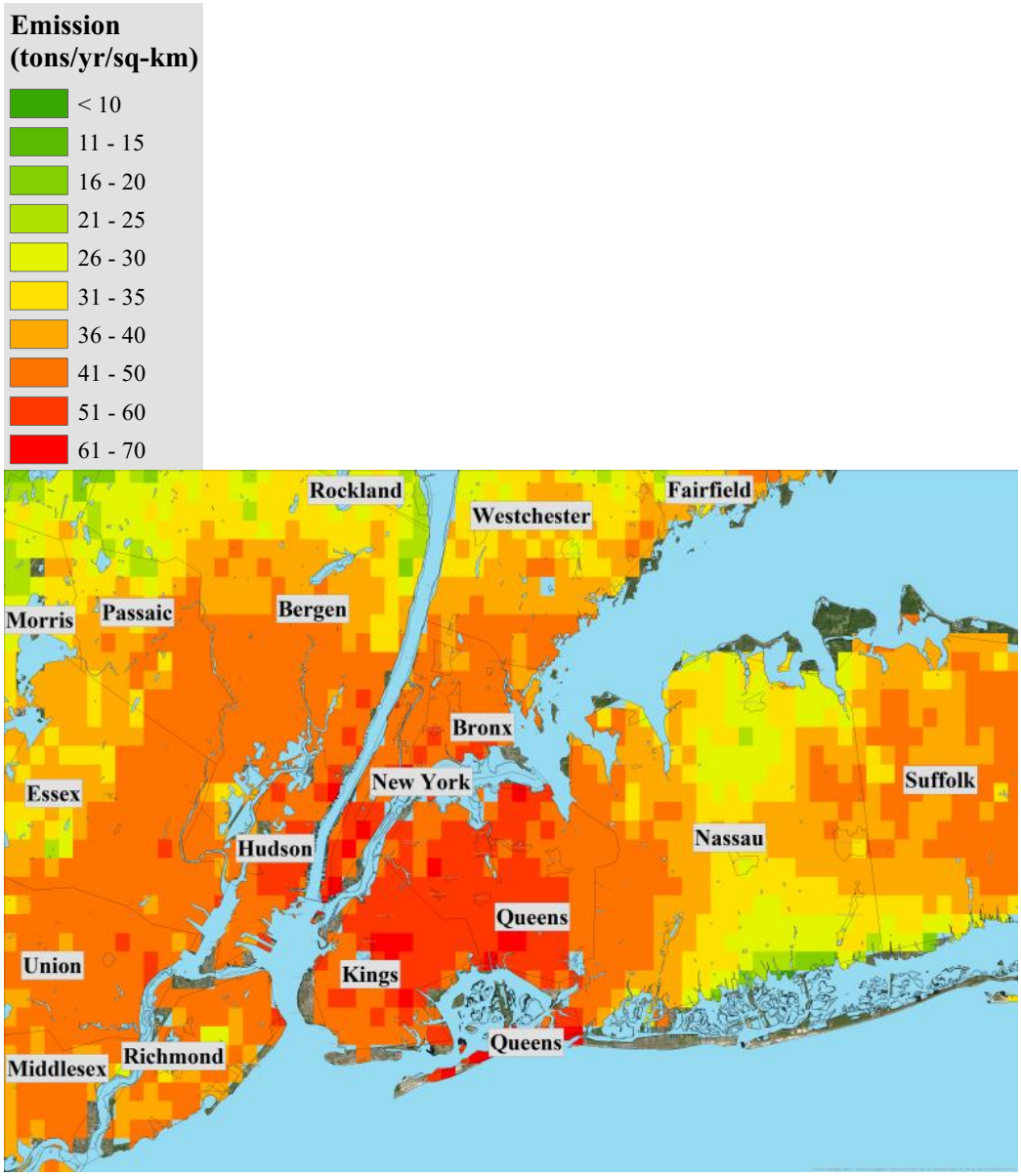


Figure 5. (e) Land cover type from NLCD 2011 database in

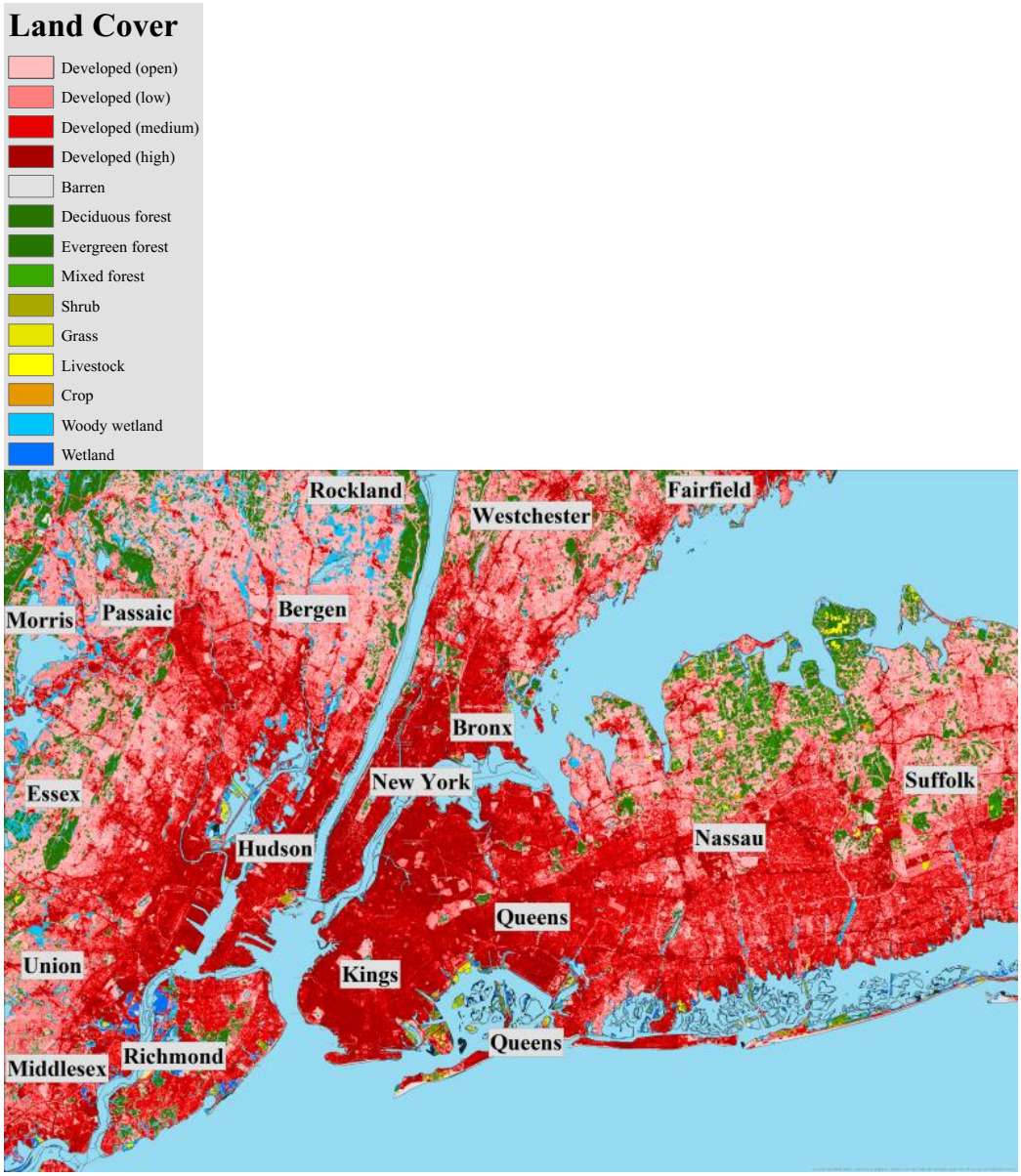


Figure 5. (f) Population of year 2000 in New York.

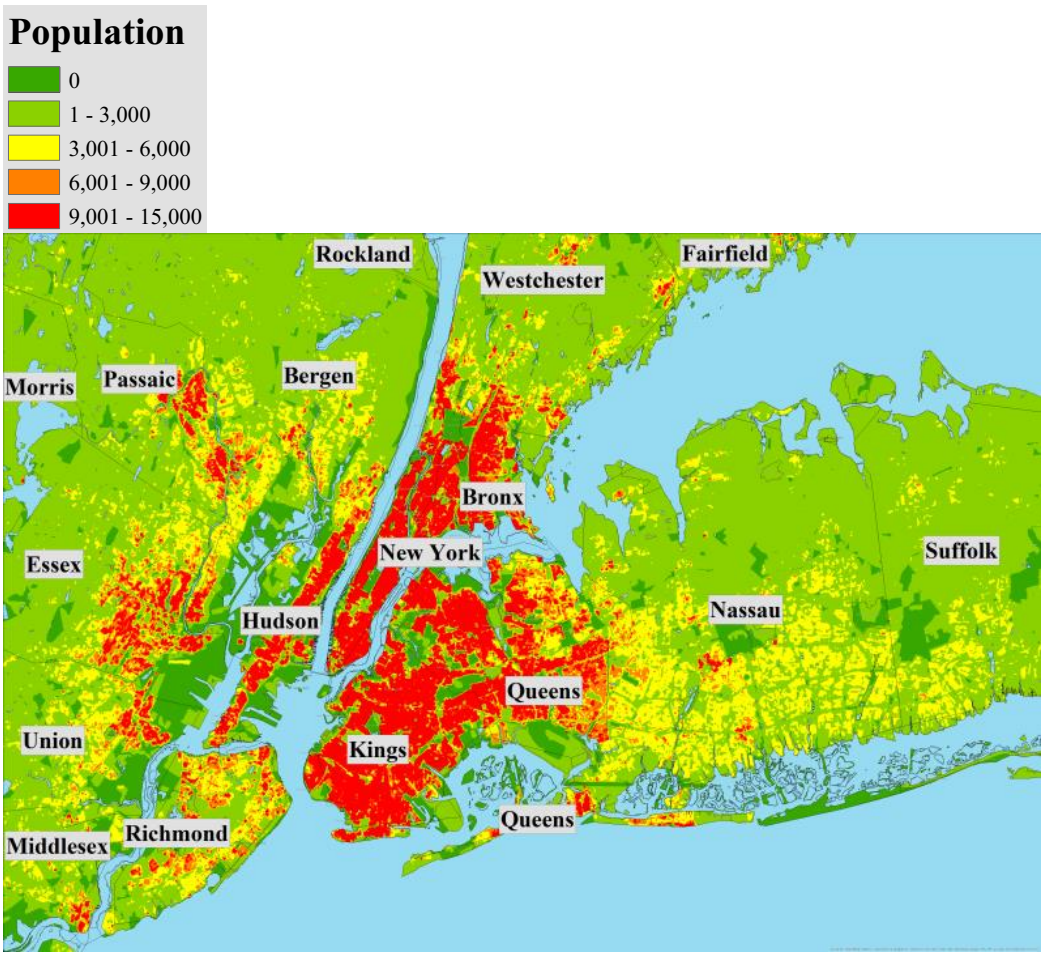


Figure 5. (g) Estimated emission in Providence.

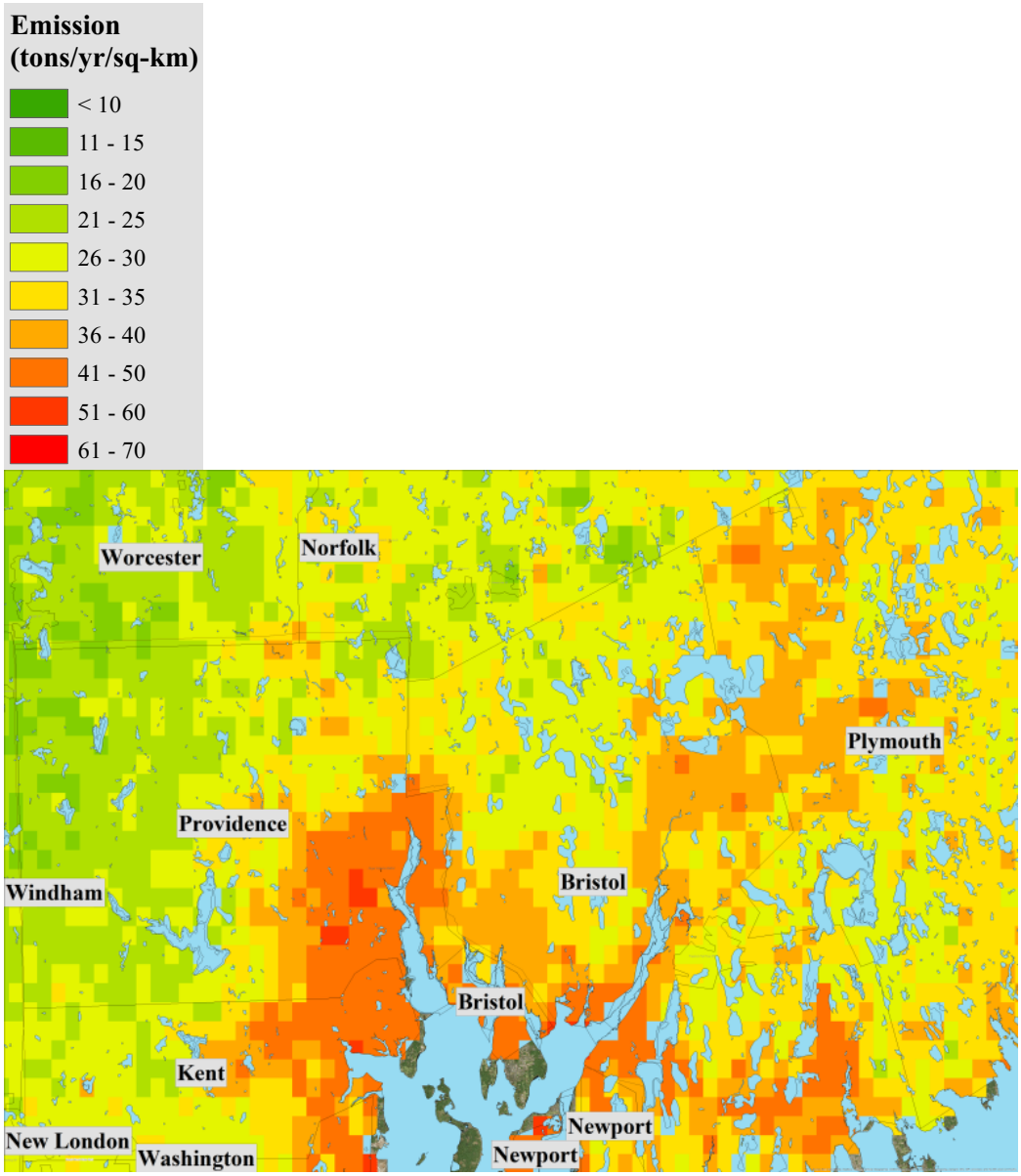


Figure 5. (h) Land cover type from NLCD 2011 database in Providence

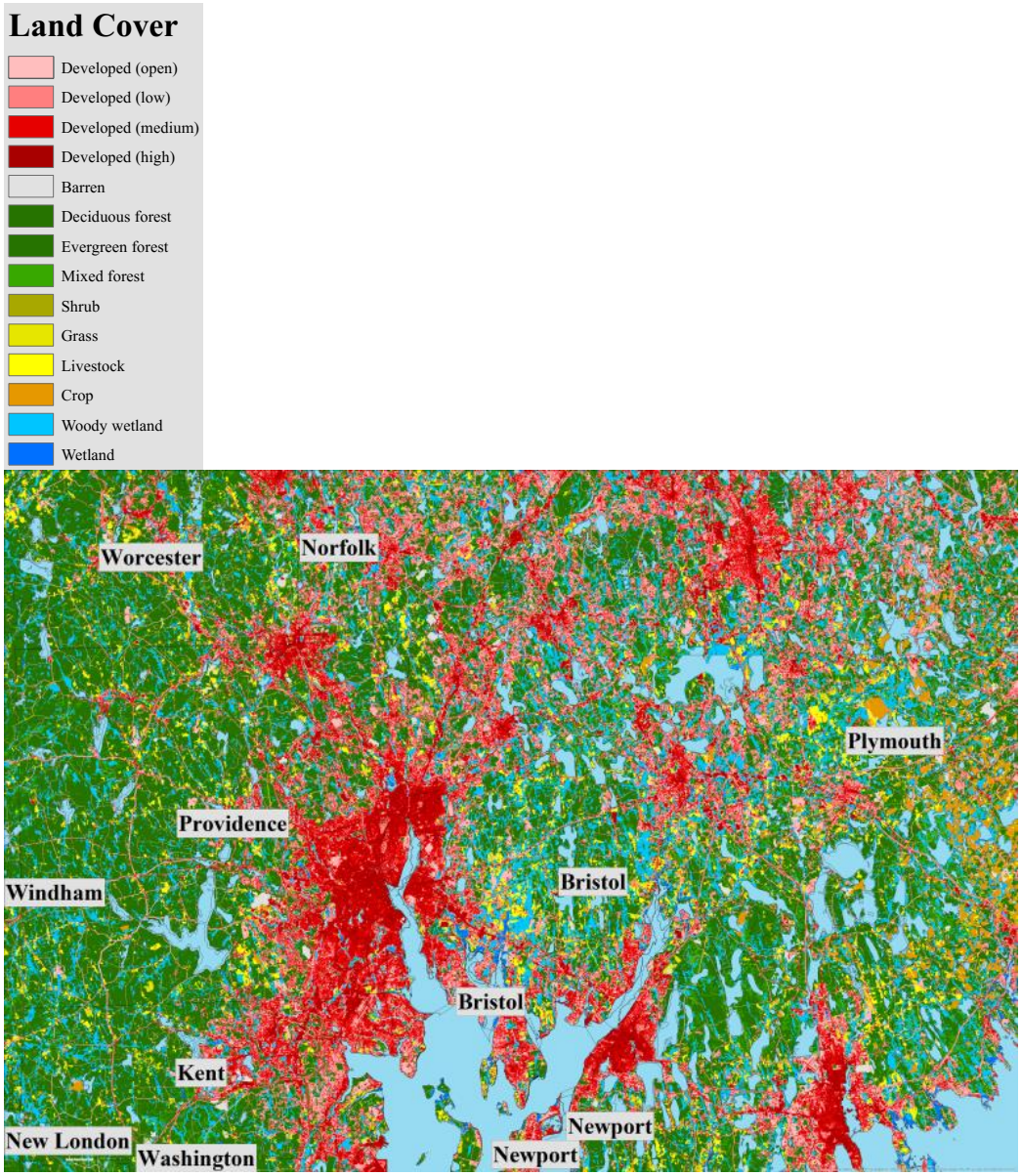
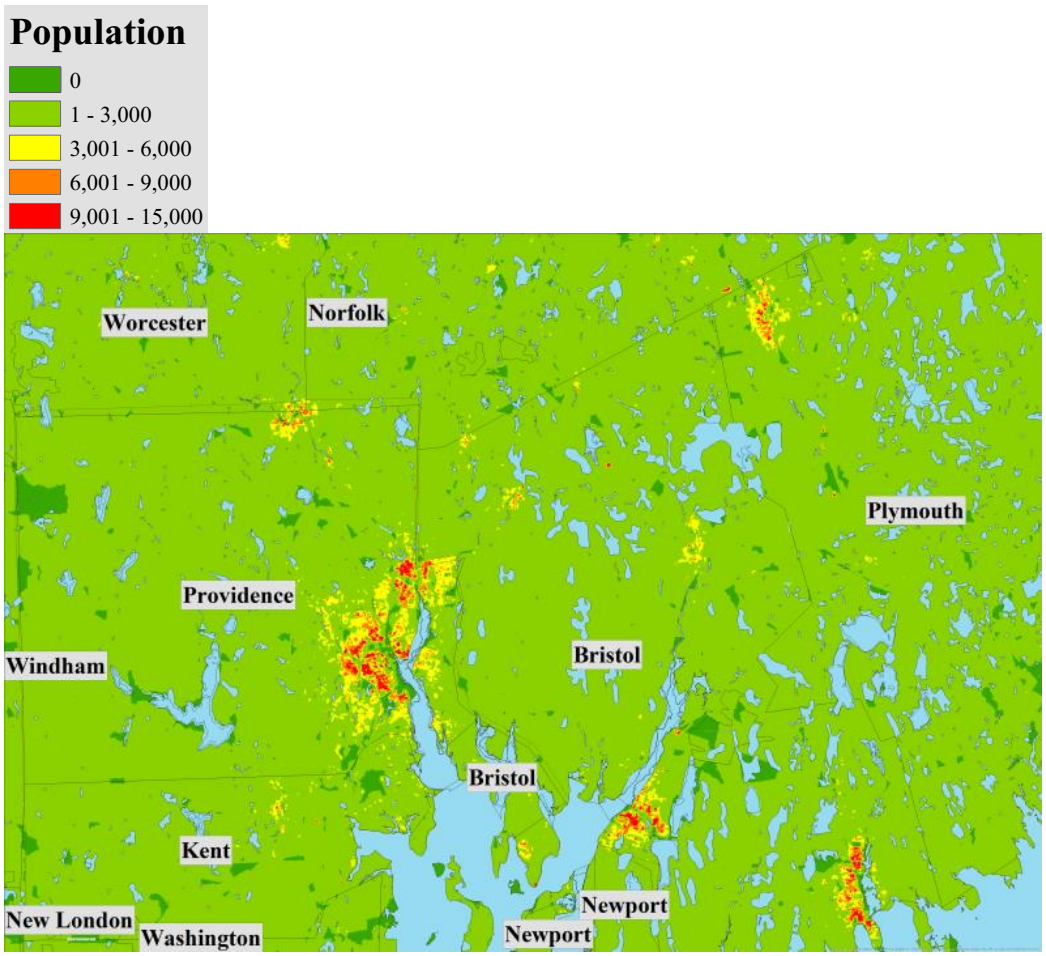


Figure 5. (i) Population of year 2000 in Providence.



Our model overestimated emissions in areas surrounded by water surfaces or wetlands, such as Providence and southeast Massachusetts (Figure 5c).For instance, the model estimated high emissions (20-30 tons/year/km^2) in some green areas of southeast Massachusetts where anthropogenic sources are scarce. The possible overestimation is likely due to bias in AOD readings, as the MAIAC algorithm was developed to retrieve AOD over land, and is less reliable for retrievals over water surfaces (Lyapustin et al. 2011). Although we excluded pixels encompassing bodies of water prior, AOD retrieval near water or the shore can still be influenced by high humidity or ocean glint. Because we estimated $PM_{2.5}$ concentrations using AOD data, biases in the AOD are likely to influence the final emission product as well. However, it is also

possible that some of the predicted emissions are due to primary or secondary particles from natural sources.

Land use regression

Predicted emissions in the northeast United States are related to land use variables ($R^2 = 0.65$). Table 2 shows the predicted land-use-specific emissions where the predominant sources include developed area (high/medium/low density and open area), traffic, and population. The land use regression model predicts that large point sources emit on average 0.58 tons/year/km^2 that is close to the averaged NEI point emission in 2008 (0.53 tons/year) and 2011(0.26 tons/year). Figure 6a shows the residuals (25th percentile = -2.7, mean = 0, 75th percentile = 2.5) of the land use regression model in the study area. As noted, the Providence and southeast Massachusetts areas have larger residuals (>5 tons/year/km^2), which suggests that either the predicted emissions are overestimated or they are not correlated to land use covariates, e.g., biogenic sources, biogenic precursors or oceanic particles. Previous studies suggest that in remote areas most particles are secondary (Kanakidou et al. 2000, Spracklen et al. 2006, Andreae 2007), which are not accounted for by land use models.

Table 2. LUR estimated emission from different land use for north east US.

Land use type/ emission sources		**Averaged emission (ton/yr/km^2)**
On road	A1-A3 Major road	0.4 ~ 28
Developed area	High density	13
	Medium density	13
	Low density	6.7
	Open area	5.6
Population	250 people/cell	0.1
	10,000 people/cell	2.7
Farms	Crop & Livestock	0.34
NEI point source	(NEI 2008 mean = 0.53)	0.58

Figure 6. (a) Residual map of land use regression (LUR) model in northeast US

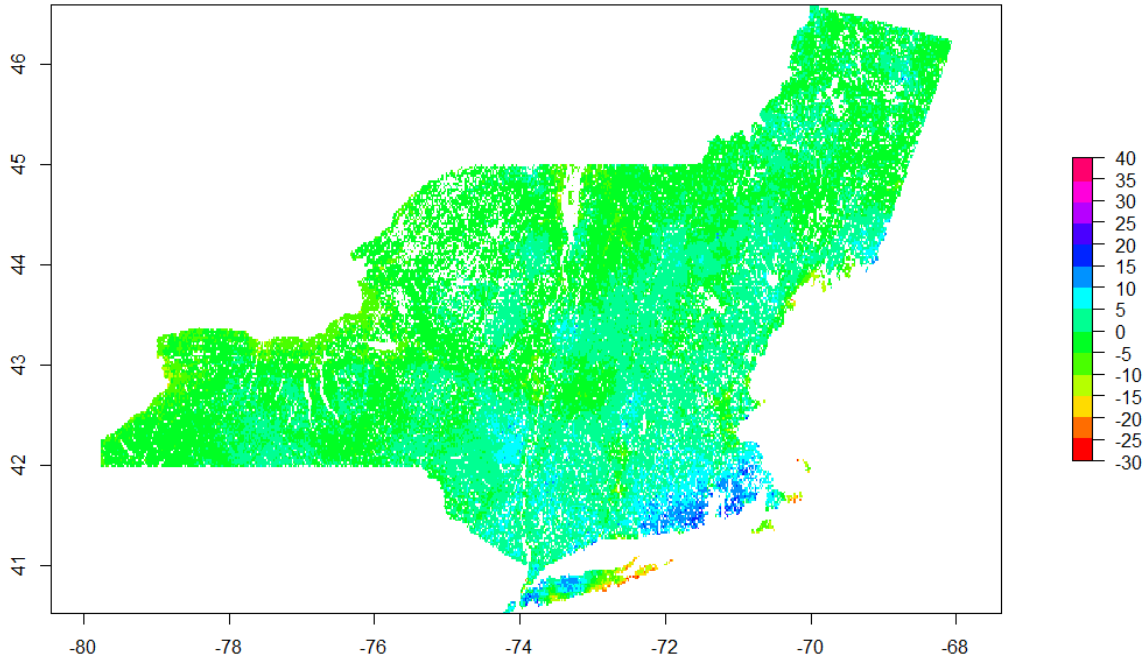
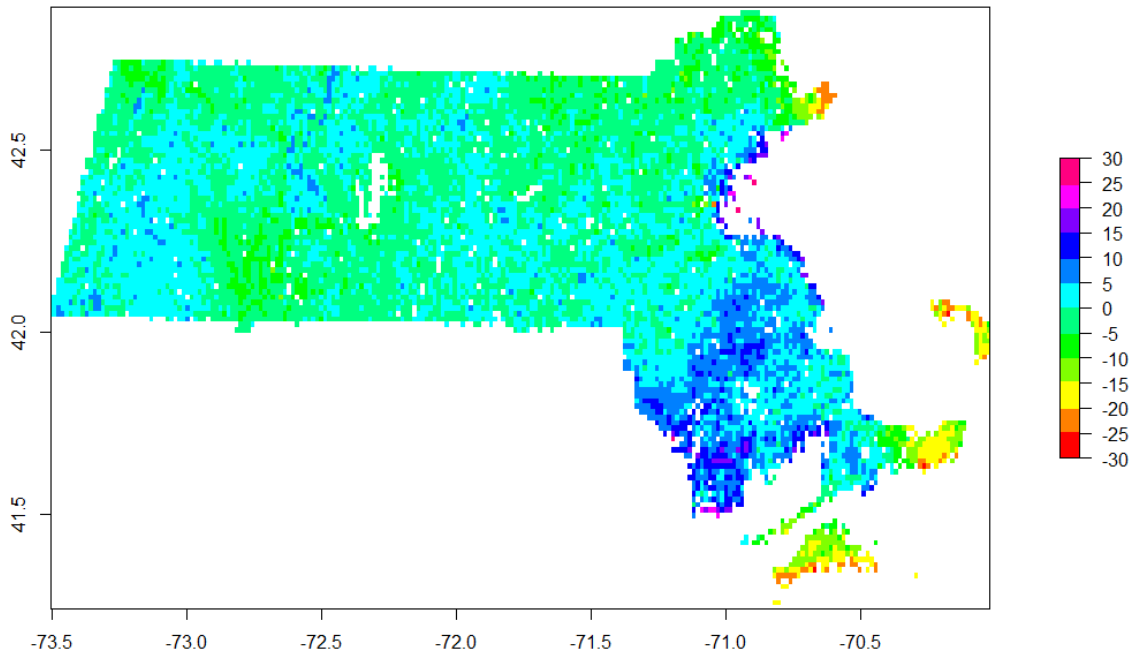


Figure 6. (b) Residual map of land use regression (LUR) model in Massachusetts.



In addition to the national land cover product, we also incorporated the land cover data classified by the Massachusetts Department of Environmental Protection (Mass DEP), and fitted a separate land use model for Massachusetts. Compared to the national model (R^2 = 0.65), the Massachusetts model has a slightly higher R^2 of 0.67. As shown in Table 3, transportation, developed area, and population are larger sources predicted by the land use model. With a finer classification, we can see the intra-urban variation related to industrial, residential, and commercial sources areas, which is not for the NEI database that are mostly county level averages (Figure 7 b and 7c). Crop and livestock farms emissions are 0.7 tons/year/km^2, which are similar to those reported by the NEI program. Cranberry bogs, nurseries, and orchards generate more emissions than other croplands. Consistent with the results of the northeast US Land use regression, we observed high residuals in the southeastern Massachusetts. As noted above, this may be due to difficulties in measuring AOD near water bodies, or possibly, to the presence of sources whose emissions are independent of land use variables.

County level evaluation

We found a moderate correlation between predicted and 2008 NEI emissions (R^2=0.52, CV=144%). The slope and intercept of the 2008 NEI county model (β_1 = 2.6 β_0 = -38.7) suggest that the model underestimated the county level emissions. Since the NEI emission data is right skewed (median = 4.9, mean = 11.8), we suspect the large error is due to extreme values (or outliers) in the NEI data. When restricting NEI data to less than 50 tons/year/km^2, we obtained a better model fit (R^2 = 0.66, CV=20.0%, β_1 = 0.65 and β_0 = -5.6). This suggests that the model may underestimate the largest emission sources. The different temporal scale among emission predictions and NEI estimates could be another source of uncertainty leading to larger error and a larger deviation of the intercept. Nonetheless, the PEIRS method has great potential in predicting emissions at finer temporal scales and would reduce the temporal inconsistency error. We found similar results when comparing model emission predictions to those reported by the 2011 NEI. The model fit was significantly improved after removing high NEI emissions from the model (R^2 = 0.65 → 0.71, CV= 105%→17.9%, β_1 = 3.4 →1.09, β_0 = -47.8 → -11.7, all data → subset). Interestingly, the slope of the county regression model shows that our model overestimated emissions in comparison to the 2008 NEI, and slightly underestimated when in comparison to 2011 NEI. This is consistent with the emission changes documented in the 2011 NEI report where emissions overall increased by 25 to 250% compared to 2008 due to resuspension of road dust and more frequent wildfires (EPA 2011).

Table 3. LUR estimated emission from different land use for Massachusetts.

Land use type/ emission sources		Averaged emission (ton/yr/km^2)
On road	A1-A3 Major road	0.4 ~ 24.5
	Public transit	2.5
Developed area	High density residential	3.6

	Medium density residential	3.5
	Low density residential	1.8
	Multifamily residential	2.1
	Open area	2.4
	Industrial	4.3
	Mining	2.4
	Commercial	3.1
	Junkyard	4.6
Population	400 people/cell	0.32
	10,000 people/cell	8
Farms	Crop & Livestock	0.7
	Cranberry bog, nursery, orchard gardens	2.2 ~ 7.8
NEI point source		1.5

Table 4 and Table 5 show the state specific county level evaluation results compared to 2008 and 2011 NEI data, respectively. We observed fair agreement for Connecticut and New Hampshire than that for states with large sources such as New York and Massachusetts. In general, the agreement was improved when outliers were removed from the analysis. Predicted emissions were weakly related to NEI data in Vermont suggesting that the model is less sensitive to low level emission variations. As shown in Figure 7, for most counties in Vermont both model predicted and NEI-reported emissions are low. The relatively high emissions predicted for Grand Isle county, which is surrounded by Lake Champlain, may be due to water bias in the AOD. In addition, we did not find a good agreement for Franklin, Chittenden, and Addison counties where the model predicts higher emissions (6 ~15 tons/year/km^2) than those reported by NEI (< 5 tons/year/km^2). However, we found out that emissions of particles precursors such as NH_3, NO_x and VOC are relatively high in these three counties according to the NEI 2008 report. Therefore, it is possible that differences are due to the formation of secondary particles and high biomass burning sources, which are accounted for by the model but not included in the NEI database. In fact, counties where predicted emissions are higher than those reported by the NEI

are locations with high precursor emissions. These include urban counties in Greater Boston, New Haven, New York City, and Long Island, as well as agriculture-driven counties in the upper west New York state.

Table 4. Evaluation results of estimated emission vs. EPA NEI emission 2008 at county level.

	R^2	CV	Slope (β_1)	Intercept (β_0)
Connecticut	0.85	6.61%	0.73	-6.8
Maine	0.65	9.11%	0.54	-5.5
New Hampshire	0.86	15.0%	1.44	-15.5
Massachusetts	0.58 (0.87)	75.4% (6.97%)	1.84 (0.47)	-26.8 (1.0)
Rhode Island	0.21	13.3%	0.20	5.03
New York	0.58 (0.67)	211.1% (26.3%)	2.99 (0.71)	-45.44 (-6.6)
Vermont	0.01	7.45%	0.02	3.40

* ( ): are model parameter when data is restricted to counties with ≤50 tons/yr NEI emission

Table 5. Evaluation results of emission estimates vs. EPA NEI emission 2011 at county level

	R^2	CV	Slope (β_1)	Intercept (β_0)
Connecticut	0.88	7.15%	1.41	-11.7
Maine	0.70	7.69%	0.76	-8.87
New Hampshire	0.89	8.13%	1.35	-14.6
Massachusetts	0.63 (0.27)	60.1% (16.1%)	2.37 (0.43)	-32.3 (3.7)
Rhode Island	0.49	10.7%	0.50	-0.47
New York	0.70 (0.80)	159.6% (21.2%)	3.80 (1.26)	-54.7 (-14.3)
Vermont	0.08	6.98%	0.11	1.96

* ( ): are model parameter when data is restricted to counties with ≤50 tons/yr NEI emission

Figure 7. (a) Estimated emission vs. EPA NEI nonpoint emission in (b) 2008 and (c) 2011 at county level.

(a) Estimated emission from the PEIRS model, downgraded to county level.

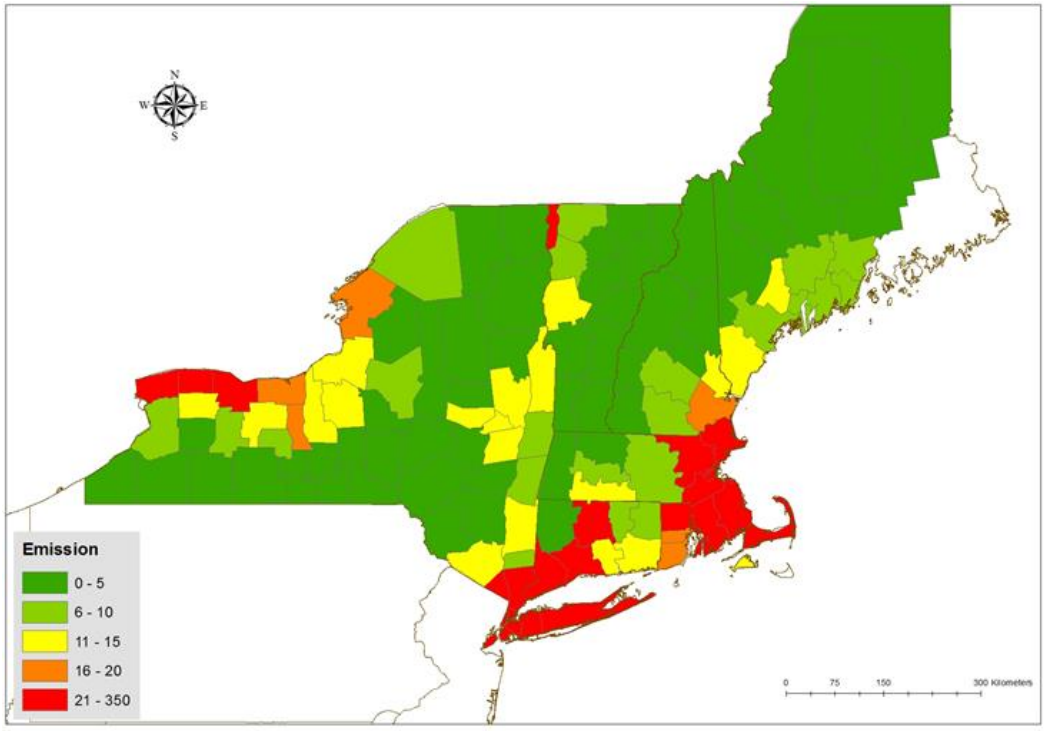
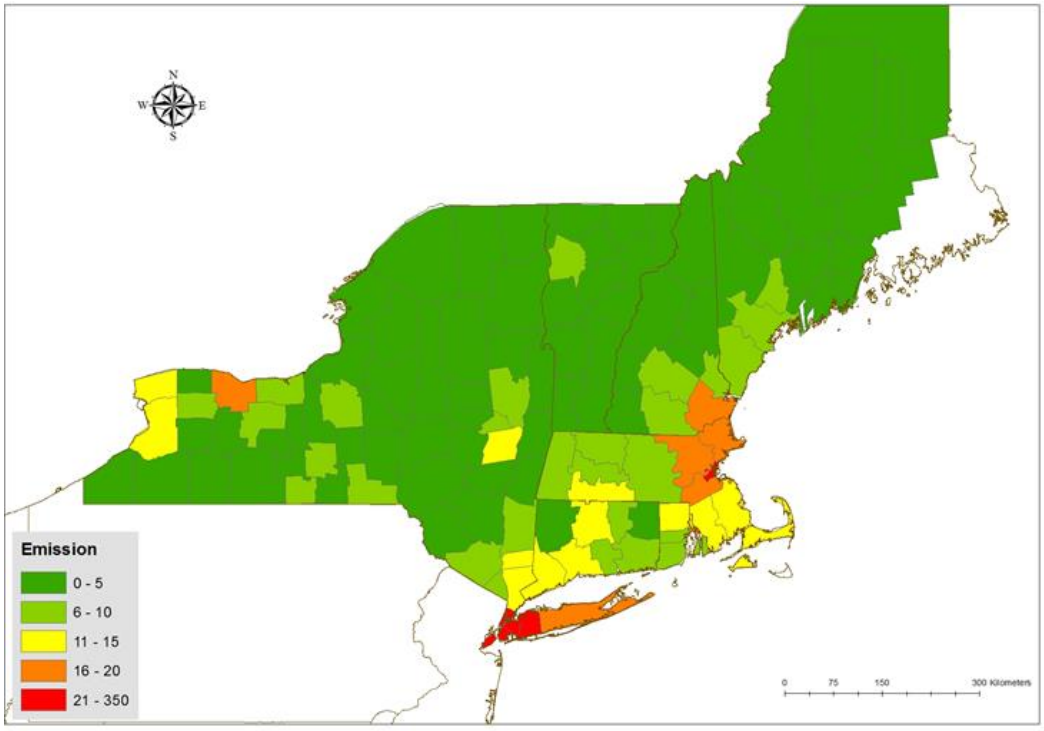
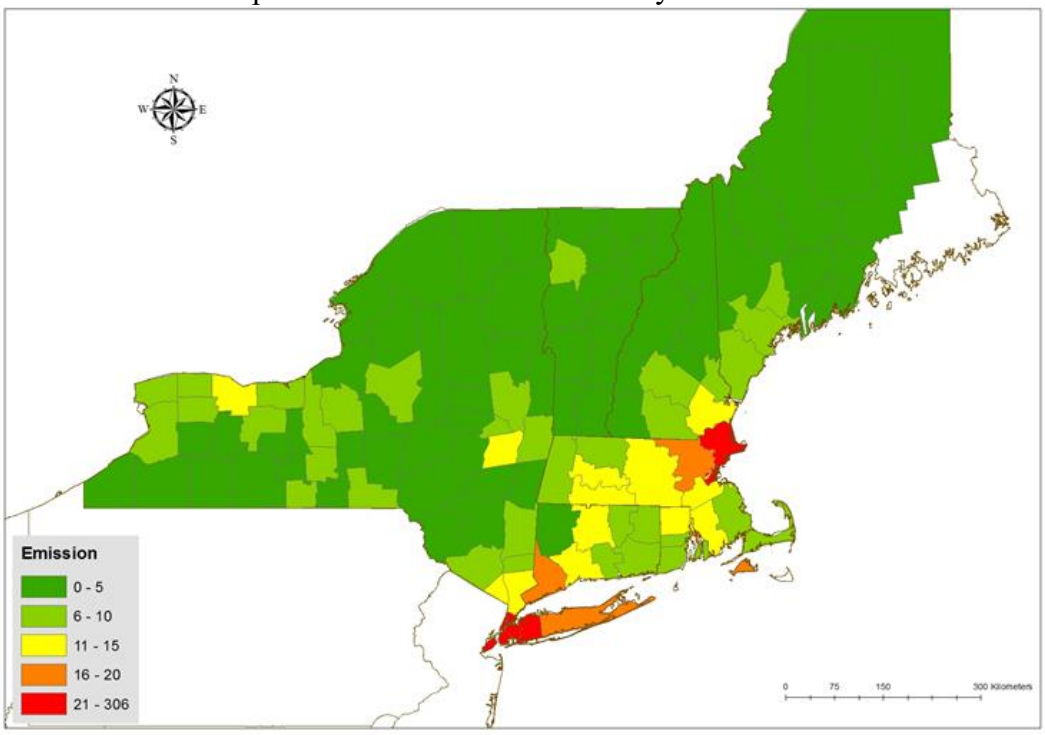


Figure 7. (a) Estimated emission vs. EPA NEI nonpoint emission in (b) 2008 and (c) 2011 at county level.
(b) US EPA NEI non-point emission in 2008 at county level.



(c) US EPA NEI non-point emission in 2011 at county level.



CHAPTER 2

TRENDS AND SPATIAL PATTERNS OF FINE RESOLUTION AOD-DERIVED $PM_{2.5}$ EMISSIONS IN THE NORTHEAST UNITED STATES FROM 2002 TO 2013

Chia Hsi Tang [a]

Brent Coull [b]

Joel Schwartz [a]

Alexei I. Lyapustin [c]

Qian Di [a]

Petros Koutrakis [a]

[a] Department of Environmental Health, Harvard T.H Chan School of Public Health

[b] Department of Biostatistics, Harvard T.H Chan School of Public Health

[c] National Aeronautics and Space Administration Goddard Space Flight Center

ABSTRACT

Clarifying the trends in quantity, location, and causes of $PM_{2.5}$ emission changes is critical for evaluating and improving emission control strategies and reduce the risk posed to human health. According to the National Emission Inventories (NEI) released by the U.S. EPA, a general downward trend in $PM_{2.5}$ emissions has been observed in the U.S. over the past decade. While this trend is representative at the national level, it lacks the precision to locate emission hotspots at a finer scale. Moreover, the changes reported in the NEI are likely confounded by periodic modification of inventory methods, and imprecision for area sources. In this regard, it is imperative to acquire emission inventories with as much spatial and temporal detail as possible to further our knowledge of particle emissions, exposure levels, and associated health risks. In this study, we employed the PEIRS (Particle Emission Inventory using Remote Sensing) approach to predict triennial averaged emissions at 1 km x 1 km resolution across the Northeast U.S. from 2002 to 2013. Notably, the PEIRS approach is able to capture both primary emission and secondary formation of $PM_{2.5}$. Regional emission trends were evaluated using quantile regression and source-oriented trends were modeled with land use regression. Our analysis found a regional decrease in $PM_{2.5}$ emissions of 3.3 tons/year/km^2 (18%) over the 12-year period. Furthermore, the rate of emission change at the extremes of the emission distribution was significantly different than that of the mean. Both quantile regression and spatial trends imply that the majority of the reduction in $PM_{2.5}$ emissions was attributable to highly developed spaces such as metropolitan areas and important traffic corridors. This urban-rural disparity was particularly apparent during the cold season. Indirect evidence suggested that the emission

decline during the warm season is primarily attributed to less secondary particle formation. These findings warrant closer investigation of the impact of seasonality on $PM_{2.5}$ emissions.

INTRODUCTION

Particulate matter with aerodynamic diameters less than 2.5 μm ($PM_{2.5}$) poses a serious public health concern. Globally, exposure to $PM_{2.5}$ contributed to approximately 3.5 million annual cardiopulmonary mortalities and 200,000 lung cancer-associated annual mortalities (Anenberg et al. 2010). In the U.S., an estimated 130,000 deaths per year were attributed to $PM_{2.5}$ exposure (Fann et al. 2012). Furthermore, a recent study reported that there is no safe threshold of $PM_{2.5}$ exposure (Shi et al. 2016). Thus, despite the significant decreases in $PM_{2.5}$ concentrations that have been achieved in the U.S.(Hu et al. 2013), there is a need to further improve air quality to reduce adverse health effects. The most effective way to improve air quality is to eliminate individual sources, and a better understanding of trends in the quantity, location, and causes of $PM_{2.5}$ emissions is of upmost importance to achieving this goal.

The U.S. EPA, which is mandated to maintain good air quality for the general public, has taken a number of steps to monitor particle emissions over the past decade. Specifically, they developed the National Emission Inventory (NEI) that contains triennially updated Criteria (CAP) and Hazardous Air Pollutants (HAP) emission estimates for a broad spectrum of source types. According to the 2011 NEI, primary anthropogenic $PM_{2.5}$ emission has decreased 53% nationally between 1990 and 2011, with the largest decline in the fuel combustion category (72%) (EPA 2011). Typically, the East Coast shows the highest $PM_{2.5}$ emission density as well as its precursor gases including sulfur dioxide (SO_2), volatile organic compounds (VOC), and

ammonia (NH_3). Most pollutants are generated in urban counties, except NH_3, which is higher in more rural areas where agricultural and pasture activities are frequent. While the NEI trends are generally representative at the national level, local-scale emission characteristics are less clear because the NEI estimates are often based on county-level models using data collected at various points in time. Furthermore, the confluence of other air pollutants may also increase the uncertainty of local-scale models. For these reasons, EPA has sought more detailed analyses of local emission assessments.

EPA deploys a source-oriented approach to construct the NEI, which estimates the unit emission per activity (or emission factor) of known sources and acquires frequencies of the emission activities to predict total emissions. The NEI is useful in answering general emission questions over broad geographic areas, but this inventory method has led to several issues in interpreting emission trends. First, the NEI only represents a fraction of the total $PM_{2.5}$ emissions, as EPA prioritizes tracking of primary particle sources. Secondary particles are currently indirectly monitored by precursor gases that facilitate particle formations are only voluntarily reported. As secondary particles play an important role in the air quality in the U.S., more comprehensive measure or estimation on this portion of the pollution would render the NEI more complete. Second, whether the NEI trends reflect a real change or is merely a consequence of the periodic adjustment of inventory methodologies is still uncertain. A slight upward trend in $PM_{2.5}$ emissions from the highway vehicle sector was reported in both the 2005 NEI (EPA 2005b) and 2008 NEI (EPA 2008). However, EPA scientists have concluded that the change is likely due to recent method and data improvements in estimating mobile sources(EPA 2008). Lastly, updating new emission factors is costly and laborious. The prospect of comprehensively researching every

existing and emerging emission source is unrealistic. Emission factors that are not regularly revised or updated may also lead to disparities in quality among source sectors. For instance, information on oil and gas operations was incomplete in the 2008 NEI, especially for nonpoint source sectors. The wide spread use of diesel engines to power hydraulic fracturing in the Marcellus shale and elsewhere means emissions from individual wells may be underestimated. Moreover, particle emissions from wildfires, prescribed fires, and biogenic sources are often excluded from the NEI due to high uncertainties of information. Since climate change is likely to increase the incidence of wildfires, this omission may produce misleading trends. Obsolete or incomplete emission information may render the NEI insufficient to accurately and comprehensively represent emission trends.

Instead of estimating individual emission factors, recently, we developed a new method (Particle Emission Inventories using Remote Sensing, PEIRS), which deploys a top-down approach by predicting the total $PM_{2.5}$ emissions directly using satellite data (Tang et al. 2016). Because satellite data has great potential to enhance the timeliness and locating accuracies for emission estimates, satellite imagery has been used to construct inventories for biomass burning or forest fire emissions (Zhang et al. 2011) and global aerosol emissions at 1º to 2º spatial resolution (Dubovik et al. 2008, Huneeus, Chevallier, and Boucher 2012). The PEIRS approach integrates state-of-the-art statistical modeling and long-term daily satellite retrievals of high-resolution 1 km x 1 km Aerosol Optical Depth (AOD) data to generate spatially and time-resolved emission inventories. It has been successfully applied to predict 12-year averaged emissions in the Northeast U.S. and the data have shown reasonable agreement with the county-level NEI. PEIRS emission estimates reflect small-scale intra-urban variations which provide crucial spatial

information for health effects studies and legislative decision making. More importantly, the PEIRS approach enables us to capture both primary emissions and secondary formation inside each 1 km x 1km cells based on a mass balance model. The limitation of the PEIRS method is that AOD data retrieval is restricted during certain weather conditions (e.g., cloudy and snow covered days). However, the PEIRS can still provide ample temporal information when more than 1 year of AOD data is used to predict emissions. Its enhanced cost-effectiveness and consistency also render the PEIRS inventory more adequate for trend analyses.

In this study, we applied the PIERS approach to estimate triennial averaged $PM_{2.5}$ emission inventories and then assessed regional temporal and spatial trends in the Northeast U.S. Calculation of multiyear emission averages is an interim strategy to compensate for weather-associated missing AOD data. Thus, our study duration consists of four 3-year periods spanning from 2002 to 2013, which corresponds to the NEI triennial update schedule. Period 1 refers to 2002–2004, Period 2 to 2005–2007, Period 3 to 2008–2010, and Period 4 to 2011–2013. Regional emission trends were examined using quantile regression models and source-oriented emission changes were predicted using land use regression. We applied the aforementioned analyses to 1) year-round, 2) warm season, and 3) cold season-specific emission estimates separately to further determine the seasonality of emission trends.

METHODS

Input Data

Satellite AOD-derived $PM_{2.5}$ concentrations. We obtained spatially resolved daily $PM_{2.5}$ concentration estimates over the period from 2002 to 2013. The concentration estimates were

derived from high-resolution (1 km x 1 km) daily AOD data from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on board the Aqua Earth Observing Satellite. High-resolution AOD data was retrieved using the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm (Lyapustin et al. 2011), which has been proven to be more robust with higher retrieval rates (Chudnovsky et al. 2013). Detailed calibration procedures and performance are described in (Tang et al. 2016).

Meteorological data. Daily averaged boundary layer height (PBL), temperature (TEMP), and wind field data during the period from 2002 to 2013 were obtained from the National Oceanic and Atmospheric Administration (NOAA) North America Regional Reanalysis (NARR) database. All NARR daily meteorological variables were linearly interpolated from the original resolution of 32 km x 32 km to a resolution of 1 km x 1 km using the scatteredinterpolant package from Mathworks (http://www.mathworks.com/help/matlab/ref/scatteredinterpolant-class.html).

Land use variables. Land use parameters often serve as surrogates of anthropogenic $PM_{2.5}$ sources; in this study, land use parameters were used to quantify source-oriented emission intensities. The percentage of land cover in a grid of 1 km x 1 km cells covering the entire Northeast U.S. was obtained from the 2011 collection of the National Land Cover Database (NLCD). Important land cover parameters used in the LUR included spaces with high, medium, and low intensity development, developed open spaces, agriculture, grass, deciduous forest, evergreen forest, and mixed forest. Major roads (A1-A3) density was gathered from the StreetMap USA database using the Feature Class Code (A1-A4) classification from the U.S.

Census Bureau Topologically Integrated Geographic Encoding and Referencing (TIGER) system. Annual averaged traffic count for major roads was obtained from the Highway Performance Monitoring System (HMPS) database. The built-in Kernel density algorithm from ArcMap was used to calculate traffic count weighted for major road density within 1 km^2. Population density was calculated within 1 km^2 from the census track database of 2000. A variable indicating the presence of industrial point sources was created by intersecting the locations of large industrial facilities and the corresponding 1 km x 1 km cell in the study domain grid.

Statistical analysis

Emission model. PEIRS is an inventory method that models the dynamics of fine particle fate and transport on a gridded domain of 1 km x 1 km cells. Three central processes are accounted for in the PEIRS model: (1) transported particles from upwind to downwind cells, (2) within-cell emissions, and (3) particle removal by air exchange. The transport process closely depends on air exchange rate (α), which is a measure of the air flow entered or exitted from a fixed space. The volume of this fixed space in our study had a base area of 1 km x 1 km and we used the boundary layer height (PBL) to estimate its height. The flow rate of this fixed volume of air was estimated by the produrct of horizontal wind speed and the cross sectional area (PBL x 1 km) of the air movement. We obtained the air exchange rate with the following formula:

$$\alpha = \frac{Flow\ rate}{Volume\ of\ air} = \frac{ws\ x\ PBL\ x\ 1\ km}{1\ km\ x\ 1\ km\ x\ PBL} \quad (1)$$

Once the air exchange rate was estimated for each grid cell daily, we then used wind direction to locate upwind cells that air masses carrying particles travelled through on the corresponding day. Additionally, temperature is included in the emission model as a surrogate for secondary particle formation. The complete model is formulated as equation 2. Detailed concepts and derivation of the emission model can be found in Tang et al., 2016.

$$C = \sum_{i=1}^{3}(C_{ui} \times Temperature) + \frac{Q}{\alpha \times PBL} \quad (2)$$

where C is the $PM_{2.5}$ concentrations in a downwind cell, C_{ui} is the $PM_{2.5}$ concentration in an upwind cell *I,* and Q is the estimated emission expressed in tons/yr/km^2). This model was fitted separately for each 1 km x 1 km grid cell across the Northeast U.S to obtain Q.

Quantile regression. Emission trends were estimated using a linear quantile regression model with a linear variable indicating the time period (*period*). Quantile regression has the advantage of estimating functional relationships for all portions of the emission distribution (e.g. percentiles) as opposed to the traditional mean estimator (Koenker and Bassett 1978). In addition quantile regression does not require any normality assumptions for variables. Quantile regression provides a more comprehensive analysis of the emission trends, specifically at the higher and lower percentiles in the distribution, where the trends in emissions may be quite different because those percentiles may have different source profiles then those that drive the center of the distribution of emissions. Quantile regression was performed from the 5th percentile to the 95th percentile using the *quantreg* package in R v3.2.2.

Land use regression. Source-specific emissions and their trends were assessed using land use terms as predictors in the regression models. The land use parameters included percent developed spaces with high (*dh*), medium (*dm*), or low (*dl*) intensity, percent developed open spaces (*dop*), percent agricultural space (*arg*), percent deciduous forest (*df*), percent evergreen forest (*ef*), percent mixed forest (*mf*), traffic count weighted for major road density within 1 km^2 (*rd*), and population density within 1 km^2 (*pop*). In addition, we also created an indicator variable (*ind*) identifying the presence of major industrial sources in each 1 km x 1 km cell. Land use terms are surrogates for emissions and their relationship to emissions may change over time as, for example, pollution controls are implemented that impact the sources of those emissions. To test for these trends, we fitted four land use models separately for the four follow-up periods in the study due to the fact that some land use terms were only measured once over time. The four sets of slopes of the land use terms represent the emission intensity of the corresponding period, and the differences over follow-up periods represent their trends. The LUR model was formulated as follows:

$$\begin{aligned} \mathrm{Q} = \beta_{0i} + \beta_{1i}(dh) + \beta_{2i}(dm) + \beta_{3i}(dl) + \beta_{4i}(dop) + \beta_{5i}(\mathrm{arg}) \\ +\beta_{6i}(df) + \beta_{7i}(ef) + \beta_{8i}(mf) + \beta_{9i}(pop) + \beta_{10i}(ind) \end{aligned} \quad (3)$$

where *i* is the study period, Q is the estimated emission, and the other predictors are as defined above. To test the significance of the LU-related emission trends (difference between slopes), we fitted the below model including emission estimates over the 4 study periods:

$$\mathrm{Q} = \beta_0 + \sum_{k=1}^{9} \beta_k(LU_k) + \sum_{k=1}^{9} \beta_k(LU_k \times Period) \quad (3)$$

where k is the k^{th} land use variable included in the LUR model (eq.3) and *Period* is a continuous variable of the period number.

RESULTS AND DISCUSSION

Regional and state-specific trends in $PM_{2.5}$ emissions

Regional $PM_{2.5}$ emission trends in the Northeast U.S. were estimated by comparing averages of year-round, cold season (Nov-Apr), and warm season (May-Oct) emission predictions (Table 6). Across the entire study period (Period 4 vs Period 1), year-round emissions decreased by 18%, which is comparable to but somewhat larger than the 11% decrease reported by the NEI during the period 2002–2011 (EPA 2011). The absolute reduction was more pronounced in the cold season, while the percent decrease was considerably larger in the warm season. The ratio of the regional mean emissions during the cold versus warm season also increased over time. These findings imply that current emission control is likely more effective in the warm season than the cold season. EPA scientists have similarly reported a generally lower effectiveness of emission controls during the winter due to strong weather interference on particle loading in the past decade (EPA 2008).

Table 6. Regional $PM_{2.5}$ Mean Triennial-averaged-emissions over the Four Study Periods

	Mean emission (tons/yr/km^2)		
	All season	**Cold season**	**Warm season**
Period 1 (2002-2004)	18.3	33.9	9.4
Period 2 (2005-2007)	17.4	31.1	7.3
Period 3 (2008-2010)	16.0	23.4	9.6
Period 4 (2011-2013)	15.0	26.4	4.4

	Mean emission changes (tons/yr/km^2 , %)		
	All season	**Cold season**	**Warm season**
Period 2 vs 1	-0.9 (-5%)	-2.8 (-8%)	-2.1 (-22%)
Period 3 vs 2	-1.4 (-8%)	-7.7 (-25%)	+2.3 (+32%)
Period 4 vs 3	-1.0 (-6%)	+3.0 (+13%)	-5.2 (-54%)
Period 4 vs 1	-3.3 (-18%)	-7.5 (-22%)	-5.0 (-54%)

$PM_{2.5}$ emission trends not only varied by season but also by location. Among seven Northeast U.S. states, Connecticut (CT) and Rhode Island (RI) exhibited the largest year-round $PM_{2.5}$ emission reduction during the study period (Table 7, Period 4 vs Period 1). Furthermore, these two formerly non-attainment states exhibited a more than 20 tons/year/km^2 decrease in $PM_{2.5}$ emissions during the cold season, which was almost twice that seen in the other five states. As the atmospheric conditions are less favorable for secondary particle formation during the cold season, the high reduction rates observed in CT and RI during cold season is likely attributable to changes in primary emissions. The significant decrease in $PM_{2.5}$ emissions during the cold season began to manifest in Period 3 (2008-2010), after the non-attainment area designation in 2006 and before the maximum attainment in April 2010. This could be suggesting that the states' control programs for primary $PM_{2.5}$ emissions were sufficiently effective to meet federal requirements. However, a number of factors may also contribute to this reduction. For instance, the demand for home heating oil in the Northeast fell by 43% between 2000 and 2012 and may have led to the declining emission during cold season(Andrews and Perl 2014). Improved insulation, furnaces and fuel switching from oil to gas may all play a role in addition to attainment designations. Furthermore, the observed reduction may be related to the economic recession during 2008 to 2010. People may tune down their thermostats lower to save money and lived in colder houses. This could also explain the increased in emission from period 3 (2008~2010) to period 4 (2011~2013) when the economy started recovering. On the other hand,

the largest long-term reduction in $PM_{2.5}$ emissions during the warm season occurred in Vermont (VT, -9.3 tons/yr/km^2, -73%) followed by New Hampshire (NH, -8.7 tons/yr/km^2, -72%). While drivers for the faster reduction rate in VT and NH during the warm season are unknown, the NEI reported larger percentage decreases in precursor gases (SO_2 and NO_x) than in primary $PM_{2.5}$ emissions in VT and NH during 2002–2011 (Table 8) which implies more rapid reduction in secondary particles than primary sources in less urbanized states.

Table 7. Regional mean $PM_{2.5}$ triennial-averaged-emission changes (tons/yr/km^2, %) by State.

Year-round							
	New York	**Vermont**	**Massachusetts**	**Connecticut**	**New Hampshire**	**Maine**	**Rhode Island**
Period 2 vs 1	-0.8 (-5%)	-2.2 (-13%)	+2.5 (+9%)	-2.7 (-9%)	-1.1 (-6%)	-0.7 (-4%)	+2.0 (+5%)
Period 3 vs 2	-1.8 (-11%)	-1.6 (-11%)	-5.1 (-18%)	-0.4 (-1%)	-1.7 (-11%)	+1.0 (+6%)	-8.5 (-21%)
Period 4 vs 3	+0.7 (+5%)	-1.7 (-12%)	-2.5 (-10%)	-5.8 (-20%)	-1.2 (-8%)	-2.7 (-17%)	-5.7 (-18%)
Period 4 vs 1	-2.0 (-12%)	-5.5 (-32%)	-5.1 (-20%)	-8.8 (-28%)	-4.0 (-23%)	-2.5 (-15%)	-12.2 (-32%)
Warm season							
	New York	**Vermont**	**Massachusetts**	**Connecticut**	**New Hampshire**	**Maine**	**Rhode Island**
Period 2 vs 1	-0.9 (-12%)	-4.5 (-35%)	-1.0 (-9%)	-1.8 (-15%)	-5.2 (-43%)	-2.7 (-30%)	+0.3 (+2%)
Period 3 vs 2	+1.4 (+20%)	+1.8 (+22%)	+2.4 (+23%)	+1.1 (+11%)	+3.6 (+52%)	+4.7 (+72%)	+1.0 (+8%)
Period 4 vs 3	-4.0 (-50%)	-6.6 (-65%)	-6.9 (-54%)	-2.8 (-25%)	-7.2 (-68%)	-6.4 (-57%)	-5.9 (-41%)
Period 4 vs 1	-3.6 (-47%)	-9.3 (-73%)	-5.6 (-49%)	-3.5 (-29%)	-8.7 (-72%)	-4.5 (-48%)	-4.6 (-36%)
Cold season							
	New York	**Vermont**	**Massachusetts**	**Connecticut**	**New Hampshire**	**Maine**	**Rhode Island**
Period 2 vs 1	-3.0 (-10%)	-2.8 (-10%)	+4.4 (+10%)	-4.9 (-9%)	+2.1 (+7%)	-7.4 (-21%)	+3.5 (+5%)
Period 3 vs 2	-6.9 (-24%)	-7.2 (-27%)	-13.3 (-28%)	-8.2 (-16%)	-11.2 (-35%)	-5.3 (-19%)	-22.9 (-34%)
Period 4 vs 3	+6.5 (+30%)	+2.0 (+10%)	-1.5 (-4%)	-11.2 (-26%)	+2.7 (+13%)	+0.3 (+1%)	-7.7 (-17%)
Period 4 vs 1	-3.4 (-11%)	-8.0 (-27%)	-10.4 (-24%)	-24.4 (-43%)	-6.4 (-21%)	-12.3 (-35%)	-27.1 (-42%)

Table 8. NEI 10-year Emission Trends for $PM_{2.5}$,* NO_x, SO_2, VOC, and NH_3 in the Northeast U.S., 2002–2011

Year-round							
	New York	**Vermont**	**Massachusetts**	**Connecticut**	**New Hampshire**	**Maine**	**Rhode Island**
$PM_{2.5}$	-25 ~ -50%	-0 ~ -25%	-25 ~ -50%	-0 ~ -25%	-25 ~ -50%	-25 ~ -50%	+25 ~ +50%
NO_x	-25 ~ -50%	-50 ~ -75%	-50 ~ -75%	-50 ~ -75%	-50 ~ -75%	-25 ~ -50%	-25 ~ -50%
SO_2	-50 ~ -75%	-50 ~ -75%	-75 ~ -100%	-50 ~ -75%	-75 ~ -100%	-50 ~ -75%	-50 ~ -75%
VOC	-25 ~ -50%	-0 ~ -25%	-25 ~ -50%	-25 ~ -50%	-25 ~ -50%	-25 ~ -50%	-0 ~ -25%
NH_3	-0 ~ -25%	-0 ~ -25%	-25 ~ -50%	-25 ~ -50%	-0 ~ -25%	-0 ~ -25%	+50 ~ +75%

* NEI $PM_{2.5}$ consists of primary sources only.

Spatial patterns and trends of $PM_{2.5}$ emissions

One of the advantages of satellite-based emission inventories is ample spatial information that allows visualization of the locations where reduction or growth occurred and consideration of appropriate emission control strategies. For year-round $PM_{2.5}$ emissions (Figure 8), urban areas like the Greater Boston area and New York City exhibited a clear downward trend while rural trends were more difficult to determine. As discussed in Tang et al., 2016, the PEIRS emission estimates in areas in the vicinity to water surfaces has larger uncertainties due to compromised AOD data. We observed potentially problematic emissions near Burlington, VT, due to interference from Lake Champlain. A similar problem occurred in Rochester, Syracuse, and areas bordering the Finger Lakes in New York. Relatively higher fraction of secondary particles involving complex reactions among precursor gases may provide another plausible explanation for the variations in emissions trends in less urbanized areas. In particular, NH_3, NO_x, and VOC emissions were reported to be high in rural areas according to the 2011 NEI report. The NEI also found that significant reductions in NO_x, VOC, and SO_2 have been achieved over time while NH_3 emissions remained fairly constant. The uneven changes in these gas pollutants could modulate emission trends and spatial patterns significantly.

Cold season $PM_{2.5}$ emissions (Figure 9) generally showed similar spatial trends to those of the year-round emissions. In addition to the apparent decrease in emissions in metropolitan areas over time, reduction in important traffic corridors became more discernible during the cold season as well (i.e., Route 90 in the middle of the New York state). This indirectly supports the

EPA inference that periodic method and data improvements used in the NEI are the main reasons for the observed increase in vehicle emissions in 2005 and 2008 NEI. In contrast, warm season $PM_{2.5}$ emissions (Figure 10) were distributed rather uniformly, with miniscule urban versus rural disparity. The onset of emissions reduction did not manifest during the warm season until the last period (2011-2013) measured. The difference between cold and warm season $PM_{2.5}$ emission trends implies a significant association between weather and total emissions and warrants further assessment of emission trends separately by season.

Figure 8. Triennial-averaged emission estimates in northeast U.S. (year-round).

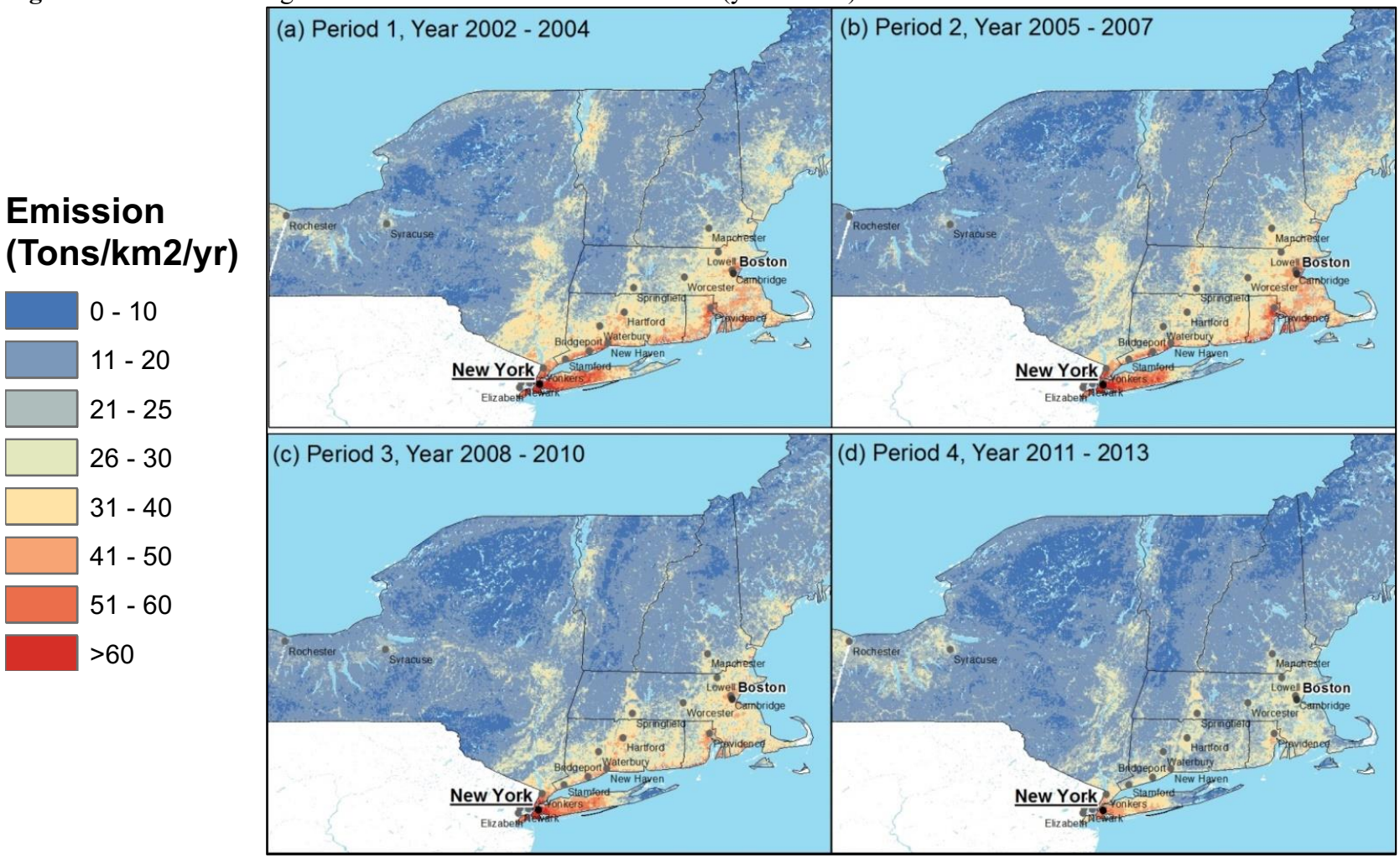


Figure 9. Triennial-averaged emission estimates in northeast U.S. (cold season).

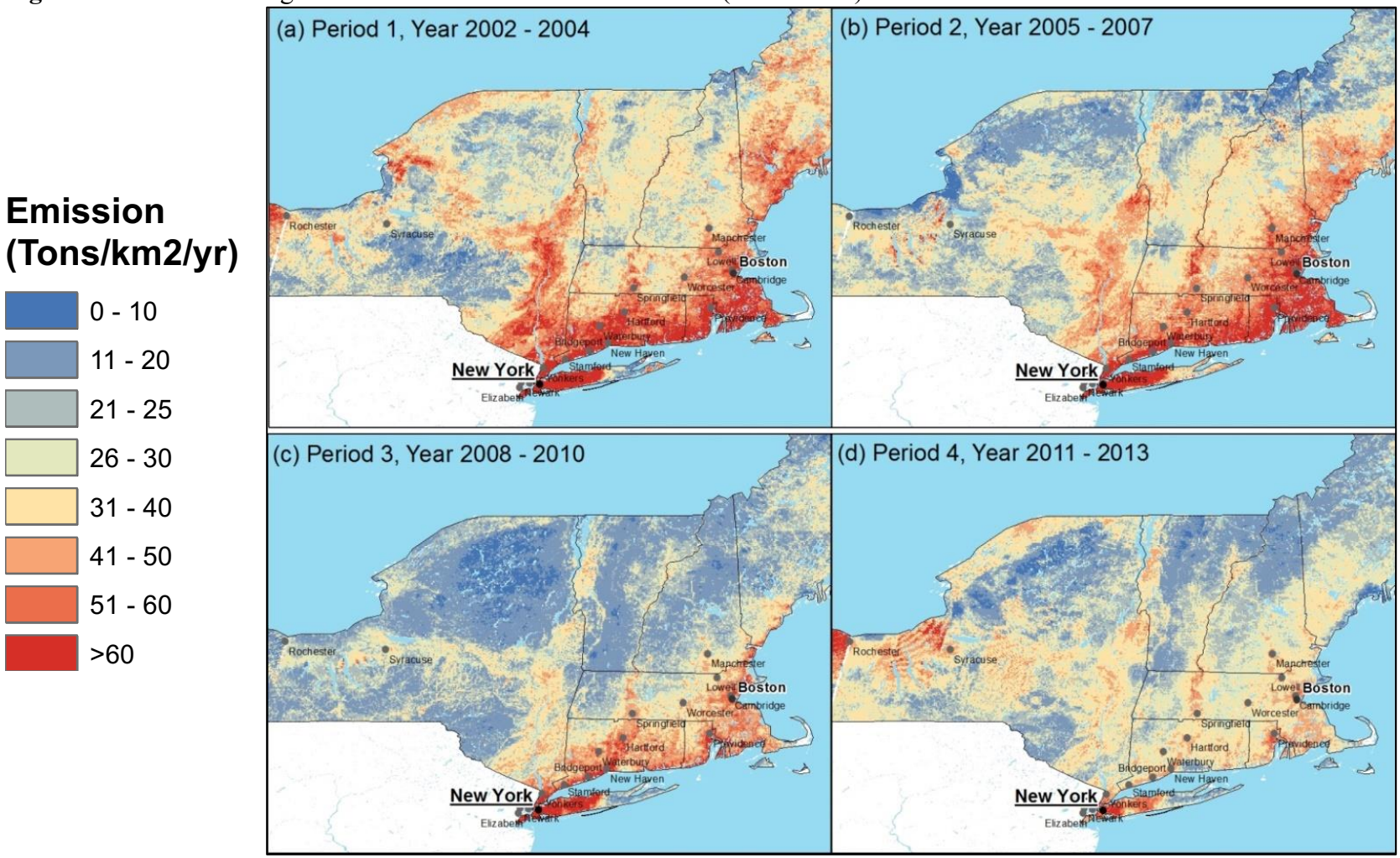
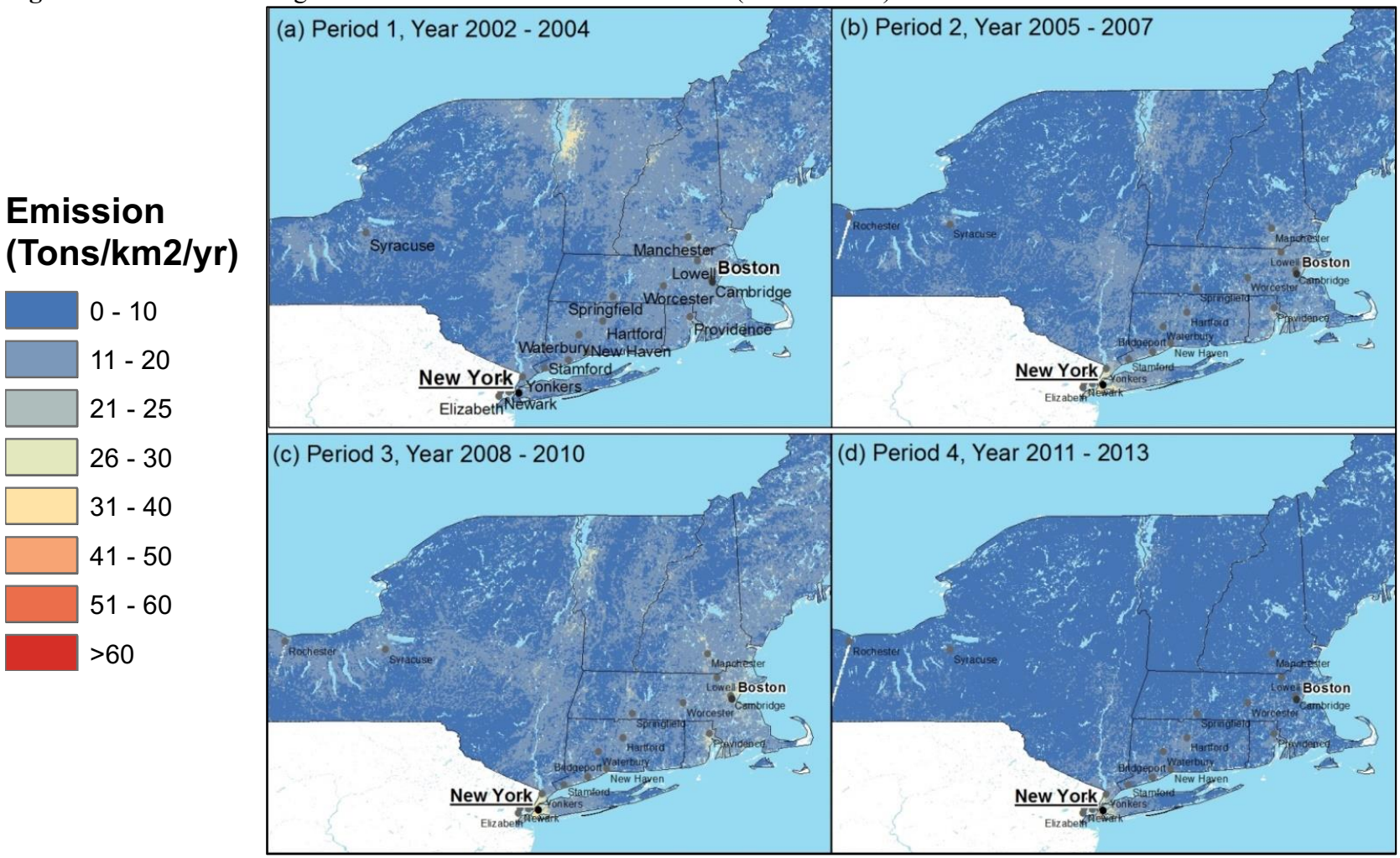


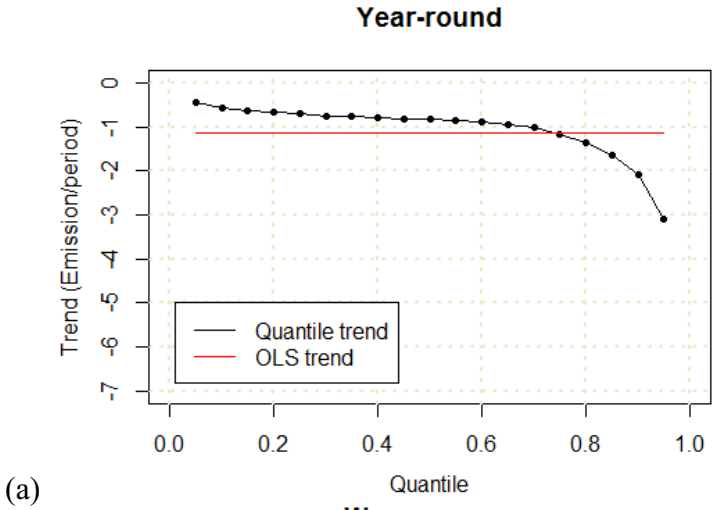
Figure 10. Triennial-averaged emission estimates in northeast U.S. (warm season).



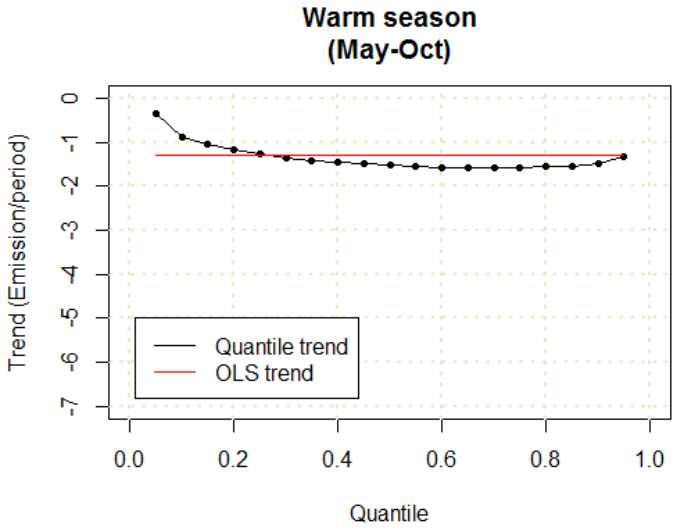
Quantile trends on $PM_{2.5}$ emission sources

To more closely investigate the large spatial variability of $PM_{2.5}$ emissions, we applied quantile regression models on emission inventories year-round (Figure 11a) and in the warm season (Figure 11b) and the cold season (Figure 11c) to quantify trends for a wider range of emission distribution. The quantile trend showed an exponential increase in the rate of reduction above the 80^{th} percentile in year-round $PM_{2.5}$ emissions (Figure 11a). This result is consistent with our qualitative evaluation of the spatial trends (Figure 8) where reduction was most apparent in urban areas (higher quantile) and gradually tapered off in suburban and rural areas (lower quantile). The cold season quantile-specific trend was similar to that of the year-round quantile trend, but with larger deviance from the mean trend overall. This implies that urban-related $PM_{2.5}$ emission sources play an important role year-round but become considerably stronger in cold weather. On the other hand, the warm season quantile regression rate of change was similar throughout the distribution except for a weakening reduction rate below the 20^{th} percentile, which given how much lower emissions were in the warm season, represents areas that are quite clean already. The uniformity suggests little urban-rural contrast during the warm season, and that the warm season reduction is likely more attributable to sources that are not particularly urban-related such as regional sources or secondary formation which are expected to be more pronounced during warm season. This is in agreement with the state-specific trend where less urban states such as VT and NH showed the most reduction in $PM_{2.5}$ emissions during warm season (Table 7).

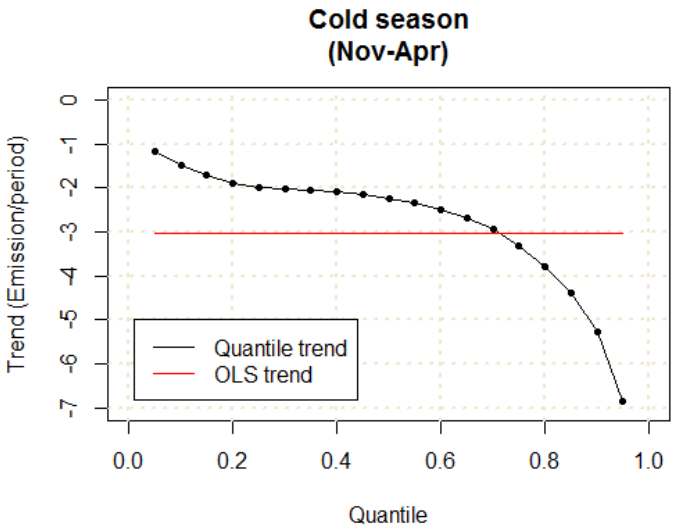
Figure 11. Quantile emission trends (a) Year-round, (b) Warm season, and (c) Cold season.



(a)



(b)



(c)

Trends in $PM_{2.5}$ emissions based on land use regression

To improve $PM_{2.5}$ emission control efficiency, it is critical to identify large contributors by source type. We fitted land use regression models to categorize our emission inventory. Table 9 depicts the land use-specific $PM_{2.5}$ emission intensities over the four study periods. Over the entire study period, substantial reduction in $PM_{2.5}$ emissions was achieved across all land use-related sources. For year-round emissions, developed spaces produced 8.4–27.3 tons/year/km^2 of particles during Period 1, which dropped to 2.5–14.7 tons/year/km^2 in Period 4, translating to a 40%–70% emissions reduction rate. The reduction in developed spaces was even more pronounced in the cold season (Table 10), possibly due to reduced heating demands, specifically heating oil consumption (Andrews and Perl 2014), as the winter in the northeastern part of the U.S. becomes warmer (Hayhoe et al. 2006) and insulation and furnaces have improved efficiency over the past decade. Source-oriented $PM_{2.5}$ emissions also decreased in the warm season in general except for high-intensity developed spaces, which almost tripled from period 1 to 3 and then decreased sharply in period 4(Table 11). High-intensity developed space contains a mixture of land use including commercial, industrial, and residential. The source profile for this geographic setting is complex, and identifying the causes of this emission trend can be difficult and require further research. Regarding the transportation sector, emission from a vehicle that travelled 10,000 miles annually went from an averaged year-round contribution of 1.58 to 0.98 tons/year/km^2 over the past 12 years (Table 9). Traffic emission trends appeared to be consistent year-round with little seasonal variance. Furthermore, emissions from large industrial point sources such as power plants have declined despite the increasing energy demands. Population related emission fell substantially during cold season, however, during warm season population input has little change over the decade –emission fell during the recession and then bounced

back. Finally, for the forest categories we found negative slopes in general which indicates a net loss of particles due to more removal of $PM_{2.5}$ by plants, including interception and absorption (Nowak et al. 2013), than their biogenic emissions or resuspensions. The slopes of the forests, specifically deciduous forest, are more negative during cold season due to fewer or no leaves on the trees and thus less biogenic emissions. The land use regression results show that the removal mechanism of the forests became gradually stronger over the study period. However, since the removal mechanism may act synergistically with the weather, meteorological variations could be a confounding factor for the reduction observed in the forest categories.

Table 9. Land Use-related $PM_{2.5}$ emissions intensity (tons/yr/km^2) year-round in the Northeast U.S., 2002–2013

	Intercept	Deciduous Forest	Mixed Forest	Evergreen Forest	Agriculture /Pasture	Industrial Points
Period 1 (2002-2004)	0.5*	-7.2*	-6.3*	-6.4*	0.1*	1.0*
Period 2 (2005-2007)	-3.7*	-5.5*	-4.5*	-5.7*	0.4*	1.7*
Period 3 (2008-2010)	7.2*	-9.5*	-5.3*	-8.2*	-0.1*	0.8*
Period 4 (2011-2013)	-1.3*	-10.2*	-8.6*	-10.7*	0.1*	0.5*
	Developed Open Space	**Developed High Intensity**	**Developed Medium Intensity**	**Developed Low Intensity**	**Major Road**	**Population**
Period 1 (2002-2004)	27.3*	19.5*	23.4*	8.4*	9.8E-05*	4.05E-04*
Period 2 (2005-2007)	26.7*	23.2*	27.6*	9.2*	8.3E-05*	4.68E-04*
Period 3 (2008-2010)	20.3*	21.4*	20.7*	9.7*	8.4E-05*	4.88E-04*
Period 4 (2011-2013)	14.7*	11.3*	11.9*	2.5*	6.1E-05*	3.35E-04*

* : statistically significant in the trend test.

Unit of Land use variables:

1. % land cover inside 1km x 1km grid --- Deciduous forest, Mixed forest, Evergreen forest, Agriculture/Pasture, Developed open space, Developed high, medium and low intensity.
2. km × no. of vehicles inside 1km x 1km grid --- Major Road
3. no. of person inside 1km x 1km grid --- Population

Table 10. Land Use-related $PM_{2.5}$ emissions intensity during the cold season (Nov-Apr) in the Northeast U.S., 2002–2013

	Intercept	Deciduous Forest	Mixed Forest	Evergreen Forest	Agriculture /Pasture	Industrial Points
Period 1 (2002-2004)	18.8*	-14.3*	-9.7*	-12.6*	-0.9	0.6*
Period 2 (2005-2007)	0.2*	-7.6*	-5.9*	-7.2*	+0.05	2.9*
Period 3 (2008-2010)	14.4*	-11.5*	-8.7*	-12.0*	+0.1	0.9*
Period 4 (2011-2013)	9.3*	-19.4*	-17.9*	-20.0*	+0.05	0.6*
	Developed Open Space	**Developed High Intensity**	**Developed Medium Intensity**	**Developed Low Intensity**	**Major Road**	**Population**
Period 1 (2002-2004)	34.7*	21.3*	32.1*	9.7*	1.7E-04*	6.63E-04*
Period 2 (2005-2007)	38.0*	23.8*	39.2*	17.8*	1.4E-04*	4.78E-04*
Period 3 (2008-2010)	27.2*	23.9*	27.5*	11.4*	1.2E-04*	6.30E-04*
Period 4 (2011-2013)	8.9*	8.8*	5.5*	-0.7*	0.5E-04*	2.72E-04*

* : statistically significant in the trend test.

Table 11. Land Use-related $PM_{2.5}$ emissions intensity during the warm season (May-Oct) in the Northeast U.S., 2002–2013

	Intercept	Deciduous Forest	Mixed Forest	Evergreen Forest	Agriculture /Pasture	Industrial Points
Period 1 (2002-2004)	3.4*	-4.0*	0.5*	-0.7*	0.3*	0.7*
Period 2 (2005-2007)	1.4*	-4.5*	-3.0*	-3.1*	0.4*	0.1*
Period 3 (2008-2010)	6.1*	-7.8*	-1.4*	-4.3*	0.0*	0.8*
Period 4 (2011-2013)	0.6*	-4.5*	-2.8*	-4.5*	0.01*	-0.2*
	Developed Open Space	**Developed High Intensity**	**Developed Medium Intensity**	**Developed Low Intensity**	**Major Road**	**Population**
Period 1 (2002-2004)	7.6*	3.4*	8.0*	4.2*	4.8E-05*	3.03E-04*
Period 2 (2005-2007)	8.8*	9.9*	10.9*	0.5*	1.4E-05*	2.20E-04*
Period 3 (2008-2010)	5.6*	11.8*	7.6*	2.6*	3.4E-07*	1.85E-04*
Period 4 (2011-2013)	6.5*	6.4*	5.2*	-1.2*	2.7E-05*	2.64E-04*

* : statistically significant in the trend test.

CHAPTER 3

PREDICTING INDOOR-OUTDOOR SULFUR RATIOS, INDOOR $PM_{2.5}$ AND BLACK CARBON LEVELS FOR GREATER BOSTON HOUSEHOLDS

Chia Hsi Tang [a]

Brent Coull [b]

Joel Schwartz [a]

Eric Garshick[c]

Stephanie Grady [a]

Petros Koutrakis [a]

[a] Department of Environmental Health, Harvard T.H Chan School of Public Health

[b] Department of Biostatistics, Harvard T.H Chan School of Public Health

[c] Brigham And Women's Hospital

ABSTRACT

Quantifying indoor air pollution levels provides more accurate exposure assessment for epidemiological studies as individuals spend most of their time indoor. Given the high cost of micro-environmental sampling, indoor exposure measurements were scarce both temporally and spatially. On days without indoor measurements, indoor exposure levels are usually estimated using outdoor concentrations. These models require knowledge of the infiltration factor which indicates the fraction of outdoor particles penetrating indoors. In this study, we developed a robust model to predict indoor-to-outdoor sulfur ratio (S_{in}/S_{out}), a common surrogate of the infiltration factor, at 95 residences of patients with Chronic Obstructive Pulmonary Disease (COPD) living in the Greater Boston Area. Subsequently, we incorporated the estimated S_{in}/S_{out} and outdoor monitoring measurements to predict indoor fine particulate matter ($PM_{2.5}$) and black carbon (BC) concentrations. The cross-validated results show that our model adequately predicted indoor-to-outdoor sulfur ratio (Out-of-sample by home and season R^2=0.89) for individual households. Our indoor-to-outdoor ratio estimates reflected behaviors that influence particle filtration rate such as window opening, forced air heating use, and purifier use. The sulfur-adjusted models used to predict indoor $PM_{2.5}$ and BC levels performed quite well (Out of sample correlations: $PM_{2.5}$, R^2 = 0.79, BC, R^2 = 0.61). This suggests that the predicted indoor-to-outdoor sulfur ratio served as a good surrogate for infiltration factor and that indoor exposures could be predicted using outdoor concentrations.

INTRODUCTION

Real-time $PM_{2.5}$ pollution measurements in indoor environments where people spend their time are the most relevant exposure metrics for health risks assessment (Morawska et al. 2001). Nevertheless, indoor particle samples are often small in number due to high cost and effort. As indoor pollution are highly correlated to those outdoors, under normal ventilation condition we can use the readily available outdoor measurements to estimate indoor exposure levels at times when indoor measurements are not available (Sarnat et al. 2002).

Typically, the mass balance equation is used to apportion indoor $PM_{2.5}$ into particles of outdoor and indoor origin. The key parameter is the infiltration factor, which may vary by housing characteristics (e.g., insulation, age of house etc.) or human activities such as open windows, use of fan, purifier and air conditioning use (Van Der Zee et al. 2016, Meng et al. 2009, Morawska et al. 2001). Overtime much effort has been put in estimating the infiltration using household and behavior characteristics (Baxter et al. 2007). However, infiltration rate reported in the literature vary significantly(Diapouli, Chaloulakou, and Koutrakis 2013). Specifically, reliable infiltration models based on questionnaire data are rather few (Wallace and Williams 2005). These exposure error associated with poorly characterized infiltration have found to bias health effects assessment as a consequence (Zeger et al. 2000).

A more quantitative approach in estimating infiltration is to incorporate a tracer element that is (1) predominantly of outdoor origin, (2) may be measured accurately and continuously in both indoor and outdoor environment, (3) is found in relatively high levels to ensure measurement accuracy, and (4) is chemically stable (Wilson, Mage, and Grant 2000). Meeting these criteria, the indoor-to-outdoor ratio of the tracer element can be used as a surrogate of the infiltration rate. Among major constituents of $PM_{2.5}$, sulfur or sulfate is generated mostly from outdoor sources

such as power plants and industrial activities (Long and Sarnat 2004). Moreover, studies have shown that indoor sulfur sources are scarce and that indoor sulfur concentrations are highly correlated with the respective outdoor concentrations (Ozkaynak et al. 1996, Koutrakis, Briggs, and Leaderer 1992). For these reasons, sulfur has been incorporated as a tracer element to assess infiltration rates previously (Wallace and Williams 2005, Sarnat et al. 2002, Gaffin et al. 2016, Geller et al. 2002, Na, Sawant, and Cocker 2004).

Nevertheless, quantifying indoor-to-outdoor sulfur ratio remains a challenging task (Diapouli, Chaloulakou, and Koutrakis 2013). For instance, uncensored or under reported indoor sulfur sources may incorrectly attribute to outdoor sources and bias the estimated infiltration rates(Wallace et al. 2006), even though there are few known indoor sulfur sources (e.g. kerosene heater use, cigarette smoking, and burning sulfur rich fuels). Specifically in the New England area, the frequent use of oil as home heating fuel could be a notable indoor sulfur source. Specifically, the Energy Information Administration reported that 31% of Massachusetts residents use fuel oil as their primary heating fuel for home heating and hot water systems, which is five times higher than the nationwide average of 6% (Energy Information Administration, 2015). The sulfur content limit for heating oil issued by the Massachusetts Department of Environmental Protection in the Regional Haze State Implementation Plan was set at 500 ppm in 2014 which means the sulfur content in heating oil could be much higher in early years. In this regard, the common assumption of zero indoor sulfur sources may be violated in homes using oil as their main heating fuel and subsequently lowered the correlation between the indoor-to-outdoor sulfur ratio and the infiltration rate.

As numerous investigations have linked increased health risks with fine particles (Dominici 2003, Simkhovich, Kleinman, and Kloner 2008, Bell, Ebisu, Peng, Samet, et al. 2009, Brauer et

al. 2012), a reliable estimation of indoor exposure levels may assist the development of effective regulation and control strategies. For this reason, we developed a robust model to estimate indoor-to-outdoor sulfur ratio, the surrogate of the main parameters governing ambient particle infiltration, in dependence of indoor activity predictors. Subsequently, we employed the estimated indoor-to-outdoor sulfur ratio to predict indoor $PM_{2.5}$ and BC concentrations.

MATERIALS AND METHODS

Study Design

The Chronic Obstructive Pulmonary Disease (COPD) Study enrolled 300 participants from the estimated pool of 2,200 patients in the Veterans Administration Boston COPD registry who live within Route 495. We collected weeklong indoor samples in each subject's home at 3 month intervals during the period of 2012 to 2014. We included 328 weeklong indoor samples collected at 95 residences to the analyses. Indoor sulfur, $PM_{2.5}$ and black carbon concentration were collected at least during two seasons. In this study, we define season by month as winter (December-February), spring (March-May), summer (June- August) and fall (September-November). Finally, outdoor concentrations were measured at a central site throughout the study.

Data Collection

Outdoor air pollution. Daily ambient $PM_{2.5}$ samples were collected at the central monitoring supersite located on the roof of the Countway Library of the Harvard Medical School in downtown

Boston throughout the study period. Details on outdoor $PM_{2.5}$ measurements at the supersite can be found in previous study (Kang, Koutrakis, and Suh 2010).

Indoor air pollution. A Harvard School of Public Health (HSPH) Micro-environmental Automated Particle Sampler (APS, figure 12) was placed in subject's home for indoor sampling. The APS includes an inertial impactor that collects $PM_{2.5}$ on Teflon filters at a low flow rate of 1.8 L/min. Teflon filters were first weighed on an electronic microbalance (MT-5 Mettler Toledo, Columbus, OH) prior to and after field measurements. Subsequently, indoor BC concentrations were analyzed by measuring filter blackness of the Teflon filter using a smoke stain reflectometer (model EEL M43D, Diffusion Systems Ltd., United Kingdom). Sulfur concentration in the indoor $PM_{2.5}$ samples was determined using X-Ray Fluorescence (XRF) spectroscopy (model Epsilon 5, PANalytical, The Netherlands) (Agency 2008).

Figure 12. Harvard School of Public Health (HSPH) Micro-environmental Automated Particle Sampler.

Questionnaires. In the initial home visit, a home environmental evaluation was conducted where participants were asked about building age, type of heat fuel, number of air conditioning units and forced air heating system in the residence. During each home visit, participants were asked to report specific activities that could generate indoor air pollution or alter penetration efficiencies of outdoor air pollution in a comprehensive questionnaire. The 7-day recall format has proved to be accurate in elder populations where the focus has been on recalling specific activities to assess energy expenditure. Our specific interest will be in recording time and type of activities at different microenvironments. The following questions were included in this analysis: (1) How many hours did you open the window during sampling session? (2) How many hours did you use an electric space heater during sampling session? (3) Did you use a purifier during sampling session and for how long? Responses to these questions were used examine the impact of human activities on the predicted sulfur ratio for each residence.

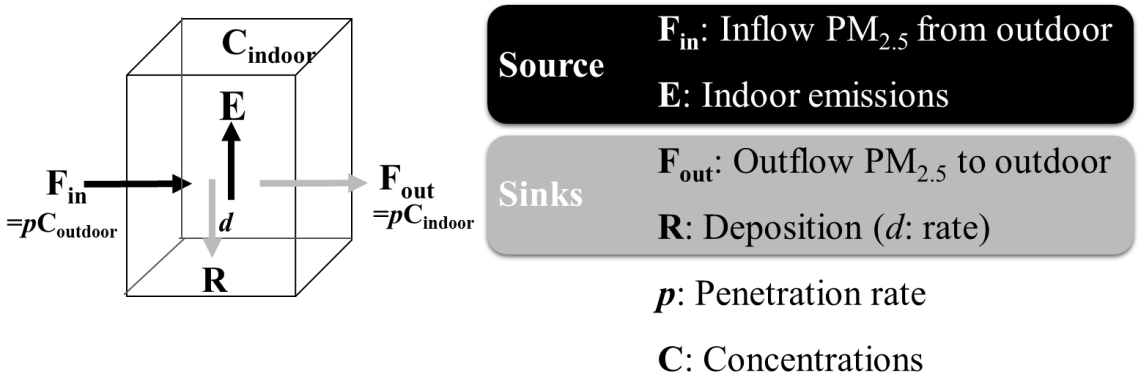
Land use parameters. Land use parameters often serve as surrogates of anthropogenic $PM_{2.5}$ sources; in this study, land use parameters were used to quantify the spatial variability of PM and

BC concentrations outside participating homes. The percentage of urban spaces in a grid of 1 km x 1 km cells covering the study area was obtained from the 2011 collection of the National Land Cover Database (NLCD). Major roads (A1-A3) density was gathered from the StreetMap USA database using the Feature Class Code (A1-A4) classification from the U.S. Census Bureau Topologically Integrated Geographic Encoding and Referencing (TIGER) system. Annual averaged traffic count for major roads was obtained from the Highway Performance Monitoring System (HMPS) database. The built-in Kernel density algorithm (Silverman 1986) from ArcMap was used to calculate traffic count weighted for major road density within 1 km^2. Population density was calculated within 1 km^2 from the census track database of 2000.

Statistical Analysis

In this study, we considered the dynamics of fine particles involving infiltration (or inflow), exfiltration (or outflow), indoor emission and deposition removal (figure 13).

Figure 13. Dynamics of inflow, outflow, emission, removal of fine particles in the indoor environment.



Based on the relationships illustrated in figure 2, we can obtain a simplified mass balance equation as follows:

$$\frac{C_{indoor}(t)}{dt} = F_{in} + E - F_{out} - R \quad (1)$$

Where C_{indoor} (t) is the indoor concentration of particle mass (ug/m^3) at time *t*, F_{in} represents particles infiltrated from outdoor, *E* is the indoor emission, F_{out} is particle exfiltration to outdoor and *R* is the indoor removal by deposition. Both inflow (F_{in}) and outflow (F_{out}) of particles are governed by the infiltration factor (p) and therefore we rearrange eq. 1 into the following form:

$$\frac{C_{indoor}(t)}{dt} = pC_{outdoor} + E - pC_{indoor} - dC_{indoor} \quad (2)$$

Where $C_{outdoor}$ is the particle concentration outside each individual household (ug/m^3), *p* is the infiltration factor and *d* is the deposition rate. Under the assumption where indoor air is well mixed and that transient changes are negligible due to sufficient length of sampling, we can solve the eq. 2 into the following:

$$C_{indoor} = \frac{E}{(p+d)} + pC_{outdoor} \quad (3)$$

From the above derivation, it is apparent that the key parameter allowing us to use outdoor particle concentrations to estimate indoor concentrations is the infiltration factor (*p*). The traditional method to acquire infiltration is to measure the building tightness of each house which could become unrealistic as the number of homes increased in the study. For this reason, the indoor-to-outdoor ratio of tracer elements is often used to quantify infiltration. In this study, we employed sulfur as the tracer element and constructed a sulfur model to estimate the indoor-to-outdoor sulfur ratio for individual households. Subsequently, we incorporated the estimated indoor-to-outdoor sulfur ratio to predict indoor $PM_{2.5}$ and black carbon concentrations based on the mass balance concept (eq. 3).

Sulfur Model. As previously described, in situations where there are no indoor sulfur emissions, we can calculate the indoor-to-outdoor sulfur ratio (eq. 4) directly and use it as a surrogate of the infiltration factor (p):

$$p = \frac{S_{indoor}}{S_{outdoor}} \quad (4)$$

However, given that indoor sulfur emissions may not be captured entirely by questionnaire data and subsequently introducing noise to the ratio calculated from eq. 1, we estimated the infiltration factor with the following model (eq. 5):

$$S_{indoor} = \alpha_{0i} + \beta_0 + (\alpha_{1i} + \beta_1)S_{outdoor} + \sum_{j=1}^{4} + \beta_{j+1}(S_{outdoor} \times I_j) \quad (5)$$

Where S_{indoor} is the sulfur concentration measured indoors, $S_{outdoor}$ is the sulfur concentration outdoor, i represents the homes included in the study and j is the season identifier. I_j is the indicator variable for season j. Since sulfur was not measured outside each individual household and it is considered as a regional pollutant with smaller spatial variation within a city, we used the sulfur concentration measured at the Countway Supersite as a surrogate of the sulfur concentration outside each participating home. Here we included fixed intercept (β_0) and random intercept by home (α_{0i}) to capture potential indoor sulfur emissions such as under reported indoor smoking. A random slope by home of the outdoor sulfur is also included to generate sulfur ratio estimates for each household individually and also the random slopes account for the spatial variability of sulfur that was not reflected by the Countway supersite measurements. In addition, we included an interaction term between outdoor sulfur concentration and season to capture the seasonal variation of the infiltration. Finally, the estimated indoor-to-outdoor sulfur ratio equals to the sum of the estimated fixed (β_1) and random slope (α_{1i}) and the slopes of the season interaction terms (β_j).

Indoor $PM_{2.5}$ and BC Model. Combining the estimated indoor-to-outdoor sulfur ratio (eq. 5) and mass balance (eq. 3) we can estimate indoor concentrations of $PM_{2.5}$ and BC with the following equation:

$$C_{indoor} = \alpha_{0i} + \beta_0 + (\alpha_{1i} + \beta_1) \cdot \frac{S_{indoor}}{S_{outdoor}} \cdot C_{outdoor} \quad (6)$$

Where C_{indoor} (ug/m^3) is the indoor concentration of particle mass, $C_{outdoor}$ (ug/m^3) is the particle concentration outside individual household, and $S_{indoor}/S_{Countway}$ (unit-less) is the predicted indoor-to-outdoor sulfur ratio from the sulfur model (eq. 5).

Nevertheless, unlike sulfur where ambient concentration is generally homogeneously distributed across space, the spatial distribution of $PM_{2.5}$ and BC can vary by location significantly even within a small metropolitan. In other words, the $PM_{2.5}$ and BC concentrations measured at the Countway supersite may not be representative of the particle levels outside participating homes. Since both $PM_{2.5}$ and BC are highly associated with the traffic volume and metropolitan activities, we employed land use variables including major road density, percent urban spaces and the distance between home and Countway supersite to predict outdoor $PM_{2.5}$ and BC concentrations in addition to the monitor measured concentrations. The final indoor model (eq. 7) is therefore formulated as below:

$$C_{indoor} = (\alpha_{0i} + \beta_0) + (\alpha_{1i} + \beta_1) \cdot (\frac{S_{indoor}}{S_{outdoor}}) \cdot C_{countway} \cdot (Road\ density + \%\ Urban + Distance\ btw\ home\ and\ Countway\ supersite) \quad (7)$$

Where $C_{countway}$ (ug/m^3) is the measured concentration at the central monitoring site, *Road density* (1/km) is the traffic density of A1-A3 roads within a 1 km^2 grid where the participating home falls in, %Urban is the percentage of urban spaces within a 1 km^2 grid surrounding the homes. In this model we included a random intercept (α_{0i}) by home (i) to account for indoor $PM_{2.5}$ and BC

emissions. Whereas the random slope (α_{1i}) by home accounts for the spatial variability that was not captured by the land use variables and the sulfur ratio.

Out-of-sample cross validation. We conducted cross validation by home and by season to evaluate the predictive power of the sulfur, indoor $PM_{2.5}$ and BC models described above. For each home, we held out one sample and predicted the held out data using the model fitted with the rest of the samples from that specific home and all data from other homes. We iterate the same procedure until all data is predicted once and examined the R^2 of the cross validated prediction *versus* the observed measurements.

RESULTS AND DISCUSSION

General Characteristics

A total of 328 weeklong sampling sessions were conducted at 95 residences and a nearby central monitoring site (Countway supersite). Table 1 displays the household characteristics and surrounding land use in this analysis. The participated households include a wide range of air conditioning and forced air heating prevalence. Moderate variability was also found in natural ventilation, specifically window opening hours. These factors are expected to influence the infiltration and exfiltration of particles and were found to lower the model predictive power in previous studies (Baxter et al. 2007). The distribution of measured indoor and outdoor concentrations is summarized in Table 2. The measured weekly $PM_{2.5}$ concentrations were generally higher than that of the outdoor measurements suggesting significant indoor sources of particles. The higher BC concentration measured at the central site is likely reflecting the larger amount of traffic emissions and biomass burning in the outdoor environment (Patel et al. 2009,

Saleh et al. 2014). It is therefore important to take into account the spatial variation of outside-home $PM_{2.5}$ and BC which is addressed by including land use parameters, distance to central site and random effects by home in our analysis. More importantly, the indoor-to-outdoor ratio of $PM_{2.5}$ and BC is found to vary considerably. Previous studies have attributed this variation to multiple factors including indoor particle emission, air exchange rate and natural ventilation (Long and Sarnat 2004). Essentially, these findings indicate high relevance of infiltration factors in the modeling process of indoor exposure levels.

Table 12. Distribution of residence characteristics, surrounding land use and questionnaire variables related to indoor air pollution.

Variables	**Mean**	**SD**	**Min**	**Median**	**Max**
%Urban within 500m radius	67%	27%	0%	74%	100%
Major road density within 500m radius (km/km^2)	2.2	1.6	0.0	2.0	11.5
Distance to supersite (km)	27.7	20.2	1.2	24.5	88.2
Building age	62.8	31.9	9.0	51.0	171.0
No. of AC	2.2	0.7	1.0	2.0	4.0
Use of Forced air heating	**Yes**	**No**			
(# subject-week)	66	262			

Questionnaire questions	**Mean**	**SD**	**Min**	**Median**	**Max**
(1) How many hours did you open the window during sampling session?	31.4	57.1	0.0	0.0	168.0
(2) How many hours did you use an electric space heater during sampling session?	5.5	6.8	0.2	3.0	24.0
(3) How many hours did you use a purifier during sampling session?	10.1	33.6	0.0	0.0	168.0
(4) What heat fuel do you use for heating?	**Gas**		**Electric**		**Other/Oil**
(# homes)	44		21		30

Table 13. Distribution of measured indoor and outdoor (at central site) $PM_{2.5}$ concentrations and its BC and Sulfur content.

	Indoor		**Central site**	
	Mean	**SD**	**Mean**	**SD**
$PM_{2.5}$ (ug/m^3)	8.8	6.5	6.5	2.2
BC (ug/m^3)	0.24	0.26	0.58	0.24

Sulfur (ug/m^3)	0.30	0.15	0.45	0.19

Predicted Indoor-to-Outdoor Sulfur Ratio

The mixed effects model (eq. 5) predicting indoor sulfur using central site measurements as outdoor proxy shows excellent predictability with high cross validated R^2 of 0.89, a non-biased slope and a negligible intercept bias when comparing the predicted to the measured indoor sulfur concentration (Table 3).

Table 4 displays the relationship between the predicted sulfur ratio and home characteristics and human activities. The indoor-to-outdoor sulfur ratio increased with higher natural ventilation (window opening) and urban density. On the other hand, purifier, air conditioning, and forced air heating use had opposite effect on the indoor-to-outdoor sulfur ratio (-0.06~-0.28). These findings are in agreement with previous investigations (Cyrys et al. 2004, Bell, Ebisu, Peng, and Dominici 2009). Homes using oil as primary heating fuel have higher sulfur ratio as opposed to those with gas or electric fuel. A plausible reason could be that outdoor furnaces emitting sulfur-rich particles added the spatial variability of sulfur concentration outdoors. Consequently, the indoor-to-outdoor sulfur ratios for these oil using homes were slightly overestimated (0.07) due to using the lower level central site measurement as outdoor sulfur surrogate (Table 4). Nevertheless, the random slopes by home in the indoor $PM_{2.5}$ and BC model (eq. 7) corrected this heating oil associated bias. Finally, the predicted sulfur ratio did not vary by electric space heater use and was borderline related to road density.

Table 14. Cross-validated R^2 and corresponding MSE between observed and predicted indoor Sulfur, $PM_{2.5}$ and BC.

	R^2	**RMSE**	**Intercept Bias**	**Slope Bias**

		(ug/m^3)	(ug/m^3)	
Indoor Sulfur Model	0.89	0.038	-0.02	1.05
Indoor $PM_{2.5}$ Model	0.79	1.695	0.38	0.90
Indoor BC Model	0.76	0.122	0.01	1.01

Table 15. Relationship between the predicted I/O S ratio and household characteristics and behaviors.

	Estimate	Std. Error	T value	P value
Heat fuel				
other/oil vs gas/electric	0.07	0.020	3.165	0.001
Window				
Open vs Close	0.10	0.020	4.340	<0.001
Road density	0.03	0.017	1.790	0.07
Electric space heater	0.003	0.030	0.096	0.92
Forced air heating				
Yes vs No	-0.06	0.030	-1.930	0.05
Land cover				
Urban vs Rural	0.20	0.043	4.477	<0.001
Purifier use				
Yes vs No	-0.28	0.095	-2.884	0.004
AC use				
Yes vs No	-0.06	0.026	-2.396	0.017

Predicted Indoor $PM_{2.5}$ and BC

Using the cross validated indoor-to-outdoor sulfur ratio as the infiltration proxy we predicted indoor $PM_{2.5}$ and BC concentrations. The cross-validated R^2 for indoor $PM_{2.5}$ and BC was 0.79 and 0.76, respectively, showing strong predictive power (Table 3). Model performance was superior to previous indoor models, specifically those without the use of sulfur tracer as infiltration surrogate (Table 5). We relaxed the assumption of none indoor sulfur sources and used fixed and random intercepts (eq. 5) to filter noise attributable to potential indoor sulfur emissions. This approach is likely a reason for our improved model performance. Furthermore,

the number of participating homes in this study is much higher than that of previous studies and therefore enhanced our model predictability with larger sample size.

We found significant indoor sources of fine particles where 54% and 72% of BC and $PM_{2.5}$ was estimated to originate from indoor sources, respectively. While $PM_{2.5}$ concentrations of indoor origin varied across households, this variability was not explained by housing characteristics except a significantly lower $PM_{2.5}$ concentration (-6.3 $\mu g/m^3$) when purifiers were used. The unexplained indoor particles could likely generated by cooking and incense burning which was found to be major sources of indoor particles in previous studies (Habre et al. 2014, Baxter et al. 2007, Pokhrel et al. 2015). On the other hand, we found significant relationship between indoor BC of outdoor origin and traffic density surrounding homes. This is consistent to the discovery that traffic pollutants including BC can easily penetrate buildings in urban areas (Dorizas et al. 2015, Kang, Koutrakis, and Suh 2010, Park et al. 2007, Maynard et al. 2007).

Potentially, our model predictability could be improved by extending the sample size via measuring daily instead of weeklong indoor concentrations or expanding the study area. We did not take into account the influence of chemical transformations, such as changes in gas-to-particle partitioning during the infiltration of volatile organic compounds, nitrate or ammonium (Lunden et al. 2008, Hering et al. 2007) which may also affect the performance of indoor PM and BC models. Furthermore, the large amount of indoor particle generation warrants further investigation regarding the sources and size distributions of these indoor particles as well as their chemical composition and biological effects. Nevertheless, the exposure models developed in this study have superior predictive power compared to the literature and therefore provide more reliable exposure data to study the relative toxicity and biological effects of particles of indoor and outdoor origins.

Table 16. Comparison among previous studies modeling infiltration and indoor particle levels and the current study.

	Current study	(Gaffin et al. 2016)	(Allen et al. 2012)	(Baxter et al. 2007)	(Wallace and Williams 2005)
Study area	95 Boston homes	30 Boston public schools	353 homes in 7 cities across US	39 Boston homes	37 North Carolina homes
Sampling Frequencies	At least 2 sampling sessions	1 sampling per season	Mostly 1 sampling session	2 sampling sessions per home	1 sampling per season
How to model Infiltration	Indoor sulfur (weeklong)	Sulfur ratio (weeklong)	Sulfur ratio (2 week)	GIS & housing characters	Indoor sulfur (weeklong)
Model type	Mixed effects By season & home	Mixed effects By school	Simple Regression	Bayesian model (without S ratio)	Simple Regression
Validation of Sulfur or ratio model	by season & home CV R^2=89.3%	10-fold CV R^2=31%	10-fold CV By city R^2=30~60%	NA	NA
Predictive power on $PM_{2.5}$	by season & home CV R^2=79.1%	10-fold CV R^2=68%	--	R^2=0.37	R^2=0.74

CONCLUSION

This chapter summarizes the main findings in this dissertation and their potential applications and implications.

In Chapter 1, we proposed a new method (the PEIRS approach) to construct spatially resolved emission inventories for fine particulate matter. The predicted emissions in the northeast US are in a reasonable agreement with those reported in the 2008 and the 2011 National Emission Inventory released by the U.S. EPA. The key feature of the PEIRS model is its capability of capturing small-scale intra-urban variations in emissions and formation of secondary particles. Although inherent biases from the satellite data, possibly due to humidity interference and ocean glint, may reduce model accuracy in predicting emissions, generally they manifest in remote areas with little population. Future research should examine whether the allegedly higher emissions are due to AOD measurement artifacts or the presence of secondary formed from biogenic precursors. Another limitation of the PEIRS method is the weather or cloud induced missing AOD retrievals which may contribute to the uncertainties of the subsequent emission estimates. Our interim solution to the problem is to predict multi-year averaged emissions including more than 2 years of daily satellite-derived $PM_{2.5}$ concentrations in the modeling process. Proper imputation or advancement in the retrieval methods of AOD are required to improve the quality of AOD data and subsequently allowing more flexible time scales for the emission estimation.

Overall, we have demonstrated the potential of satellite AOD data to predict $PM_{2.5}$ emissions at a fine spatial scale (1km x 1km). In contrast to conventional methods collecting emission data for known sources, we developed a model, based on the physical properties of

particles, that captures emissions from all sources, including primary and secondary particles, within a specific area. We can take full advantage of the breadth of satellite based remote sensing data to predict emission with a high spatial and temporal resolution at low cost. Ultimately as satellite remote sensing improves, more robust data with better temporal and spatial resolution will become available for predicting emissions not only for particles but for different gaseous air pollutants.

In Chapter 2, as an implication of the PEIRS approach described in Chapter 1, we constructed and analyzed triennial averaged trends of satellite-based $PM_{2.5}$ emission inventories from 2002 to 2013 in the northeastern region of the United States. Our trend analyses suggest that $PM_{2.5}$ emissions in the North East region has declined over the past 12 years, with major reductions achieved for almost all land use-related sources. Results from the quantile regression are in agreement with the spatial trends where most reductions were identified in urban areas or along important traffic corridors, particularly during the cold season. Nevertheless, we found that emission reduced in a much faster rate during the warm seasons as opposed to the cold seasons even though the absolute amount of reduction is more in the later.

Seasonal variations in $PM_{2.5}$ emissions were significantly distinguishable in all trend analyses performed in this study, warranting future efforts to elucidate the underlying mechanisms for these seasonal differences in order to improve the efficacies of emission control strategies. Many studies have demonstrated that temperature variations could lead to human behavioral changes and environmental adjustment that could affect both particle emission or formation rates directly. For instance, High $PM_{2.5}$ concentrations in winter were often found in neighborhoods with high heating demands or a large proportion of oil and wood burning boiler usage (Ross et al. 2013, Noonan et al. 2012), and mobile vehicles were found to produce more

$PM_{2.5}$ emissions as temperature decreased (EPA 2010). The dependence of different $PM_{2.5}$ sources on temperature and other weather parameters is complex and thus previous sensitivity analyses of these relationships often have high uncertainty (Aw and Kleeman 2003, Kleeman 2007, Tai, Mickley, and Jacob 2010). The impact of the variation of weather parameters on particle emissions should be evaluated more carefully with the newly developed spatially- and temporally-resolved emission inventories. This information will enable policy makers to consider impact of weather on $PM_{2.5}$ emission rates and relax the common assumption of a constant meteorological scenario over time. Last but not least, as satellite-based emission inventories were built for other pollutants, relationships between $PM_{2.5}$ emissions and precursor gases emissions such as NO_2, NH_3, and SO_2 and concurrent pollutant such as O_3, could be examined to further our understanding of their complex chemical reactions.

In Chapter 3, This study highlights methodology and data sources that strengthen the estimation of infiltration factor, indoor $PM_{2.5}$ and BC concentrations for individual residences in the Greater Boston Area. Indoor-to-outdoor sulfur ratio, serving as a surrogate of the infiltration factor, was estimated using filter-based indoor and outdoor sulfur concentrations and seasonality within a mixed effects model. The estimated sulfur ratio reflects household characteristics and human activities such as open window, use of purifier, and AC. More importantly, we demonstrated that indoor $PM_{2.5}$ and BC concentrations can be reliably quantified through the combination of outdoor monitoring observations, land use parameters and the sulfur tracer method. Finally, our models show promise in improving indoor $PM_{2.5}$ exposure assessments for future health effects studies and could serve as exposure model framework more generally for other locations.

One of the advantages of our approach estimating indoor-to-outdoor sulfur ratio is to relax the assumption of absent indoor sulfur sources and include a random intercept to account

for potential sulfur emissions inside homes. This way, uncensored indoor sources would not be attributed to ambient particle penetrated indoors and by doing so we lower the risk of generating biased infiltration rates for predicting indoor particle exposure levels. However, the treatment is only adequate given the high demand of oil as heating fuel in the New England region and may not be necessary in other locations where gas or electric fuel are of primary heating source. The random slopes on the other hand would be broadly applicable for it takes into account the variation of infiltration between households. Finally, model performance and the interpretations of our findings may still be limited by the small sample size available in this study. More cost-effective indoor sampling devices or protocols would still greatly enhance our ability to assess indoor exposures more accurately. Nevertheless, the presented models show promise for improving exposure assessment for health effects analyses. We also demonstrated applications of statistical modeling techniques when resources are restricted.

In summary, this dissertation presents methodologies to assess $PM_{2.5}$ exposures and address important knowledge gaps in both the ambient and indoor environment that were not pursued previously. The emission inventory method for ambient particles is broadly applicable to other regions of the world and many different pollutants using various chemical composition satellite data. Whereas the series of models predicting infiltration from outdoor to indoors and indoor particle concentrations can be serve as a framework to determine the relationships between outdoor and indoor air quality more generally. Altogether, this work illustrates novel applications of statistical approaches in quantifying particle exposure from a macro to micro perspective.

BIBLOGRAPHY

Agency, U.S. Environmental Protection. 2008. EPA positive matrix factorization (PMF) 3.0 fundamentals and user guide. edited by U.S. EPA Office of Research and Development. http://www.epa.gov/ttnamti1/files/ambient/inorganic/mthd-3-3.pdf.

Allen, R. W., S. D. Adar, E. Avol, M. Cohen, C. L. Curl, T. Larson, L. J. Liu, L. Sheppard, and J. D. Kaufman. 2012. "Modeling the residential infiltration of outdoor PM(2.5) in the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air)." *Environ Health Perspect* 120 (6):824-30. doi: 10.1289/ehp.1104447.

Andreae, M. O. 2007. "Atmosphere. Aerosols before pollution." *Science* 315 (5808):50-1. doi: 10.1126/science.1136529.

Andrews, A., and L. Perl. 2014. The Northeast Heating Oil Supply, Demand, and Factors Affecting Its Use. edited by Committees of Congress.

Anenberg, S. C., L. W. Horowitz, D. Q. Tong, and J. J. West. 2010. "An estimate of the global burden of anthropogenic ozone and fine particulate matter on premature human mortality using atmospheric modeling." *Environ Health Perspect* 118 (9):1189-95. doi: 10.1289/ehp.0901220.

Aw, Jeremy, and Michael J. Kleeman. 2003. "Evaluating the first-order effect of intraannual temperature variability on urban air pollution." *Journal of Geophysical Research: Atmospheres* 108 (D12):n/a-n/a. doi: 10.1029/2002JD002688.

Baxter, L. K., J. E. Clougherty, C. J. Paciorek, R. J. Wright, and J. I. Levy. 2007. "Predicting residential indoor concentrations of nitrogen dioxide, fine particulate matter, and elemental carbon using questionnaire and geographic information system based data." *Atmos Environ (1994)* 41 (31):6561-6571. doi: 10.1016/j.atmosenv.2007.04.027.

Bell, M. L., K. Ebisu, R. D. Peng, and F. Dominici. 2009. "Adverse health effects of particulate air pollution: modification by air conditioning." *Epidemiology* 20 (5):682-6. doi: 10.1097/EDE.0b013e3181aba749.

Bell, Michelle L, Keita Ebisu, Roger D Peng, Jonathan M Samet, and Francesca Dominici. 2009. "Hospital admissions and chemical composition of fine particle air pollution." *American journal of respiratory and critical care medicine* 179:1115–1120.

Brauer, M., M. Amann, R. T. Burnett, A. Cohen, F. Dentener, M. Ezzati, S. B. Henderson, M. Krzyzanowski, R. V. Martin, R. Van Dingenen, A. van Donkelaar, and G. D. Thurston. 2012. "Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution." *Environ Sci Technol* 46 (2):652-60. doi: 10.1021/es2025752.

Chudnovsky, A., C. Tang, A. Lyapustin, Y. Wang, J. Schwartz, and P. Koutrakis. 2013. "A critical assessment of high-resolution aerosol optical depth retrievals for fine particulate matter predictions." *Atmospheric Chemistry and Physics* 13 (21):10907-10917. doi: 10.5194/acp-13-10907-2013.

Clougherty, J. E., I. Kheirbek, H. M. Eisl, Z. Ross, G. Pezeshki, J. E. Gorczynski, S. Johnson, S. Markowitz, D. Kass, and T. Matte. 2013. "Intra-urban spatial variability in wintertime street-level concentrations of multiple combustion-related air pollutants: the New York City Community Air Survey (NYCCAS)." *J Expo Sci Environ Epidemiol* 23 (3):232-40. doi: 10.1038/jes.2012.125.

Cyrys, J., M. Pitz, W. Bischof, H. E. Wichmann, and J. Heinrich. 2004. "Relationship between indoor and outdoor levels of fine particle mass, particle number concentrations and black smoke under different ventilation conditions." *J Expo Anal Environ Epidemiol* 14 (4):275-83. doi: 10.1038/sj.jea.7500317.

Diapouli, E., A. Chaloulakou, and P. Koutrakis. 2013. "Estimating the concentration of indoor particles of outdoor origin: A review." *Journal of the Air & Waste Management Association* 63 (10):1113-1129. doi: 10.1080/10962247.2013.791649.

Dominici, F. 2003. "Airborne Particulate Matter and Mortality: Timescale Effects in Four US Cities." *American Journal of Epidemiology* 157:1055-1065. doi: 10.1093/aje/kwg087.

Dorizas, P. V., M. N. Assimakopoulos, C. Helmis, and M. Santamouris. 2015. "An integrated evaluation study of the ventilation rate, the exposure and the indoor air quality in naturally ventilated classrooms in the Mediterranean region during spring." *Sci Total Environ* 502:557-70. doi: 10.1016/j.scitotenv.2014.09.060.

Dubovik, O., T. Lapyonok, Y. J. Kaufman, M. Chin, and P. Ginoux. 2008. "Retrieving global aerosol sources from satellites using inverse modeling. Atmospheric Chemistry and Physics." *Atmospheric Chemistry and Physics* 8 (2):250.

Duncan, Bryan N., Ana I. Prados, Lok N. Lamsal, Yang Liu, David G. Streets, Pawan Gupta, Ernest Hilsenrath, Ralph A. Kahn, J. Eric Nielsen, Andreas J. Beyersdorf, Sharon P. Burton, Arlene M. Fiore, Jack Fishman, Daven K. Henze, Chris A. Hostetler, Nickolay A. Krotkov, Pius Lee, Meiyun Lin, Steven Pawson, Gabriele Pfister, Kenneth E. Pickering, R. Bradley Pierce, Yasuko Yoshida, and Luke D. Ziemba. 2014. "Satellite data of atmospheric pollution for U.S. air quality applications: Examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid." *Atmospheric Environment* 94:647-662. doi: 10.1016/j.atmosenv.2014.05.061.

EPA, U.S. 2005a. "2005 National Emissions Inventory Data & Documentation."

EPA, U.S. 2005b. 2005 National Emissions Inventory Data & Documentation.

EPA, U.S. 2008. 2008 National Emissions Inventory:Review, Analysis and Highlights.

EPA, U.S. 2010. MOVES Sensitivity Analysis: The impact of temperature and humidity on emissions.

EPA, U.S. 2011. Profile of the 2011 National Air Emissions Inventory.

Fann, N., A. D. Lamson, S. C. Anenberg, K. Wesson, D. Risley, and B. J. Hubbell. 2012. "Estimating the national public health burden associated with exposure to ambient PM2.5 and ozone." *Risk Anal* 32 (1):81-95. doi: 10.1111/j.1539-6924.2011.01630.x.

Fine, Philip M., Constantinos Sioutas, and Paul A. Solomon. 2008. "Secondary Particulate Matter in the United States: Insights from the Particulate Matter Supersites Program and Related Studies." *Journal of the Air & Waste Management Association* 58 (2):234-253. doi: 10.3155/1047-3289.58.2.234.

Gaffin, J., C. Petty, M. Hauptman, C. Kang, B. Coull, J Schwartz, D. Gold, P. Koutrakis, and W. Phipantanakul. 2016. "Modeling Indoor Particulate Exposure in Inner City School Classrooms." *(in submission)*.

Geller, M., M. Chang, C. Sioutas, B. Ostro, and M. Lipsett. 2002. "Indoor outdoor relationship and chemical composition of fine and coarse particles in the southern california deserts." *Atmos Environ (1994)* (36):1099-1110.

Gupta, Pawan, and Sundar A. Christopher. 2009. "Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach." *Journal of Geophysical Research* 114 (D20). doi: 10.1029/2008jd011497.

Habre, R., B. Coull, E. Moshier, J. Godbold, A. Grunin, A. Nath, W. Castro, N. Schachter, A. Rohr, M. Kattan, J. Spengler, and P. Koutrakis. 2014. "Sources of indoor air pollution in New York City residences of asthmatic children." *J Expo Sci Environ Epidemiol* 24 (3):269-78. doi: 10.1038/jes.2013.74.

Hayhoe, Katharine, Cameron P. Wake, Thomas G. Huntington, Lifeng Luo, Mark D. Schwartz, Justin Sheffield, Eric Wood, Bruce Anderson, James Bradbury, Art DeGaetano, Tara J. Troy, and David Wolfe. 2006. "Past and future changes in climate and hydrological indicators in the US Northeast." *Climate Dynamics* 28 (4):381-407. doi: 10.1007/s00382-006-0187-8.

Hering, S. V., M. M. Lunden, T. L. Thatcher, T. W. Kirchstetter, and N. J. Brown. 2007. "Using Regional Data and Building Leakage to Assess Indoor Concentrations of Particles of Outdoor Origin." *Aerosol Science and Technology* 41 (7):639-654. doi: 10.1080/02786820701368026.

Hoff, Raymond, and Sundar Christopher. 2009. "Remote Sensing of Particulate Pollution from Space: Have We Reached the Promised Land?" *Journal of the Air & Waste Management Association* 59 (6):645-675. doi: 10.3155/1047-3289.59.6.645.

Hooyberghs, J., C. Mensink, G. Dumont, F. Fierens, and O. Brasseur. 2005. "A neural network forecast for daily average PM concentrations in Belgium." *Atmospheric Environment* 39 (18):3279-3289. doi: 10.1016/j.atmosenv.2005.01.050.

Hu, X., L. A. Waller, A. Lyapustin, Y. Wang, and Y. Liu. 2013. "10 yr spatial and temporal trends of PM_{2.5} concentrations in the southeastern US estimated using high-resolution satellite data." *Atmospheric Chemistry and Physics Discussions* 13 (10):25617-25648. doi: 10.5194/acpd-13-25617-2013.

Hu, Xuefei, Lance A. Waller, Alexei Lyapustin, Yujie Wang, Mohammad Z. Al-Hamdan, William L. Crosson, Maurice G. Estes, Sue M. Estes, Dale A. Quattrochi, Sweta Jinnagara Puttaswamy, and Yang Liu. 2014. "Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a

two-stage model." *Remote Sensing of Environment* 140:220-232. doi: 10.1016/j.rse.2013.08.032.

Huneeus, N., F. Chevallier, and O. Boucher. 2012. "Estimating aerosol emissions by assimilating observed aerosol optical depth in a global aerosol model." *Atmospheric Chemistry and Physics* 12 (10):4585-4606. doi: 10.5194/acp-12-4585-2012.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, Roy Jenne, and Dennis Joseph. 1996. "The NCEP/NCAR 40-Year Reanalysis Project." *Bulletin of the American Meteorological Society* 77:437-471. doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kanakidou, Maria, Kostas Tsigaridis, Frank J. Dentener, and Paul J. Crutzen. 2000. "Human-activity-enhanced formation of organic aerosols by biogenic hydrocarbon oxidation." *Journal of Geophysical Research* 105 (D7):9243. doi: 10.1029/1999jd901148.

Kang, C. M., P. Koutrakis, and H. H. Suh. 2010. "Hourly measurements of fine particulate sulfate and carbon aerosols at the Harvard-U.S. Environmental Protection Agency Supersite in Boston." *J Air Waste Manag Assoc* 60 (11):1327-34.

Kleeman, Michael J. 2007. "A preliminary assessment of the sensitivity of air quality in California to global change." *Climatic Change* 87 (1):273-292. doi: 10.1007/s10584-007-9351-3.

Kloog, Itai, Alexandra A. Chudnovsky, Allan C. Just, Francesco Nordio, Petros Koutrakis, Brent A. Coull, Alexei Lyapustin, Yujie Wang, and Joel Schwartz. 2014. "A new hybrid spatio-temporal model for estimating daily multi-year PM2.5 concentrations across northeastern

USA using high resolution aerosol optical depth data." *Atmospheric Environment* 95:581-590. doi: 10.1016/j.atmosenv.2014.07.014.

Kloog, Itai, Petros Koutrakis, Brent A. Coull, Hyung Joo Lee, and Joel Schwartz. 2011. "Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements." *Atmospheric Environment* 45 (35):6267-6275. doi: 10.1016/j.atmosenv.2011.08.066.

Koenker, R., and G. Bassett. 1978. "Regression Quantiles." *Econometrica* 46 (1):33. doi: 10.2307/1913643.

Koutrakis, Petros, Susan LK Briggs, and Brian P Leaderer. 1992. "Source apportionment of indoor aerosols in Suffolk and Onondaga Counties, New York." *Environmental Science & Technology* 26 (3):521-527.

Lee, H. J., B. A. Coull, M. L. Bell, and P. Koutrakis. 2012. "Use of satellite-based aerosol optical depth and spatial clustering to predict ambient PM2.5 concentrations." *Environ Res* 118:8-15. doi: 10.1016/j.envres.2012.06.011.

Lepeule, Johanna, Francine Laden, Douglas W Dockery, and Joel David Schwartz. 2012. "Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009."

Levy, J. I., D. Diez, Y. Dou, C. D. Barr, and F. Dominici. 2012. "A meta-analysis and multisite time-series analysis of the differential toxicity of major fine particulate matter constituents." *Am J Epidemiol* 175 (11):1091-9. doi: 10.1093/aje/kwr457.

Liu, Y., C. J. Paciorek, and P. Koutrakis. 2009. "Estimating regional spatial and temporal variability of PM(2.5) concentrations using satellite data, meteorology, and land use information." *Environ Health Perspect* 117 (6):886-92. doi: 10.1289/ehp.0800123.

Long, Christopher M., and Jeremy A. Sarnat. 2004. "Indoor-Outdoor Relationships and Infiltration Behavior of Elemental Components of Outdoor PM2.5for Boston-Area Homes." *Aerosol Science and Technology* 38 (sup2):91-104. doi: 10.1080/027868290502281.

Lunden, Melissa M., Thomas W. Kirchstetter, Tracy L. Thatcher, Susanne V. Hering, and Nancy J. Brown. 2008. "Factors affecting the indoor concentrations of carbonaceous aerosols of outdoor origin." *Atmospheric Environment* 42 (22):5660-5671. doi: http://dx.doi.org/10.1016/j.atmosenv.2008.03.017.

Lyapustin, A., Y. Wang, I. Laszlo, R. Kahn, S. Korkin, L. Remer, R. Levy, and J. S. Reid. 2011. "Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm." *Journal of Geophysical Research* 116 (D3). doi: 10.1029/2010jd014986.

Martin, Randall V. 2008. "Satellite remote sensing of surface air quality." *Atmospheric Environment* 42 (34):7823-7843. doi: http://dx.doi.org/10.1016/j.atmosenv.2008.07.018.

Maynard, D., B. A. Coull, A. Gryparis, and J. Schwartz. 2007. "Mortality risk associated with short-term exposure to traffic particles and sulfates." *Environ Health Perspect* 115 (5):751-5. doi: 10.1289/ehp.9537.

Meng, Q. Y., D. Spector, S. Colome, and B. Turpin. 2009. "Determinants of Indoor and Personal Exposure to PM(2.5) of Indoor and Outdoor Origin during the RIOPA Study." *Atmos Environ (1994)* 43 (36):5750-5758. doi: 10.1016/j.atmosenv.2009.07.066.

Mesinger, Fedor, Geoff DiMego, Eugenia Kalnay, Kenneth Mitchell, Perry C. Shafran, Wesley Ebisuzaki, Dušan Jović, Jack Woollen, Eric Rogers, Ernesto H. Berbery, Michael B. Ek, Yun Fan, Robert Grumbine, Wayne Higgins, Hong Li, Ying Lin, Geoff Manikin, David

Parrish, and Wei Shi. 2006. "North American Regional Reanalysis." *Bulletin of the American Meteorological Society* 87 (3):343-360. doi: 10.1175/bams-87-3-343.

Morawska, L., C. He, J. Hitchins, D. Gilbert, and S. Parappukkaran. 2001. "The relationship between indoor and outdoor airborne particles in the residential environment." *Atmos Environ (1994)* (35):3463-3473.

Na, Kwangsam, Aniket A. Sawant, and David R. Cocker. 2004. "Trace elements in fine particulate matter within a community in western Riverside County, CA: focus on residential sites and a local high school." *Atmospheric Environment* 38 (18):2867-2877. doi: 10.1016/j.atmosenv.2004.02.022.

Noonan, C. W., W. Navidi, L. Sheppard, C. P. Palmer, M. Bergauff, K. Hooper, and T. J. Ward. 2012. "Residential indoor PM2.5 in wood stove homes: follow-up of the Libby changeout program." *Indoor Air* 22 (6):492-500. doi: 10.1111/j.1600-0668.2012.00789.x.

Nowak, D. J., S. Hirabayashi, A. Bodine, and R. Hoehn. 2013. "Modeled PM2.5 removal by trees in ten U.S. cities and associated health effects." *Environ Pollut* 178:395-402. doi: 10.1016/j.envpol.2013.03.050.

Ozkaynak, H., J. Xue, J. Spengler, L. Wallace, E. Pellizzari, and P. Jenkins. 1996. "Personal exposure to airborne particles and metals: Results from the particle team study in Riverside, California." *Journal of Exposure Analysis and Environmental Epidemiology* 6 (1):57-78.

Park, S. K., M. S. O'Neill, B. J. Stunder, P. S. Vokonas, D. Sparrow, P. Koutrakis, and J. Schwartz. 2007. "Source location of air pollution and cardiac autonomic function: trajectory cluster analysis for exposure assessment." *J Expo Sci Environ Epidemiol* 17 (5):488-97. doi: 7500552 [pii]

10.1038/sj.jes.7500552.

Patel, Molini M., Steven N. Chillrud, Juan C. Correa, Marian Feinberg, Yair Hazi, Deepti Kc, Swati Prakash, James M. Ross, Diane Levy, and Patrick L. Kinney. 2009. "Spatial and Temporal Variations in Traffic-related Particulate Matter at New York City High Schools." *Atmospheric environment (Oxford, England : 1994)* 43 (32):4975-4981. doi: 10.1016/j.atmosenv.2009.07.004.

Peng, Roger D., Michelle L. Bell, Alison S. Geyh, Aidan McDermott, Scott L. Zeger, Jonathan M. Samet, and Francesca Dominici. 2009. "Emergency Admissions for Cardiovascular and Respiratory Diseases and the Chemical Composition of Fine Particle Air Pollution." *Environmental Health Perspectives* 117:957-963. doi: 10.1289/ehp.0800185.

Pokhrel, Amod K., Michael N. Bates, Jiwan Acharya, Palle Valentiner-Branth, Ram K. Chandyo, Prakash S. Shrestha, Anil K. Raut, and Kirk R. Smith. 2015. "PM2.5 in household kitchens of Bhaktapur, Nepal, using four different cooking fuels." *Atmospheric Environment* 113:159-168. doi: 10.1016/j.atmosenv.2015.04.060.

Pope, C. A., 3rd, R. T. Burnett, M. J. Thun, E. E. Calle, D. Krewski, K. Ito, and G. D. Thurston. 2002. "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution." *Jama* 287 (9):1132-41.

Ren, Cizao, Gail M Williams, and Shilu Tong. 2006. "Does Particulate Matter Modify the Association between Temperature and Cardiorespiratory Diseases?" *Environmental Health Perspectives*. doi: 10.1289/ehp.9266.

Ross, Z., K. Ito, S. Johnson, M. Yee, G. Pezeshki, J. E. Clougherty, D. Savitz , and T. Matte. 2013. "Spatial and temporal estimation of air pollutants in New York City: exposure assignment for use in a birth outcomes study."

Saleh, Rawad, Ellis S. Robinson, Daniel S. Tkacik, Adam T. Ahern, Shang Liu, Allison C. Aiken, Ryan C. Sullivan, Albert A. Presto, Manvendra K. Dubey, Robert J. Yokelson, Neil M. Donahue, and Allen L. Robinson. 2014. "Brownness of organics in aerosols from biomass burning linked to their black carbon content." *Nature Geosci* 7 (9):647-650. doi: 10.1038/ngeo2220

http://www.nature.com/ngeo/journal/v7/n9/abs/ngeo2220.html#supplementary-information.

Sarnat, A., C. Long, P. Koutrakis, B. Coull, J. Schwartz, and H. Suh. 2002. "Using Sulfur as a Tracer of Outdoor Fine Particulate Matter." *Environmental Science & Technology* 36 (24):5305-5314. doi: 10.1021/es025796b.

Schwartz, J. 2001. "Is there harvesting in the association of airborne particles with daily deaths and hospital admissions?" *Epidemiology (Cambridge, Mass.)* 12:55-61.

Schwartz, J., and D. W. Dockery. 1992. "Increased mortality in Philadelphia associated with daily air pollution concentrations." *The American review of respiratory disease* 145 (3):600-4.

Schwartz, Joel, Francine Laden, and Antonella Zanobetti. 2002. "The concentration-response relation between PM(2.5) and daily deaths." *Environmental health perspectives* 110:1025-1029.

Shi, L., A. Zanobetti, I. Kloog, B. A. Coull, P. Koutrakis, S. J. Melly, and J. D. Schwartz. 2016. "Low-Concentration PM2.5 and Mortality: Estimating Acute and Chronic Effects in a Population-Based Study." *Environ Health Perspect* 124 (1):46-52. doi: 10.1289/ehp.1409111.

Simkhovich, Boris Z., Michael T. Kleinman, and Robert A. Kloner. 2008. "Air Pollution and Cardiovascular Injury." *Journal of the American College of Cardiology* 52:719-726. doi: 10.1016/j.jacc.2008.05.029.

Spengler, J.D., J.M. Samet, and J.F. McCarthy. 2001. *Indoor air quality handbook*: McGraw-Hill.

Spracklen, D. V., K. S. Carslaw, M. Kulmala, V. M. Kerminen, G. W. Mann, and S. L. Sihto. 2006. "The contribution of boundary layer nucleation events to total particle concentrations on regional and global scales." *Atmospheric Chemistry and Physics* 6 (12):5631-5648. doi: 10.5194/acp-6-5631-2006.

Streets, David G., Timothy Canty, Gregory R. Carmichael, Benjamin de Foy, Russell R. Dickerson, Bryan N. Duncan, David P. Edwards, John A. Haynes, Daven K. Henze, Marc R. Houyoux, Daniel J. Jacob, Nickolay A. Krotkov, Lok N. Lamsal, Yang Liu, Zifeng Lu, Randall V. Martin, Gabriele G. Pfister, Robert W. Pinder, Ross J. Salawitch, and Kevin J. Wecht. 2013. "Emissions estimation from satellite retrievals: A review of current capability." *Atmospheric Environment* 77:1011-1042. doi: 10.1016/j.atmosenv.2013.05.051.

Tai, Amos P. K., Loretta J. Mickley, and Daniel J. Jacob. 2010. "Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change." *Atmospheric Environment* 44 (32):3976-3984. doi: 10.1016/j.atmosenv.2010.06.060.

Tang, C., B. A. Coull, J Schwartz, A. Lyapustin, Q Di, and P. Koutrakis. 2016. "Developing Particle Emission Inventories Using Remote Sensing (PEIRS)." *Journal of the Air & Waste Management Association* (in review).

Turner, Michelle C, Daniel Krewski, C Arden Pope, 3rd, Yue Chen, Susan M Gapstur, and Michael J Thun. 2011. "Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers." *American journal of respiratory and critical care medicine* 184:1374-1381. doi: 10.1164/rccm.201106-1011OC.

Van Der Zee, S. C., M. Strak, M. B. Dijkema, B. Brunekreef, and N. A. Janssen. 2016. "The impact of particle filtration on indoor air quality in a classroom near a highway." *Indoor Air*. doi: 10.1111/ina.12308.

Vehkamaki, H., and I. Riipinen. 2012. "Thermodynamics and kinetics of atmospheric aerosol particle formation and growth." *Chem Soc Rev* 41 (15):5160-73. doi: 10.1039/c2cs00002d.

Wallace, L., and R Williams. 2005. "Use of personal indoor outdoor sulfur concentrations to estimate the infiltration factor and outdoor exposure factor for individual homes and persons." *Environ. Sci. Technol.* 39:1707-1714.

Wallace, L., R. Williams, J. Suggs, and P. Jones. 2006. "Estimating contributions of outdoor fine particles to indoor concentrations and personal exposures: effects of household characteristics and personal activities." *U.S Environmental Protection Agency*.

Wilson, W. E., D. T. Mage, and L. D. Grant. 2000. "Estimating separately personal exposure to ambient and nonambient particulate matter for epidemiology and risk assessment: Why and how." *Journal of the Air and Waste Management Association* 50 (7):1167-1183. doi: 10.1080/10473289.2000.10464164.

Zanobetti, Antonella, and Joel Schwartz. 2006. "Air pollution and emergency admissions in Boston, MA." *Journal of epidemiology and community health* 60:890-895. doi: 10.1136/jech.2005.039834.

Zeger, S., D. Thomas, F. Dominici, J. Samet, J Schwartz, D. Dockery, and A. Cohen. 2000. "Exposure measurement error in time series studies of air pollution: concepts and consequences." *Environmental Health Perspectives* 108:417-426.

Zhang, J. H., F. M. Yao, C. Liu, L. M. Yang, and V. K. Boken. 2011. "Detection, emission estimation and risk prediction of forest fires in China using satellite sensors and simulation models in the past three decades--an overview." *Int J Environ Res Public Health* 8 (8):3156-78. doi: 10.3390/ijerph8083156.