University of Cincinnati

Date: 4/24/2013

I, Harikishan C Perugu , hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Civil Engineering.

It is entitled:

Integrating Advanced Truck Models into Mobile Source PM2.5 Air Quality Modeling

Student's name: **Harikishan C Perugu**

This work and its defense approved by:

Committee chair: Heng Wei, Ph.D.

Committee member: Hazem Elzarka, Ph.D.

Committee member: Mingming Lu, Ph.D.

Committee member: Ala Tabiei, Ph.D.

3557

Integrating Advanced Truck Models into Mobile Source $PM_{2.5}$ Air Quality Modeling

A dissertation submitted to the
Division of Research and Advanced Studies of the University of Cincinnati
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in the School of Advanced Structures of
College of Engineering and Applied Science
April 2013

By
Harikishan Perugu

M.S., University of Cincinnati, Ohio, 2007

Committee

Dr. Heng Wei, Chair
Dr. Hazem Elzarka
Dr. Mingming Lu
Dr. Ala Tabiei

Abstract

The U.S. Environmental Protection Agency is concerned about fine particulate matter (also called as $PM_{2.5}$ as the average particle size is less than 2.5 μm) pollution and its ill effects on public health. About 80 percent of the mobile-source $PM_{2.5}$ emissions are released into the urban atmosphere through combustion of diesel fuel by trucks and are composed of road dust, smoke, and liquid droplets. To estimate the regional or local air quality impact of $PM_{2.5}$ emissions and also to predict future $PM_{2.5}$ concentrations, we often utilize atmospheric dispersion models. Application of such sophisticated dispersion models with finer details can provide us the comprehensive understanding of the air quality problem, including the quantitative effect of pollution sources. However, in the current practice the detailed truck specific pollution estimation is not easily possible due to unavailability of a modeling methodology with applied supporting data to predict the link-level hourly truck activity and corresponding emission inventory.

In the first part of this dissertation, we have proposed a methodology for estimating the disaggregated link-level hourly truck activity based on advanced statistics in light of the AERMOD based dispersion/pollution modeling process. This new proposed truck model consists of following sub models: (a) The **S**patial **R**egression and **O**ptimization based **T**ruck-demand (**SROT**) model is developed to predict truck travel demand matrices using the spatial regression model-output truck volumes at control locations in the study area. (b) The hourly distribution factor model to convert daily truck volumes to hourly truck volumes (c) The Highway Capacity Manual (HCM) based highway assignment model for assigning the hourly truck travel demand matrices. In the second part of dissertation, we have utilized the link-level hourly truck activity

to predict the typical 24-hour and maximum 1-hr $PM_{2.5}$ pollution in urban atmosphere. In this AERMOD based dispersion/pollution modeling process, the gridded hourly emission inventories are estimated based on bottom-up approach using link-level hourly truck activity and emission factors from MOVES model. The proposed framework is tested using the observed $PM_{2.5}$ concentrations for four different seasonal weekdays in the analysis year 2010 and these observations are collected at one of the air quality monitoring stations located within Cincinnati urban area. The comparison with default results has revealed that the proposed models anticipate higher $PM_{2.5}$ emission contribution from the heavy duty trucks.

The innovation of the current research will be reflective of the following aspects: (a) An enhanced comprehensive truck-related $PM_{2.5}$ pollution modeling approach and also consistent estimation of heavy-duty trucks apportionment in urban air quality (b) More reliable estimation of spatial and temporal truck activity which takes care of peak hour congestion through application of advanced modeling techniques (c) The gridded emission inventory is better estimated as detailed truck activity and emission rates are used as part of the bottom-up approach (d) Better ground-truth prediction of $PM_{2.5}$ hot-spots in the modeling area (e) A transferable methodology that can be useful in other regions in the Unites States.

Acknowledgements

My deepest gratitude is to my advisor, Dr. Heng Wei. I have been incredibly fortunate to have an advisor who gave me the freedom to explore on my own and at the same time provide inspirable guidance to be on right path. He has helped me out to grow both professionally and personally; supported me in times of hardship; motivated me to do my best always. During my graduate study at UC, he did not only lead me to jump over big hurdles in creating and accomplishing the research tasks, but also helped me to achieve more additional credits to my resume through advising me in wining several paper competitions. My utmost respect, greatest admiration, and sincere thanks to you, Dr. Wei.

I would also like to thank Dr. Hazem Elzarka, Dr. Mingming Lu and Dr. Ala Tabiei for accepting to be part of my dissertation committee and providing invaluable advice in my research. A lot of thanks go to the University Research Council for its support on my preliminary study. Further, the student paper awards from the Citi Labs Inc. and the Great Lakes ITE also immensely motivated to make a good progress in this research. I am very thankful to my fellow research student Yao Zhuo for his enthusiastic discussions on my research ideas and for being such a good friend. I would also like to express my sincere gratitude towards my former supervisor Mr. Andrew Rohne, Manger, Transportation Modeling, OKI Regional Council of Governments and my present supervisor Mr. Vernon Hughes, Atmospheric Modeling and Support, California Air Resources Board, for their professional support and encouragement.

Last but not the least; I would like to thank my parents, my in-laws and most importantly my loving wife, Hasmitha Paturu for being there always with love and understanding.

Table of Contents

Table of Tables

Table of Figures

Nomenclature

VMT	Vehicle Miles Traveled
TDM	Travel Demand Model
HPMS	Highway Performance Management Systems
AQS	Air Quality System by US-EPA
SIP	State Implementation Plan
HCM	Highway Capacity Manual
BPR	Bureau of Public Roads
AERMOD	American Environmental Protection Agency Regulatory Model
MOVES	Mobile Vehicle Emissions Simulator
ACS	American Community Survey
FAF	Freight Analysis Framework
QRFM	Quick Response Freight Manual
CO	Carbon Monoxide
NOx	Nitrogen Oxides
PSD	Prevention of Significant Deterioration
Panel Data	A dataset containing observations for multiple years and multiple variables
Contiguity	Neighboring locations based on certain pattern
PUMS	Public Usable Micro data Samples
AERMAP	Terrain processor of AERMOD model
Cross Section Data	A dataset observations for multiple variables
ArcGIS	Geographical Information System Software from ESRI

Shapefile	Arc GIS's type of file which contains geographical feature data
Geo-database	The database which stores records with corresponding spatial features
FAZ	Freight Analysis Zone
Ppm	Parts per million

Chapter 1: Introduction

1.1 Problem Statement

The U.S. Environmental Protection Agency is concerned about fine particulate matter (also called as $PM_{2.5}$ as the average particle size is less than 2.5 μm) pollution and its ill effects on public health (US-EPA 2011, US-EPA 2012). The primary sources of $PM_{2.5}$ pollution in the urban areas are on-road mobile sources. About 80 percent of these emissions are released into the atmosphere through the combustion of diesel fuel by trucks and they are composed of road dust, smoke, and liquid droplets (Frey 2008, Kanaroglou 2008, Fraser 1999). The increased trend of truck activity and related congestion worsen the $PM_{2.5}$ pollution (US-EPA 2012). However, in the current practice the detailed truck specific pollution estimation is not easily possible due to unavailability of a modeling methodology with applied supporting data to predict the link-level hourly truck activity and corresponding emission inventory.

To estimate the regional or local air quality impact of $PM_{2.5}$ emissions and also to predict future $PM_{2.5}$ concentrations, we often utilize atmospheric dispersion models. The dispersion models use mathematical equations to simulate how pollutants disperse in the ambient atmosphere. These are typically employed to determine whether existing or proposed new industrial or transportation facilities are or will be in compliance with the air quality standards set by national/state environmental regulatory agencies (US-EPA 2009).

Theoretically, the dispersion models could predict pollution concentrations for areas as small as 2500 square meters to as big as a few hundred square kilometers. Application of such sophisticated dispersion models with finer details can provide us the comprehensive understanding of the air quality problem, including the quantitative effect of pollution sources. In fact, modeling urban truck-related $PM_{2.5}$ concentrations alone can help us to evaluate the effectiveness of different travel-demand management strategies and policies since trucks are the main cause for $PM_{2.5}$ pollution in cities (US- EPA Oct 2004). To model the contribution of heavy duty diesel trucks to urban fine particulate pollution in a reliable manner is the primary motivation problem for this dissertation research.

Currently, to estimate source apportionments we calculate each pollution source's (industrial, biogenic, on-road, and non-road, etc.) emission dispersion in urban atmosphere independently and adjust them using measured total concentrations; then their relative contribution is determined using these adjusted concentrations. The estimation of each pollution source's emission dispersion consists of multiple steps. Each of these steps is equally important, and they are explained in detail in coming sections. Further, in the current practice there are some serious shortcomings in the aforementioned process which would affect the pollution source apportionment results; such drawbacks are also pointed out in the following discussion.

1.2 Modeling Transportation Caused Pollution in Urban Areas

Dispersion modeling for any pollution source is a very complex procedure, and for mobile sources it is much more convoluted. Traditionally, transportation-caused air pollution modeling is conducted in the following steps:

1) Estimate the detailed traffic activity data such as vehicle miles traveled and speeds.
2) Predict emission inventory using detailed traffic activity, meteorology, fuel data, and vehicle age information.
3) Allocate regional/county-level emission inventories to the grid defined for modeling domain both spatially and temporally.

4) Modeling pollutant dispersion and concentrations at established receptors using detailed gridded, temporal, speciated[1] emission inventory, terrain, meteorology data (Bachman 2000).

The reliability of the dispersion model output depends upon how much emphasis is put on first three steps in the above procedure. Fig 1-1 details the development of gridded/spatial, temporal, and speciated emission inventory to be used in air quality models (Markakis 2012). Based on the available data resources, steps 2 and 3 are modified as shown in the schematic diagram (Fig 1-1). Traditionally there are two methods for allocating county level inventories to grid-level: (a) Top-down approach- uses vehicle miles traveled and number of trips as spatial surrogates (b) Bottom-up approach- uses multiplies emission rates with vehicle miles traveled. In either case, it is critical to estimate the accurate vehicle activity in terms of Vehicle Miles Traveled by vehicle type (VMT mix) and average speed profiles.

The bottom-up approach (which is more accurate and efficient than other method) of calculating gridded emission requires detailed link-level vehicle activity with the most possible accuracy (Lindhjem 2010). For vehicle activity inputs, the emission models often rely on traditional travel demand models, but these models are mostly calibrated and validated using total vehicle volumes and travel times as opposed to detailed vehicle-type-based volumes and travel times (Bhat 2003). Since, we know from the past studies that truck activity is the major contributor of mobile source $PM_{2.5}$ emissions in urban areas, simply using the output from traditional four-step travel demand models may be inappropriate.

---

[1] This is the process of disaggregating inventory pollutants into individual chemical species components or groups of species (e.g. all organic related compounds into one category). This process depends upon purpose of inventory such as dispersion modeling photochemical modeling, air toxics inventories, chemical mass balance modeling, and visibility modeling

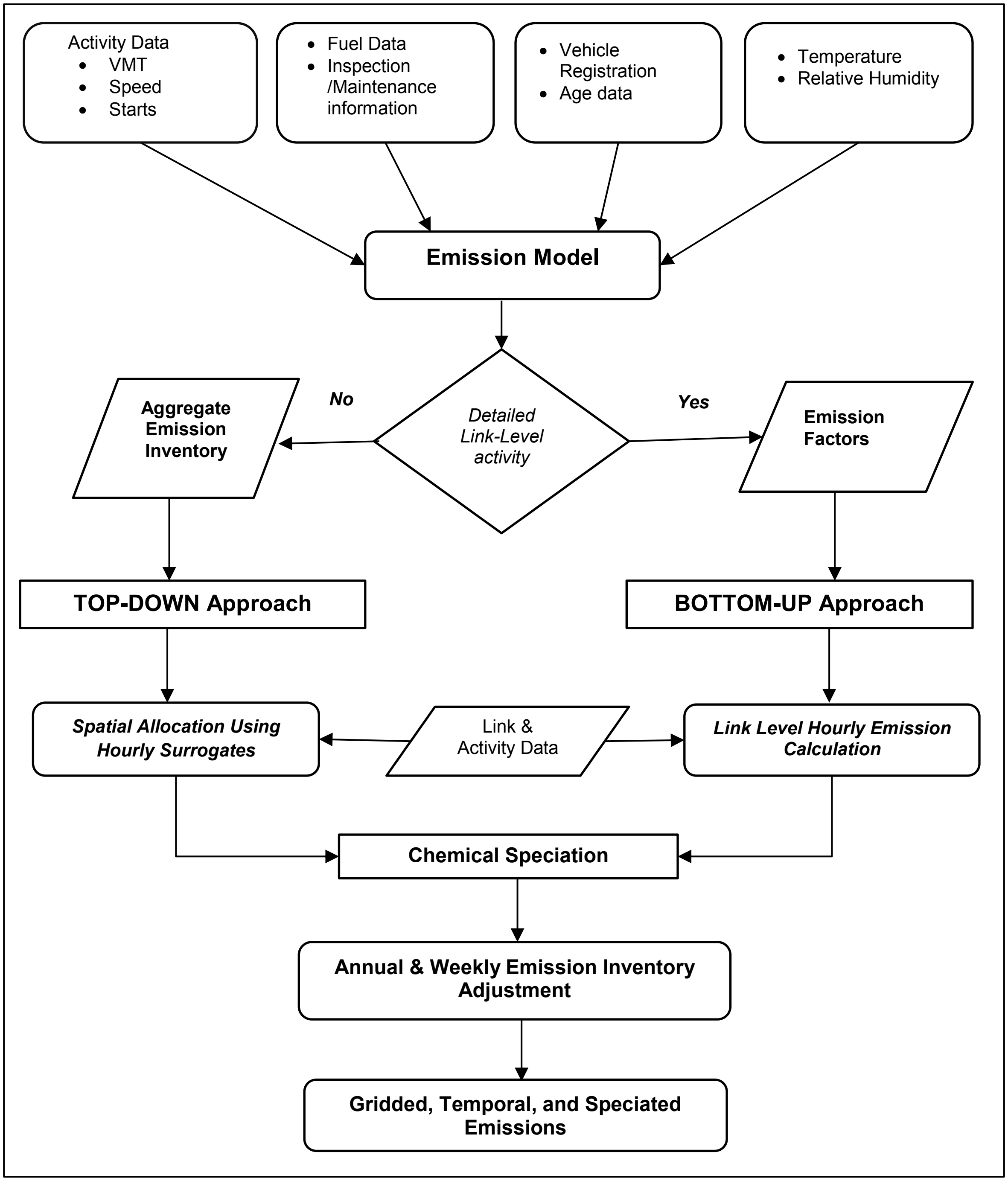

Figure 1-1: The Process of Preparing Mobile Source Emission Inventory Inputs to Air Quality Models

Further, it has been confirmed that the truck miles traveled and their average link-level speeds are the most important activity inputs for better estimation of vehicle exhaust $PM_{2.5}$ emissions (Bai 2007). In the proposed research, we have emphasized in developing the reliable truck activity data inputs like total miles traveled and average link-level speeds. The details are explained in the next section.

1.3 Important Activity-Related Inputs for Truck Emission Modeling

Recently, many researchers have tried to establish the importance of accurate vehicle activity inputs for emission purposes. They found that improved prediction of VMT by vehicle type or relative vehicle activity distribution (also called as VMT mix) has increased mobile source emission inventory by 25-40 percent depending on the pollutant under consideration and regional traffic pattern (Frey, et al., 2006).

It is evident that the vehicle miles traveled data is the most important input for emission models. Applying the same analogy to truck activity mix, it makes sense that improved truck activity estimation may predict truck-exhausted $PM_{2.5}$ emissions in a better manner. However, currently there is no consistency between model-predicted truck activity mix and the observed data. For example, if we compare the truck miles predicted from the

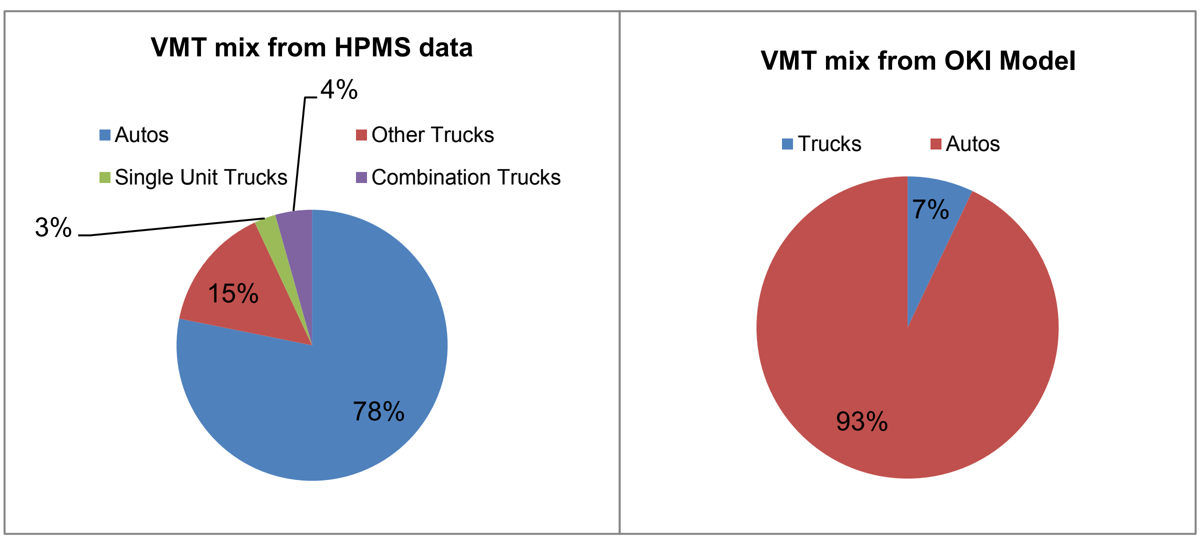

Figure 1-2: VMT Mix Comparison between OKI Model and HPMS Data for Freeways

Ohio-Kentucky-Indiana regional council's (OKI) travel demand model and the Highway Performance Monitoring Systems' (also called as HPMS[2]) data for freeways in the region, the former showed under prediction (Figure 1-2). On the contrary, the HPMS data alone cannot be directly used in emission analysis since it contains factored data for most highway types like minor arterials, collectors and locals. So, the existing travel demand models cannot predict truck miles traveled (TMT) and its constituent mix. We have also found that most of the regional agencies (usually better data sources for local planning data) have limited traffic count data that can be used for truck activity estimation. As part of this research effort, we need to develop a reliable methodology to estimate truck activity with limited traffic data available from local planning organizations.

Other research studies found that the next important activity-related input into the emission-models is: vehicle-type specific average speed profiles. The mobile source $PM_{2.5}$ emissions are highly sensitive to truck hourly speeds (US-EPA, 2003). However, most of the present travel demand models estimate average daily speeds rather than hourly speeds, thus the relative speed distribution may not be representative of peak hour congestion. The OKI travel demand model is somewhat better compared to most of the travel demand models since it can predict four different speeds in a day, i.e. morning peak period, evening peak period, mid-day, and off-peak period. The comparison of these two different average speed distributions for the Cincinnati area revealed that there is much congestion in the case of four time period speeds

[2] The HPMS is a national level highway information system that includes data on the extent, condition, performance, use and operating characteristics of the nation's highways. The traffic count data in HPMS is developed through sample data for arterial and collector functional systems.

compared to daily speeds, and it can lead to incorrect estimation of emissions if former type of speeds is used (Fig. 1-3).

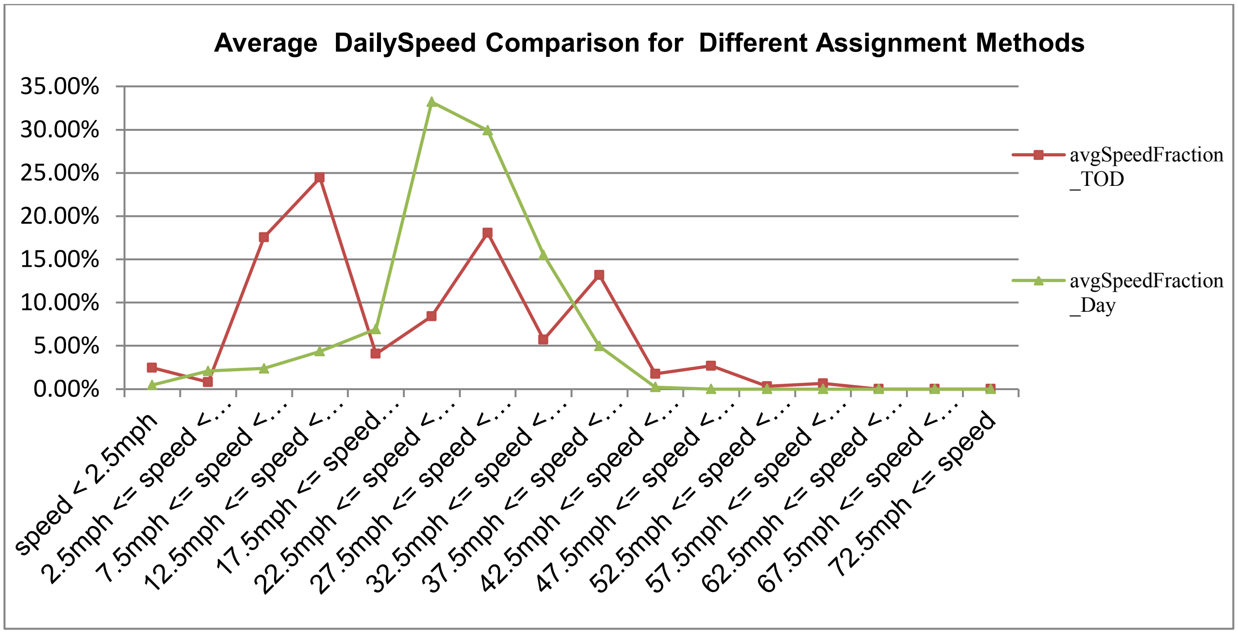

Figure 1-3: Average Speed Distribution based on Daily Assignment and Time of Day Assignment

In fact, the traffic flow pattern in each hour of a day is different, and related average speeds are also different. Ideally, hourly truck-type-specific speeds should be used in emission modeling (US-EPA, 2007). In the current practice to improve speed data, we do post-process the daily volumes and speeds using advanced highway assignment models. But this methodology has inherent problems of over estimation as it uses daily volumes (Bai, et al., 2007). If the hourly assignment procedure is applied, not only more accurate hourly truck activity can be estimated but also peak-hour-congestion representative truck speeds can be predicted. This improvement can impact the spatial and temporal emissions for the dispersion models in a positive manner. So, there is an opportunity to use hourly assignment to estimate better link-average speeds. To maximize the efficiency of these advanced hourly assignment techniques we need precise truck travel information and the details are explained in next sections.

1.4 Modeling Link Level-Truck Activity and Presence of Spatial Autocorrelation

To model the spatial truck activity, many truck models were proposed in the past. Truck models range from simple growth factor models to very complex GPS-data-based truck activity/tour models. We present an elaborate discussion about such models in the literature review chapter. The growth factor models could not capture change in truck activity due to zonal land use and demographic changes. On the other hand, the complex activity/tour-based models need expensive truck travel surveys. For this kind of situation, regression models can be a viable alternative and, even the variation of regression models such as econometric models would be more suitable. These are used for prediction of econometric variables.

The econometric models are very efficient in capturing social and economic changes even at smaller domain level, and they can be used to model link truck volumes. However, the traffic data samples used in these model developments are collected within a single large regional area. They are expected to suffer from spatial correlation among dependent variables due to homogeneity (Anselin 2002). If there is a spatial correlation (first-order serial correlation), it has been observed that the actual standard error will be larger if we specify and estimate such models using classic regression theory (Kapoor 2007). In general, ignoring spatial dependence tends to underestimate the real variance in the data, thus models are mis-specified (Gleditsch, 2007). Before using any advanced spatial regression modeling theory, we should confirm the presence of spatial autocorrelation in truck traffic count data, which is explained in the next section.

1.4.1. Determination of Spatial Autocorrelation in Truck Traffic Data

To deal with the spatially correlated data, a separate type of econometric modeling is used, namely *spatial econometrics* (Anselin, 2002). Consider truck volumes at location i depend on similar to truck volumes at locations j, where $j \neq i$. It should be noted that the dependence may be among several observations, as the index i can take on any value from i = 1 to N. This dependence may be due to similar characteristics of surrounding locations or from unobservable latent variables that are spatially correlated. Such models can be formulated as indicated below.

$$y_i = f(y_j) + X_i \beta + c \qquad \textbf{Eq. 1}$$

To explain the spatial autocorrelation denoted by $f(y_j)$ different statistics were proposed in the past, among them Moran's I is the prominent (Elhorst, 2001). Essentially, this statistic is the cross product of spatial proximity between observations and similarity of values for a particular attribute. The functional form of this statistic is:

$$I = \frac{N}{\sum i \sum j W ij} \frac{(\sum i \sum j W_{ij})(y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2} \qquad \textbf{Eq. 2}$$

Where N is the number of spatial units indexed by i and j; y is truck factor; $\bar{y}$ is the mean of y; and W_{ij} is an element of a matrix of spatial weights. We have estimated Moran'I static for single unit (2-4 axles), combination (more than 4 axles), and total truck volumes assuming the null hypothesis to be there is spatial autocorrelation among truck data. The Moran'I values range from 0.017 to 0.083, which means there is a strong spatial dependency among the data. Further, based on these values, the alternative hypotheses are selected, which means we must consider spatial auto correlation while modeling link-level truck volumes for better prediction. Another important consideration in this modeling procedure is that most empirical volume or activity prediction models were not able to perform well under rigorous validation.

Table 1-1: Spatial Autocorrelation Estimated for Different Aggregated Truck Types

Dependent Variable	Moran's I	Expectation	p-Value	Alternate Hypothesis	Standard Deviate
Single Unit Trucks	0.0174	-0.00119	0.1273	Greater	1.139
Combination Trucks	0.0836	-0.00127	1.07E-07	Greater	5.1877
Total Trucks	0.0295	-0.00142	0.02921	Greater	1.8925

1.5 Modeling Temporal Truck Activity

As explained earlier, modeling truck volumes and speeds by the hour is very important for accurate estimation of $PM_{2.5}$ dispersion in urban atmosphere. Currently, most travel demand models could not estimate link-level hourly truck volumes, thus, hourly distribution factors (a type of fractional responses) are being used to get hourly truck volumes (Bai 2007). These distribution factors are developed through some crude empirical methods, such as the simple averaging method (i.e. calculating average hourly factors for each highway type). However, some past studies identified that the hourly variation of truck activity depends upon surrounding land use, truck haul type, and congestion period of the day. So to model link-based, truck-specific temporal disaggregation factors, we should consider the above-mentioned variables. But the interdependency of hourly factors brings much complexity into modeling. To deal such complexities, logit or probit[3] type regression modeling is very popular. Application of such advanced methodologies to predict 24-hour activity distribution is desirable and had not been attempted before. Currently, the advanced highway assignment models like HCM model require hourly link volumes, and modelers supply the average hourly volumes (daily volume divided by 24). Another advantage of developing hourly distribution factors would be preparing hourly volumes to be used in highway assignment models to estimate realistic hourly speeds.

1.6 Identified Problems

In summary, the problems of the current truck-related air pollution modeling can be reiterated as:

1) It is very difficult to model heavy-duty trucks related $PM_{2.5}$ air pollution alone in the urban setting using current methodologies as they are only able to model all on-road mobile sources together.
2) Most of the regression based truck activity models can predict only average daily truck activity and none of them can estimate hourly activity due to limitations associated with the modeling methodology used.
3) In current practice, the vehicle-specific link-average speeds are estimated using post-processing method which has inherent drawback of over-estimation as it is not possible to incorporate peak-hour congestion in such method.

---

[3] The logit and probit are both S-shape forming functions with a domain between 0 and 1 and the y are also inverses of the cumulative distribution function (CDF) of a probability distribution. This property makes them readily applicable for any variable which has fractional values and have minimum ,maximum values of 0 and 1, respectively

4) From literature review it is evident that the regression based daily truck activity/volume models were not validated using observed traffic counts. Thus, they may not useful or reliable for other regions and future years.

1.7 Goals and Objectives of the Research

From the above discussion it is clear that, the bottom-up approach provides us the reliable and ground-truth transportation related emission inventory output at grid-level. However, to get the precise emission rates that are used in the bottom-up approach, we need consistent inputs like truck miles traveled and average link-level truck speeds. As it is mentioned in the problem statement, the primary goal of this dissertation research is to develop a methodology to estimate the contribution of heavy-duty trucks to urban air quality using an improved truck model that produces detailed and reliable output compared with traditional travel demand model. To accomplish this goal and also improve the existing methodology, we have divided the modeling process into following achievable objectives:

1) Applying the spatial regression modeling methodology for daily truck volume prediction as we have found that there is strong spatial autocorrelation among traffic data.
2) Since most of agencies do not have the luxury of large traffic data sample size, apply an optimization model using the spatial regression model results as control totals to improve the quality of model prediction.
3) Develop a separate hourly distribution model to predict hourly truck volumes as it is not possible to model them using any traditional spatial regression theory; and apply those hourly truck volumes in HCM highway assignment to estimate consistent truck

speed profiles that reflect peak-hour congestion as existing post-processing technique over-estimate them.

4) Apply rigorous calibration and validation approach for the truck activity models so that the methodology is reliable and applicable to other modeling areas.
5) Finally, converting the outputs from one constituent model to match the input needs of another constituent model to facilitate seamless integration in a comprehensive framework.

1.8 Scope of the Data Used for the Model Development

In this research, the author has used vehicle-classified traffic count data, socio-economic data, highway network data, and land-use data collected in the Greater Cincinnati area. The data is obtained from OKI regional council of governments for the counties of Boone, Butler, Campbell, Clermont, Dearborn Hamilton, Kenton, and Warren. Selection of the traffic count locations for analysis should meet the following criteria:

1) The Average Annual Daily Traffic data is aggregated at least by hourly time period;
2) The traffic count data collected between years 2003 and 2009
3) The types of trucks modeled as part of this research are following:

Table 1-2: Type of Trucks Modeled in the Present Research

ID	Truck Type	Description
51	Refuse Truck	Trucks primarily used to haul refuse to a central location.
52	Single Unit Short-haul Truck	Single unit trucks with more than four tires with a range of operation of up to 200 miles.

53	Single Unit Long-haul Truck	Single unit trucks with more than four tires with a range of operation of over 200 miles.
54	Motor Home	Trucks whose primary functional design is to provide sleeping quarters.
61	Combination Short-haul Truck	Combination tractor/trailer trucks with more than four tires with a range of operation of up to 200 miles.
62	Combination Long-haul Truck	Trucks whose primary functional design is to provide sleeping quarters.

4) The socio-economic data is based on 2005 American Community Survey and 2001 Census data; and,
5) Meteorology and vehicle-age-distribution data is collected in the year 2010;

1.9 Case Study Area

In this research, the modeling domain is 25 km in length and 23 km in width and it includes the City of Cincinnati, City of Covington, and City of Newport. Even though the latter two cities are not part of same county, their proximity to Cincinnati downtown and similar air quality problems made us consider them in this study. Further, it also provides us a complete picture of urban air dispersion since the study areas are similar in land-use characteristics and are separated by just a river. A complete description of study area is carried out in the data description chapter.

1.10 Significance of the Research

1) *Truck modeling with limited data* - The proposed methodology is two-step modeling process which is based on spatial regression and optimization techniques. The optimization step is particularly useful for modeling with limited data like current

case study where the traffic count data is available for less than 3% links in the region.

2) *Modeling truck speeds that reflect peak-hour congestion* - Most of the current travel demand models cannot predict hourly speeds since the trip matrices are day-based thus, the congestion time period is not conventionally taken into account. In the proposed methodology hourly trip matrices are used which would improve the reliability of truck speeds.

3) *Modeling truck related air quality impacts independently* - Not many previous (if not none) studies independently modeled the air quality impact of heavy duty trucks on urban atmosphere since reliable truck-activity is not readily available from regional travel demand models.

4) *Transferability of the proposed research* - The methodology used in this study is easily adoptable by any other metropolitan areas in the United States as the data used in this research is available from most of the planning agencies and US-EPA (measured air quality data). Most importantly, using proposed methodology the policy makers can evaluate different travel demand management scenarios to reduce fine particulate pollution.

5) *Application of more effective bottom-up approach for grid level emissions* - In current practice, to model urban scale dispersion of pollutants the county level emission inventories are distributed using surrogates like population or land use type; which do not represent actual mobile activity. My proposed methodology directly uses gridded emissions (aggregated link level emissions) as area sources in air dispersion model, thus the pollutant dispersion is predicted much better.

1.11 Organization of the Dissertation

The dissertation is organized as follows: Chapter 2 presents a comprehensive review of pertinent literature of the mobile source-related $PM_{2.5}$ pollution. It also enumerates different currently available truck activity models, emission models, and dispersion models with their corresponding merits and demerits. Chapter 3 proposes a comprehensive approach to model urban $PM_{2.5}$ pollution caused by heavy duty trucks. It also proposes the SROT Model and Hourly Fractions Model for predicting link-level truck volumes and their temporal activity. Chapter 4 discusses the study area and data used in the research. Chapter 5 presents the detailed explanation of the results from truck activity models with related validation studies. Chapter 6 details the emission modeling procedure for accurate estimation of truck-related emissions using MOVES model and activity estimated from the proposed models. Chapter 7 presents the detailed approach of how the $PM_{2.5}$ emissions exhausted from heavy duty trucks disperse in urban atmosphere using the emission inventory from MOVES model and local meteorology data. In this chapter, we also compared the real-world $PM_{2.5}$ concentrations with modeled values to estimate the contribution of heavy-duty trucks to urban air quality. Chapter 8 wraps up the dissertation by drawing conclusions from the present research.

Chapter 2: Literature Review

2.1 Apportionment of Mobile Source $PM_{2.5}$ Emissions in Ambient Air Quality

According to US-EPA's studies more than 70% of urban $PM_{2.5}$ pollutant quantities are contributed by the heavy duty diesel vehicles (Ketzel et al. 2008). Subsequently, the $PM_{2.5}$ emissions from trucks clearly influence government policy and program decisions, but the relationship is less direct. However, the on-road $PM_{2.5}$ related studies have been very influential in shaping up our current air quality policy and diesel emission reduction programs in the United States.

As part of the State Implementation Plans (SIPs) designed by environmental protection agencies to control air pollution, local planning agencies are required to develop transportation emission inventories and transportation-related control strategies to reduce the air pollution. Due to direct perilous impact of particulate matter on human health, it has become very crucial to estimate their pollution accurately. As the $PM_{2.5}$ emission rates are so uncertain, and also there is very limited detailed PM profile data available; the air-quality modelers have been struggling to estimate the fine grained level $PM_{2.5}$ emission inventories (Facanha et al. 2006). To measure actual air quality in urban areas, EPA has designed the AQS monitoring system, which continuously monitors ambient air quality. Traditionally the dispersion of emission quantities is compared to monitoring values to find out source apportionment. But, it has been historically very difficult to apportion the particulate matter back to specific vehicle classes and stationary source categories as it is very important to design air pollution control measures. To determine each source's contribution to urban air quality can be possible if the modeled fine particulate concentrations are close enough to the measured concentrations. Thus, the high-resolution emission data is critical in conducting source-apportionment especially for $PM_{2.5}$ pollution (Loretta et al 2008).

2.2 Mobile Source Emission Estimation Models

The first effort of mobile source emission modeling in United States started with the first edition (1978), Volume II of AP-42 which contained of all available information about mobile source emission factors, including the source code listing in the MOBILE1 model, the first highway vehicle emission factor model. Since then, numerous developments in vehicle technology, testing procedures and instruments have resulted in subsequent model developments in MOBILE series. MOBILE6.2 is the last version in MOBILE series which had improved particulate matter emission and carbon monoxide (CO) estimation when compared with its predecessors. In 2010, US-EPA has released its state-of-art emission estimation model MOVES (US-EPA 2007). This model has flexibility to maintain huge databases, can accommodate second by second speed and acceleration changes and also employs different criteria for start and running emissions in comparison with MOBILE6.2. US-EPA stipulates the use of MOVES 2010a for air conformity of transportation projects in accordance with State Implementation Plan, Regional Transportation Conformity and Air quality Hot-Spot Analyses. Emission modeling is very important step in transportation planning since for project funding conformity is an essential. Further in the congestion mitigation and subsequent transportation demand management strategies processes, the emission modeling serves as guidance.

In addition to MOVES; CSIRO model, ITEM, CMEM. EMFAC and IVE models provide us with various options in emission estimation in United States. The IVE and EMFAC models are activity based models and very similar to MOBILE in some respects. On contrary, F-factor method and CSIRO model (used in UK and Australia) were engine and fuel specific and cannot be used for platoon of vehicles. There is wide variety of emission models developed by US-EPA based on different principles and intended to use for different purposes. A comprehensive review of such models is presented in the Table 2-1.

Table 2-1: Different Emission Inventory Estimation Models

Model	Description	Application
MOVES	EPA's current official model for estimating air pollution emissions from cars, trucks and motorcycles. In the future the model will also cover non-road emissions. It is developed using extensive vehicle tailpipe emission data.	This model is used to calculate emission inventories for all State Implementation plans and transportation conformity. It can estimate ozone, PM and mobile source air toxics at national, state, region, county and project domains.
MOBILE	This model is predecessor of MOVES; it was first released in 1978 and had been updated continuously since then. The latest version of this model is MOBILE6.2 and the "basic emission rates" are developed from driving a "Federal Test Procedure" (FTP) driving cycle under specific laboratory conditions.	MOBILE is used to calculate current and future emission inventories at the national and local level. Inventories based on MOBILE are used to meet the federal Clean Air Act standards through State Implementation Plan (SIP) and transportation conformity processes, and are sometimes used to meet requirements of the National Environmental Protection Act (NEPA).
NONROAD	EPA has developed this model for modeling non road mobile sources based on laboratory and industry data. Fuel types included in the model are: gasoline, diesel, compressed natural gas, and liquefied petroleum gas.	It calculates past, present, and future emission inventories (i.e., tons of pollutant) for all non-road equipment categories except commercial marine, locomotives, and aircraft.
NMIM (National Mobile	This is a consolidated emissions modeling system for EPA's	It was developed to produce, in a consistent and automated way, national, county-level mobile

Inventory Model)	MOBILE6.2 and NONROAD models.	source emissions inventories for the National Emissions Inventory (NEI) and for EPA rule making.
GEM (Greenhouse Gas Emissions Model)	EPA has created this model for the purpose of vehicle compliance demonstration, thus, it is less configurable than similar commercial products and its only outputs are GHG emissions and fuel consumption.	This model is designed to predict truck GHG emissions.
EMFAC (Emission Factor Model)	EMFAC is California's model for estimating emissions from on-road vehicles operating in California. It is built on decades of vehicle testing and analysis, and is informed by DMV registration data, the Smog Check program, and many other data sources.	EMFAC is used as a starting point for developing plans to meet air quality standards, and for assessing the impact of motor vehicle emissions regulations on emissions and air quality.
Fuels Models (multiple)	EPA has developed several fuel models for estimating emissions impacts to changes in fuel properties and composition.	*Heavy-Duty Diesel Fuel Analysis Program :* Used for non-road and highway heavy-duty diesel engines *RFG Program:* to determine whether gasoline complies with RFG and anti-dumping emissions performance standards *Fuels Emulsions Analysis Program:* looking at the effects that emulsions have on NOx, PM, HC, and fuel economy. *Biodiesel Emissions Analysis Program:* Used for diesel engines those have not been specifically modified to operate on biodiesel.

IVE (International Vehicle Emissions Model)	The International Vehicle Emissions (IVE) Model is specifically designed to have the flexibility needed by developing nations in their efforts to address mobile source air emissions.	The model makes estimates of local air pollutants (criteria pollutants), greenhouse gas emissions, and toxic pollutants. The model is intended to help cities and regions develop emissions estimates to: Focus control strategies and transportation planning on those that are most effective; Predict how different strategies will effect local emissions; and Measure progress in reducing emissions over time.
CMEM (Comprehensive Modal Emissions Model)	It is microscopic model which predicts second-by-second tailpipe emissions and fuel consumption based on different modal operations from in-use vehicle fleet. This model uses a physical, power-demand approach based on a parameterized analytical representation of fuel consumption and emissions production.	This model is used for evaluating emissions benefits of project-level or corridor-specific transportation control measures (e.g. HOV lanes), intelligent transportation systems (ITS) implementations (e.g. electronic toll collection), and traffic flow improvements (e.g. traffic signal coordination).
VT-Micro (the Virginia Tech Microscopic energy and emissions model)	This is a regression based model which was developed from experimentation using instantaneous speed and acceleration levels as independent variables. Linear, quadratic, cubic, and quartic terms of speed and acceleration were tested using chassis dynamometer data collected at the Oak Ridge National Laboratory (ORNL).	This model can be utilized for the evaluation of operational-level transportation projects such as re-timing signals, modeling toll plazas, and modeling highway sections.
CSIRO-Power based model (Australia)	This model estimates exhaust emissions for CO, NOx and hydrocarbons (HC) for each hour and road link for an average weekday and weekend	The power based model generates road-specific vehicle emission fluxes for the purpose of near-road impact modeling.

DMRB (UK)	The volume 11 of DMRB contains emission rates for CO, NOx, PM, HC and CO2 which are based on assumed vehicle fleet for years till 2020 for regions in UK.	Emission rates are applied using weighted method for all links with assumed vehicle fleet.
COPERT (European Union)	The COPERT 4 is developed by European Environmental Agency to estimate emissions from road transport to be included in official annual national inventories for use in Europe. The emissions are consistent with the 2006 IPCC Guidelines for the calculation of greenhouse gas emissions.	This model allows for a transparent and standardized, hence consistent and comparable data collecting and emissions reporting procedure, in accordance with the requirements of international conventions and protocols and EU legislation.

Of all these models, activity based models are of primary importance because of their ease for use (directly submitting basic traffic characteristics to estimate emission factors) and most importantly their ability to predict and fit in dispersion models. Recent development in emission modeling focused on dynamic modeling with capabilities for second-by-second emission estimation of various pollutants. Examples include the Comprehensive Modal Emissions model (CMEM) and the Virginia Tech microscopic (VT-Micro) model (Nemalipuri 2010). However, they are particularly useful for smaller domains and demand very large size input data.

2.3 Critical Input Data for $PM_{2.5}$ Emission Inventory Modeling

As part of literature review effort we have verified different studies in which the researchers have established the importance of different input data items for better estimation of $PM_{2.5}$ emission quantities. We already knew that most part of $PM_{2.5}$ quantities are caused by heavy duty trucks from EPA research. So, in this section we have summarized the important truck related inputs and their effect on $PM_{2.5}$ emission estimation (Facanha et al. 2005).

Table 2-2: Truck-Related Input Data Items affecting $PM_{2.5}$ Emission Estimation

Input Parameter for Emission Model	Geographic Scale	Impact on $PM_{2.5}$ Emission Factor	Uncertainty in Present Methodologies
Truck Miles Traveled	Regional and Local	High	Medium/High
VMT Share by Truck Type	Regional and Local	High	Medium/High
VMT Share by Time of Day	Regional and Local	Medium/High	Medium/High
Age Distribution	Regional and Local	High	Medium
Mileage Accumulation	Regional and Local	Medium/High	Medium
Distribution Emission Control Technology	Regional and Local	Low/Medium	Medium
Truck Fuel Type	Regional and Local	Medium	Low/Medium
Average Speed	Regional and Local	Medium	Medium
Driving Cycles	Local	Medium	High
Road Grade	Local	Low	High
Emission Factors	Regional and Local	High	Medium
Classification of Truck Type	Regional and Local	Medium	Low
Empty Miles	Regional and Local	Medium	High

The truck miles traveled (TMT), the VMT share by truck type, empty truck miles and the VMT share by time of day are estimated from truck travel demand model. The vehicle registration data contains other important information like age distribution and mileage accumulation. State governmental agencies control emissions through Inspection and maintenance programs, thus maintain truck fuel data and emission control technology data. The driving patterns and road grade have very less impact on regional truck related emission estimation, thus the literature pertaining to those inputs are not covered in this literature review.

2.4 The Truck Activity Inputs into Emission Models

Typically, truck activity data is estimated using traditional travel demand modeling methodology. Researchers argue that truck demand modeling is more complex than passenger transport modeling, because freight flows are under the control of many decision-makers who interact in a dynamic environment (Cambridge Systematics 1997; Jonnavithula2004). There is vast number of studies to model truck activity at regional level. A comprehensive review of those methodologies has been enumerated in the Table 2-3. This list is not exhaustive; however it covers all varieties of truck models proposed till date. The table shows different freight/truck models with details like type of methodology used; the model area; advantages and disadvantages. This analysis provided the author with vital information such as the most suitable methodology for present problem and available opportunities etc.

Table 2-3: List of Different Models Used to Estimate Freight/Truck Activity in United States

Authors	Methodology	Application and Advantages	Drawbacks
Quick Response Freight Model, (1996)	Truck trip generation rates were estimated from the number of jobs in employment sectors associated with commodity shipments. The default rates provided by the manual were taken from a survey in Phoenix, Arizona.	This Model is applied by many metropolitan organizations in united states to develop their truck models. Since this approach borrows truck trip generation rates from QRFM, it is very handy.	Applications of this model demonstrate that the default parameters like truck trip generation rates are not easily transferable between different regions.
Murat Celik et al. (2002)	This study expanded previous spatial interaction models of commodity flows by incorporating new variables into the model, using a flexible Box-Cox functional form, and applying the analysis to all manufacturing commodities.	The 1993 U.S. Commodity Flows Survey is used for estimating state-to-state flow models for 16 commodity groups. It could able to consider important independent variables and spatial interaction and have served as the precursor for later studies	This state level model and it also could not estimate detailed level truck vehicle volumes which are required by emission model estimation.
Aruna Siva Kumar et al., (2002)	The approach estimates the fraction of commodity consumed at each destination zone that originates from alternative production zones	TRANSEARCH Freight data, U.S. Census data, the Regional Economic Information System (REIS) data were used for state of Texas. The fractional split model for commodity flow	This model is very difficult to validate and actual volume splits of trucks based on engine technology is impossible.

		distribution is more general in structure than the typical gravity model used.	
Krishnan Viswanathan et al. (2003)	This paper provides a method to disaggregate FAF^2 (Freight Analysis Framework) data to county level geography.	This method is based on simple linear regression equations. The methodology is applied to Florida state freight analysis zones data into county data freight attraction production quantities.	This model can predict county level freight quantities. It is very difficult synthesize the truck volumes using this model.
Wang et al. (2010)	They proposed a hierarchical model by coupling the truck traffic and weight data with socio-economic variables to establish a freeway-level statistical model for freight demand estimation and prediction.	The model can predict the growth of freight demand in a freeway network, thus allowing the identification of potential freight transport bottlenecks.	The model cannot be useful more accurate analysis such as congestion and emission estimation since the prediction is more aggregated.
Maria Boilé et al. (2006)	This paper employs linear regression algorithms in order to train models under the presence of limited training data.	These models are most naïve truck models and can be easily applied to any area. This methodology is used in New Jersey for state level model.	In these models some serial and spatial autocorrelation in data is expected. They could not predict hourly truck variation.
David Novak et al. (2008)	They have provided an insight into different variable transformation techniques, evaluate the	This is the one of the significant freight modeling studies where a spatial regression modeling	The proposed methodology is applicable only for national level data.

	use of spatial regression variables, and apply them to correct for spatial autocorrelation.	methodology used for freight models. They have applied methodology for national level in United States.	
Kangarolu et al. (2008)	The approach estimates the commercial vehicles volumes using survey data, tube counts and turning movement data.	In this study contribution of commercial vehicle activity to urban emissions is estimated in Hamilton, Canada. This model actually estimated the truck assignment through converting commercial vehicle into corresponding passenger car equivalents	The main disadvantage of the model is it predicts only peak hour commercial vehicle activity.
José Holguín-Veras,et al.., (2006)	This approach only proposes new mathematical formulations to model commercial vehicle empty trips	This methodology is very important since all of the previous freight and truck models ignored the empty trips. These proposed models have been applied to Guatemala and Dominican Republic	The methodology can only partially useful since they estimated only empty trips.
Feng Liu et al. (2006)	In this study, statistical models were developed to forecast TMT growth of four facility categories at the county and statewide levels.	These models are simple and straight forward and also incorporate both socioeconomic and transportation system supply variables. This model is applied to Pennsylvania state data	This model is more likely to estimate growth in trucks on freeways and cannot predict total truck split and temporal variation of demand.

In aforementioned studies the freight/truck activity was estimated by using varied types of models from commodity flow models to regression models. The commodity and freight models are data intensive and it is very expensive to develop detailed inputs to these models (Sivakumar et al 2002, Wang et al.2010). Whereas the linear regression models are very sensitive (due to large unexplained effects) to the input data could give unreliable truck volumes.

There are several concerns about estimating TMT from travel demand models or truck counts. First, the traditional trip based models which identify the number of trips between each pair of spatially defined zones in the model, could not provide the trip information based on truck categories required for emission models, which is very important for emission estimation. Second, when used for forecasting TMT, travel demand models often do a poor job of representing the complex trip generation and trip distribution patterns of commercial vehicles. Third, the accuracy of average speed at the link level is questioned given that it is not measured directly, but rather estimated from vehicle volume and road capacity. Finally, a high number of time periods is necessary to properly capture the speed variations throughout the day, which increases the computation requirements substantially (Facanha et al 2006, Nam et al 2007, and Wang et al 2010).

Regression based modeling for truck volumes prediction is also very relevant and viable option to develop inputs for emission models. Current methodologies to develop such models suffer from data autocorrelation errors. For example if the traffic count data is in panel format (traffic count data for multiple years and different cross-sections form panels), these models contain some unexplained effect due to spatial and temporal correlation. Such effects should be well explained in the new generation regression or econometric models.

2.5 The Temporal Travel Activity Inputs into Emission Models

In 2003, EPA has conducted multiple rigorous studies to analyze direct correlation between diurnal traffic activity and corresponding $PM_{2.5}$ emissions. They have observed $PM_{2.5}$ emissions in five big cities[4] in United States for more than a year. The results in this study showed that the temporal traffic activity has direct impact on hourly PM2.5 emissions. Following figure shows one of the observations in this study in Detroit, Michigan (US-EPA 2003).

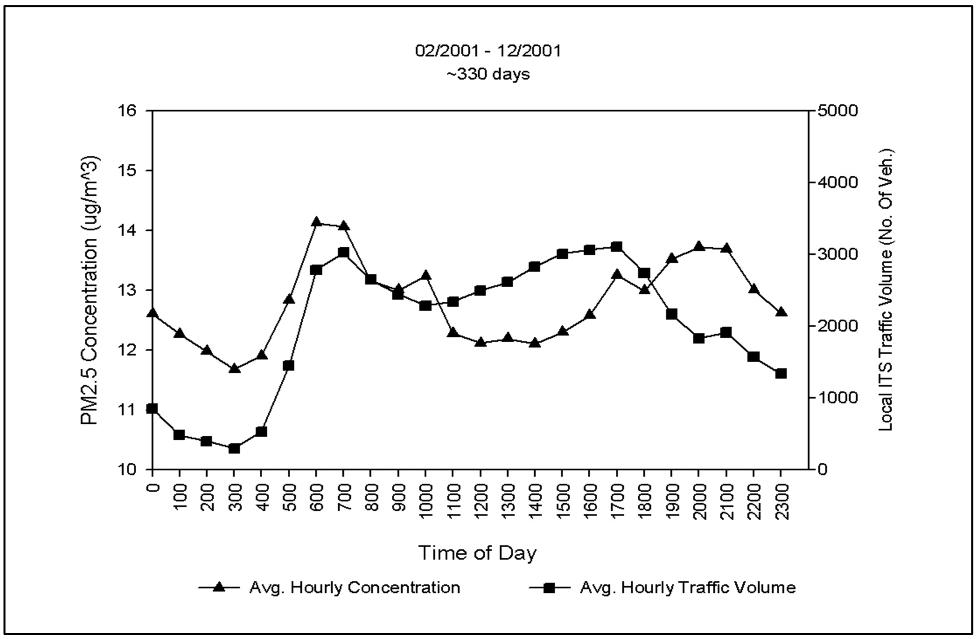

Figure 2-1: Correlation between $PM_{2.5}$ and Hourly Traffic Volumes in Detroit, Michigan (US-EPA 2003)

In addition to diurnal truck traffic activity, hourly temperature also shows direct impact on emission factors thus it is very important to develop temporally disaggregated truck activity mix as an input for emission models. It is also very important that in view of the non-attainment problems faced by several metropolitan areas, such an accurate modeling may be necessary for

[4] New York, Detroit, Pittsburgh, Atlanta and Baltimore

future conformity determinations (US-EPA, 2001). Typically most of the emission models allow truck activity mix to be inputted at the hourly level for accurate modeling purpose.

The first rigorous effort to predict the link level hourly vehicle activity mix (or diurnal activity) was done by Bhat and Nair (2001). As MOBILE6 requires hourly VMT mix inputs, compared to the 24-hour averages that MOBILE5 required, they have proposed a fraction split model to predict hourly VMT mix for each highway link in Austin, Texas metropolitan region. Their model predicts fractional split on links as a function of the following variables: R*oadway classification of the link, Physical attributes of the link, Operating conditions of the link and Attributes of the traffic analysis zone in which the link lie.*

The Bhat and Nair's hourly VMT mix model could forecast the VMT mix for six vehicle classifications (autos, sports utility vehicles, pickups and vans, motorcycles, buses, and trucks). One of the major drawbacks of this methodology is, it can predict an average hourly VMT mix for whole region for each vehicle type, which does not represent the real world. So, the improved hourly VMT prediction methodology should be able to do at the link level such that estimated emission inventory is more accurate than using regional average hourly distribution

2.6 The Meteorology Inputs into Emission Models

The Meteorological data is very important for accurate prediction of temporal emission factors and the meteorological measurements includes wind speed, wind direction, temperature, and humidity. However, temperature and humidity are the most important influencers for emission estimation. There were multiple studies conducted by EPA's Office of Transportation and Air Quality and other agencies to evaluate the effect of these parameters on diurnal emission rates and it is found that NO_x and PM emission rates are directly and significantly influenced by ambient temperatures (Benjay et al. 2002). The Kansas City Light-Duty Vehicle Emissions

Study (KCVES) has estimated the effect of temperature on PM emissions and results from the study are shown in Fig. 2-2. In general, PM emissions doubled for every 20 °F drop in ambient temperature and were independent of vehicle model year.

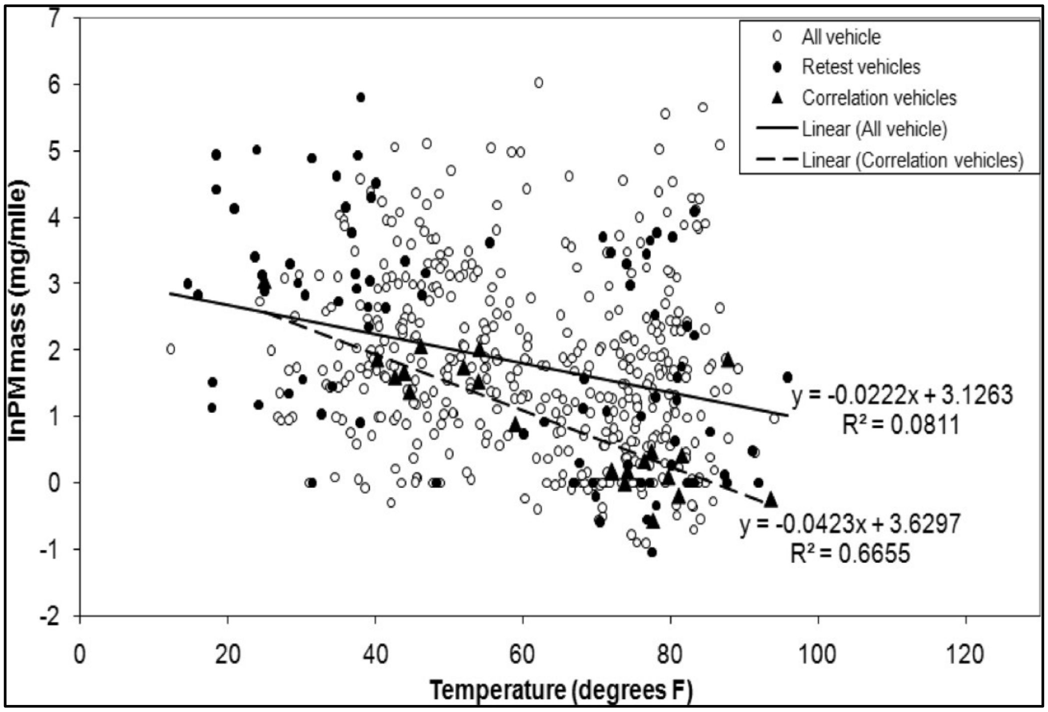

Figure 2-2: Results from Kansas City Light-Duty Vehicle Emissions Study (US-EPA 2003)

From this study they have also found that the effects of temperature on vehicle emissions were most pronounced during the initial start-up of the vehicle (cold start phase) when the vehicle was still cold, leading to operation under fuel-rich conditions, inefficient combustion, and inefficient catalyst operation. Through these important findings it has been corroborated that the diurnal activity (since temperature changes during day) is also very important for accurate estimation of $PM_{2.5}$ emission rates (Nam et al, 2006).

2.7 The Impact of Highway Assignment on Emission Modeling

All of the mobile source pollutants depend on vehicle speeds, thus, emission estimation models require detailed estimated or observed speeds by functional class of roadway as inputs. Although many urban areas track level of service (LOS) for arterials and freeways for peak and off-peak periods, EPA does not believe that agencies will generally have available robust databases of observed speeds (US-EPA 2001). Further, LOS classifications cannot be directly used as inputs to any of the emission models in place of speed distributions. Therefore, effective procedures are needed to estimate speeds from travel demand output or traffic count data (Bai et al 2007).

The highway assignment step in travel demand models uses calculated speeds and route choices to minimize travel time. The effect of speeds on assignments is evaluated through validation of travel demand models such as comparison of assigned traffic versus count data. Historically, the disagreement between Travel Demand Model speeds and observed speeds was mostly due to the following reason: the input trip tables were time period based and simulation results provide a single assignment representative of that period. In other words, traditional travel demand models cannot describe hourly variation in congestion and speeds since they can vary within the time periods (US-EPA 2001). To solve this drawback of traditional travel demand models, post processing techniques are available that use HCM procedures and the BPR curve to calculate hourly congested speeds. Their general approach is:

1. Distribute link-level volumes by hour of day using user input temporal distributions, which are developed from count data sets;
2. Calculate v/c using either link-specific capacities or lookup tables;
3. Apply the BPR curve/HCM model, using link-specific free flow speeds or lookup tables, to estimate the hourly congested speeds.

An extensive research in application of postprocessors for estimating accurate speeds for emission model purposes was done by Bai et al (2005). They have compared different post processing models part of study and arrived at conclusion detailed HCM based models could predict more accurate speeds and these speeds affect emission inventories at large. A serious caveat with this approach is: since volumes and speeds are interdependent, the initial input of average daily traffic assignment for post processing may not be accurate considering fact that hourly assignment may vary with congestion and speeds.

2.8 Other Important Inputs into Emission Models

The other important inputs to the emission models that would really affect the results are: link driving schedules, operating mode distribution/ vehicle specific power distribution, link grade and emission factor estimation methodology. But from previous studies it has been realized that these inputs would play important in project level peak period analysis, thus we did not emphasize in developing such inputs in our current research (Facanha et al, 2006).

2.9 High Resolution Modeling of Mobile Source Emissions Inventory

Recently, some researchers have successfully completed the high resolution mobile source emissions modeling in terms of spatio-temporal fine granularity. One of the most comprehensive studies in the mobile source air quality field was conducted by Wang et al (2009). They developed a framework to incorporate different models for predicting mobile source contribution to urban air pollution. In this methodology, they have used EMFAC (Emission Factor estimation model) to estimate emission quantities using four time period traffic volumes from SACMET (Sacramento Metropolitan Travel Demand Model). The DTIM (Direct Travel Impact Model) could predict gridded hourly emission quantities to be inputted into ISCST (the atmospheric dispersion model). Finally they could able to find out the on-road sources' contribution through comparison with the Air Quality System monitoring stations observations. Using the similar framework rather different constituent models, Guo et al. (2008) conducted a study the City of

Hangzhou, China for high resolution modeling of emissions. They have used the International Vehicle Emissions (IVE) Model for emission factor estimation and the CALPUFF models to estimate the dispersion of emissions. The other important feature of this study is they have detailed different methodologies to develop accurate inputs if they are not available already.

The study conducted by Wang et al. (2009) have only considered the ISCST model, which is a steady state dispersion model specifically designed for point and volume sources with considerably above the ground level. But lately, US-EPA has mandated using AERMOD (the steady state air dispersion model) for atmospheric dispersion modeling of emissions from all sources. Guo et al. (2008) in their study did not use sound travel prediction model for vehicle activity; whereas they have predicted whole region's travel activity based on data collected at few highways. There were many other studies which have tried fine granular emission inventory estimations instead of air quality. Buchman et al. (2006) is one of those important studies mentioned above, and in this study they have modeled gridded emissions in a geographical framework. Most importantly, they have allocated start and evaporative emissions to those grids using vehicle registration data. However, they could not actually model dispersion of such emissions and also did not compare with monitored data. Similarly, Kanaroglou et al.(2006) have also modeled link level NO_x, hydrocarbon and particulate emissions, only using more reliable commercial vehicle survey data. However, the major issue with these studies could not provide the full picture of the impact of mobile emissions on urban air quality since they are not comprehensive.

2.10 Air Dispersion Modeling of Mobile Source Emissions

Air pollution/dispersion modeling can provide the comprehensive understanding of the air quality problem, including contributing factors such as pollution sources, meteorological processes, and physical-chemical changes. Specifically, air pollution models are the only type of models that can quantify the deterministic relationship between emissions and concentrations/depositions, and can provide guidance to determine appropriate mitigation strategies (Air quality Modeling, 2005). US-EPA recommend use of air quality models for State Implementation Plans (SIP), revisions for existing sources and to new source reviews, including prevention of significant deterioration (PSD).

Air quality modeling procedures are in general categorized into four broad categories: Gaussian, numerical, statistical or empirical, and physical. Within the statistical models based on how the plume parcels trajectories move in the atmosphere, there are two types of models namely: the Lagrangian model which uses a moving three dimensional Cartesian grid as frame of reference whereas the Eulerian model uses a fixed frame of reference. All of the US-EPA recommended air quality models are enumerated in Table 2-4 with their respective underlying methodologies and applications.

Table 2-4: Different Air Dispersion Models used for Mobile Source Emissions

Model	Methodology	Application
AERMOD/ AERSCREEN	Based on planetary boundary layer turbulence structure and scaling concepts, including treatment of both surface and elevated sources, and both simple and complex terrain.	State Implementation Plan (SIP) revisions for existing sources and for New Source Review (NSR) and Prevention of Significant Deterioration (PSD) programs. The Screening model will produce estimates of "worst-case" 1-hour concentrations for a single source.

ISC3/ISCST/ SCREEN3/ TSCREEN/ ASPEN	A steady-state Gaussian plume model used to assess pollutant concentrations from a wide variety of sources associated with an industrial complex.	ISC3 model is predecessor of AERMOD, SCREEN3 is screening version and TSCREEN is used to screen toxics. ASPEN is an alternate dispersion model based on ISC3 methodology.
CALINE/ CAL3QHC/ CAL3QHCR/	A steady-state Gaussian dispersion model	CALINE is designed to determine air pollution concentrations at receptor locations downwind of highways located in relatively uncomplicated terrain. CAL3QHC is a CO model with queuing and hot spot calculations at signalized intersection.
CALPUFF	A multi-layer, multi-species non-steady-state puff dispersion model that simulates the effects of time- and space-varying meteorological conditions on pollution transport, transformation and removal.	It is applied to estimate sub grid scale effects (such as terrain impingement), as well as, longer range effects (such as pollutant removal due to wet scavanging and dry deposition, chemical transformation, and visibility effects of particulate matter concentrations).
COMPLEX1/ VALLEY	A steady-state, complex terrain, univariate Gaussian plume dispersion model.	VALLEY model for estimating either 24-hour or annual concentrations resulting from emissions from up to 50 (total) point and area sources. COMPLEX1 is a multiple point source screening model.
OCD (Offshore and	A straight line Gaussian model.	Used to determine the impact of offshore emissions from point, area or

Coastal Dispersion Model)		line sources on the air quality of coastal regions.
ADAM (Air Force Dispersion Assessment Model)	A modified box and Gaussian dispersion model which incorporates thermodynamics, chemistry, heat transfer, aerosol loading, and dense gas effects.	The emission release scenarios include continuous and instantaneous, area and point, pressurized and unpressurized, and liquid/vapor/two-phased options.
HYROAD (Hybrid Roadway Model)	A Lagrangian puff formulation model.	This model is used to determine hourly concentrations of carbon monoxide (CO) or other gas-phase pollutants, particulate matter (PM) and air toxics at receptor locations that occur within 500 meters of the roadway intersections.
DEGADIS	A Gaussian dispersion model	It simulates the atmospheric dispersion at ground-level of area source dense gas (or aerosol) clouds released with zero momentum (with gravity driven flow) into the atmospheric boundary layer over flat, level terrain.
HOTMAC/ RAPTAD	HOTMAC is a 3-dimensional Eulerian model for weather forecasting; RAPTAD is a 3-dimensional Lagrangian random puff model for pollutant transport and diffusion.	These models are used for prediction of transport and diffusion processes for complex terrain, coastal regions, urban areas, and around buildings where conventional models fail.
Panache	An Eulerian (and Lagrangian for particulate matter), 3-dimensional	This model can simulate continuous and short-term pollutant dispersion in the atmosphere, in simple or complex

	finite volume fluid mechanics model.	terrain.
PLUVUEII	An Eulerian dispersion model	The model predicts the transport, dispersion, chemical reactions, optical effects and surface deposition of point or area source emissions.
SCIPUFF (Second-order Closure Integrated PUFF Model)	A Lagrangian puff dispersion model that uses a collection of Gaussian puffs to predict three-dimensional, time-dependent pollutant concentrations.	SCIPUFF provides a prediction of the statistical variance in the concentration field resulting from the random fluctuations in the wind field.

According to US-EPA's - Guidelines on Air Quality Models "*The extent to which a specific air quality model is suitable for the evaluation of source impact depends upon several factors. These include: (1) The meteorological and topographic complexities of the area; (2) the level of detail and accuracy needed for the analysis; (3) the technical competence of those undertaking such simulation modeling; (4) the resources available; and (5) the detail and accuracy of the data base, i.e., emissions inventory, meteorological data, and air quality data*" (US-EPA, 2005). However, for most of regulatory applications in United States it is mandatory to use AERMOD as air quality model, this the most sophisticated model which can model air pollution coming from all three important types of sources such as point, volume and area. AERMOD also combines geophysical data such as terrain elevations and land use with the meteorological data to derive boundary layer parameters such as Monin-Obukhov length, mixing height, stability class, turbulence, etc. to predict more accurate dispersion of solid pollutants like $PM_{2.5.}$ As we are trying to model accurate and detailed dispersion of $PM_{2.5}$ it should be absolutely necessary to use more robust model like AERMOD for our analysis.

2.11 Summary of Literature Review

A review of the extensive $PM_{2.5}$ air pollution modeling literature contributes toward the identification of many issues associated with accurate modeling of truck activity. The literature result confirmed the need for soliciting innovative modeling methods to well consider those issues into the air quality models.

The literature review also indicates that the regression methodology is one of the best suited to develop accurate and uncomplicated truck activity models. A sound regression based truck volume model should estimate accurate hourly TMT with corresponding true individual speed profiles. Applying more sophisticated and state-of-the-art emission and air dispersion models like MOVES and AERMOD combined with statistically robust truck activity prediction models could provide us with better air quality estimates. The next chapter describes the development of the econometric models of truck activity and age distribution based on regional data that can accommodate variations due to socio-economic, roadway and land use characteristics.

Chapter 3: Methodology

3.1 Mobile Source Air Quality Modeling

To better quantify the impacts on air quality by on-road motor vehicles or assess the effectiveness of control strategies, accurate high-resolution emission inventories are needed as the inputs to air dispersion models. However, due to limited modeling methodologies of motor vehicle activities (including Vehicle Miles Traveled distributions, vehicle fleet age, type and technology distributions etc.), most importantly emissions from trucks in US cities are not well quantified. Furthermore, there is no established quantification method to project future emissions. As a result, decision-makers are unable to design effective control strategies to improve air quality caused by the urban area truck emissions and medical researchers are not able to accurately evaluate $PM_{2.5}$ exposure impacts on human health.

As an important step in the air quality modeling process, we use emission factor models to estimate local emission inventories. The emissions factor models (like EMFAC, MOVES) require several traffic-related inputs, including travel speeds, vehicles miles of travel, on-road operating conditions (operating mode of vehicles, environmental conditions, existence of inspection/maintenance programs, etc.), vehicle age distribution by vehicle class, and vehicle mileage accumulation rates by vehicle class. These inputs are used to calculate emissions factors (in grams per mile of vehicle travel for each pollutant) for different vehicle classes. The vehicle-class and pollutant specific emissions factors are then applied to the VMT accumulated by each of the vehicle classes, and finally aggregated to obtain total emission quantities by pollutant type. But the US-EPA requires that the non-attainment regions (for ozone and $PM_{2.5}$ pollution) should estimate their mobile source emissions using network-based transportation models.

Ideally, the link-specific emissions are estimated based on a) link VMT, b) vehicle speed on the link, c) the vehicle class-specific emissions factors, and d) VMT mix fractions. These variables are estimated from travel demand models. However, in practice the most of agencies involved in either in transportation conformity or in air pollution modeling does not perform at such detailed level. So, in this research we have proposed detailed and accurate truck activity estimation for better modeling of $PM_{2.5}$ air pollution.

A methodology for estimating the disaggregate truck activity through an advanced truck model is developed during the first part of this research. The Hourly Truck Vehicle Miles Traveled and the Hourly Truck Speed distribution values outputted from the proposed truck model; they are used as mobile source activity inputs into the emission model. Specific details of proposed truck model are discussed in next section. A separate truck age distribution model based on spatial panel modeling methodology is used to develop inputs for the emission model. In this research, the emission inventory is estimated using the sophisticated and US-EPA recommended MOVES (Mobile Vehicle Emission Simulator) model. The output from emission model is post- processed to obtain hourly gridded mobile source emission inventories, which are regarded as area sources to model their corresponding atmospheric dispersion. Through specifying appropriate receptor grid network and using the hourly gridded emission inventory coupled with local meteorological and terrain data, the typical 24-hour $PM_{2.5}$ pollution in urban atmosphere can be modeled. The conceptual framework for modeling the contribution of $PM_{2.5}$ to urban air quality has been shown in Fig. 3-1.

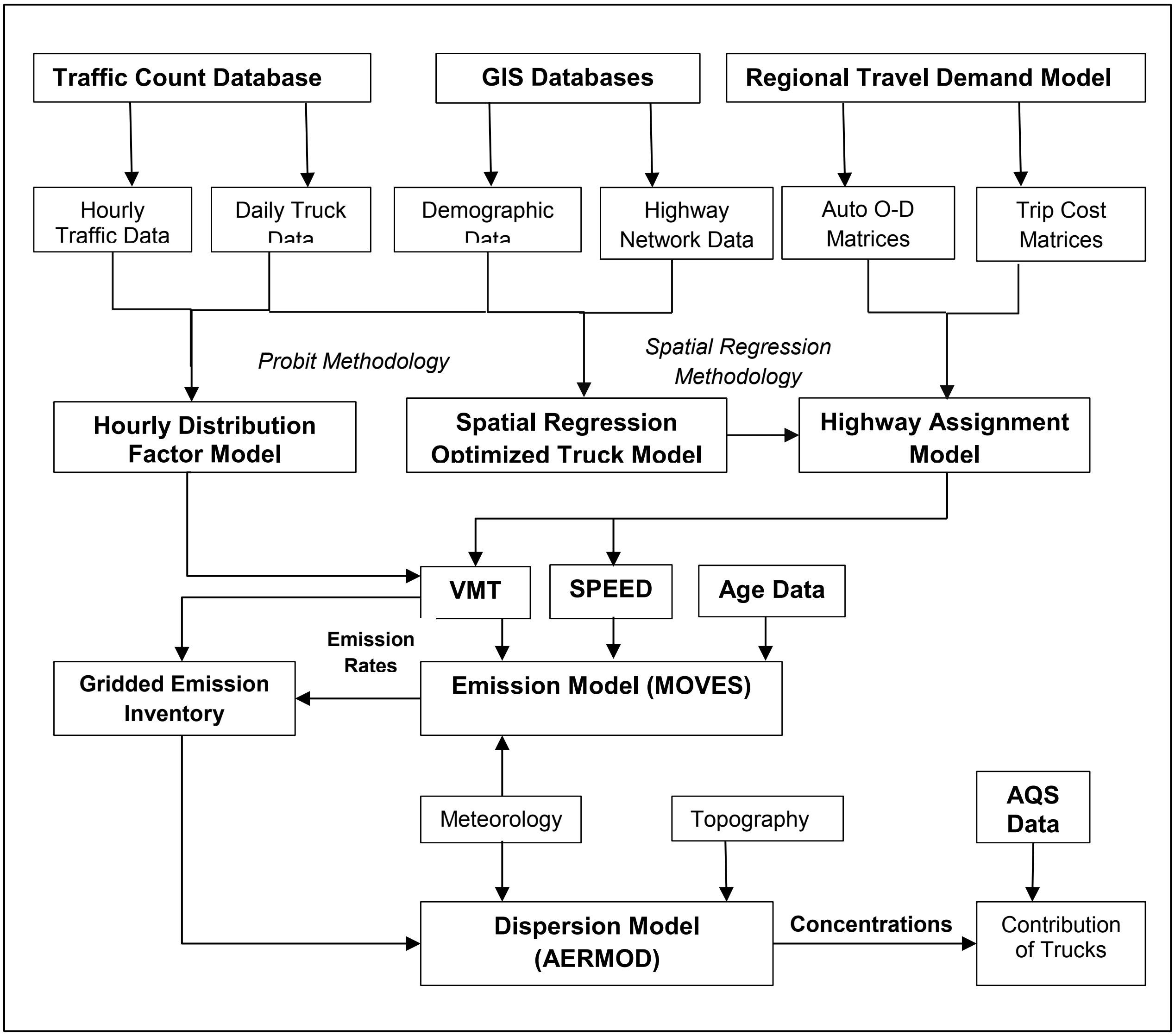

Figure 3-1: Conceptual Framework to Estimate Contribution of Trucks to PM2.5 Pollution in the Urban Atmosphere

3.2 Modeling Link –level Truck Activity

As mentioned above the typical mobile source emission estimation model requires accurate truck activity (TMT), truck speeds and truck age distribution data as input to predict the $PM_{2.5}$ emission rates for the model area. Even if some researchers proposed to use regression models, they could not validate such models due to instability and sensitivity of such models. To counteract these problems we have proposed a two-stage model; first stage is to create a spatial regression model based on the training dataset and second stage is to optimize the truck demand obtained from the spatial regression model output at selected control locations. This demand adjustment is very critical to forecast future year changes. The detailed depiction of data flow among different sub models in the truck demand model are shown in Fig. 3-2.

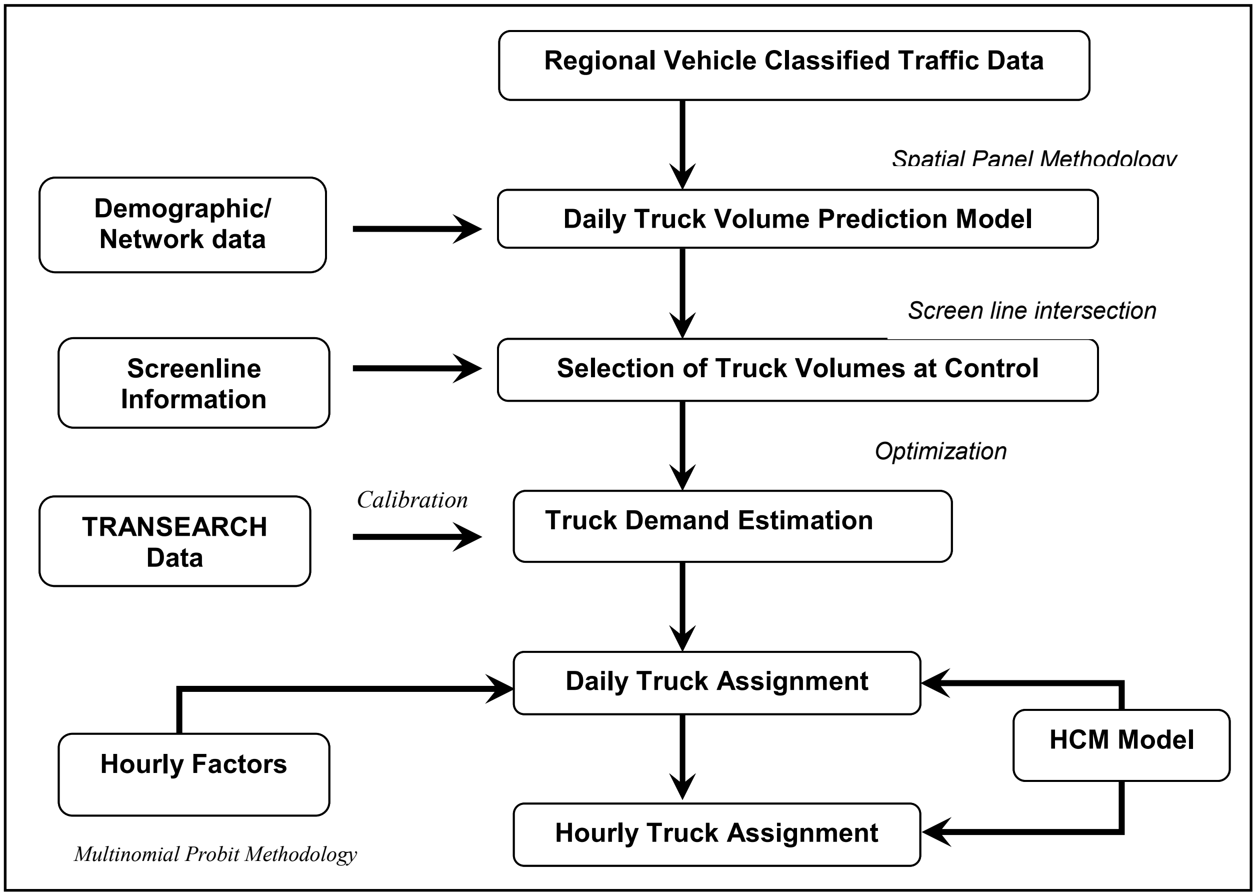

Figure 3-2: Detailed Spatial Regression and Optimization based Truck Model Development

The truck demand matrices are assigned to highway network in two steps: In the first step they are assigned for whole day at once, then the daily volumes are disaggregated using hourly distribution factors developed independently. In the second step, the hourly disaggregated truck volumes are reassigned to the highway network. As a result of this hourly reassignment, we can estimate the truck speeds more realistically. For highway assignment we have used advanced HCM model, and the details are discussed in coming sections.

In this whole process, we have used two important travel demand modeling concepts: one is screen lines (or control locations) and travel demand matrices. Screen lines are imaginary lines defined by features such as railroads, creeks, and rivers. Since all roadways are not reflected in the travel demand model, these types of features serve to funnel traffic into corridors so that all trips can be analyzed where crossing of these features is possible. Traditionally these lines are used for validating travel demand model output. These lines are parallel to geographic area boundaries. Control locations are the actual highway links which are intersecting with screen-lines. These locations are the most important links on highway network that can be used for optimizing the traffic flows among traffic analysis zones. Travel demand matrices are the matrices which contain the number of trips originating in one zone and ending in another zone. These travel demand matrices can be assigned to the highway networks based on different assignment algorithms like all-or-nothing-based, equilibrium-based or HCM based.

3.3 Two Stage Spatial Regression and Optimization Model: Spatial Regression

As discussed in the introduction, to establish the presence of spatial dependency among the count data we have estimated the spatial autocorrelation parameter. To specify the spatial regression model we have used advanced spatial panel regression modeling theory proposed by Kapoor et al so that it takes into account spatial autocorrelation among dependent variables. The details about model specification and related estimation procedure for model parameters are explained in the following sections.

3.3.1. Spatial Regression Model Specification

Assuming that the truck volume on a particular highway link i (1 to N) and for a year t (1 to T) is denoted by y_{it}, which can be modeled using set of independent variables x_{it} and the corresponding coefficients are given by β. It is also assumed that a spatial relationship exists among the variables. The spatial weighting matrix (row normalized) is denoted as W, which is $N \times N$ dimensions having zero value diagonal elements and its entries are typically declining with distance. This matrix does not change over the time horizon (Baltagi, 2008). The spatial correlation among the data can be quantified by spatial autoregressive parameter ρ (Elhorst, 2010). The unobserved effect can be explained using the spatial weight matrix, spatial autoregressive factor and unexplained observation specific error ε. The model can be represented by following equation:

$$y_{it} = x_{it}\,\beta + [I_T \otimes (\mathrm{I} - \rho\mathrm{W})^{-1}]\varepsilon_i \qquad \textbf{Eq. 3}$$

Where

y_{it} = NT x 1 vector of observations on the time period t

x_{it} = NT x K matrix of observation on K exogenous variables.

β = NT x K matrix of coefficients

I_T = Identity matrix of size T x T

$\otimes$ = tensor multiplication operator (used in the context of vector or matrix multiplication)

I = Identity matrix of size $N\ x\ N$

W = spatial weight matrix of size $N\ x\ N$

ρ = spatial autocorrelation

ε_i = NT x 1 vector of unexplained observation specific error for i

3.3.2. Spatial Panel Model Estimation

The error components due to spatial autocorrelation and observation specific autocorrelation are independent and identically distributed with distributions $N(0, \sigma_v^2)$, $N(0, \sigma_1^2)$ respectively. For estimation purpose we need the composite error covariance matrix (Ω_ε) in terms of σ_v^2 (spatial error variance) and σ_1^2 (observation error variance). We can use the standard transformation matrices Q_0 and Q_1 to convert spatial error variance matrix and observation error matrix to same size of composite (unexplained) error matrix. The transformation matrices should repeat the observation specific errors for all time periods using selector matrix J_T which is size of T x T with unit elements (Kapoor et al., 2007).

$$\Omega_\varepsilon = \sigma_v^2 Q_0 + \sigma_1^2 Q_1 \qquad \textbf{Eq. 4}$$

Where

$$Q_0 = (I_T - \frac{J_T}{T}) \otimes I_N \qquad \textbf{Eq. 5}$$

$$Q_1 = \frac{J_T}{T} \otimes I_N \qquad \textbf{Eq. 6}$$

In many previous studies, panel models were estimated using Pooled Ordinary Least Square (OLS) estimation methods. Nevertheless, it has been identified that the common error component over individuals induces correlation across the composite error terms, thus OLS estimation is inefficient. For estimating this model feasible generalized least squares (FGLS) estimators are used, since they are computationally simple and much reliable compared with OLS estimators. The feasible GLS estimator of β is given by the following equations:

$$\hat{\beta}_{FGLS} = \{X^{*'}[\Omega_\varepsilon^{-1}]X^*\}^{-1}X^{*'}[\Omega_\varepsilon^{-1}]y^* \qquad \textbf{Eq. 7}$$

Where

$$X^* = [I_T \otimes (I_N - \breve{\rho}\, W)]X \quad \textbf{Eq. 8}$$

$$y^* = [I_T \otimes (I_N - \breve{\rho}\, W)]y \quad \textbf{Eq. 9}$$

$$\Omega_\varepsilon{}^{-1} = \sigma_v^{-2} Q_0 + \sigma_1^{-2} Q_1 \quad \textbf{Eq. 10}$$

$\breve{\rho}$ = Estimated spatial autocorrelation parameter

3.4 Two Stage Spatial Regression and Optimization Model: Optimization

Application of spatial regression model can yield link based truck volumes by truck type, however the output from such a link based prediction model cannot be used in future years unless the model is validated. As regression models being sensitive and less stable, we need to modify such models. In present case we introduced an optimization step, which synthesizes travel demand matrices from link volumes predicted using spatial regression model and then optimizes using independent truck flow data. To optimize the truck trip distribution/demand matrices, we have used TRANSEARCH freight data.

To estimate the truck trips or truck travel demand T_{ij} between TAZs i and j, we use a derived truck demand optimization model (Vaughn et al, 2010), which optimizes the predicted link volumes and convert them into truck demand matrices. This model is given by following equation.

$$T_{ij} = a_i\, b_j\, c_{ij}^{\alpha}\, e^{-\beta c_{ij}} \prod_K X_K \quad \textbf{Eq. 11}$$

Where

a_i, b_j = Model parameters which depends on productions at TAZ i and attractions at TAZ j

c_{ij} = cost function between TAZs i and j

α, β = Generalized Cost Function parameters between TAZs i and j

X = estimated truck traffic volume

K = number of highway links between TAZs i and j

$\prod_K X_K$ = the product of all estimated link truck traffic volumes between TAZs i and j

We assume in this model that truck trips also follow similar distribution as autos for initial matrix formation purpose and later we adjust truck trips based on the link volumes. Since the truck trip distribution model used in this methodology is linear equation, any standard statistical estimation procedure can be used to find out the model parameters. Here we have used a statistically rigorous function called "Most likelihood objective function"; which is given by following equation.

$M = \sum_H \lambda_H \ H - \lambda_H H \log(\lambda_H h)$ **Eq. 12**

H = Link Truck Volume estimated from Truck Spatial Panel model

$h =$ Truck trips implied by the distribution model between TAZs i and j

$\lambda_H =$ The confidence level associated with link truck volumes

This procedure is an iterative procedure start with model parameter values as 1 and it can be implemented any statistical software. The output from this model is optimized truck traffic matrices for each truck type.

Reebie Associates® is a consulting company which develops very reliable freight/truck flow data using independent truck surveys for all counties in United States. This database is called TRANSEARCH Database and it consists of all modes of freight data by different commodity groups. Other useful feature of this database is: it also includes empty truck trips, which are very important for validation and calibration of any truck model. The truck trip matrices prepared using the truck demand optimization model, are aggregated into county based truck matrices and compared with TRANSEARCH data. As part of optimization/calibration process the truck demand optimization model parameters such as a_i, b_j are adjusted to match the truck flows with TRASEARCH data, in other words the model becomes more reliable.

According to traditional travel demand forecasting theory, the truck demand matrices obtained from proposed truck spatial panel model are added to auto travel demand matrices to get total travel demand among traffic analysis zones. These matrices are assigned to highway network using HCM model (details in the section 3.7) to get daily traffic assignment (volumes) and speed profiles. However, the daily traffic assignment cannot take peak hour congestion into account thus the link level truck speeds and related emissions factors may be incorrect. For this reason, we distribute the daily truck volumes into hourly truck volumes and reassign them to estimate hourly speeds and the corresponding details are discussed below.

3.5 Modeling Hourly Distribution of Truck Activity

In current practice, the hourly traffic volume disaggregation process is performed in a crude way (i.e. applying only handful types of factors). In reality, the hourly variation of traffic is dependent on many factors starting from type of surrounding land use to socio economic characteristics of TAZ under consideration. Typically, the current emission models also need hourly distribution factors as separate input since they use them for diurnal vehicle activity disaggregation and adjusting proportions of vehicle starts during a typical day for emission rate calculation. In this methodology we have proposed very relevant temporal truck volume disaggregation model.

The hourly distribution of traffic can be classified as fractional response variables which are dependent upon various independent variables associated with land use, socio-economic, and transportation network data (Perugu, 2009). Modeling of these variables cannot be done using classic statistical models (Bhat et al., 2003). Since these fractional response variables (e.g. market shares, regulation compliance rates) are included in the econometric models, a similar methodology is employed for modeling hourly volume distributions.

3.5.1. Model Specification and Estimation

Let us assume, i = 1 to N is traffic analysis TAZ (TAZ) in the region; j = 1 to 24 is hour id in a day and t = 1 to T is the analysis year. Then y_{ijt} is the hourly traffic distribution of hour j in traffic analysis zone i for year *t*. For the most of the analysis the year index *t* is not necessary and the variable is simplified to y_{ij} .. Then, hourly distribution factors are bounded, $0 \leq y_{ij} \leq 1$, and therefore cannot be modeled as a linear function of the covariates (Papke and Woodbridge, 2008). The generalized form to estimate the conditional mean of hourly distribution using a non-linear function of an index of covariates is given by

$$E\,(y_{ij}|\, x_{ij}, c) = F(x_{ij}\beta + c_{ij}) \qquad \textbf{Eq. 13}$$

Where,

x_{ij} are the strictly exogenous covariates,

β is the vector of parameters

c is unobserved effect

$F(.)$ is a non-linear function of the index of covariates

There are two popular procedures available to solve this function, the fractional logit and the fractional probit models. Since the dataset is multi-dimensional (i.e. panel data set), the fractional probit is better suited (Papke and Woodbridge, 2008). The fractional probit modeling specifies the non-linear function as Φ (.) which is the standard normal cumulative distribution function. These strictly exogenous covariates include travel demand model output, network variables, land use variables and demographic variables. Being the data is in panel format the unobserved effect c can be due to unobservable hourly specific and TAZ specific effects. Then, the conditional mean of unobservable TAZ specific effects, c_i, and hour specific effects c_j are in linear relationship with the mean value of the covariates

$c_i = \Gamma\, \bar{x}_i + u_{ci}$ and $c_j = \Lambda\, \bar{x}_j + u_{cj}$ **Eq. 14**

Where,

Γ, Λ are vectors of parameters,

$\bar{x}_i, \bar{x}_j$ are average of vector of covariates

u_{ci}, u_{cj} are residuals.

Using equations (11) and (12) the conditional mean of hourly distribution factors can be written as:

$E(y_{ij} \mid x_{ij}, u_{ci}, u_{cj}) = \Gamma\, \bar{x}_i + \Lambda\, \bar{x}_j + x_{ij}\beta + u_{ci} + u_{cj}$ **Eq. 15**

Furthermore, the residuals are assumed to follow normal distribution $\{u_{ci}, u_{cj} \sim N\,(0, \sigma_c^2)\}$. Using mixing properties of normal distribution, Equation (10) can be written as:

$E(y_{ij} \mid x_{ij}, u_{ci}, u_{cj}) = \Gamma_u\, \bar{x}_i + \Lambda_u \bar{x}_j + x_{ij}\beta$ **Eq. 16**

In the above equation the coefficients are scaled down to take care of residuals as indicated by subscript u. In probit based modeling it is important to find out the partial effect of the continuous covariates to explain change in the expected hourly distribution, which is given by:

$$\frac{\partial E(y_j|x_j,c)}{\partial x_{ij}} = \beta_i \phi(x_j\beta + c) \qquad \textbf{Eq. 17}$$

This shows that the partial effects depend on the level of covariates and unobserved effect. To deal with this situation, the average the partial effects for hours and for TAZs were used in this study. These average partial effects are taken into account as explanatory variable in the estimation process, thus the model estimated is robust compared to the linear form. This heterogeneity of partial effects is an advantage of the fractional probit model over the standard linear model, which predicts the same partial effect across hours and links (Gleditch 2007). To estimate actual hourly distribution fractions the covariates should be rescaled by the factor $\frac{1}{\sqrt{1+\sigma_c^2}}$.

The proposed hourly fraction model gives us hourly distribution factors to estimate hourly vehicle volumes in a typical day based on socio economic, highway network information. Such hourly volumes used to determine vehicle speed profiles in our proposed framed work. Typically all of the current emission models also do need hourly distribution factors as input since they use them for temporal vehicle activity disaggregation and adjusting proportions of vehicle starts during a typical day.

As discussed in the literature review, the average roadway segment travel speeds play an important role in estimating stabilized running vehicle emissions. Especially $PM_{2.5}$ emissions are highly variable with truck speeds. For accurate estimation of truck average speeds on highway links we have used the hourly assignment approach which is explained in the next section.

3.6 Highway Assignment Procedure

Currently the mobile running emissions are computed based on speeds produced during the travel demand modeling process. The conventional highway assignment procedures do not typically generate sufficiently resolved or accurate enough vehicle speed profiles (Stopher and Fu, 1998). Due to recent advances in traffic assignment methodology we can predict realistic speed and volumes on roadway segments, but those methods are not useful for regional scale emissions modeling due to significant data and computational requirements. Frequently post-processing techniques are seen as the most cost-effective means of improving the accuracy of the speed estimates. Although that the speed profiles cannot be accurate enough for emission estimation unless they are predicted based on hourly volumes or assignment.

In this proposed methodology we have already accurately disaggregated daily vehicle (including different truck types and other vehicle) volumes into hourly volumes. To estimate the reliable vehicle activity distribution in terms of average speeds, we have borrowed the traffic flow model from Highway Capacity Manual and customized regional traffic flow characteristics (HCM 2010, Bai et al 2008, and US-EPA 2006). The proposed model uses the travel speed – traffic relationship equations to calculate loaded travel time on roadway links. The travel speed – traffic relationship equations are used to calculate the degradation in free-flow speed (i.e. the congested speed) that results from non-zero traffic volumes. There are five equations developed for this model, one for each group of facility type. Mathematically the relationship may be expressed as

$$T_i^h = T_i^o * [1 + a * \left(\frac{V_i^h}{C_i}\right)^b] \quad \textbf{Eq. 18}$$

Where,

T_i^h = the loaded travel time in hour h on link i.

T_i^o = the free-flow speed on link i.

V_i^h = the traffic volumes in hour h on link i.

C_i = the hourly capacity on link i

a, b = Model parameters or the coefficients.

The coefficients/model parameters are different for different roadway type. This allows for more realistic speed degradation for a particular type of highway. The data used to calibrate the equations are generated primarily based on the procedures of Highway Capacity Manual (HCM). The derived values for the coefficient are shown in Table 3.1.

Table 3-1: Coefficients in Travel Speed – Traffic Volume Relationship Equations

Group	Facility Type	a	b
1	freeways, ramp controlled expressways	0.200	8.00
2	expressways, freeway-to-freeway ramps, on-ramps, rural arterials	0.195	8.16
3	arterials with four-way stop	0.198	4.67
4	urban major roads, off-ramps	0.196	7.18
5	minor roads	0.259	6.12

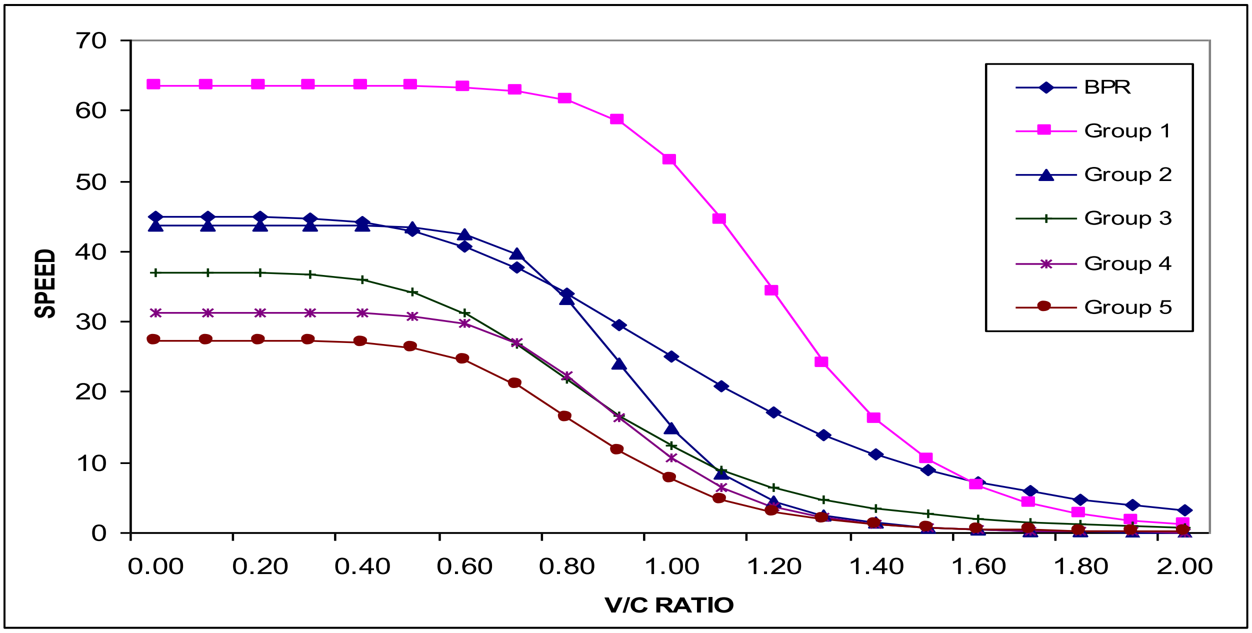

Figure 3-3: Travel Speed – Traffic Volume Relationships

In the Fig. 3-3, we have compared average traffic speed and traffic flow on different types of roadway links. It is also shown in the figure that the traditional BPR (Bureau of Public Road) equation used in travel demand modeling; which aggregates all types of roadways into single group and also overestimates link level speeds. Application of proposed modeling procedure iteratively, we can calculate more realistic hourly link level traffic speeds for different vehicle types. The vehicle activity eventually distributed among different average speed bins to be used with emission estimation model.

3.7 The Software Framework for Implementing Proposed Methodology

The proposed methodology is going to be implemented in a standard travel demand modeling software framework. We use the travel demand modeling software suite named "Cube" which is developed by Citi labs. This software suite has can efficiently handle different steps in the traditional travel demand modeling as well as most of the advanced models can be implemented using its built-in matrix processing programs. The truck model with proposed hourly assignment and truck age prediction models are implemented in this software and to prepare emission model ready outputs. An example of proposed methodology in Cube's model flowchart interface is shown in the Fig. 3-4.

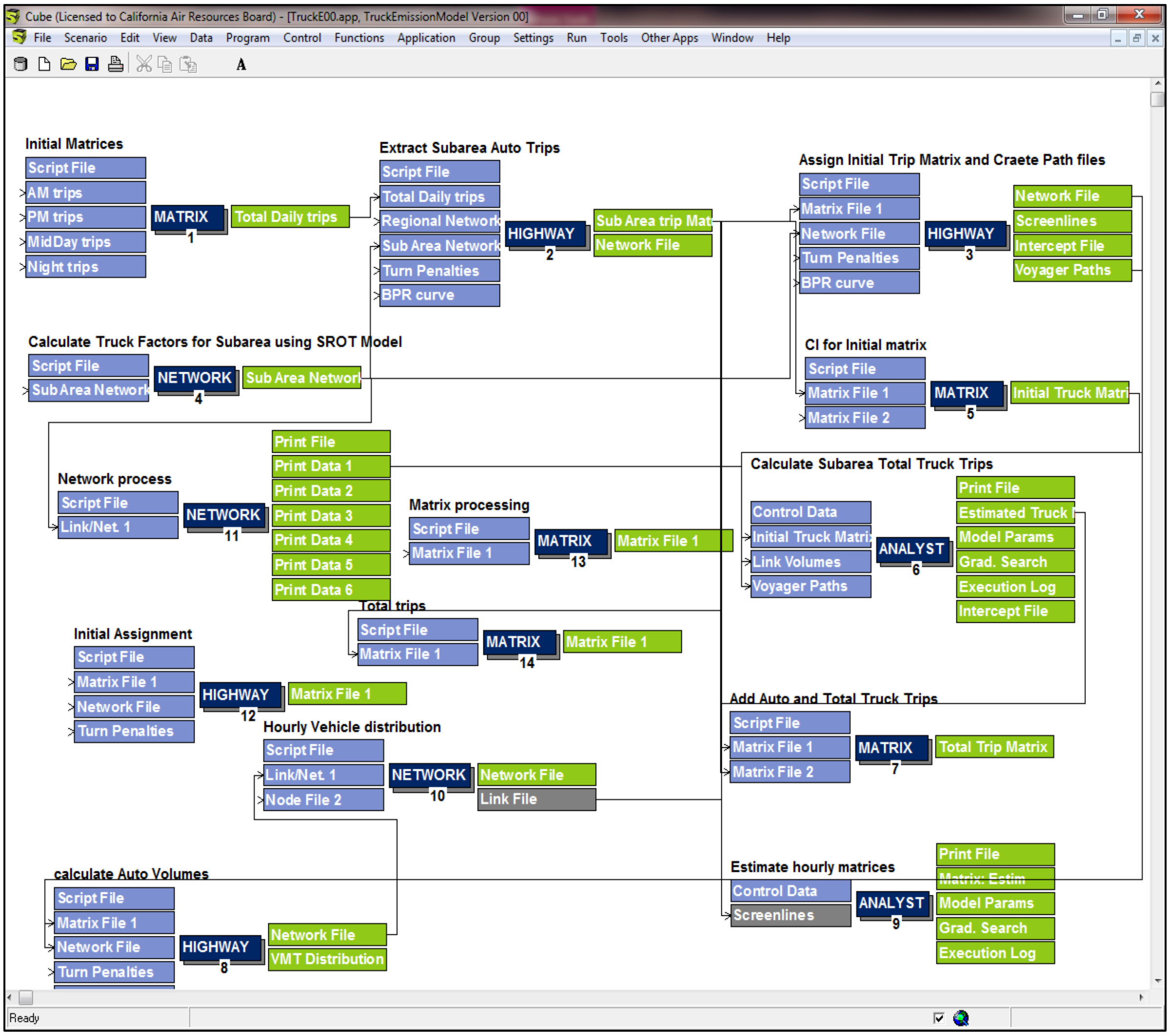

Figure 3-4: Integrated Model Developed in Cube Framework

3.8 Validation of Proposed Models

As a final step in truck activity modeling, we need to validate the proposed models. We have used regional traffic, socio economic, vehicle registration, land use and network data before a base year (2009) for all of the models since the base year is always chosen with consistent data.

For validation we are going to use the vehicle classified traffic counts and registration data for year 2010. As part of validation the standard travel model validation parameters like R^2 and root Mean Square Errors are estimated.

3.9 Modeling Link Level Emission Inventory

To model very detailed $PM_{2.5}$ mobile source emission inventory we have used the latest emission model called MOVES (Mobile Vehicle Emissions Simulator). It is developed by US-EPA through extensive research and recommended by FHWA for all regional and state level on-road mobile related air quality estimation purposes (MOVES User Guide, 2010). This model takes different types of inputs such as the VMT mix distributions, meteorological data, fuel characteristics, speed distribution based on Vehicle Hours Traveled, vehicle population etc. at very detailed level.

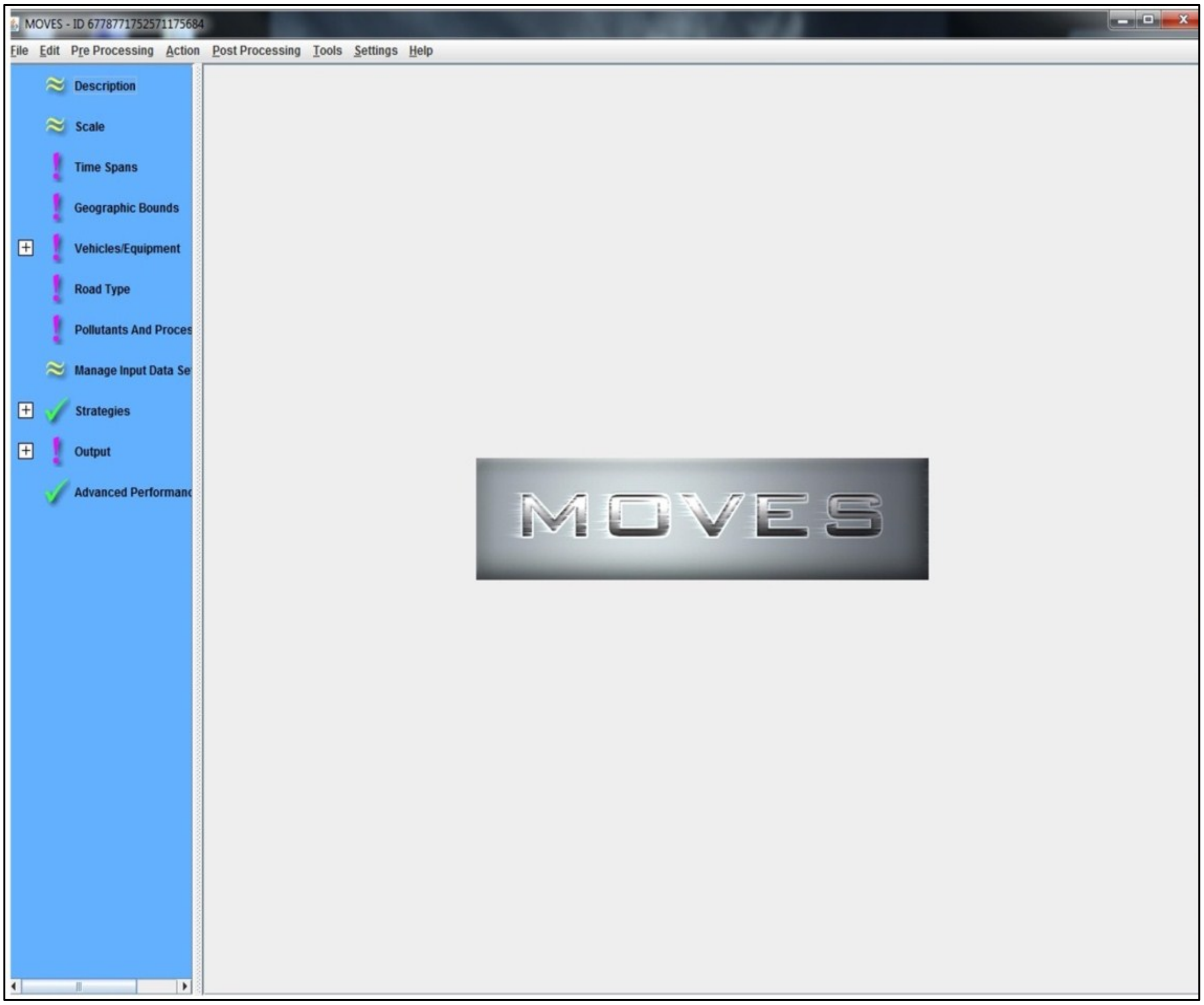

Figure 3-5: MOVES Graphical User Interface

Application of proposed models results in the required input data for the MOVES model at the required level of detail. The $PM_{2.5}$ emission rates are disaggregated by process, road type, vehicle type, link type, average speed (bin) and temperature (MOVES User Guide, 2010). In addition to rates per distance (running emissions), rate per vehicle (start and idling emissions) (no $PM_{2.5}$ evaporative emissions) are also obtained from the MOVES model runs (US-EPA 2010).

3.9.1. Temporal Allocation of Running Emission Inventory

The emission model could only provide us with emission rates based on link type, hour, average speed, vehicle type etc. To find the ambient air quality using dispersion models we have to supply spatial and temporal details of emission inventory as input. The link level hourly mobile source running emission inventories are obtained multiplying emission rates with corresponding activity by vehicle mix which is obtained from travel demand model.

3.9.2. Temporal Allocation of non-Running Emission Inventory

In this study, an alternative method was developed to allocate non running emissions to link level emission inventory due to starts and evaporation

$$M_{ij} = \sum_v \frac{VMT_{vij}}{VMT_{vj}} (S_j + E_j) \qquad \textbf{Eq. 19}$$

Where,

M_{ij} = Non running emission inventory for link i at hour j

VMT_{vij} = vehicle miles traveled by vehicle type v on link i during hour j

VMT_{vj} = vehicle miles traveled by vehicle type v in whole study area during hour j

S_j = Total start emission inventory in whole study area during hour j

E_j = Total brake wear and tire wear emission inventory in whole study area during hour j

3.10 Dispersion Modeling of Mobile Source $PM_{2.5}$ in Ambient Urban Atmosphere

To estimate pollutant concentration in urban atmosphere it is necessary to simulate its dispersion in air. Since 2005, the EPA adopted the AMS/EPA Regulatory Model (AERMOD) as the regulatory model for pollutant concentrations, which can estimate/predict the downwind concentration of air pollutants/toxins emitted from sources such as industrial plants, vehicular traffic and accidental chemical releases using the mathematical simulation. These regulatory models are typically employed to determine whether existing or future air quality is or will be in compliance with the National Ambient Air Quality Standards (NAAQS) in the United States and other nations. The AERMOD model replaces earlier the Industrial Source Complex (ISC) model for air dispersion modeling (US-EPA 2005).

AERMOD is a steady-state plume model that incorporates air dispersion based on planetary boundary layer turbulence structure and scaling concepts and can include treatment of both surface and elevated sources and both simple and complex terrain. But AERMOD does not account for chemical interactions or secondary formation; however it can be used to estimate the relative contributions of particular sources to the receptor-estimated ambient concentrations. AERMOD also allows for the estimation of the concentration impacts of potential control strategies on sources without the need to re-run the emission reductions through the model. It generates daily, monthly as well as annual concentrations of pollutants in ambient air. AERMOD is actually a modeling system with three separate components: AERMOD (AERMIC Dispersion Model), AERMAP (AERMOD Terrain Preprocessor), and AERMET (AERMOD Meteorological Preprocessor). AERMOD requires two types of meteorological data files, a file containing surface scalar parameters and a file containing vertical profiles. These two files are produced by the AERMET meteorological preprocessor when provide the typical daily meteorological data obtained from Nation Weather Service for study area AERMAP terrain-preprocessor can be used to generate hill height scales as well as terrain elevations for all receptor locations. The modeling AERMOD is carried out using different files called pathways.

Major pathways in this process are Control pathway, Source pathway, Receptor pathway, Meteorology pathway and Output pathway.

3.10.1. Conversion of Link-based Emission Inventory to Area Source Emissions

For air quality modeling purposes, the mobile emissions needs to be converted into gridded emissions rather than leaving them as line source emissions, since AERMOD can only take area source emissions as input. The trip-end (starts/parks) emissions are assigned to the surrounding links, as discussed above. Since link level emissions include running and non-running emissions combined in present methodology, the gridded emissions are more accurate compared to other methodology which allocate non running emissions to TAZ-centroid (Niemeier and Zheng, 2004). In this proposed process, the emissions from links can be assigned to the appropriate grid cells, given the coordinates of link nodes and other location information from regional highway network. Specifically, the proposed methodology is better compared to other methods since temporal emission inventory distribution has been calculated before converting them to gridded emissions.

Generally, it would be ideal that TAZs and grid cells have a comparable size. Although TAZs in suburban and rural areas are much larger than those in urban areas and central business districts (CBDs), actually the urban TAZs contribute more to urban air pollution. In addition, most of TAZs do not have a regular geometric shape and, thus, the length of one side is not necessarily larger than 1 km although the area of this TAZ is possibly much larger than sq. km (all of the dispersion models use distance units as km). Therefore, the 1x1 km grid cells are reasonable in terms of resolution, and the spatio-temporal $PM_{2.5}$ emissions at this grid level will be generated. For this purpose a geo processing script is developed which would calculate spatial and temporal $PM_{2.5}$ emission inventory for AERMOD purpose.

3.10.2. Meteorological and Terrain Data for Dispersion Model

As discussed earlier, to process the typical daily meteorology data and convert it into AERMOD's required format, the preprocessor AERMET is used. In fact we have to supply an annual cycle of local or regional meteorological information to predict the pollutant dispersion. The typical meteorological data set can be developed using data from National Weather Service (NWS) for the model area. This meteorological dataset provides the following hourly inputs to AERMOD: the hour of day, wind direction, wind speed, ground-level ambient temperature, atmospheric stability class etc. In addition to this we need some more meteorological input data like albedo (solar reflection), Bowen ratio (moisture term) and roughness height (surface roughness) which can inputted using technical guidelines provided by EPA.

It is to be noted that the terrain will affect air quality concentrations at individual receptors/source, and thus the AERMAP model first determines the base elevation at each receptor and source. For complex terrain situations, AERMOD captures the essential physics of dispersion in complex terrain and therefore needs elevation data that convey the features of the surrounding terrain. The AERMAP model needs standardized computer input files of terrain data and this proposed modeling methodology we use the Digital Elevation Model (DEM) format.

In the current research we are going to use the commercial version of AERMOD software, which is called AERMOD View. This software is released by Lakes Environmental Software LLC, and they have created very user friendly Graphical User Interface as shown in the following figure. As part of software we also obtained licenses for software like AERMAP, AERMET View which are used to process and view terrain and meteorological data. There were options to create input files from processed data and the detailed steps are discussed in "Dispersion Modeling" chapter.

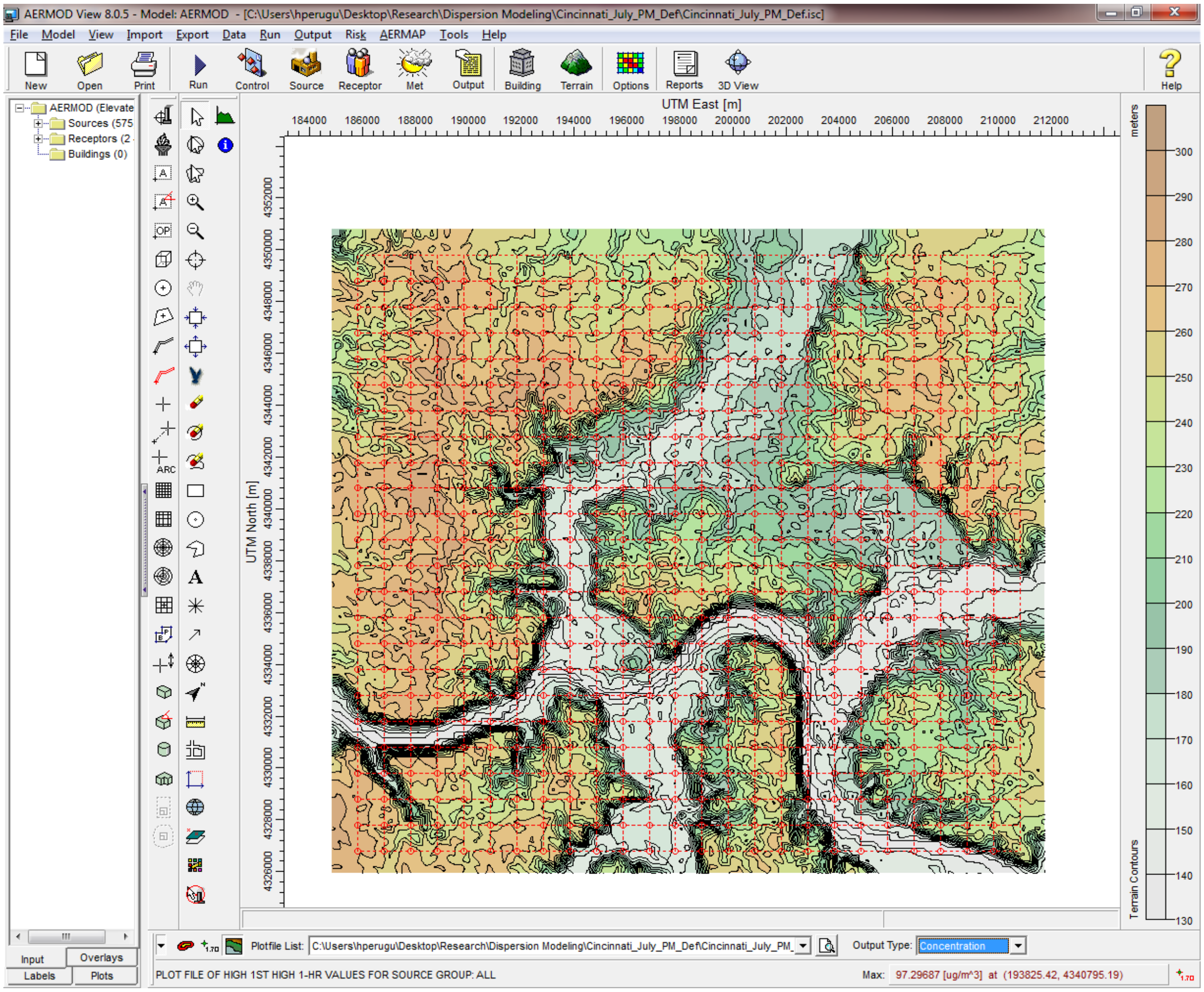

Figure 3-6: AERMOD View Graphical User Interface

3.11 Estimating Contribution of Trucks to $PM_{2.5}$ Pollution in Urban Atmosphere

The US-EPA measures hourly pollution data for monitoring stations throughout the country (US Environmental Protection Agency, 2007b). In this methodology we plan to use observed data from such air quality stations in the modeling domain to represent the urban $PM_{2.5}$ pollution levels. The ambient concentrations are calculated based on the receptor measured data for model year. Thus, we can compare the mobile source-based $PM_{2.5}$ concentrations with the ambient measurements and eventually identify the contributions of truck activity to urban air pollution. However, there are some limitations with respect to the receptor dataset; e.g., not every station has data for all pollutants, some stations do not have any data, and some measured data are not good quality. So, some aberrations are expected in this process.

In summary, we discussed the methodology to model detailed hourly truck activity and corresponding emissions in the first part of this chapter. Later, we have explained how to model dispersion of truck related $PM_{2.5}$ pollution using AERMOD model.

Chapter 4: Data Description

4.1 Study Area

To test and validate the proposed methodology we considered Cincinnati as a case study. The Greater Cincinnati area includes eight different counties. For travel modeling and census data collection purposes, it is called the Ohio-Kentucky-Indiana Metropolitan Statistical Area. This region is comprised of around 50 different jurisdictions and is the largest metropolitan area in Ohio. The total area of Cincinnati is 2,619 square miles, and the population density is 756 persons per s square mile.

Large industries like GE Aviation, AK Steel, and other big logistics services hubs contribute to a lot of truck activity in the region. According to Federal Highway Administration (FHWA) estimates, I-75 is one of the busiest trucking routes in North America with truck traffic approaching six billion miles annually. I-75 goes through the middle of the region, contributing to frequent traffic congestion and deteriorating air quality around city of Cincinnati. Within the city, industrialized areas like Evendale, Batavia, Hebron, and Middletown generate lot of truck traffic. I-75 and I-71 merge near the Brent Spence Bridge, creating a major bottleneck in the region and further worsening traffic congestion. The total annual vehicle miles traveled in the region are 17.5 billion miles. According to the Mobility Report, the travel time index of the region is 1.17 and one of the lowest in the Midwest region (ranked 45th in the nation). Other important interstates like I-74, I-71, and I-275 also go through the urban area.

Cincinnati is the 15th highest $PM_{2.5}$ emitting city in United States, according to US-EPA's National Emission Inventory. The average per capita $PM_{2.5}$ emission of Cincinnati was approximately two tons each year. The transportation sector contributes 75 percent of the total $PM_{2.5}$ emissions in the city According to US-EPA's designation; most of the region is under

$PM_{2.5}$ maintenance. However, the latest observation of $PM_{2.5}$ concentrations shows an upward trend. The county and regional boundaries of the area are shown in the following map.

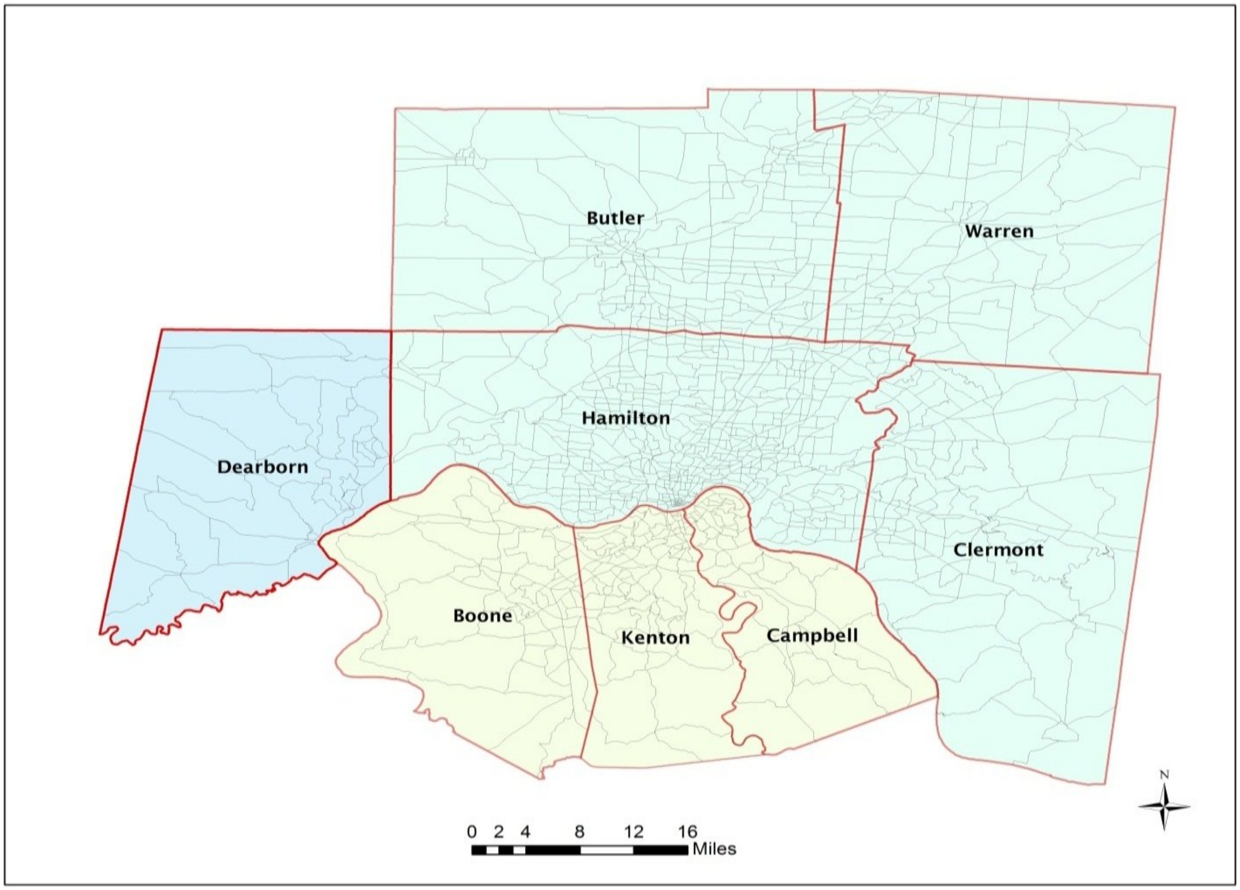

Figure 4-1: Greater Cincinnati Area

To model regional travel activity (including truck travel), OKI has developed a traditional four-step travel demand model for the region, and they continually improve the model. The region was divided into 1352 traffic analysis zones for the travel demanding modeling purposes as shown in the map above. The existing model also predicts link truck volumes. For this purpose the synthetic truck matrix-based approach is followed. The synthetic matrix method synthesizes a pseudo truck trip matrix from old traffic data. However, the model is validated through comparing the total daily traffic output with annual average daily traffic volumes. The output from the travel demand model is used in air quality estimation, transportation conformity, the congestion mitigation process, environmental justice processing, etc. To estimate emission inventory for transportation conformity and State Implementation planning, the emission rates from MOVES are multiplied with corresponding vehicle miles traveled. The emission totals are estimated at county-level aggregation.

As discussed in the methodology chapter, we have used the spatial regression and optimization truck modeling methodology to predict link truck volumes. The next step in this process is to use a separate hourly VMT distribution model to disaggregate the daily truck miles traveled into hourly values. Then, the hourly link level $PM_{2.5}$ emission quantities are calculated by looking up corresponding $PM_{2.5}$ emission rates from MOVES output database. The link level emission quantities are aggregated into grids for dispersion modeling purposes. Regional meteorology and detailed terrain data are also used in the dispersion modeling step. In the present research, we have also prepared an emission inventory using existing truck model output. Finally, we have compared the results using the default truck model methodology with proposed SROT truck model with observed $PM_{2.5}$ concentration at two different monitoring stations.

In this research, we have focused on estimating the $PM_{2.5}$ pollution in the urban atmosphere. So the selection of the study domain must cover only the urban area. Further, the domain selection has to be a regular shape so that it is straightforward to supply area source emission inventory input to the pollution dispersion model. The selected domain is 25 km X 23 km rectangular domain as shown in the Fig. 4-2. To develop a strong spatial regression model, we need sufficient traffic count samples to cover the domain. However, due to insufficient traffic count samples for the analysis time period, we have used traffic counts outside the domain.

4.2 Traffic Count Data Description

The Ohio-Kentucky-Indiana Regional Council maintains comprehensive traffic count geo-databases for the purpose of travel-demand validation and independent traffic count requests. As mentioned earlier, the traffic count data is stored in geo-databases in the form of Annual Adjusted Daily Traffic data at hourly time resolution. The AADT values are factored through applying day-of-week, month-of-year, and other factors, developed by the Ohio Department of Transportation, since the short period traffic counts are typically collected during only a single week (Monday-Thursday). The type of spatial database used by OKI is: ArcGIS personal geo-database type. In the geo-databases, location information is stored as a feature dataset and traffic count data as tables. These tables form the relational databases. In other words, these tables are interconnected, and we can perform spatial queries on this data.

There were actually 8,000 unique location data in the database, and there is at least one AADT count available for each of these stations. Only the past 10 years of data is available in this database. As mentioned earlier, since there were not enough locations within the modeling domain, we have used traffic count data from whole region. Even though there are 8,000 spatially unique locations in the databases, they do not have continuous data. Since,\ we need temporally continuous data for the spatial regression model development, we have selected only 1,000 locations. We have also taken care to ensure that there is enough coverage of the modeling domain. In the current study, we have used only data from the years 2004-2008. The total sample size in this study is 3,000 (i.e. 1,000 locations x 3 analysis years). The coverage of traffic counting locations is shown in the Fig. 4-2.The counts covered all the functional roadway classes, and a combination of land-uses, with intent to obtain a sample representative of the TMT mix in the region. About 75 percent of these locations are in urban or suburban areas. Some of the data had 15-minute time resolution, and we have converted all of the samples to one type time resolution.

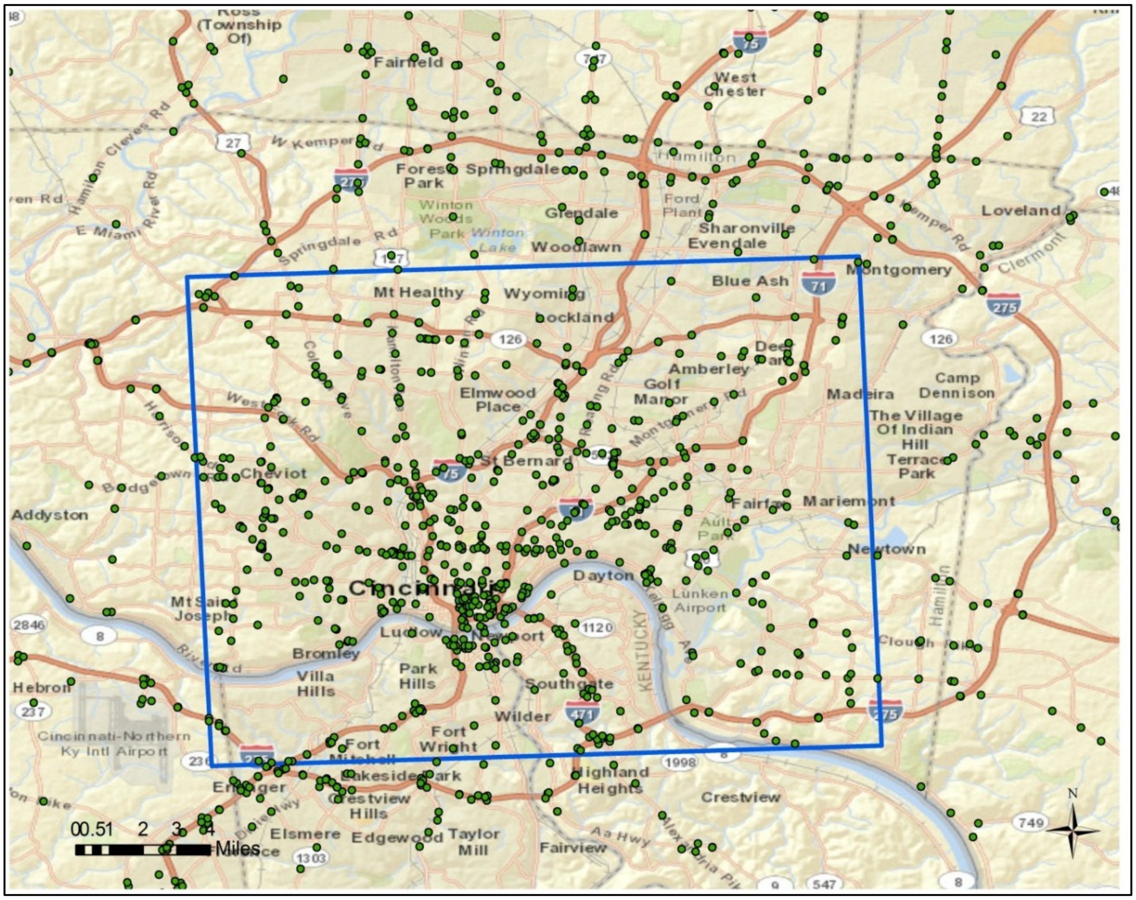

Figure 4-2: Traffic Count Locations Used in the Study

4.3 Traffic Count Data Processing

In this study, we have used short-term coverage traffic counts and permanent station traffic data. These counts are primarily used to supply input into the Highway Performance Management System, and thus use the FHWA's 13-type vehicle classification as shown in the Fig. 4-3. However, as mentioned in the first chapter, the emission models require data by different type truck classification. To model the emission model ready truck activity, we need to process the FHWA type vehicle classifications into US-EPA's MOVES truck classification. Unfortunately, until now, we have not found relevant literate to convert FHWA classes into MOVES types. To work around this, we have used FHWA versus MOBILE and MOBILE versus MOVES crosswalk tables to come up with a crosswalk between FHWA and MOVES classifications (Trevor, et al).

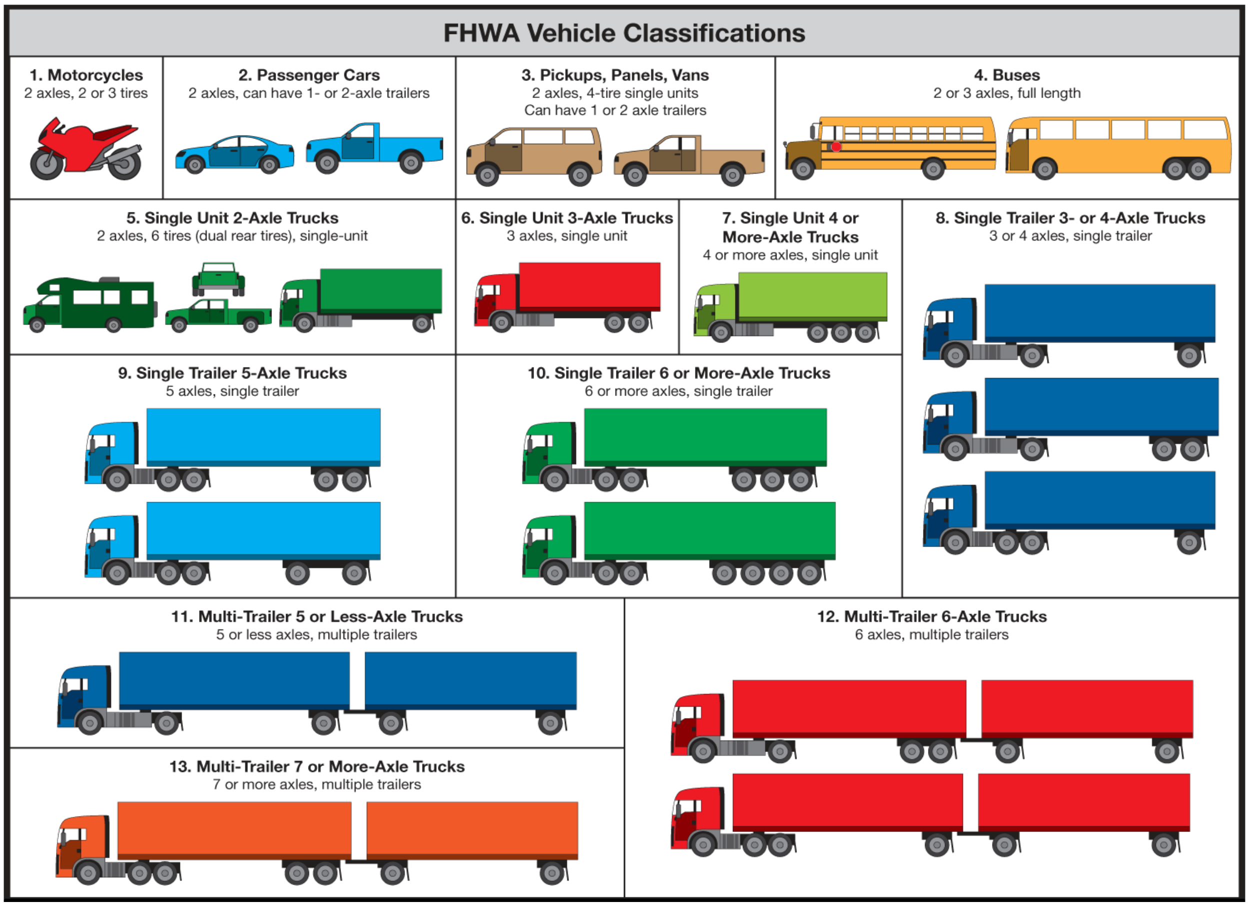

Figure 4-3: FHWA 13-Type Vehicle Classification

Even though both the classifications are axle based, there is no direct corresponding truck type for US-EPA's motor home type trucks in the FHWA classification. For this purpose, we have developed a corresponding fraction using MOVES default data. The relative fractions are shown in Table 4-1.

Table 4-1: Crosswalk between MOVES Source Types and FHWA Classification

MOVES Type	FHWA Types								
	5	6	7	8	9	10	11	12	13
Refuse Truck	0.853	0.006	0	0	0	0	0	0	0
Single Unit Short-Haul Truck	0.097	0.944	0.135	0	0	0	0	0	0
Single Unit Long-Haul Truck	0	0	0.121	0.006	0.006	0.006	0.007	0.007	0.007
Motor Home	0.05	0.05	0.491	0.282	0.121	0.006	0	0	0
Combination Short-Haul Truck	0	0	0.253	0.712	0.873	0.891	0.09	0.09	0.09
Combination Long-Haul Truck	0	0	0	0	0	0.097	0.903	0.903	0.903

Table 4-2 provides the descriptive statistics of these traffic counts over different years, and these data are the basis for the spatial regression modeling pursued in the present study. As expected, on average, the automobile fractional split is highest, followed by the fraction of pick-ups and vans. The average regional percentage of trucks is between 3 to 5 percent. However, at an individual link level, the truck percentage is as high as 26.3 percent. The fraction of buses and motorcycles in the vehicle stream is relatively low. The percentage of observations for which the fractional mix of trucks, buses, and motorcycles is at or very close to the boundary value of zero is rather high. In particular, the truck fraction is less than 0.01 for 33 percent of observations and the bus fraction is less than 0.01 for 99 percent of observations. The single unit and combination short-haul trucks are the most prevalent truck types in the region and comprise about 60 percent of total trucks found on the highways in a day. Motor homes are the least prevalent trucks in the region and they comprise of only 2 percent of total trucks on highways.

Table 4-2: Summary Statistics of Explanatory and Dependent Variables

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
Number of Lanes	3000	1.66	0.8415	1	5
Capacity	3000	876.22	276.94	480	2000
Functional Class	3000	14.81	3.15	2	19
Speed	3000	29.56	8.62	15	70
Population	3000	7893.94	39501.81	0	45803
Employment	3000	3551.98	3483.36	0	17752
Accessibility	3000	0.00021	0.00026	1.00E-05	0.00216
Single Unit Short-Haul Trucks	3000	480.10	822.86	15	1661
Single Unit Long-Haul Trucks	3000	135.32	219.31	1	2287
Motor Homes	3000	24.45	39.65	1	377
Combination Short-Haul Trucks	3000	450.62	1357.40	1	8563
Combination Long-Haul Trucks	3000	56.88	157.85	1	2167
Refuse Trucks	3000	277.21	306.57	1	966
Years	2004, 2006, 2008				

As mentioned earlier, five varieties of explanatory variables are included in truck volumes on links, and this selection is based on guidelines provided in Bhat, et al. These variable groups are: a) link functional classification, b) link physical attributes, c) link posted speed variables, d) accessibility of the zone (in which link lies), and e) zonal land-use characteristics. We started with a large number of variables within each of the five variable classes, and the final set of variables that were included in the model was determined based on a systematic process of eliminating them through covariance analysis in the previous specifications. In the description below, we briefly highlight the characteristics of the variables in each of the five sets of variables that were retained in the final model specification.

4.4 Demographic, Land Use, and Network Data Description and Processing

In this study, we have used different socio-economic, land-use, and highway network attributes as explanatory variables for model estimation. The data used to develop the independent variables is extracted from other geo-databases. The demographic information is developed from Public Use Micro Data Sample database and land-use data collected from different county offices. As part of travel demand model maintenance, OKI also collects latest roadway attribute data. Following is the final list of predictor variables used in the statistical analysis with their designation name in the database in parentheses:

Population (POPDEN): Using PUMS and American Community Survey data, the regional demographer compiles population data for future years. The population is aggregated to TAZ level to facilitate the trip generation in the travel demand modeling practice. Fig. 4-4 shows the map of population distribution among the modeling domain. During covariance analysis, we observed that the correlation between population and other important variables like employment is significant whereas the population density and employment density variables are independent. Thus, we have chosen population density as a predictor variable. We have geo-processed the TAZ-level employment density to join link-specific traffic counts.

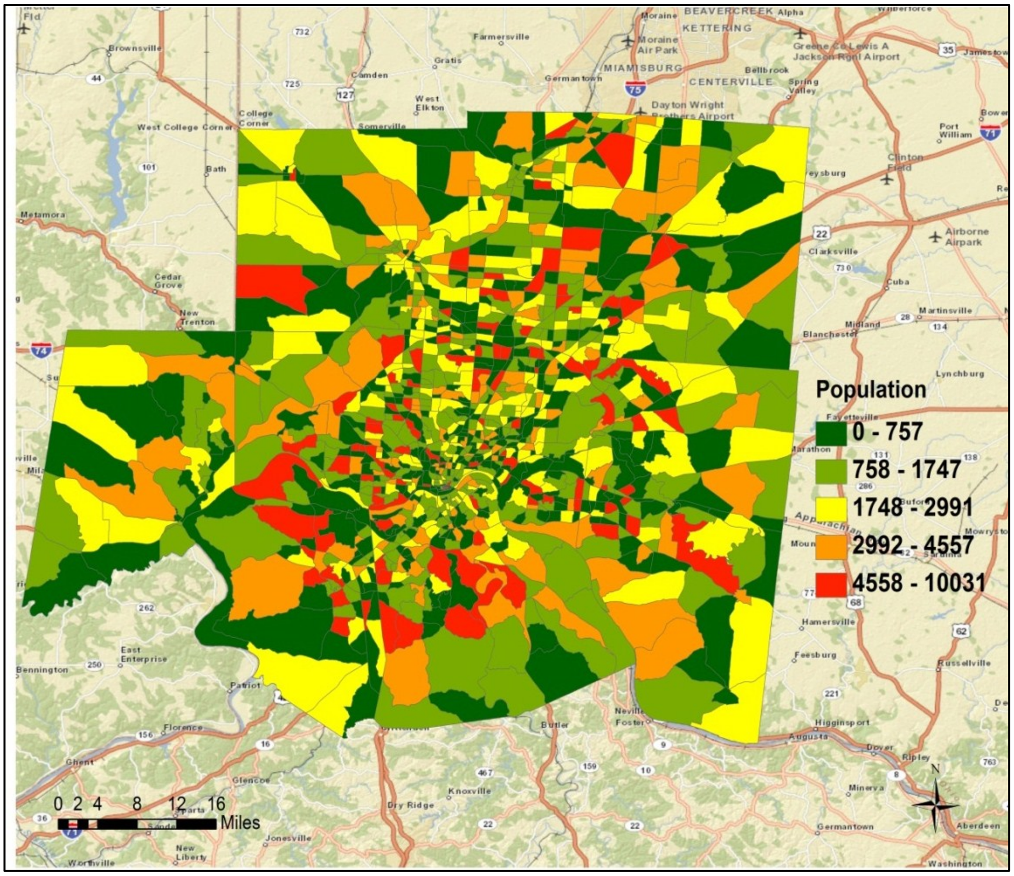

Figure 4-4: The Population Distribution in the Study Area

Employment (EMPDEN): Similar to earlier variable, employment is a critical demographic variable which is directly proportional to total trips in a TAZ. The employment data also extracted from ACS and PUMS data; it is projected for future years based past trends. The employment distribution in the region is shown in the Fig. 4-5. The population and employment in a TAZ are auto correlated, thus, we have chosen employment density as predictor variable. Further, from other regression based truck models it is evident that employment in a zone is positively correlated to the truck inflow and out flows of the zones.

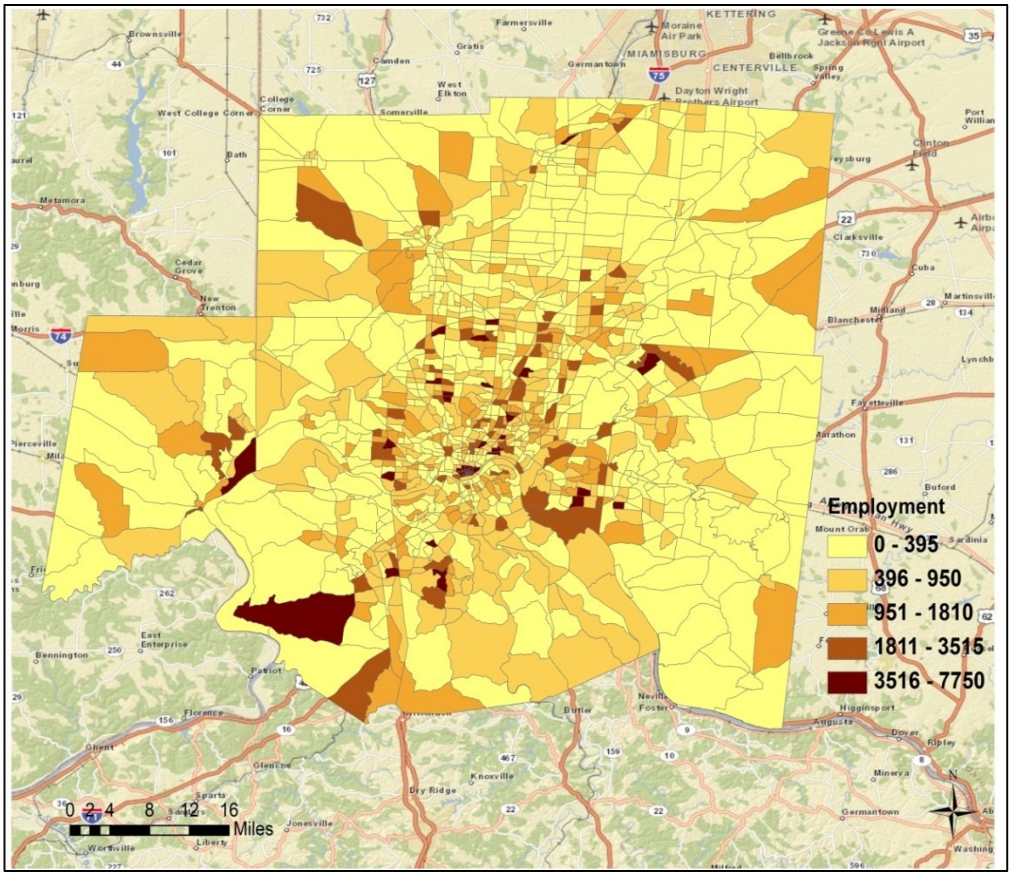

Figure 4-5: The Employment Distribution in the Study Area

Accessibility (ACCS): Some of the earlier vehicle volume regression models used TAZ accessibility as a predictor variable. However, they have used nominal values for these variables. In this research, we have calculated accessibility of zone using TAZ area and total highway network length in the zone. To estimate this variable, the data from two different geo-databases were used. The first one is demographic geo-database, which provided the TAZ/zonal area, and second one is highway network geo-database, which provided highway link length in the corresponding TAZ. The highway geo-database is developed based on data collected from Ohio DoT and other county engineer offices. This geo-database contains information like posted speed limits, number of lanes, lane width, type of intersection control, lane configuration, etc. for all highway links in the region. The highway network data map is shown in Fig. 4-6. The

Metropolitan Area (MA) includes more than 40,000 unique roadway links to represent the roadway network.

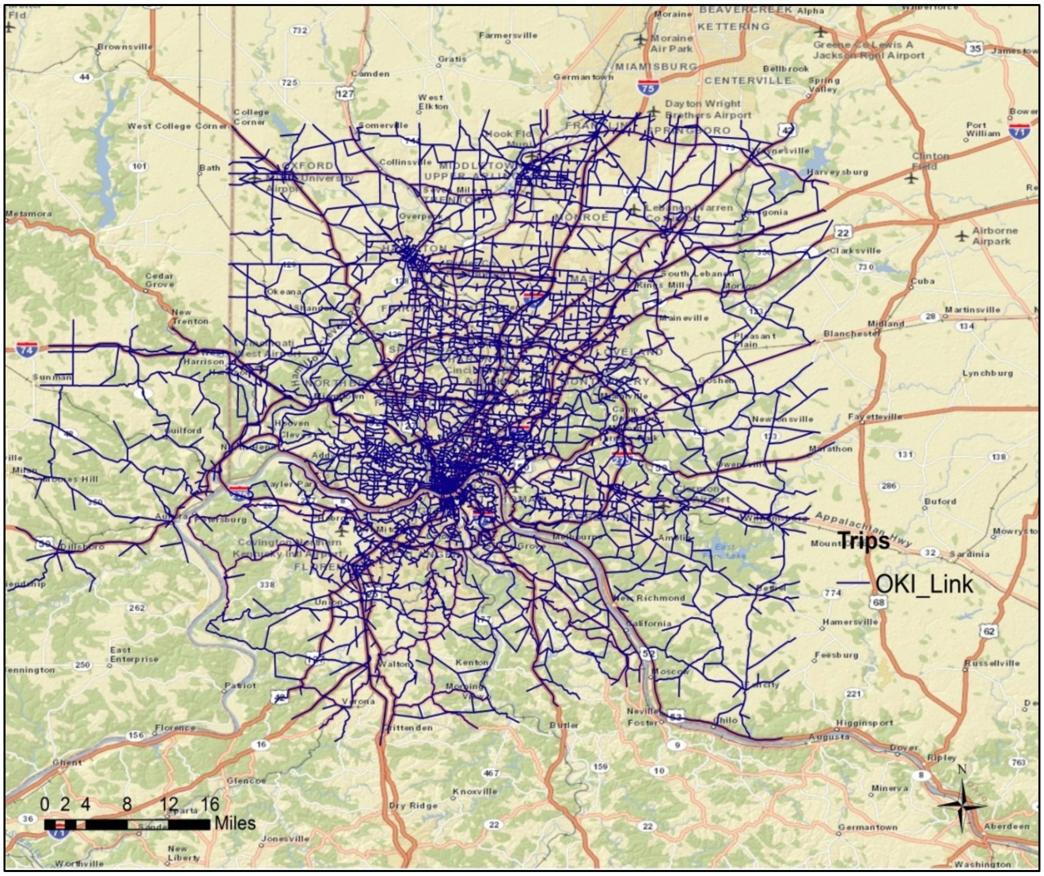

Figure 4-6: Highway Network Information used in the Study

Speed (SPD): The posted speed limits or free speeds on links used for regression modeling are extracted from the highway network geo-database. It has been observed from initial analysis that trucks use the highway facilities that have higher posted speed limits. Typically, posted speed limits in the region range between 15 mph and 70 mph. The mean speed in the region is 29.56 mph. Forty percent of the total links have posted speed limits less than 35 mph At least 25 percent of links have coded speeds greater than or equal to 55 mph. The rest of the samples have speeds varying between 35 mph and 55 mph.

Capacity (CAP): From a traffic engineering point of view, it is anticipated that trucks may be using the highway facilities with higher capacities. During covariance analysis, we also observed that there is no auto correlation between speed and capacity. This information is also obtained from the highway network geo-database. Link capacity is estimated based on various factors like area type, facility type, control type, lane width, etc. The capacity values for links range between 480 and 2,000, and the mean value is 276.94. Freeways and expressways have the highest capacity, and this independent variable does not have any correlation with other independent variables.

Number of Lanes (LANES): From a traffic engineering perspective, trucks need more room for better maneuverability thus the link physical attributes like the numbers of lanes are important determinants of truck volumes on the link. The highway network database also contains the link's number of lanes data. The link information is by direction. A majority of links (54.6 percent) in the sample have two lanes; 10.2 percent of links have one lane; 24.2 percent have three lanes; and 11 percent have four lanes.

4.5 Truck Age Distribution Data

For emission modeling purposes, we need accurate enough truck age distribution fractions. Age distributions can be extracted using the registration data from the Department of Motor Vehicles. In the present study, we have used Hamilton County registration data. The DMV also provided an appropriate conversion table to convert different truck varieties into MOVES truck types. The detailed truck age distribution and population information used in this study is discussed in the emission inventory modeling section.

4.6 Fuel Data

For accurate estimation of emission inventory, we need precise fuel composition and type of fuel usage by season. Such data is obtained from the Ohio EPA. The most important aspect of this data is that it actually provides the chemical constituents of diesel fuel used for trucks. The fuel types available during different months of a year are different, thus their chemical constituents also change. All of these details are available in the fuel database. Specific details about the data used for this research are discussed in the emission inventory modeling section.

4.7 Meteorology Data

A number of different sets of surface meteorological data were available for the modeling period, including data from the National Weather Service (NWS), high schools, and state agencies. Due to inconsistencies in the data records, data quality issues, and reformatting concerns eliminated several of these sources. Data from two NWS sites at local airports (Lunken Airport and CVG Airport) the Cincinnati area had good coverage and instrumentation. However CVG Airport did not fall in the modeling. So we used data collected at Lunken Airport in the study. The final selection of the parameters collected is shown in the following list. The abbreviations are:

WS = wind speed

WD = wind direction

ST = sigma theta

T = temperature

CH = ceiling height

DP = dew point temperature

P = precipitation

SP = surface pressure

CC = cloud cover

CH = ceiling height

Upper air data is used in AERMET to define the mixing heights and wind fields above the surface. The nearest NWS upper air observation site is also at the Lunken Airport, and the data collected includes wind direction, wind speed, temperature, pressure, and geopotential height. The geophysical data is used by AERMET to adjust the observed met data from the six surface sites and extend the data to produce wind fields that cover the entire domain. The geophysical data includes terrain elevations and land use. The terrain elevation data helps characterize the wind patterns as they meet varying topography within the domain, including the up-slope flows during daytime surface heating, and down-slope flows that occur with nighttime surface cooling. Classifications of land use (for example, forest, residential housing, water bodies) allow the model to adjust flows to reflect the mixing and turbulence effects near the surface and to characterize the vertical mixing that varies with rising air masses from different levels of surface heating.

4.8 Terrain Elevations

The terrain data were derived from the USGS digital elevation model (DEM) with 7.5- meter horizontal resolution. These elevation data were used as inputs to AERMET, and as elevations for the receptors used in AERMOD. Other details about how terrain was processed for area sources and receptor grid are explained in Chapter 7.

In this chapter, we have explained how we collected and processed the following data: the traffic count data, demographic and highway network used for truck models; truck age distribution and other fuel information data for emission model; and meteorology and terrain data for dispersion models.

Chapter 5: Modeling Link Level Hourly Truck Activity

As explained in the introduction, we have observed spatial auto correlation among traffic count data using Moran' I method. Further, the hourly traffic flow variation can only be modeled using the advanced fractional response modeling methodology. To estimate hourly link-level truck vehicle miles traveled, we use Spatial Regression and Optimization (SROT) Model output, corresponding link lengths, and hourly activity fraction model output. In this chapter we have presented detailed steps of the SROT Model and hourly fraction model application and corresponding empirical results for modeling domain.

5.1 Spatial Regression and Optimization Truck Model Application

To model the link daily truck volumes, we used the two-stage spatial regression and optimization modeling methodology. The first stage is developing the statistical model using regional data and applying the model to estimate truck volumes at the "control locations." The second stage is estimating link truck volumes using trip cost matrix and controlling totals. To apply the SROT model for the current study area and validate the results, the following tasks are carried out.

1. Synthesizing the control locations/links using screen line information.
2. Development of the regional spatial regression models for all six truck types as needed by the emission factor model.
3. Applying the spatial regression models to estimate "control truck totals" at the control locations.
4. Estimating truck trip matrices using trip cost matrices and the control totals in the optimization model.

5. Calibrating the truck trip matrices using the Freight Analysis Zone[5] level truck matrices.
6. Assigning truck trip matrices using equilibrium assignment, and
7. Validating the truck models through comparing with latest individual link-level truck counts.

In this study, the screen lines are drawn / imagined to represent the boundaries of the domain. In addition to these four screen lines, we have also considered three more cordon lines to increase number of control locations. The screen lines are shown in the following map (Fig. 5-1). Using ArcGIS software geo-processing tool (Intersection), we have synthesized the control locations as shown in the Fig. 5-2. However, the redundant control locations are consolidated using another ArcGIS tool. The truck volumes at these locations are estimated using the spatial regression model

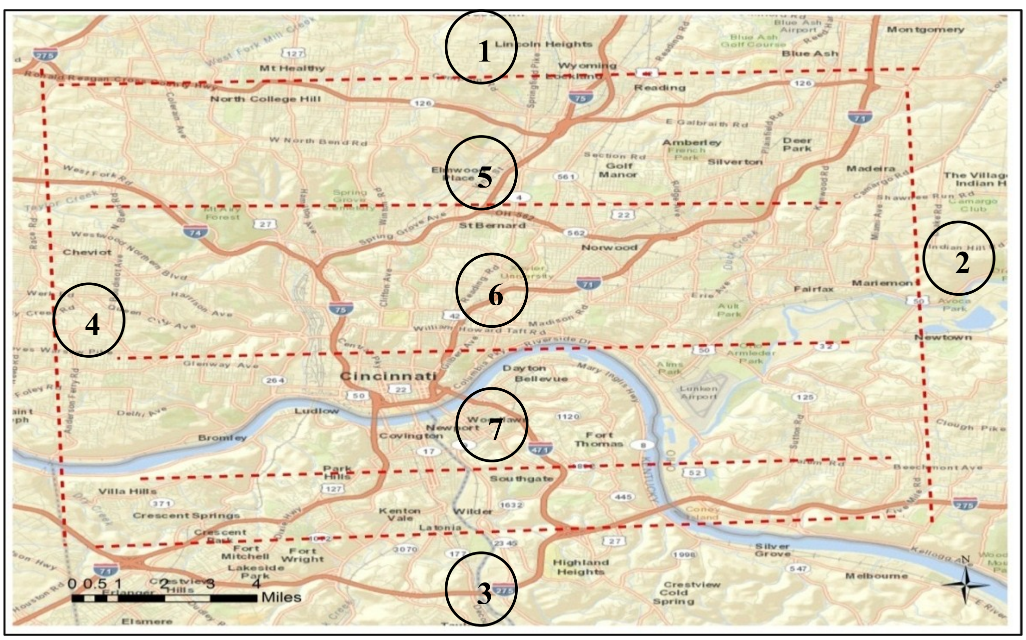

Figure 5-1: Screen-line Information used in the Study

---

[5] Freight Analysis Zone is agglomeration of Traffic Analysis Zones for freight modeling purpose.

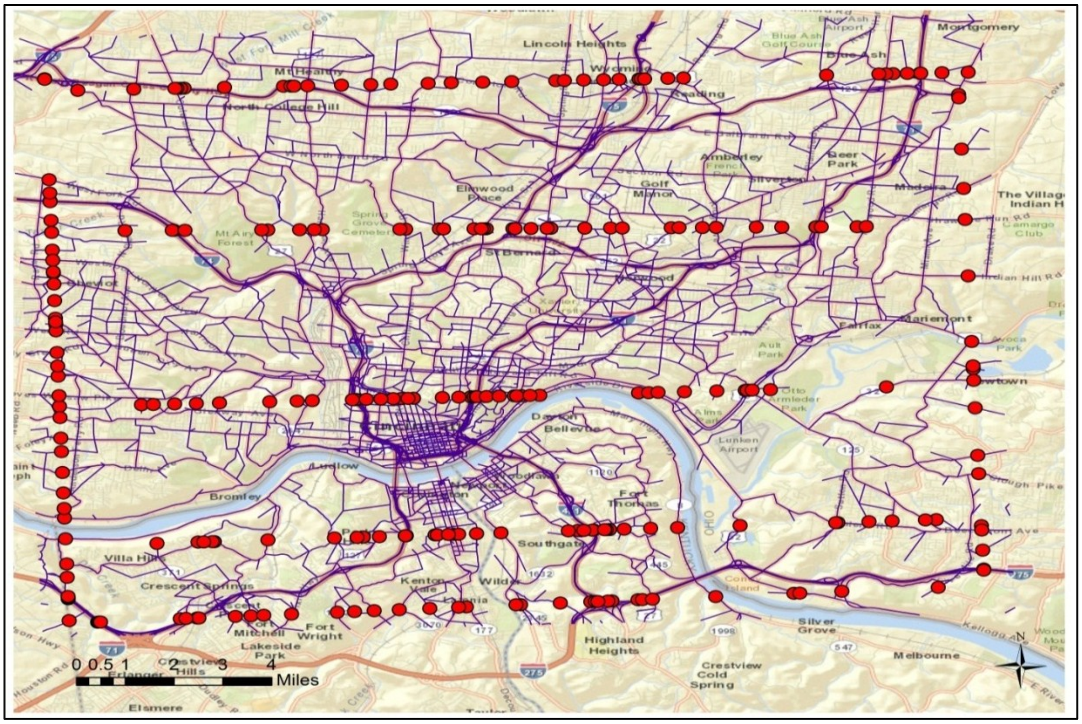

Figure 5-2: The Control Locations used in the Truck Models

5.2 Empirical Results of the Spatial Regression Model

As discussed in the methodology chapter, the statistical models have been specified using the spatial panel methodology used by Kapoor, et al. To estimate model parameters we have also used their methodology. The important parameters to be estimated in this modeling methodology are: the spatial auto correlation, unit-specific error, and time-specific error. The estimation is performed using Stata® software through a user-developed program (Prucha 2011).

We have developed six different models for six truck types Viz. refuse trucks, motor homes, single unit short-haul, single unit long-haul, combination short-haul, and combination long-haul truck types. For estimation purposes, we need the spatial weightage matrix, which has been created using the ArcGIS geo-processing tool. All of the variables have been converted into matrices for regression-process facilitation. The code used for this estimation has been included in Appendix II. For model parameter estimation, we used Ordinary Least Square and Feasible Generalized Least square methods, and the comparison results are shown in Table 5.1 -Table 5.6.

Table 5-1: Spatial Regression Model for Refuse Trucks

Refuse Trucks		
Variable	**OLS Estimator**	**FGLS Estimator**
Constant	-5.50E+01	3.71E+01
Number of Lanes	7.27E+01	6.75E+01
Capacity	2.04E-01	1.73E-01
Speed	-2.67E+00	-2.88E+00
Employment Density	1.53E-04	1.95E-04
Population Density	6.09E-03	8.87E-03
Accessibility	8.41E+04	6.70E+04
RMSE	647.87	1.74
Spatial Auto Correlation Parameter	-	0.608
Time-Specific Error Variance	-	102837
Observation-Specific Error Variance	-	146727.8

Table 5-2: Spatial Regression Model for Motor Homes Type of Trucks

Motor Homes		
Variable	**OLS Estimator**	**FGLS Estimator**
Constant	2.11E+00	1.14E+01
Number of Lanes	7.31E+00	7.47E+00
Capacity	1.77E-02	1.25E-02
Speed	1.99E-01	1.91E-01
Employment Density	-5.58E-05	-5.07E-05

Population Density	-1.48E-03	-1.19E-03
Accessibility	9.40E+03	6.97E+03
RMSE	65.95	1.78
Spatial Auto Correlation Parameter	-	0.494
Time-Specific Error Variance	-	593.8
Observation-Specific Error Variance	-	2492.9

Table 5-3: Spatial Regression Model for Single Unit Short-Haul Trucks

	Single Unit Short-Haul Trucks	
Variable	**OLS Estimator**	**FGLS Estimator**
Constant	-3.15E+02	-9.10E+01
Number of Lanes	1.09E+02	1.00E+02
Capacity	8.15E-01	6.19E-01
Speed	-2.92E+00	-4.11E-01
Employment Density	-2.03E-03	-1.70E-03
Population Density	-3.46E-02	-2.64E-02
Accessibility	2.15E+05	1.48E+05
RMSE	1334.98	1.78
Spatial Auto Correlation Parameter		0.556
Time-Specific Error Variance	-	298812.1
Observation-Specific Error Variance	-	799983.8

Table 5-4: Spatial Regression Model for Single Unit Long-Haul Trucks

	Single Unit Long-Haul Trucks	
Variable	**OLS Estimator**	**FGLS Estimator**

Constant	-1.08E+02	-2.28E+01
Number of Lanes	5.62E+01	5.23E+01
Capacity	1.19E-01	8.45E-02
Speed	8.33E-01	1.38E+00
Employment Density	-6.15E-04	-5.37E-04
Population Density	-6.45E-03	-5.16E-03
Accessibility	3.03E+04	1.54E+04
RMSE	363.37	1.83
Spatial Auto Correlation Parameter	-	0.603
Time-Specific Error Variance	-	16049.584
Observation-Specific Error Variance	-	66569.388

Table 5-5: Spatial Regression Model for Combination Short-Haul Trucks

	Combination Short-Haul Trucks	
Variable	**OLS Estimator**	**FGLS Estimator**
Constant	1.64E+02	2.97E+02
Number of Lanes	7.10E+01	6.71E+01
Capacity	5.48E-01	4.61E-01
Speed	-3.02E+00	-1.47E+00
Employment Density	-1.10E-03	-8.80E-04
Population Density	-2.99E-02	-2.07E-02
Accessibility	-7.75E+04	-1.61E+05
RMSE	2322.07	1.74
Spatial Auto Correlation Parameter	-	0.312
Time-Specific Error Variance	-	1257789.1

Observation-Specific Error Variance	-	2614277.3

Table 5-6: Spatial Regression Model Combination Long-Haul Trucks

Combination Long-Haul Trucks		
Variable	**OLS Estimator**	**FGLS Estimator**
Constant	-2.21E+01	-1.01E+00
Number of Lanes	3.64E+01	3.51E+01
Capacity	4.63E-02	3.43E-02
Speed	1.19E-01	4.04E-01
Employment Density	-3.42E-04	-2.91E-04
Population Density	-5.03E-03	-4.28E-03
Accessibility	4.43E+04	3.03E+04
RMSE	264.82	1.77
Spatial Auto Correlation Parameter	-	0.431
Time-Specific Error Variance	-	9952.4
Observation-Specific Error Variance	-	42275.6

5.2.1. Discussion of Empirical Results

In the present analysis, we have estimated model parameters using two different methods: Ordinary Least Square Method and Feasible Generalized Least Square Estimation. These estimations provide us with enough confidence about the results. The Root Mean Square Error values also offer information about better parameters. Since we have used spatial regression modeling methodology in this research to support the claims, we have also calculated spatial auto correlation associated with each type of .trucks. The variance of error is an indicator of robustness of the regression model, thus we have estimated time-specific error and unit- specific error variances.

The lane configuration variable has a strong relation with all types of trucks. The number of lanes variable on the links is introduced with the single lane being the base category. The results indicate an increase in the single unit short-haul truck volumes on roadways that have a higher number of lanes. Interestingly, the number of lanes has the least effect on single unit long-haul trucks. For the rest of the categories, the effect of lane configuration is tempered, but not weak. The results of the effect of link physical attributes indicate an increase in truck volumes on highway facilities with several lanes.

The link capacity coefficients show fewer single unit long-haul trucks, combination long-haul trucks, and motor homes on low-capacity links relative to other types of trucks. In the case of refuse and single unit short-haul trucks, their relative presence on low-capacity links is high. As the motor homes and single unit-long haul trucks volumes are very similar. Their variable coefficients are similar.

The refuse trucks are mostly found on minor arterials or roadways with lower speeds from the model estimation. A similar kind of behavior is observed from combination short-haul trucks. Link speed has less effect on single unit short-haul trucks, and they are prevalent on higher speed facilities as well as medium speed facilities. Generally speaking, combination trucks and single unit trucks should be more prevalent on higher speed links than on lower speed links, and the model results also prove the same.

The coefficients of the variables characterizing accessibility of a zone show lower truck volumes on the links than those that exist in less-accessible zones. The single unit short-haul truck volumes are found to be higher compared to other truck types in dense (in terms of more roadway length per square mile) zones. The same is true for the single unit short-haul and combination short-haul trucks, except that this effect is weaker than long-haul trucks. The same, though more tempered, negative trend exists for combination short-haul trucks. The results also indicate that the motor homes are the least affected of the zone's accessibility since they are primarily used for recreational purposes and tend to use mostly expressways.

The final sets of variables are the land-use variables Viz. population density and employment density. The coefficients for employment density showed a strong implicit relationship between the average annual daily truck volumes and the employment. Specifically, both types of single unit and combination short-haul truck volumes are higher in the zones where employment density is higher. The motor homes and combination long-haul trucks had the least effect due to employment since these types are used for recreation and longer trips. The refuse trucks on links in low-employment density areas have higher volumes. This relationship can be explainable since garbage trucks tend to be prevalent in highly populated areas. So the population density parameters show higher refuse trucks in more populated areas. The single unit and combination short-haul trucks are fewer in populated areas, which are the opposite of dense employment areas. Motor homes are the least affected by population density due their longer trip lengths. Finally, population density has the least effect on single unit and combination long-haul trucks since they primarily serve manufacturing plants and warehouses.

The results show that single unit long-haul trucks have a stronger spatial auto correlation since the warehouses tend to be located closer together and those trucks are primarily used for conveyance. On the other hand, combination short-haul trucks have the least auto-correlated data, since combination short-haul trucks tend to travel expressways. When compared to OLS estimation, FGLS estimation showed less RMSE values, which indicates that the error associated with model output is very small. The model for motor home truck type has a very small variance for both time specific and observation, i.e. the model would not change spatially as well as temporally. The refuse trucks, single unit, and combination short-haul truck models showed high variability in model output, which means the output can change spatially and temporally.

5.3 Optimization Stage of SROT Model

As mentioned in the methodology, we used the optimization module to estimate the truck O-D matrices using truck volumes predicted from spatial regression. This model is implemented using Cube Analyst. In this software, we can set our optimization criteria such as the relative gap between consecutive estimated trip matrices and this criterion set as 0.1. Since the proposed model applies empirical model results and optimizes the O-D matrices, we have developed new computer scripts to automate the process every time we run the model.

The TRANSEARCH data consists of all modes of freight data by different commodity groups, which is collected using independent truck surveys for all counties in the United States. Another useful feature of this database is that it also includes empty truck trips, which are very important for validation and calibration of any truck model. The truck trip matrices prepared using the proposed truck demand optimization model are aggregated into freight analysis zone (bigger than traffic analysis zone) level matrices and compared with TRANSEARCH data. As part of calibration process the truck demand optimization model parameters such as a_i, b_j are adjusted to match the truck flows with TRANSEARCH data. The calibration process was carried out until the objective level confidence was achieved. The number of iterations for this step was relatively low per say 50. The implementation is particularly straight-forward using a GIS platform.

5.4 Highway Assignment Results

The results of the optimization module are in the form of trip matrices by the following types: refuse trucks, motor homes, single unit short-haul trucks, single unit long-haul trucks, combination short-haul trucks, and combination long-haul trucks. According to the traditional travel demand forecasting theory, these truck demand matrices obtained from the proposed optimization model are added to auto travel demand matrices. The total sets of matrices are assigned to the highway network to get a daily traffic assignment (volumes) by truck class. We have used the latest HCM methodology for multi-class equilibrium-based highway assignment. The analysis year 2010 is used in the truck volume prediction.

As mentioned earlier, the parameters in assignment model are borrowed from the OKI model. These parameters have been calibrated and validated using speed data from years 2002-2008. This step has been implemented using Cube ®'s highway assignment module. We have used the relative gap between consecutive assignments as 0.01, and highway assignment iterations are carried out until this criterion has been achieved. The model-predicted truck volumes can be used in emission models. These predicted truck volumes need to be validated according to FHWA standards. The procedure is discussed below.

5.5 Validation of SROT Model Results

The traditional travel demand model validation involves screen-line comparison and estimation of coefficient of determination (R^2) values. To validate the proposed truck models, we have carried screen-line comparison and R^2 estimation using the traffic count data collected for the years 2009 and 2010.

5.5.1. Screenline Validation

Ohio Department of Transportation sets travel demand model validation standards for all the MPOs in Ohio, and the standards are derived from NCHRP Report 255. For our study area, we have established screen lines across the four sides of the domain as major rivers in region, and other screen lines across major travel corridors. These screen lines are drawn in such a way to be able to assess travel patterns east to west and north to south across a wide area of the modeling area. These are developed across corridors where major travel is or expected to occur. Fig. 5-1 shows the location and orientation of the screen lines used in the validation.

In this analysis, we have summed all the link volumes that cross each of the screen lines and compared them with corresponding truck counts. The comparison proved that the proposed truck models could predict the link-level truck volumes at a confidence level of about 90 percent, which meets FHWA guidelines for travel model validation (Federal Highway Adminstration 2005). However, the screen- line comparison for motor homes truck type yielded inconsistent results, because of the inconsistency in the classification. We have noticed that the motor home axle sizes vary. From the comparison, it is also evident that the TMT mix is mostly consistent as the relative changes in truck volumes are much less. Table 5.7 summarizes the screen-line comparison for seven different screen lines. Actually, most of the travel-demand models as of now compare total vehicles. However in the present study, we compared by truck type. On screen-line No. 4, there were many collector and local streets compared with other screen lines, so the error percentages are very high. The screen lines (like No.5) go through higher freeways at a good fit as the tuck volumes are higher.

Table 5-7: Percentage Errors of by Truck Types from Screen-line Comparison Analysis

Screen Line No	Refuse Trucks	Single Unit Short-Haul	Single Unit Long-Haul	Motor Homes	Combination Unit Short-Haul	Combined Unit Long-Haul
1	4.22%	5.86%	12.64%	-9.22%	4.62%	6.48%
2	-5.29%	-7.15%	2.98%	-12.53%	8.66%	-4.92%
3	-3.06%	7.41%	5.26%	-10.75%	11.69%	7.30%
4	6.79%	-9.96%	-9.88%	11.52%	-7.7%	-8.12%
5	4.09%	8.09%	6.69%	6.51%	9.17%	6.46%
6	-6.18%	3.33%	-7.27%	5.72%	-2.3%	-4.21%
7	11.09%	8.09%	6.27%	2.51%	7.1%	-6%

5.5.2. Validation Scatter Plots

NCHRP Report 255 recommends comparing individual link-level assigned truck volumes to observed counts, which is the most useful tool for highway assignment validation. This tool checks assigned volume and count disparities at individual link and also by magnitude of volume. Usually, ODOT link-level validation is conducted in two steps: 1) maximum allowable error at individual link level using the ODOT standard maximum allowable error curve 2) scatter-plot-based comparison with R^2 checks. Since the truck volumes are on the order of hundreds and the maximum allowable curve is very steep for volumes of less than 10,000 vehicles, the error curve cannot be used.

We have plotted different scatter plots for all six types of trucks as shown in Fig.5-3-Fig.5-5. The modeled truck volumes are pretty well correlated with observed truck counts. The link-level motor home volumes are much higher compared to the number of registered motor homes in the region. This may be because of through movement of such vehicles. On the contrary, the combination long-haul trucks registered in the county are much higher than average link volumes and may be due their pattern of traveling to external faraway destinations. The comparison for different truck types shows that the model performs at a satisfactory level and all the points are close enough to the trend line The motor homes that are fewer than 250 vehicles/day are mostly under predicted with an error of 10-15 percent by the model since observed counts cover local streets to freeways. After threshold 250 vehicles/day, the model performed very well and consistently. In the case of refuse trucks, the model is consistent between 300-1100 values. For smaller and larger volumes, the model is pretty inconsistent. The reason for this discrepancy may be the schedule of refuse trucks, which is not taken into account in the model.

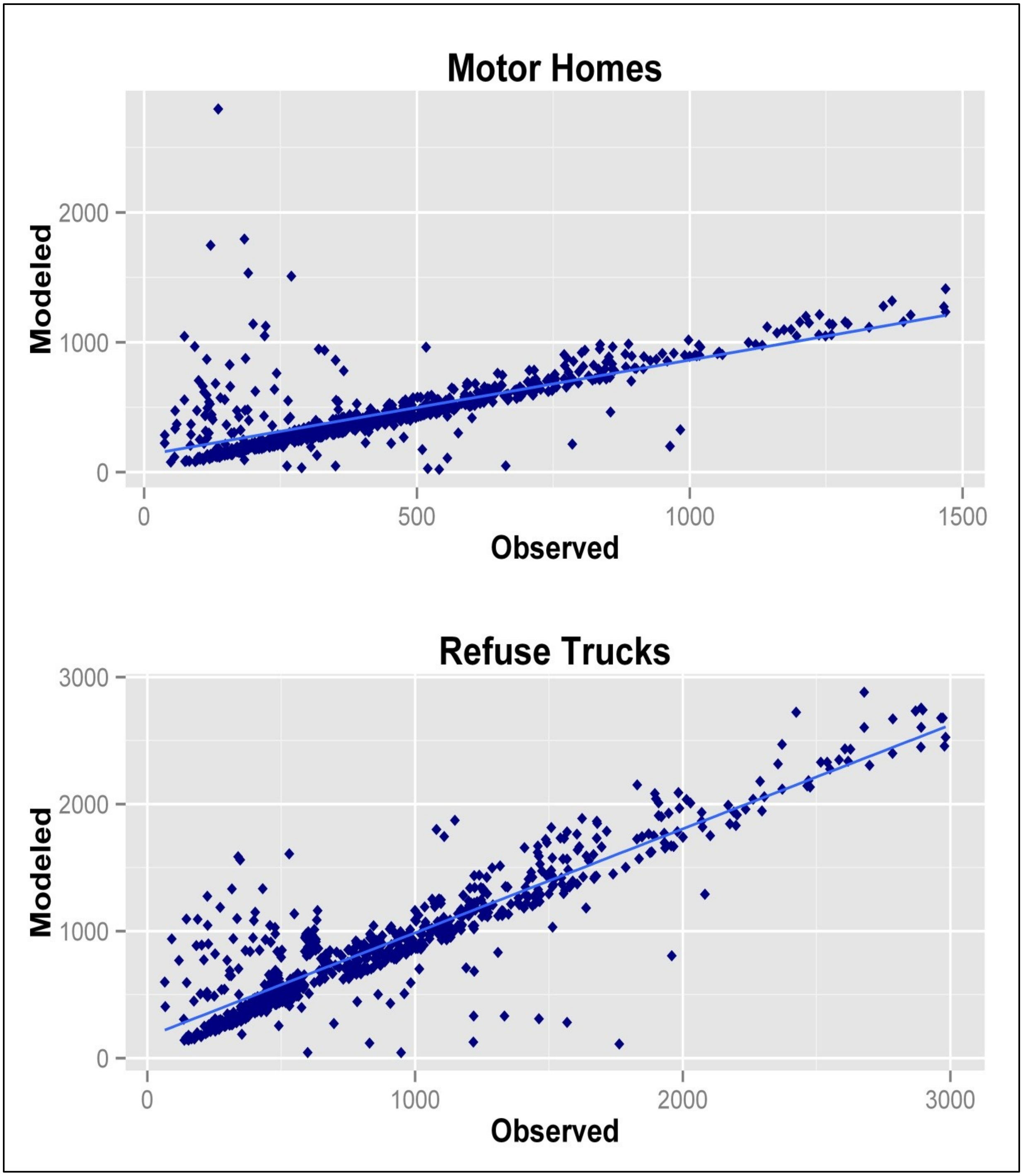

Figure 5-3: SROT Model Validation Results for Motor Home and Refuse-Type Trucks

The single unit long-haul trucks were predicted accurately enough by the model as the values were scattered along the trend line. The model prediction for single unit short-haul trucks is mostly accurate as observed from the scatter plot. The observed combination short-haul trucks are mostly null values for minor arterials and local streets. If we ignore zero volumes, both types of combination trucks were predicted reasonably well from the model. Based on the "Model Validation and Reasonableness Checking Manual," the data points on the regression plots should be close to the 45^o line and the R^2 should be greater than 0.8. Except for motor-home- type trucks, for all other truck types the data points were close to 45^o line. We have also verified the R^2values for all types of trucks as shown in the Table16.

As previously mentioned, the motor homes have the least fit and their corresponding R^2 value is 0.5103 and t-statistic is 28.475. The regression statistics showed that the model is somewhat unstable. For the rest of the models, the coefficient of determination values range between 0.87-0.97, which are well above the FHWA and ODOT standards. The motor homes model has the highest standard error in predicted values when compared to other models.

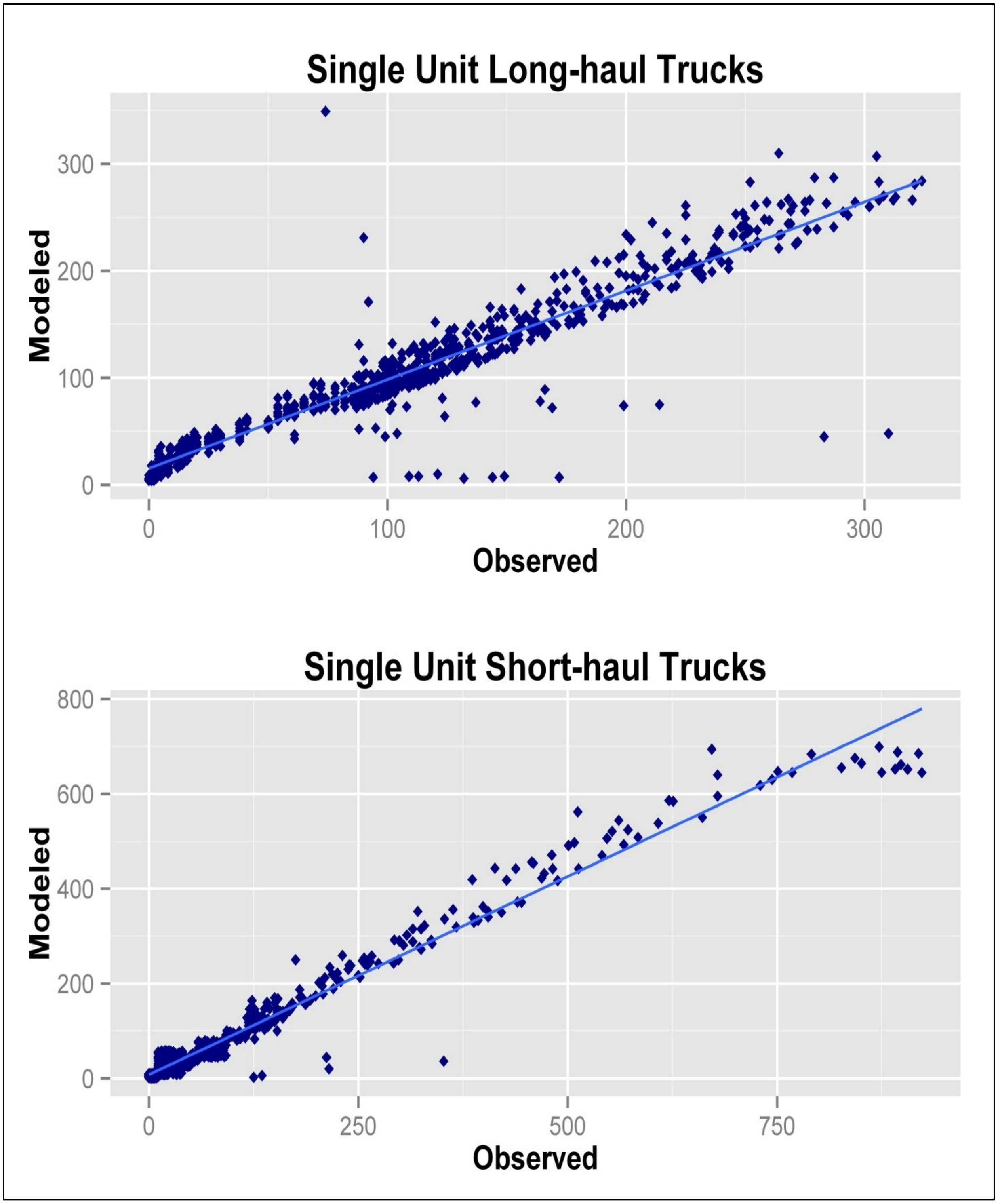

Figure 5-4: SROT Model Validation Results for Single Unit Type Trucks

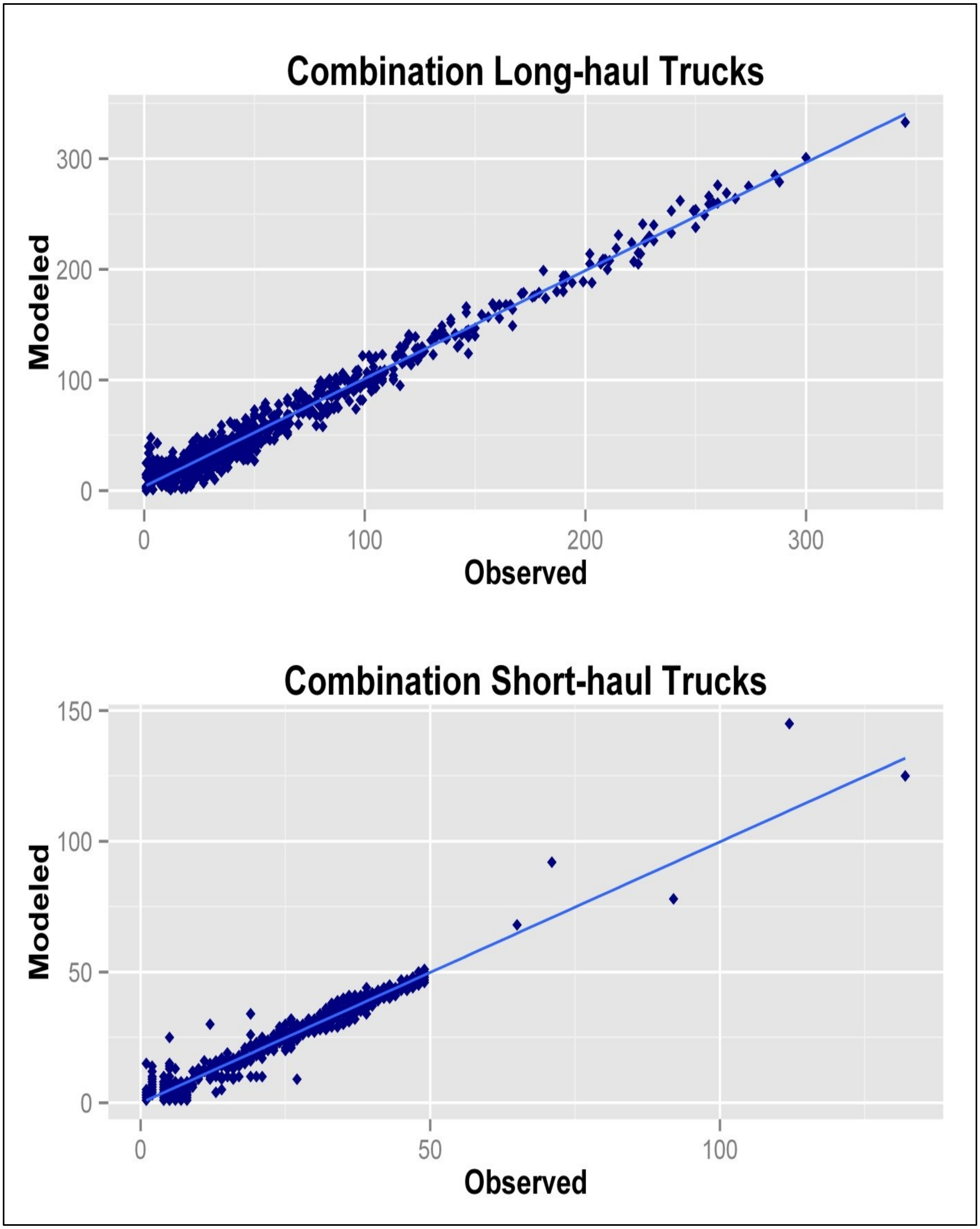

Figure 5-5: SROT Model Validation Results for Combination-Type Trucks

Table 5-8: Model Reliability Statistics

Model	R square	t-Value	Standard Error
Refuse Trucks	0.8197	59.46	0.01685
Single Unit Short-Haul	0.9665	149.927	0.00771
Single Unit Long-Haul	0.8732	73.198	0.01438
Motor Homes	0.5103	28.475	0.02493
Combination Unit Short-Haul	0.9281	119.309	0.008075
Combined Unit Long-Haul	0.9718	163.007	0.005938

To estimate the link-level hourly truck miles traveled (TMT), we have developed a multinomial probit model using the same set of demographic and geometric variables. Ideally, the hourly activity of each type of truck is expected to follow a different pattern. However, for simplification purposes, we have aggregated all refuse, single unit short-haul, and single unit long-haul trucks as one group and the rest of the truck types as one group. For the sake of emission model convenience, we have aggregated all arterials, collectors, and local streets as unrestricted highways. Since our case study area is considered an urban area according to regional area type classification, we have developed hourly distribution factors for only two types.

The model estimated results for the two road types and two types of trucks using Greater Cincinnati regional data are shown in Tables 5-9 to 5-12. As specified in the model, we have considered the covariates and their respective means for estimation. In these models, the hour number is used as a dummy variable (i.e. 24 different variables) and not shown in the tables. Since these are fractional response models, the logarithm of large explanatory variables was considered. For parameter estimation, the generalized estimating equations method is used through Stata ® software (Stata Corp., 2011).

Table 5-9: Single Unit Type Trucks Hourly Distribution Model for Urban Unrestricted Access Highways

Urban Unrestricted Access Roadway				
Variables	**Coefficient**	**Standard Error**	**z**	**Pr<|z|**
Logarithm of Capacity	4.07E-02	2.15E-02	1.89	0.059
Average of Logarithm of Capacity	-1.73E-02	4.30E-02	-0.4	0.689
Speed	1.45E-05	6.82E-05	0.21	0.832
Average Speed	2.06E-04	2.12E-04	0.97	0.333
Employment Density	5.72E-05	6.38E-05	0.9	0.37
Average Employment Density	-2.91E-04	4.18E-04	-0.7	0.486
Population Density	2.25E-04	1.51E-04	1.49	0.135
Average Population Density	1.07E-03	1.11E-03	0.96	0.337
Constant	-2.546	0.093	-27.36	0
Log Pseudo Likelihood	-3949.93			

Table 5-10: Single Unit Type Trucks Hourly Distribution Model for Urban Unrestricted Access Highways

Urban Restricted Access Roadway				
Variables	**Coefficient**	**Standard Error**	**z**	**Pr<|z|**
Logarithm of Capacity	6.72E-03	1.80E-03	3.74	0
Average of Logarithm of Capacity	5.85E-04	3.38E-03	0.17	0.862
Speed	-9.33E-06	3.66E-05	-0.25	0.799

Average Speed	-2.39E-03	4.83E-03	-0.5	0.62
Employment Density	1.32E-05	1.48E-05	0.89	0.373
Average Employment Density	-1.31E-03	3.73E-03	-0.35	0.726
Population Density	-6.44E-05	9.03E-05	-0.71	0.476
Average Population Density	1.88E-04	9.83E-05	1.91	0.056
Constant	-2.4	0.036	-67.39	0
Log Pseudo Likelihood	-1743.33			

Table 5-11: Combination Type Trucks Hourly Distribution Model for Urban Restricted Access Highways

Urban Restricted Access Roadway				
Variables	**Coefficient**	**Standard Error**	**z**	**Pr<|z|**
Logarithm of Capacity	3.43E-02	1.81E-02	1.59	0.05
Average of Logarithm of Capacity	-1.46E-02	3.63E-02	-0.34	0.58
Speed	1.22E-05	5.75E-05	0.18	0.70
Average Speed	1.74E-04	1.79E-04	0.82	0.28
Employment Density	4.83E-05	5.38E-05	0.76	0.31
Average Employment Density	-2.46E-04	3.53E-04	-0.59	0.41
Population Density	1.90E-04	1.27E-04	1.26	0.11
Average Population Density	9.03E-04	9.37E-04	0.81	0.28
Constant	-2.15	7.85E-02	-23.08	0.00
Log Pseudo Likelihood	-3332.56			

Table 5-12: Combination Type Trucks Hourly Distribution Model for Urban Restricted Access Highways

Urban Restricted Access Roadway				
Variables	**Coefficient**	**Standard Error**	**z**	**Pr<|z|**
Logarithm of Capacity	5.12E-03	1.37E-03	2.85	0.00
Average of Logarithm of Capacity	4.46E-04	2.58E-03	0.13	0.66
Speed	-7.11E-06	2.79E-05	-0.19	0.61
Average Speed	-1.82E-03	3.68E-03	-0.38	0.47
Employment Density	1.01E-05	1.13E-05	0.68	0.28
Average Employment Density	-9.99E-04	2.84E-03	-0.27	0.55
Population Density	-4.91E-05	6.88E-05	-0.54	0.36
Average Population Density	1.43E-04	7.49E-05	1.46	0.04
Constant	-1.83	2.74E-02	-51.38	0.00
Log Pseudo Likelihood	-1329.11			

The capacity of the roadway has a positive impact on urban unrestricted highways. On the other hand, the hourly combination of trucks on restricted highways is relatively less affected by capacity of highways since their capacities are higher. Similarly, posted speed limit on unrestricted highways has the same kind of impact on hourly distribution of single-unit and multi-unit trucks. On the contrary, the hourly activity of multi-unit trucks on urban restricted highways are negatively impacted by posted speed limits. The coefficients of land-use variables show the same kind of impact on hourly combination of trucks on urban unrestricted compared to restricted highways.

The z-scores and their respective probabilities have shown very high quality of goodness-fit for the models. We have also observed that the log pseudo likelihood values are pretty consistent through the iterations and also relatively low (lower values means the model is robust). The standard errors reveal the variability of the model, and in this case, they are relatively low. We have carried reasonableness checks with the default hourly distribution used by the MOVES model for hourly disaggregation of activity.

5.6 Compared Results of the Hourly Distribution Model

The hourly distribution factors used by different vehicle types in the default mode of the MOVES are very similar. The default mode has one set of distribution factors for weekdays and another set for weekends. Within the weekday factors, the default set only contained five sets of hourly distributions based on roadway types. The roadway-specific values are used for all vehicle types. In this research, we have only estimated weekday hourly combination of two types of trucks. We have compared both the default hourly distribution and the output from the proposed model with observed values during 2010. The charts are shown below.

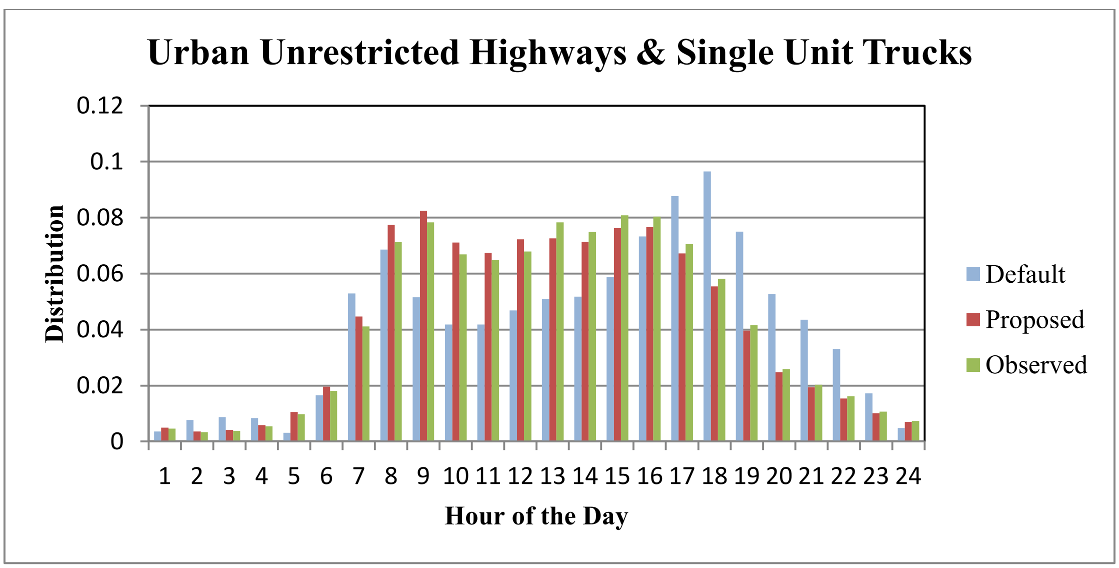

Figure 5-6: Comparison of Observed vs. Predicted Hourly Factors for Urban Unrestricted Highways and Single Unit Type Trucks Group

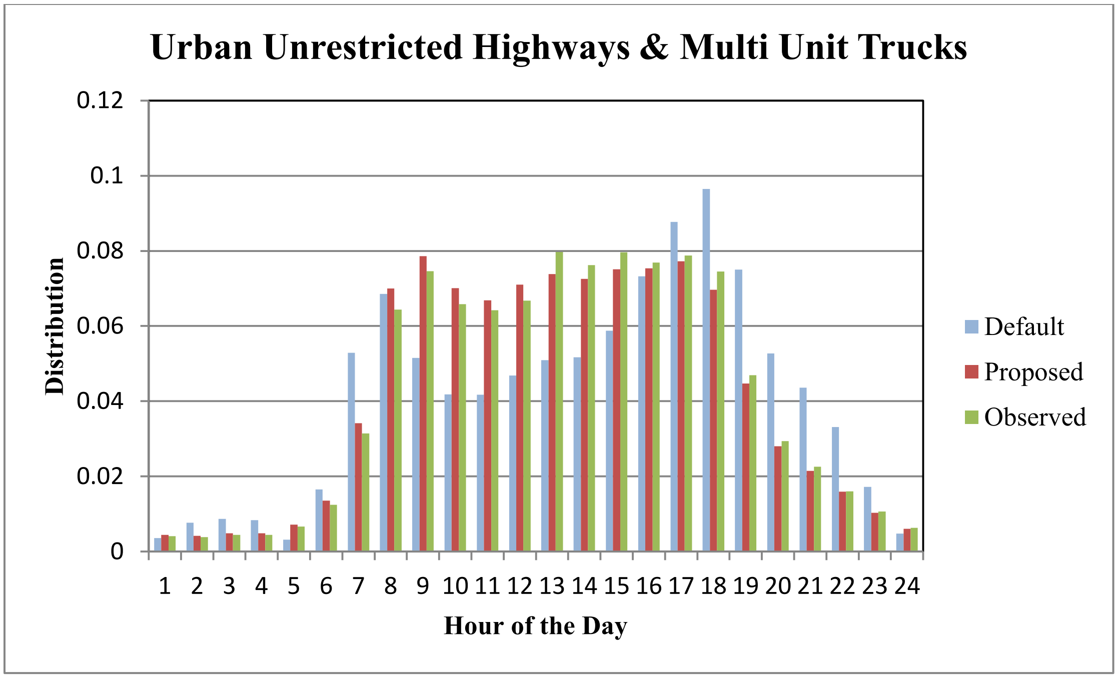

Figure 5-7: Comparison of Observed vs. Predicted Hourly Factors for Urban Unrestricted Highways and Combination Type Trucks Group

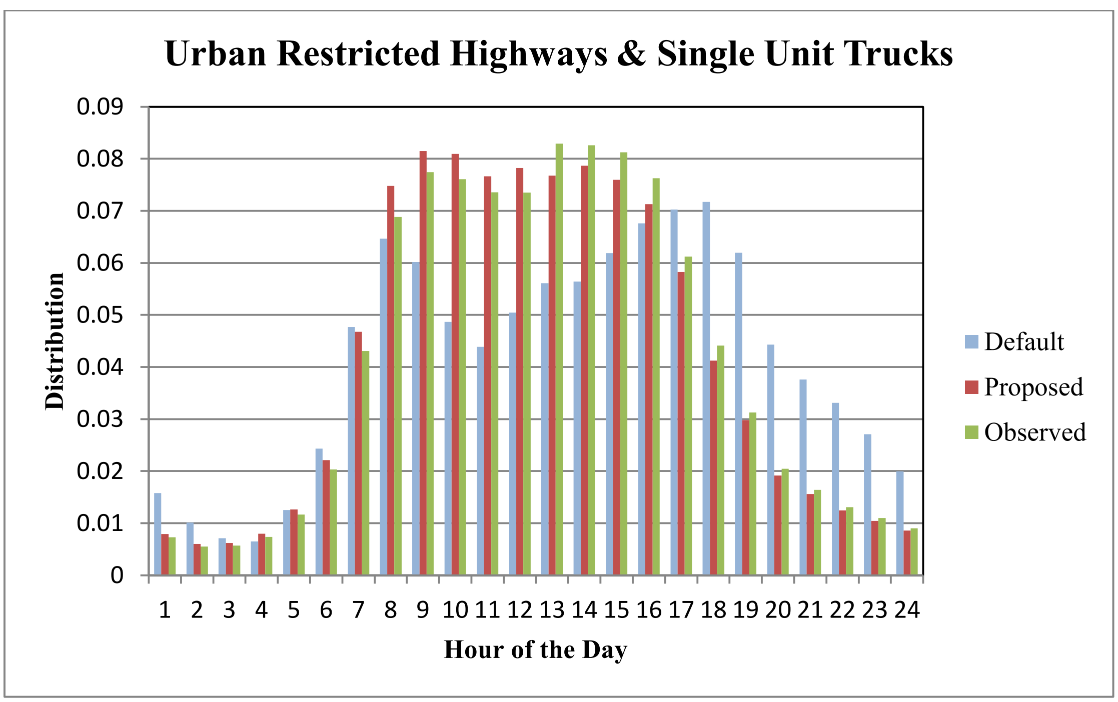

Figure 5-8: Comparison of Observed vs. Predicted Hourly Factors for Urban Restricted Highways and Single Unit Type Trucks Group

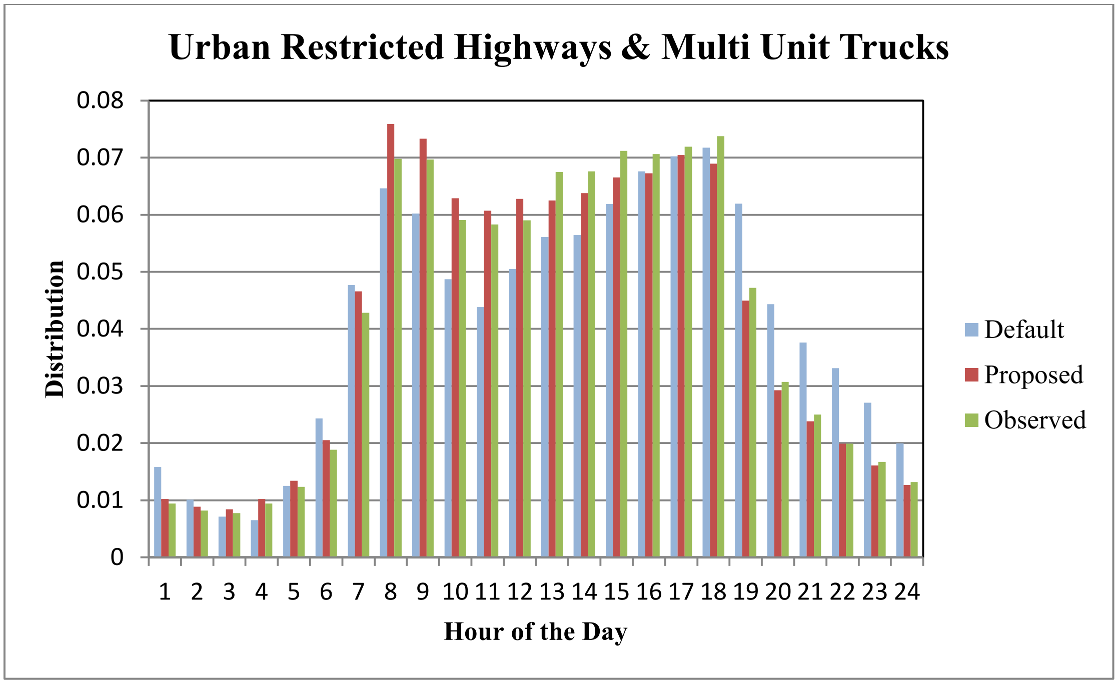

Figure 5-9: Comparison of Observed vs. Predicted Hourly Factors for Urban Restricted Highways and Combination Type Trucks Group

The default values were predicted higher in the evening peak hours, and predicted values are consistent throughout the mid-day for single unit trucks. The observed values showed same trend with $\pm$ 5 percent error of the predicted model for both types of trucks. The default distributions are the same for both types of trucks, which seems to be untrue if the real behavior of the trucks is considered. This comparison gives us enough confidence about the predicted values. The output of these models can be used as follows:

$$V_{i,hr,trk} = HF_{RT,HR,TRK} \; x \; DV_{i,trk} \quad \textbf{Eq. 20}$$

Where,

$V_{i,hr,trk}$ = Hourly truck volume for link i, hour hr and truck type trk

$HF_{RT,HR,TRK}$ = Hourly distribution factor for hour HR (1 to 24), roadway type i belongs to (i.e. restricted to unrestricted), and truck type trk belongs to (i.e. single- unit truck or multi-unit truck)

$DV_{i,trk}$ = Daily truck volume on link i of truck type trk

5.7 Link Level Hourly Activity Estimation

For emission models, the activity data should be in terms of vehicle/truck miles traveled and average hourly speeds. We have developed hourly truck volumes using the above-discussed models. However, other vehicle activity such as cars, buses, motorcycles, and pickup trucks, etc. is also needed for any modal-based emission model. The daily (typical weekday) assignment volumes for other auto and transit modes are obtained from OKI's travel demand model. They also provided us with average daily link-level speed data along with the link description by road type and link capacity. Each link's ratio of volume to capacity was used to estimate the average speed for that hour. Before processing the speeds, we summed up all types of vehicle volumes. We have used the HCM algorithm to estimate link-congested speeds, and free-flow speeds were adjusted based on latest speed data. Even though this step can be easily achieved using spread sheet software, we preferred Cube to deal with peak-hour congestion.

The estimation of truck miles travelled is a pretty straightforward step. The link-level hourly volumes of each truck type are multiplied with corresponding link lengths. The activity processing is also carried within Cube environment since the highway network data also includes true lengths of links. In this process, only average annual daily truck volumes or truck miles traveled are estimated. However, for accurate dispersion modeling of air pollution, we need each day's activity. For this purpose, we used monthly and day-of-week factors. The details are explained in Chapter 7.

In summary, we applied the Spatial Regression and Optimization Modeling methodology to model six different types of trucks. At first, using greater Cincinnati data, the author developed a set of spatial regression models. The models are used to predict daily truck volumes at control locations. Utilizing trip cost matrices to study area traffic analysis zones and predicted truck volumes; we synthesized a new set of truck trip matrices. The trip matrices totals were adjusted/optimized using independent truck flow data. Finally, the truck trip matrices were added to transit and personal travel matrices for highway assignment step. A separate set of hourly distribution models was developed using the same set of hourly data. Hourly truck volumes were obtained using the daily truck volumes and hourly distribution factors.

Chapter 6: Modeling Link-Level Hourly Truck-Related $PM_{2.5}$ Emission Inventory

6.1 Emission Model Inputs and Processing

To model atmospheric dispersion of vehicle exhaust pollutants in nonattainment areas, air quality modelers need to prepare mobile emissions inventories to represent actual activity in the region. We have already verified from literature that truck emissions directly impact air quality, thus the spatial and temporal truck emission inventories are vital in the air quality modeling. To predict the link-level emission inventories, the emission model would be able to predict emission rate at such a detailed level. In the present study, we used MOVES 2010 to predict link-level emission rates. Applying MOVES 2010 emission rates to link-level activity data required a substantial revision to the emissions analysis approach previously used with the MOBILE model because of the new requirements and features of MOVES 2010.

6.1.1. Processing Link-Level Activity Data

The link-level activity information is the most critical among the input to the MOVES 2010 model. In the last chapter, we discussed how we have modeled truck activity and average speeds for each link by hour of day in the study area. As a first step in the emission inventory estimation, we processed all highway links in the study area to put them into different link type bins. These bins are classified based on characteristics such as the highway type, average speed on link, area type, and grade of the link. There are two different types of highways used by MOVES model i.e. restricted (expressways and freeways) and unrestricted (includes arterials, collectors, and local streets). The MOVES model uses the bin approach to treat vehicle speeds. Sixteen speed bins are used in the model. The average speed bin description is shown in the following table.

Table 6-1: Average speed bins used in MOVES model

Average Speed Bin ID	Average Bin Speed	Average Speed Bin Description
1	2.5	speed < 2.5 mph
2	5	2.5 mph <= speed < 7.5 mph
3	10	7.5 mph <= speed < 12.5 mph
4	15	12.5 mph <= speed < 17.5 mph
5	20	17.5 mph <= speed <22.5 mph
6	25	22.5 mph <= speed < 27.5 mph
7	30	27.5 mph <= speed < 32.5 mph
8	35	32.5 mph <= speed < 37.5 mph
9	40	37.5 mph <= speed < 42.5 mph
10	45	42.5 mph <= speed < 47.5 mph
11	50	47.5 mph <= speed < 52.5 mph
12	55	52.5 mph <= speed < 57.5 mph
13	60	57.5 mph <= speed < 62.5 mph
14	65	62.5 mph <= speed < 67.5 mph
15	70	67.5 mph <= speed < 72.5 mph
16	75	72.5 mph <= speed

We defined six bins for grades between -6 percent to +6 percent. In total, 192 link type bins are used (16 speed bins x 2 road types x 6 grade bins) in this study. Each link in the modeling domain should belong one of this bins. Links are denoted by their starting and ending node numbers. The link average speeds are obtained from the post-processing step discussed in the earlier chapter.

6.1.2. Truck Population and Age Distribution

Vehicle population by source (vehicle) type is a key input to MOVES2010. Vehicle populations are created from the vehicle registration data. In this work, the truck population was determined from Ohio DMV's Vehicle Identification Number (VIN) data. Matching VIN classification with US-EPA's source type classification is a cumbersome process. However, the state department of transportation provides guidelines for such a conversion process. In the following table, we have enumerated the truck population data we used for emission rate runs.

Table 6-2: Truck Population used in Emission Modeling

Truck Type	Truck Population
Refuse truck	140
Single Unit Short-Haul Truck	232
Single Unit Long-Haul Truck	194
Motor Home	1650
Combination Short-Haul Truck	2063
Combination Long-Haul Truck	2365

The DMV data also contains the year and make of each registered vehicle in the county. Based on the year the vehicle was first registered, we can bin the vehicle population into 30 different year bins. All of the trucks more than 30 years old are put in 30th bin. MOVES model expects the age data in terms of relative distribution as shown in the following table.

Table 6-3: Truck Age Distribution Data

Year	Refuse truck	Single Unit Short-Haul Truck	Single Unit Long-Haul Truck	Motor Home	Combination Short-Haul Truck	Combination Long-Haul Truck
0	0.049837	0.00542	0	0.073713	0.084252	0.166845
1	0.039756	0.04878	0.006211	0.045616	0.067209	0.133094
2	0.034049	0.062331	0.037267	0.07393	0.057562	0.113989

3	0.076702	0.070461	0.009317	0.048698	0.050629	0.11399
4	0.092621	0.086721	0.02795	0.060515	0.0693	0.118554
5	0.060388	0.04336	0.034161	0.060804	0.056228	0.080365
6	0.054352	0.04336	0.018634	0.044092	0.048773	0.064287
7	0.024289	0.054201	0.018634	0.040781	0.037878	0.040271
8	0.069582	0.054201	0.012422	0.031984	0.045255	0.030358
9	0.062471	0.075881	0.015528	0.04417	0.053548	0.03154
10	0.051437	0.02168	0.021739	0.060195	0.056029	0.03196
11	0.07302	0.04065	0.037267	0.056322	0.054994	0.029048
12	0.061045	0.078591	0.009317	0.05743	0.059676	0.007988
13	0.079566	0.054201	0.031056	0.044695	0.05284	0.00866
14	0.044175	0.051491	0.021739	0.050087	0.048713	0.011471
15	0.047907	0.067751	0.037267	0.053065	0.040034	0.006193
16	0.0145	0.03252	0.021739	0.036308	0.016687	0.001319
17	0.016871	0.00813	0.034161	0.02215	0.014696	0.001057
18	0.01562	0.01626	0.012422	0.012724	0.013335	0.003519
19	0.004048	0.00271	0.018634	0.001733	0.017959	0.001198
20	0.004314	0.00813	0.024845	0.013844	0.011167	0.000994
21	0.004336	0	0.037267	0.019059	0.00904	0.000621
22	0	0.00271	0.018634	0.0267	0.009914	0.001019
23	0.009217	0.00271	0.024845	0.016931	0.003839	0
24	0.002736	0.01355	0.006211	0.004455	0.004789	0.000851
25	0.007047	0	0.037267	0	0.004765	0.000283
26	0.000116	0	0.015528	0	0.004	0.000339
27	0	0	0.018634	0	0.003639	0
28	0	0.00271	0.021739	0	0.002622	0.000189
29	0	0	0.018634	0	0.000628	0
30	0	0.051491	0.350932	0	0	0

6.1.3. Trip Data

Lastly, the trip starts and ends data by TAZ and time period was used to allocate vehicle population for start and parked vehicle emission processes. The relative number of trip starts and ends for each TAZ compared to the total number in each hour was used to allocate the vehicle population to the TAZ for start and parked emissions. This information is obtained from the previous truck trip matrix estimation step. The trip information is used for the post processing of emission rates and link-level activity to obtain link-level emission inventory.

6.1.4. Meteorology and Fuel data

Ideally, the temperature and relative humidity data should be used at link level for more accurate estimation of emission rates. However, the overall spatial variability of temperatures is less in our study area. Further, the $PM_{2.5}$ emission rates are not highly variable with temperatures. So, in this study we have used one set of hourly temperatures for a day. As mentioned above, the temperature and relative humidity data for the month of July 2010 was obtained from the Weather Station at Lunken Airport. The temperatures ranged between 64.8°F and 85.5°F in the month. The relative humidity varied in between 40.6 percent to 81.1 percent in the month. These values were directly used in MOVES2010 to generate emission rates. In the present study, we used only one type of diesel fuel throughout the month. The diesel fuel contains 11 ppm of sulfur and carbon 2 percent by weight. The chemical composition of the diesel fuels is also very important for estimation of accurate $PM_{2.5}$ emission rates. Since there are no inspection and maintenance programs in practice for study area, we have not considered that in the analysis.

6.1.5. *Generating Lookup Emission Rates*

The MOVES2010 model has two output options: one is inventory, which would provide us the aggregate inventory totals, and another is emission rates based on processes. As we discussed earlier, to estimate the link-level emission inventory, the emission rate mode is very suitable. MOVES model generates three separate lookup tables called "RatePerDistance," "RatePerVehicle," and "RatePerProfile." They collectively contain all the emission factors for each emission process. Table 6-4 describes the emission processes incorporated in each lookup table for $PM_{2.5}$ emission estimation.

Table 6-4: MOVES 2010a Emission Factors Tables

Process Name	RatePerDistance (g/mile)	RatePerVehicle (g/vehicle/hour)
Running Exhaust	X	-
Crankcase Running Exhaust	X	-
Brakewear	X	-
Tirewear	X	-
Start Exhaust	-	X
Crankcase Start Exhaust	-	X
Extended Idle Exhaust	-	X
Crankcase Extended Idle Exhaust	-	X

The "RatePerDistance" table can be directly related to on-road truck activity as the emission rates are by speed bin, road type, and truck type. The "RatePerVehicle" and "RatePerProfile" tables contain parked and start truck emission rates as shown in the table. It should also be noted that the “RatePerVehicle” and “RatePerProfile” tables provide emission rates in units of gram per vehicle for each hour of day, while “RatePerDistance” is in gram per mile. The “RatePerVehicle” table includes all of the parked vehicle emissions processes with the exception of an “evaporative fuel vapor venting” process. The emission rates are provided for each hour of day so that the emission rates are different for each hour of day under the same meteorological conditions (Lindhjem 2010).

Using the truck activity information such as truck miles traveled and average speed distribution of trucks and other important inputs like the meteorological data, fuel data, truck population, and age data, we have set up a MOVES run specification for the entire month of July. This run specification provided us with three different emission rate tables in MySQL format. The county-level geographic domain was chosen for this model run.

6.2 Post- Processing the Emission Rate Model Results

To calculate the link-level emissions inventory from emission rate output, rates in “RatePerDistance” values need to be multiplied by the appropriate Truck Miles Travelled (TMT), and rates in “RatePerVehicle” and “RatePerProfile” must be multiplied by appropriate trip starts values. However, the Truck Miles Travelled obtained from the SROT model are valid for an average day. To estimate a month-long emission inventory, we have to apply month and day-of-week factors. These factors can be obtained from the state department of transportation, and they are year specific. As a first step in emission model output post processing, we calculated the truck activity by hour for each weekday (Monday-Friday).and weekend day (Saturday-Sunday) in the month of July. We assumed that the truck hourly patterns during weekends should also be similar to weekends (weekdays).

Applying corresponding emission rates to activity data requires a little bit of care since it requires multiple considerations. The emission rates are by bins, and individual link activity cannot be directly related to that bin as it has range. For more accurate emission estimation, we need to perform appropriate interpolation as rates applied. For example, to calculate the total emissions from 100 truck miles/hr with an average speed of 37 mph, a simplistic approach would be to multiply all of the activity by the emission rate for speed bin 7 (32.5 mph-37.5 mph). However, this approach is too sensitive to even very small changes in speed. To reduce these boundary issues, instead interpolate between speed bins (in this example, between the rates for speed bin 7(32.5 mph-37.5 mph) and speed bin 8(37.5 mph-42.5 mph)) based on the average speeds for those speed bins.

The link-level $PM_{2.5}$ running emissions are calculated by looking up corresponding emission rates from MOVES output database for each link in the preprocessed activity database based on the link type. The evaporative and start emissions are calculated multiplying the hourly starts and parked vehicles in zone. Finally, we have aggregated the entire link-level $PM_{2.5}$ running and start emission inventories for the whole study area. Table 6-5 contains the daily aggregated emission quantities generated using the default and the new input data in the MOVES model.

Table 6-5: Comparison of Emission Output for Different Truck Activity Estimation Models

Source use/Truck types	Daily emissions using default inputs (Kg)	Daily emissions using new truck and hourly fraction models (Kg)
Refuse Trucks	5.50	11.79
Single Unit Short-Haul	95.85	205.73
Single Unit Long-Haul	12.77	329.35

Motor Homes	4.11	49.81
Combination Unit Short-Haul	202.87	351.49
Combined Unit Long- Haul	321.79	620.28

Using the default input yielded fewer daily emission totals for all of the source-use types, since the corresponding TMT values were also smaller. Further, the impact of hourly TMT distribution is also significant while estimating the aggregated emission inventory. Above totals when compared with National Emission Inventory for year 2010 were 1763 Kg/day, whereas estimated emissions were 1618 Kg for a weekday in July. A detailed discussion about possible reasoning for the discrepancy between default and proposed model outputs is in the following sections.

In case of start emissions, we have used the surrogate methods to allocate them to the links. As discussed in the methodology chapter, in this method the total zonal start emissions are distributed among the links in the zone using their respective TMT as a surrogate. Since the start emissions portion in total emissions is relatively low, this assumption can yield legitimate results. This assumption can also be useful to improve spatial distribution of emission inventory.

6.2.1. Discussion on Hourly Emission Variation

To verify the importance of hourly TMT distribution on daily total emission quantities, we analyzed the hourly emission variations of different trucks on July 17, 2010. For comparison, we have used the hourly emission output using default hourly distribution and proposed hourly distribution model output. The preliminary analysis of results showed that if we use default hourly distribution, the hourly emission variation would be identical for both road types, which may not be true in the real world. On the contrary, the hourly emission estimates using the new hourly TMT distributions are quite varied by road type and truck type, which is reasonable since the hourly speeds and the activity of each truck type is different during a typical weekday. The higher hourly emissions in the proposed method can also be attributed to the added start emissions.

In Fig. 6-1 and 6-2, we have shown the emission quantities released by all truck types on urban unrestricted highways (includes arterials and collectors) and urban restricted highways (includes freeways and expressways) during July 17, 2010. The motor homes and refuse trucks emit fewer than 0.1 kilo grams during a.m. peak hour on all types of urban roadways. The morning peak hour emissions estimated using the proposed TMT hourly distributions sum up to 8 kilo grams in the study area, since much recreation vehicle activity is expected during summer time in the region. In the case of single unit short-haul and combination short-haul trucks, the emissions are under-predicted in the default case as they tend to experience a lot of congestion during peak hours and have to traverse hilly terrain in the region.

Even though very few single unit long-haul trucks are registered in the region, more emission quantities are estimated due to the external-external activity of this truck type. The combination long-haul trucks use restricted highways since they tend to haul freight longer distances compared to other types of trucks, thus the related hourly emissions are also high. However, the emission output due to default hourly distribution showed that they emit the same amount of emissions on both types of roadways. Most interestingly, the default emissions were higher during evening peak hour, but the proposed model output was almost flat during midday, which is more reasonable since truck drivers try to avoid congested hours in major cities when their trips are longer. As another method to evaluate the proposed models, we ranked the highways with high emissions and plotted their maps. It revealed interesting results, which are discussed in the next section.

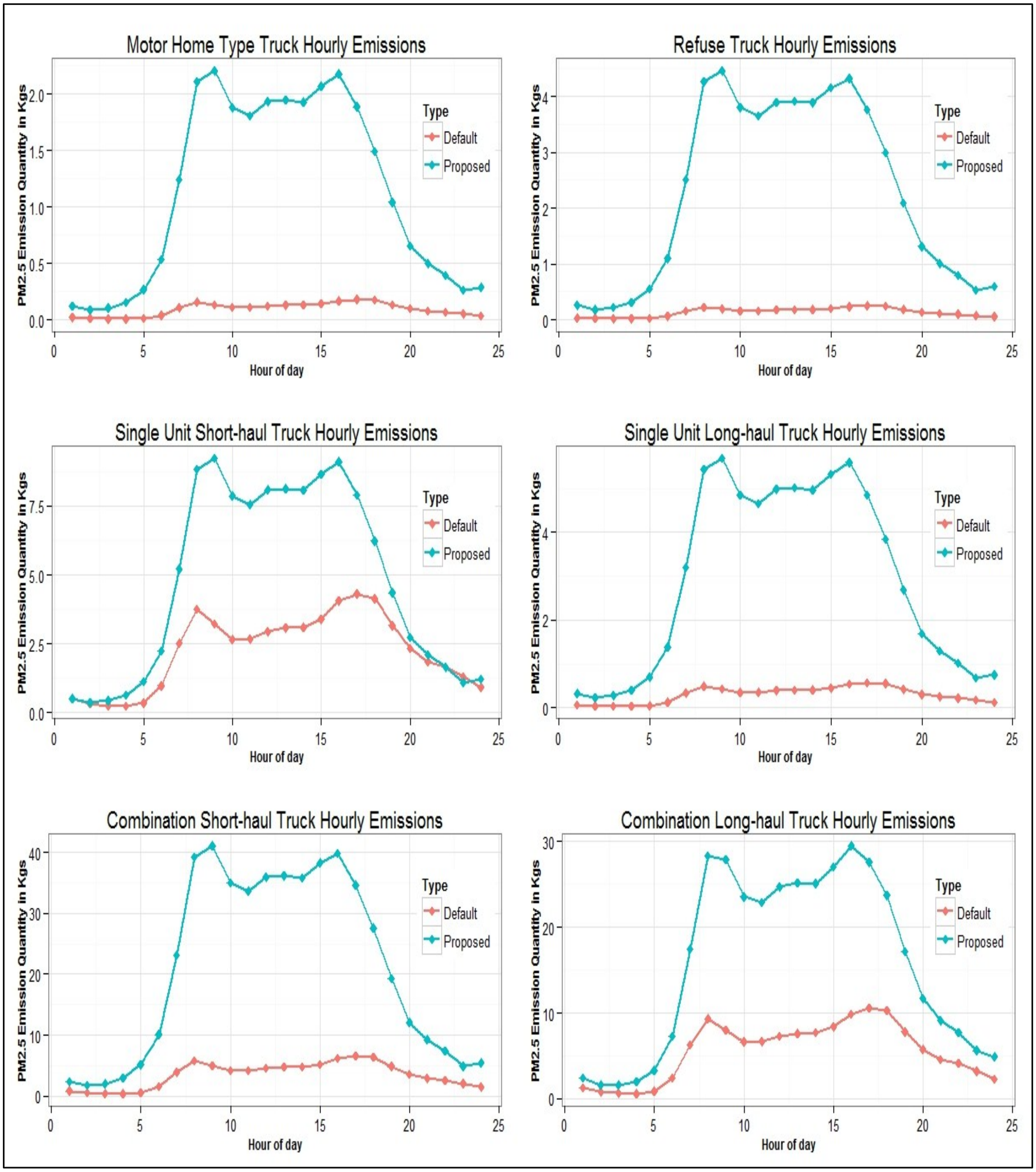

Figure 6-1 Different Temporal Emission Inventories Estimated on Arterials and Collectors using Default and Proposed Hourly Activity Distributions on July17, 2010

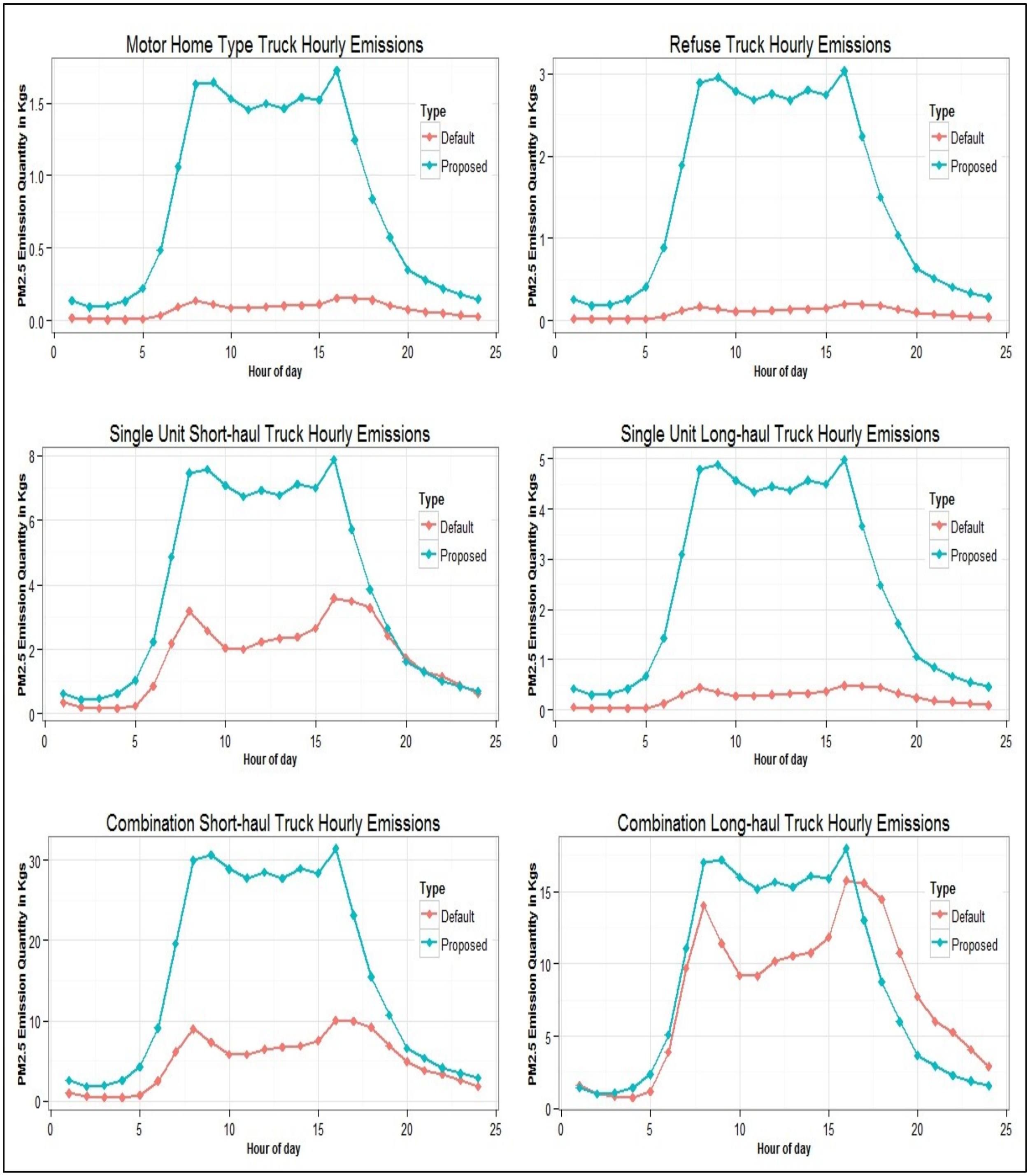

Figure 6-2: Different Temporal Emission Inventories Estimated on Freeways using Default and Proposed Hourly Activity Distributions on July 17, 2010

6.3 Link Level Emission Prediction

The link-level total daily $PM_{2.5}$ emissions were calculated using the link TMT and corresponding emissions rates and adding them up with the start emissions, which were distributed to links using TMT surrogates. For comparison purposes, we calculated the default link-level emissions using county-level emission inventory and default truck activity using following equation.

$$e_i = E \cdot \frac{T V_i \cdot L_i}{\sum_i T V_i \cdot L_i} \quad \textbf{Eq. 21}$$

Where,

e_i = link level $PM_{2.5}$ emission in grams

E = Total county level $PM_{2.5}$ emissions in grams

$T V_i.$ = The truck volume on link i estimated using default methodology

L_i = The link length in miles

We have used the same emission quantity scales for plotting the map as shown in Fig. 6-3 and Fig. 6-4. The default map indicates that it has a higher emission prediction for I-71 in the study area. From regional knowledge and the truck activity plot created using HPMS data and latest traffic count data (shown in Fig. 6.5), I-71 is predominantly used by cars and passenger trucks since it is the convenient freeway for the large commuting population who live in the most popular surrounding cities such as Blue Ash and Mason. Surprisingly, there are fewer emissions on the same highway after the Dana Avenue exit, and there are no industries located near the exit to explain the truck travel behavior. The other major aberration we observed is in the Sharonville area where many industries, warehouses, and General Electric are located. The default methodology predicted fewer emissions on I-75 and other arterials surrounding the Sharonville area. The default methodology predicted truck congestion on the I-275 Bridge and less congestion at I-71and I-75 interchange. In reality and also based on HPMS data (Fig. 6-5), the congestion situation is vice versa, thus different spatial variation of emissions.

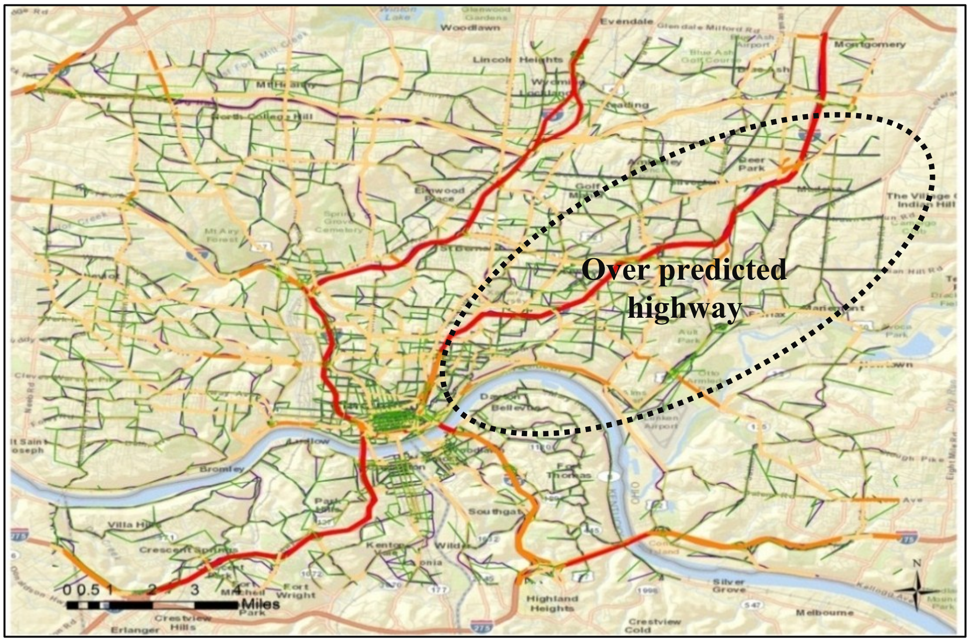

Figure 6-3: The Link-based PM2.5 Emissions Estimated using Default Truck Activity Methodology

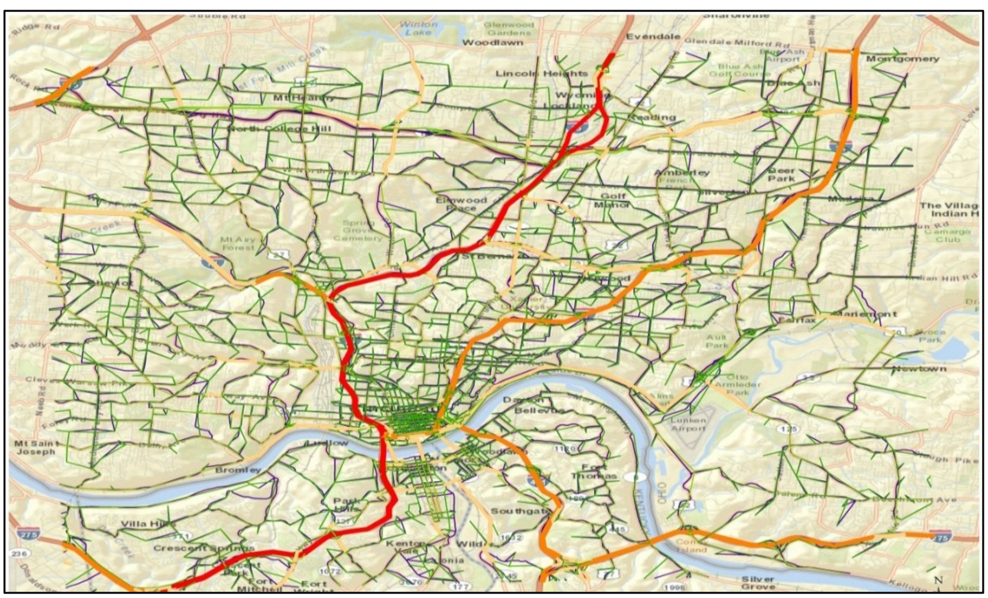

Figure 6-4: The Link-based PM2.5 Emissions Estimated using Proposed Truck Activity Methodology

Figure 6-5: Truck activity based on HPMS data

6.4 Relative Source Apportionment Results

The contribution of different trucks to the regional emission inventory is an interesting topic for transportation and air quality modelers. This information would be useful for the transportation modelers to develop particular travel demand strategies to curb the $PM_{2.5}$ emissions released by particular truck types. Air quality modelers would be interested to know the contribution of each truck type to the regional air quality. We conducted a source (in this case truck type) apportionment analysis. The aggregated $PM_{2.5}$ emission analysis showed that almost 50 percent of regional emissions are emitted by combination long-haul trucks in both outputs. The motor homes predicted only 2 percent in the default method even though they actually represent 17 percent in the total study area's truck population. On the other hand, the proposed method predicts 8 percent contribution to total emissions. The default method over predicts the contribution from refuse trucks to almost18 percent whereas they actually represent 1.83 percent of the total population. For other truck types, the proposed models predict their actual contribution when their population and relative activity are considered.

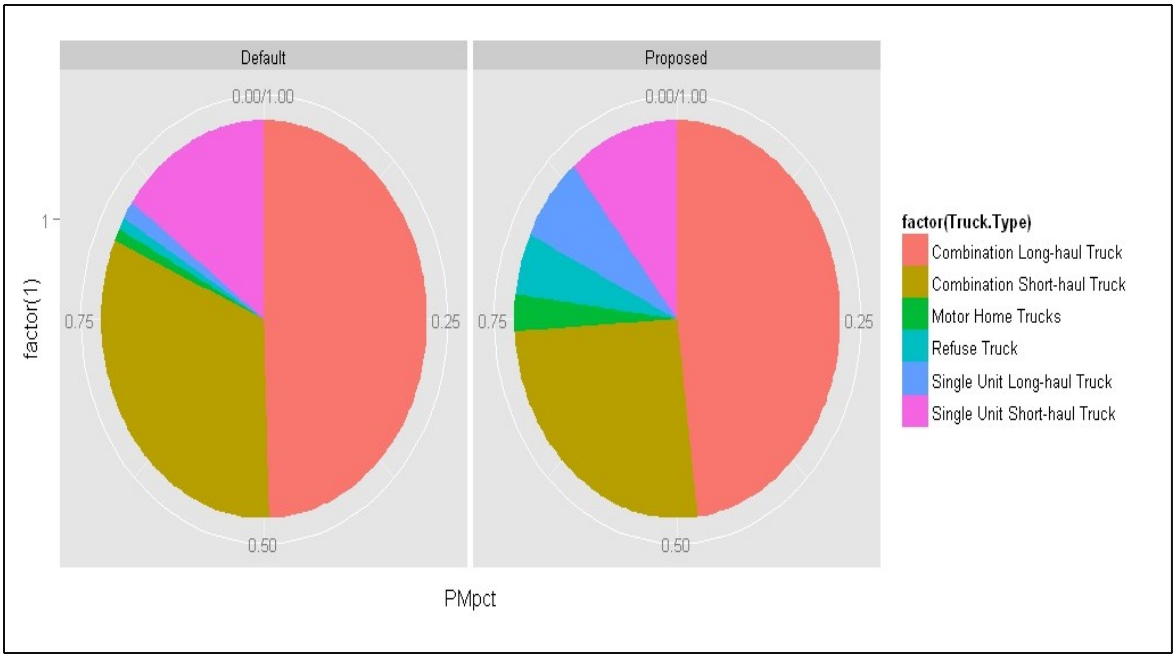

Figure 6-6 : Emission Contribution from Different Types of Trucks Estimated using Different Modeling Methodologies

The analysis of results from this case study provided us enough confidence that proposed truck models would give us better truck activity information than default data since the models are calibrated and validated. The important outcome of the case study is the ability to obtain the hourly emission inventory, which is very important for any standard air dispersion model

Chapter 7: Modeling Truck-Related $PM_{2.5}$ Pollution in Urban Atmosphere

7.1 Steps in Air Dispersion Modeling

There are multiple dispersion models available for modeling dispersion of pollution in urban atmosphere. Usually to model transportation-related air pollution models like CALINE4, CAL3QHC or CALPUFF are used. These models are Steady State Gaussian plume models, which describe the transport of a pollutant from a source to a receptor. They can model the concentration near road ways effectively. However, to model the pollution dispersion at the urban level, these models fall short since they do not consider road way grade, atmospheric chemistry (chemical reactions between different pollutants), and they can model only line sources. The urban pollution model, US-EPA's AERMOD, is very useful due its capability for modeling all types of source complex terrains and larger areas. Other specific advantages of using AERMOD for this particular research problem have already been explained in the literature review chapter.

There are multiple modeling and data processing steps involved in regional air quality modeling. The steps involved in the current study are enumerated as:

1. Converting spatial and temporally detailed emission inventory as area sources or line sources
2. Extracting and processing meteorological data for study domain
3. Extracting and processing terrain data for the study area
4. Creating appropriate and accurate receptor grid for improved prediction of concentrations
5. Analyzing emission concentration and creating Isopleths or contour maps for visual representation of emission concentration

The detailed description of each step to model urban-level, truck-related $PM_{2.5}$ emission concentration estimation for the Cincinnati area is explained in the following sections.

7.2 Conversion of Link-Level Emissions into Gridded Emissions

In current practice, to model mobile sources for the project-level analysis in AERMOD, the highways are considered as rectangular area sources. This assumption introduces a lot of complex geo-processing into the modeling process as all the highways need to be converted into rectangles of length equal to highway link length and highway width. The emissions may not be motionless in the atmosphere. So we can conveniently simplify source-shape modeling by considering all the link emissions in a cell (of predefined size) as single source instead of individually representing those links as thin rectangles. The next step in the modeling procedure is to convert the hourly link emission inventory into hourly, gridded emission inventories to address temporal and spatial distributions of truck emissions.

Selection of optimum cell size is very important in dispersion modeling. For country-level modeling, US-EPA uses 12kmx12km cell size and for state-level modeling domain, California uses 4kmX4km grid cell size. For the regional level, we noticed that most of the modelers use 1kmx1km cell size. Further, in the Cincinnati region, 98 percent of TAZs are larger than the 1X 1 km grid cell resolution. Generally, it would be ideal for TAZs and grid cells to have a comparable size. In addition, most TAZs do not have a regular geometric shape and, thus, the length of one side is not necessarily larger than 1 km, although the area of this zone is possibly much larger than 1 km^2. Therefore, the 1X1 km grid cells are reasonable in terms of resolution, and, accordingly, emissions at this grid level will be generated. In summary, we divide the modeling domain into 575 (=25X23) grid cells at a 1 km X 1 km resolution, according to the Universal Transverse Mercator (UTM) coordinate system. The UTM coordinates of southwest corner of modeling domain are 185825.42m east and 4326795.19 m north.

The link-level emissions need to be summarized based on grid cell size, and care should be taken about the links that extend beyond the cell boundaries. We have geo-processed the links such that the longer links are divided into smaller links to fit within the cell boundaries. Through this process, we have ensured that emissions are allocated where they belong. Since we have already summed up the running and start emissions at link level, we did not perform another extra step of allocating starting emissions to the grid cells. As stated earlier, our air quality modeling period is the whole month of July. The day of month and day of week factors were applied after gridding process.

We have carried some intuitive checks during this emission gridding process, and the important check among them was to look at the spatial patterns of daily gridded emissions as well as during peak hours like 7 a.m.-9 a.m., 3 p.m.-5 p.m. Fig. 7-1 and Fig.7-2 indicate spatial patterns of gridded $PM_{2.5}$emissions during the July 17, 2010, and based on the summary process explained above. The patterns are consistent with the regional industrial pattern (see Fig. 4-4) and transportation networks (see Fig. 4-8). For example, the grids along I-75 have the highest emissions because of a lot of freight movement through the region. Obviously the major emissions stretch along freeways and arterials leading to major industries such as GE Aerospace, etc. The comparative analysis spatial distribution provided us much confidence in the proposed methodology as it predicted high emissions near truck- activity-dominant areas. The default model's $PM_{2.5}$ spatial prediction was much higher than anticipated near the University of Cincinnati area as Light Duty Gas Vehicles are predominant in this area. Similarly, grid cells near Rookwood and Kenwood were also allocated large amount (i.e. 1800 gm. /sq. km) of $PM_{2.5}$

emissions when compared to surrounding industrial areas like Norwood. We suspect this may be due to the drawback in truck volume prediction, which is based on personal travel activity. Another very important location where the emissions do not match real-world activity was in the Hebron area. This area is a major employment and warehouse center. Emissions were around 600 gm. /sq. km/day, and the proposed methodology predicted above 1800 gm/sq km/ day. For better analysis of these spatial disparities, we have also calculated grid cell level emission differences between the two methods as shown in the Figure 7-3. We have observed a big spatial difference in Sharonville, Hebron, Erlanger, and the west side of Cincinnati, which are major industrial and employment areas. This quality-control process gave us enough confidence for our gridded emission inventory.

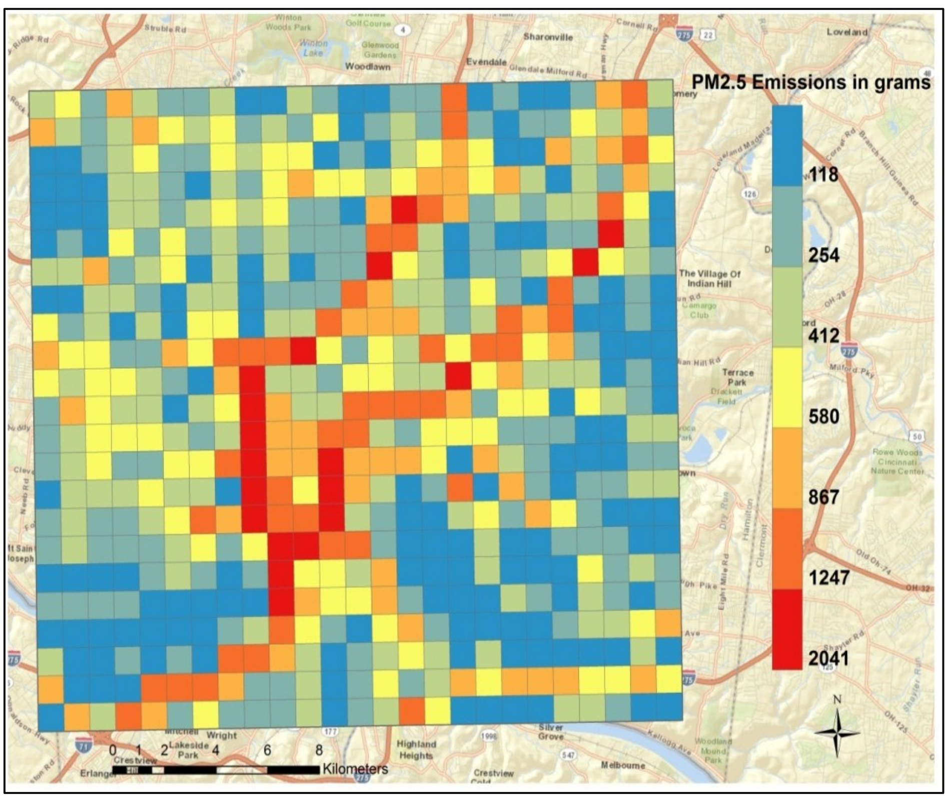

Figure 7-1: Gridded PM2.5 Mobile Source Emission Inventory Prepared using Default Methodology

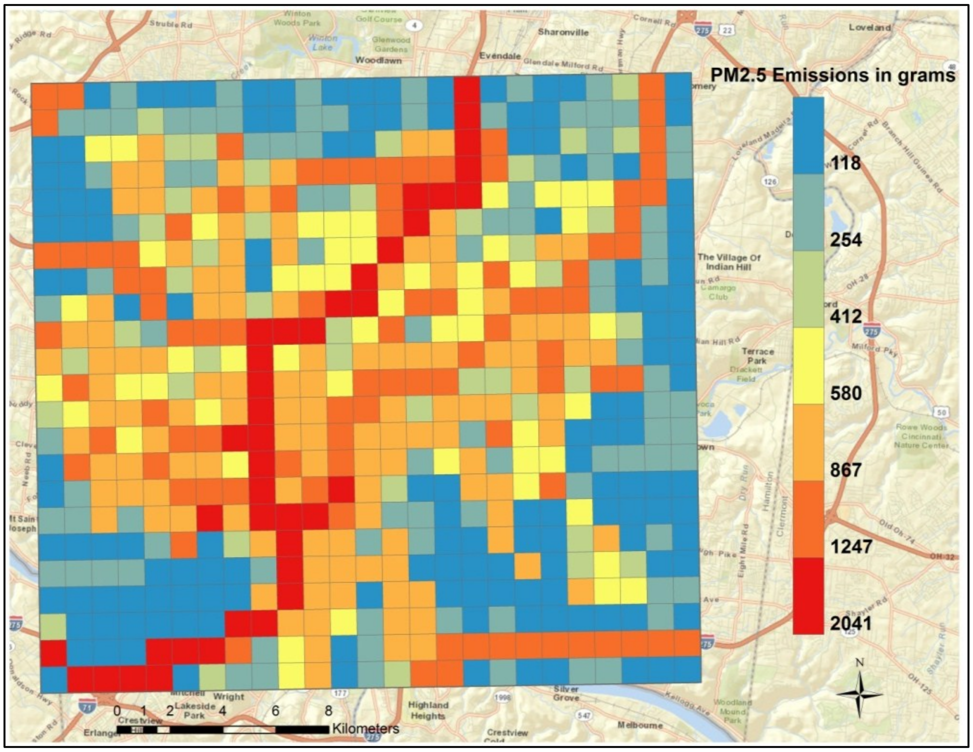

Figure 7-2: Gridded PM2.5 Mobile Source Emission Inventory Prepared using Proposed Methodology

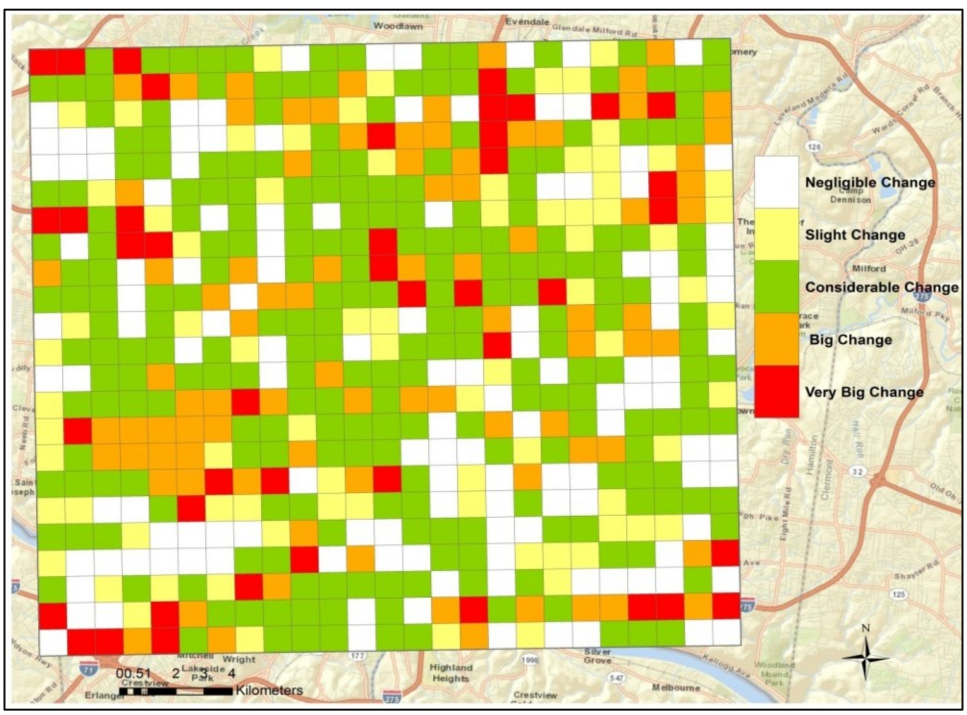

Figure 7-3: Differences in Gridded PM2.5 Mobile Source Emission Inventory Prepared using Default and Proposed Methodologies

7.3 Wind Data Processing

AERMET produces gridded, hourly-varying, three-dimensional wind field data using the wind direction data recorded at the weather observation station. For the current study, we have used the weather data from the NWS weather station (Station No: 93812) located at Lunken Airport. The latitude and longitude of the location are 39.103° and -84.418°, respectively, and the elevation of the station is 490 ft. above sea level. The location of weather station is shown in the following figure. Because of the significant changes in terrain elevation and land use across the domain, a grid resolution of 1000 meters (same as area source resolution) was chosen. The output from a metrological preprocessor would be able to predict accurate enough data for the dispersion model. Nine vertical layers (each around 300m) were chosen at 0-3000 m for this analysis. Based on elevations, AERMET calculates average predicted wind speeds at different grids. An example of a wind field for the month of July 2010 from AERMET is shown in Fig. 7-4.

The wind rose plot shows that the winds flows predominantly (45 percent of the time) toward the southwest during the whole month of July. Another major direction of flow of the wind is northeast, and 28 percent of samples were observed in that direction. We also analyzed the wind speeds during the month, and it can be considered as moderate as we have observed slower wind speeds. Figure 7-5 shows the relative distribution of wind speeds. Most of the samples were within 0-4 knots/hr. range. The wind speed and direction highly influence the pollutant transport and dispersion in atmosphere. The output from AERMET can be directly used in AERMOD without any further processing.

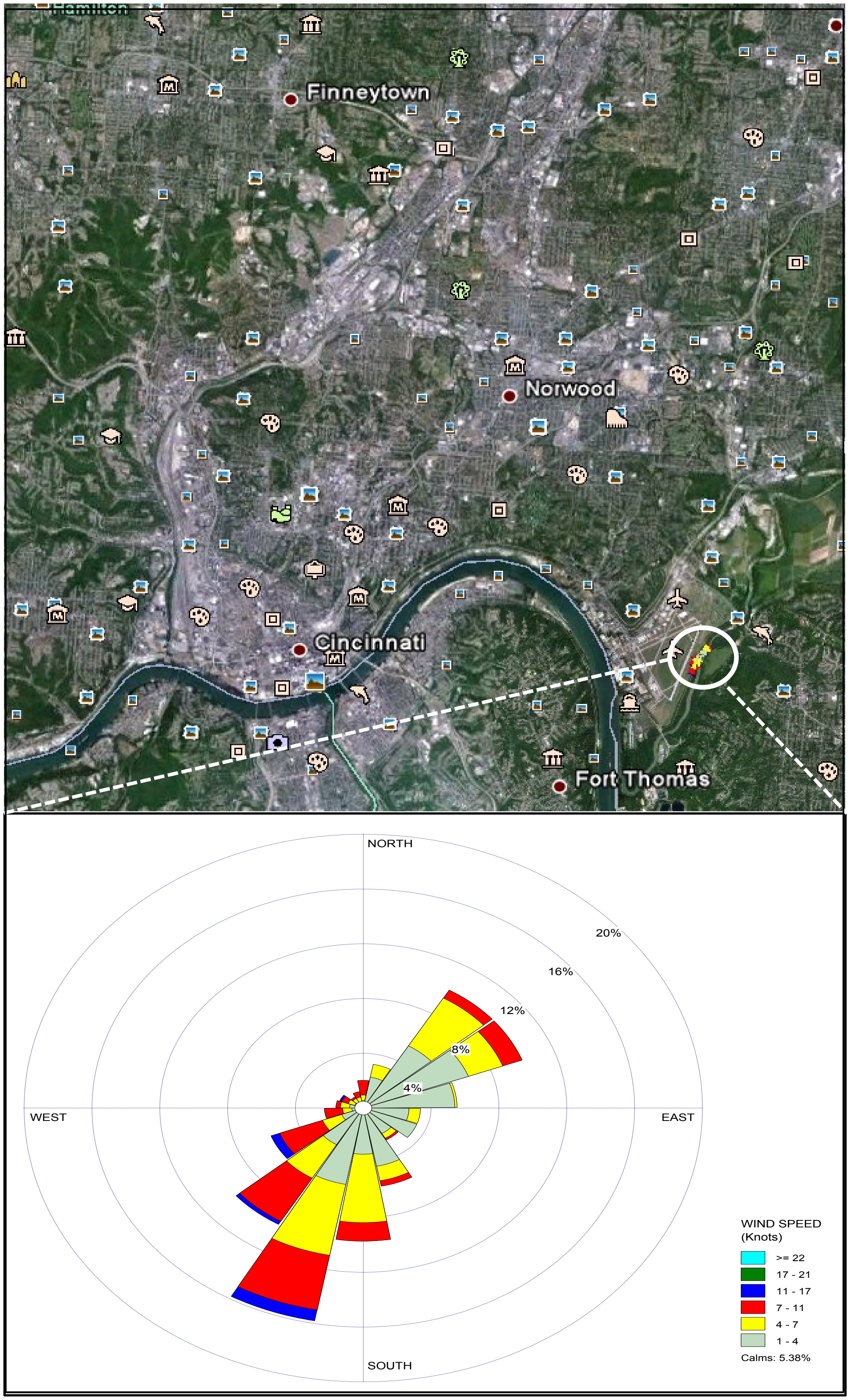

Figure 7-4: Wind direction distribution for the month of July 2010 at Lunken Airport, Cincinnati

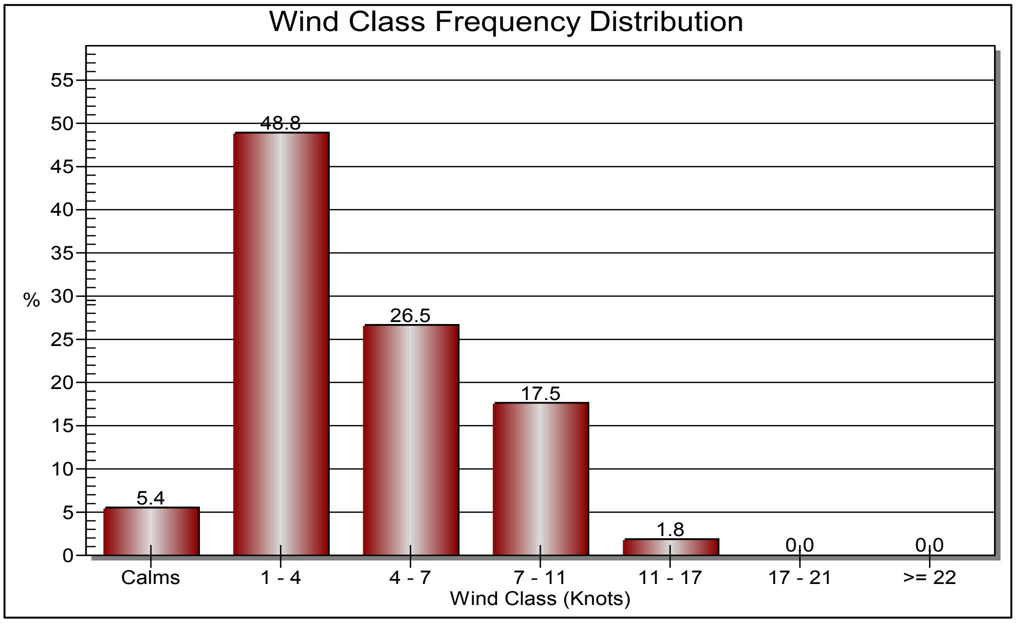

Figure 7-5: Wind speed distribution during month of July 2010

7.4 Terrain Data Processing

The terrain (geophysical data) is used by AERMET for meteorological processing as well as finding source and receptor elevations. As mentioned earlier, the terrain elevation data helps to characterize the wind patterns, such as the up-slope flows during daytime surface heating, and down-slope flows during nighttime surface cooling. The terrain data were derived from the USGS digital elevation model (DEM) with 7.5-meter horizontal resolution. Figure 7-6 shows the terrain of the present study area with distinct topographic sub areas, including the west side, the valley areas of Ludlow, Newport and Covington, the relatively level areas of the Northern Cincinnati and hills of Mt. Adams and Price Hills. Land-cover data was obtained from the U.S. Geological Survey (USGS). The AERMAP model is used to process the DEM data and create suitable terrain data for AERMOD. Another model AERSURFACE is used to process land-cover data for the dispersion model. The output from these models can be directly used in the AERMOD model. Figure 7-7 shows the 3-D view of processed terrain data for our modeling domain.

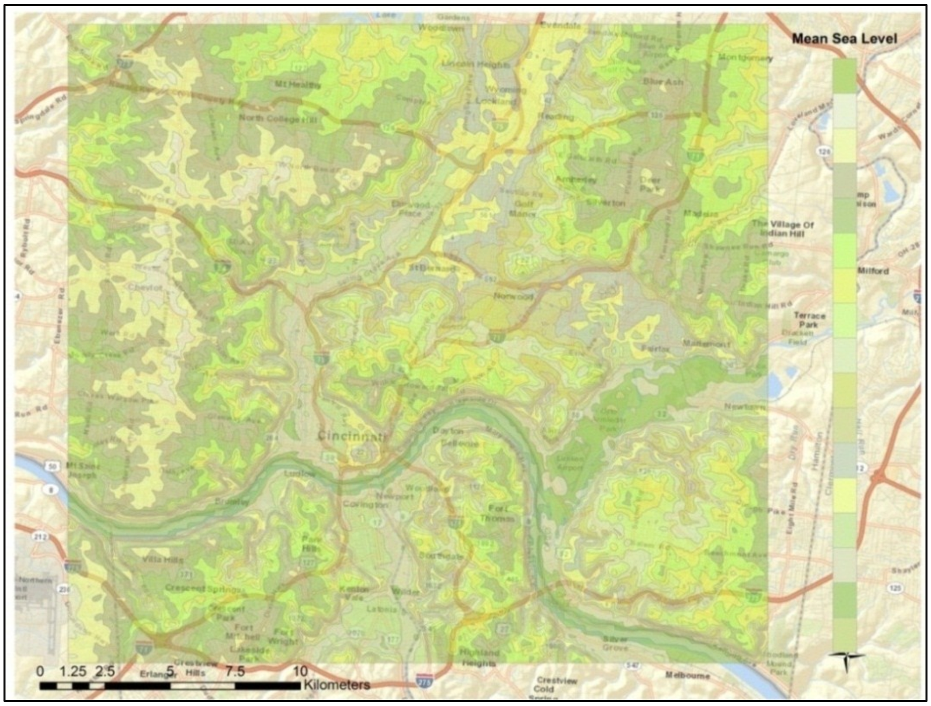

Figure 7-6: Processed Terrain Data for Modeling Domain

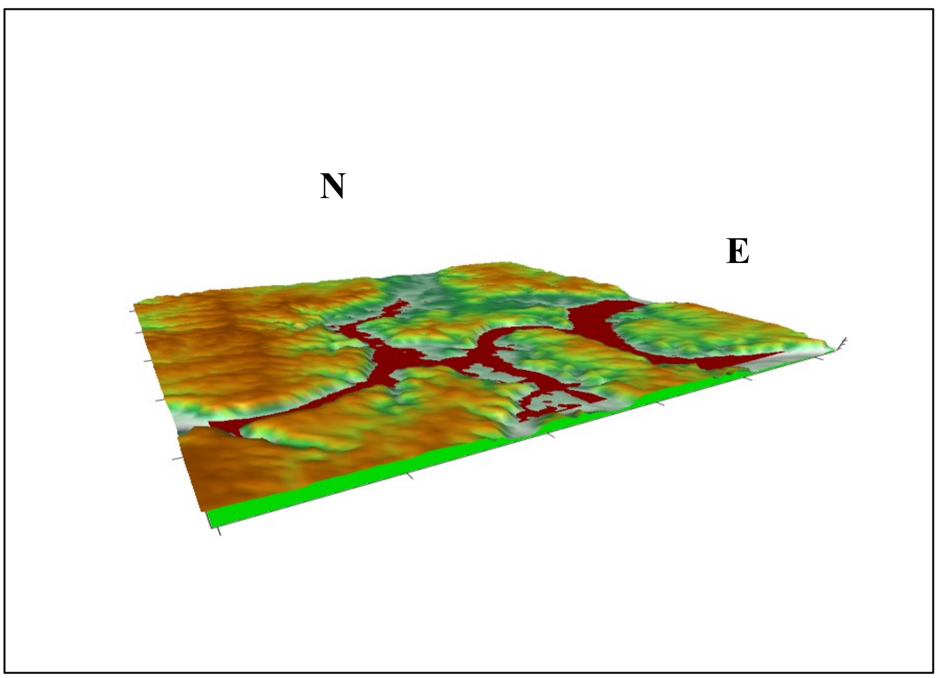

Figure 7-7: 3-D View of Terrain

7.5 Receptor Data

The AERMOD model predicts concentration at specific points or receptors, which are established by the user in the modeling domain. With the assumption that terrain will affect air quality concentrations at individual receptors, AERMAP first determines the base elevation at each receptor and source. For complex terrains, the terrain preprocessor searches for the terrain height and location that has the greatest influence on dispersion for each individual receptor. Both the base elevation and the hill elevations that influence dispersion are produced by AERMAP as a file or files that can be directly inserted into an AERMOD input control file. In this research, these receptors were located at all the vertices of grid cells, since these matches with source grids. In addition, receptors were also located at the three ambient monitoring sites to facilitate the model-to-monitor comparisons that are part of the model-validation process. There are a total of 579 receptors located across the modeling domain. Elevations for the centroids, or receptors, were taken from the DEM elevation data mentioned above. Fig. 7-8 shows locations of the established receptors in the modeling domain.

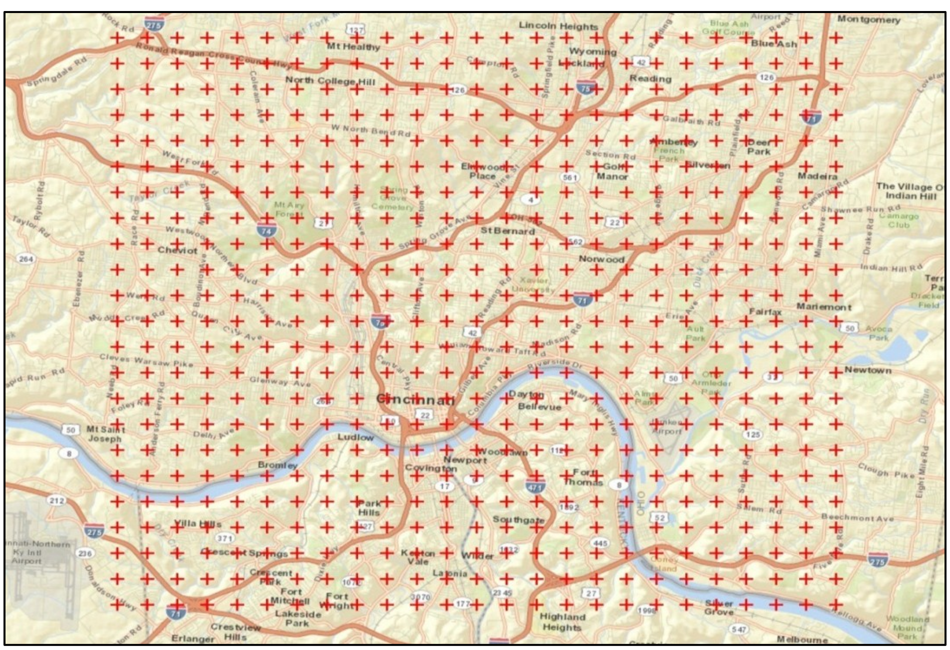

Figure 7-8: Receptor Grid

7.6 Dispersion Modeling

The modeling AERMOD is carried out using different files called pathways. Major pathways in this process are Control pathway, Source pathway, Receptor pathway, Meteorology pathway, and Output pathway. Control pathway is the most important input where we need to specify the pollutant considered, the concentration units, domain details, etc. in a standard format. AERMOD expects the source data in a standard fixed tab format in which we need to supply X and Y coordinates, SW corner of cells (area sources), and lengths in each direction, and their respective elevations should be supplied. Since the emissions vary during hour of day and day of the month, the emission inventories should be provided in another file. For each area source, information on the effective height, base elevation, and initial vertical dispersion coefficient (σ_z) were input. The latter (σ_z) takes into account traffic-induced mixing near the roadway as well as canyon effects, to a certain extent. Values for (σ_z) around 3 m (and up to 30 m) are commonly used for traffic-dispersion modeling. A value of 3 m for (σ_z) on all roadways was used in the current study, except in downtown Toronto where a value of 10 m was used. All these option are specified in "SO" pathway. As mentioned earlier, all the information about receptors including base elevations and hill height scale are supplied in Receptor or "RO" pathway. This pathway is output from the AERMAP model. Meteorology pathway contains the wind speed data for dispersion of the pollutant. Output pathway specifies the type of output and detailed level of output. In this file, we have specified the time aggregation, ranking of concentrations. We ran the model 30 times for 30 days in the month of July. The output from the model was stored and subsequently analyzed.

7.7 Analysis of Dispersion Results

The output contains the estimated concentration at each receptor in the grid as well as discrete receptors. It has been found that the July 8 had an overall 24-hour maximum dispersion and one-hour maximum $PM_{2.5}$ concentration in the region. Initially we analyzed the percentile distributions of the maximum values of 24-hour averages for default and proposed activity models. The proposed model has predicted higher concentrations, and the corresponding curve is linear for 70 percent of the receptors. Only 8 percent of receptors were predicted more than 4 $\mu g/m^3$ of concentrations. The default inputs resulted in concentration predictions much less than proposed and also the curve is not linear i.e. non-uniform predictions as shown in Figure 7-9.

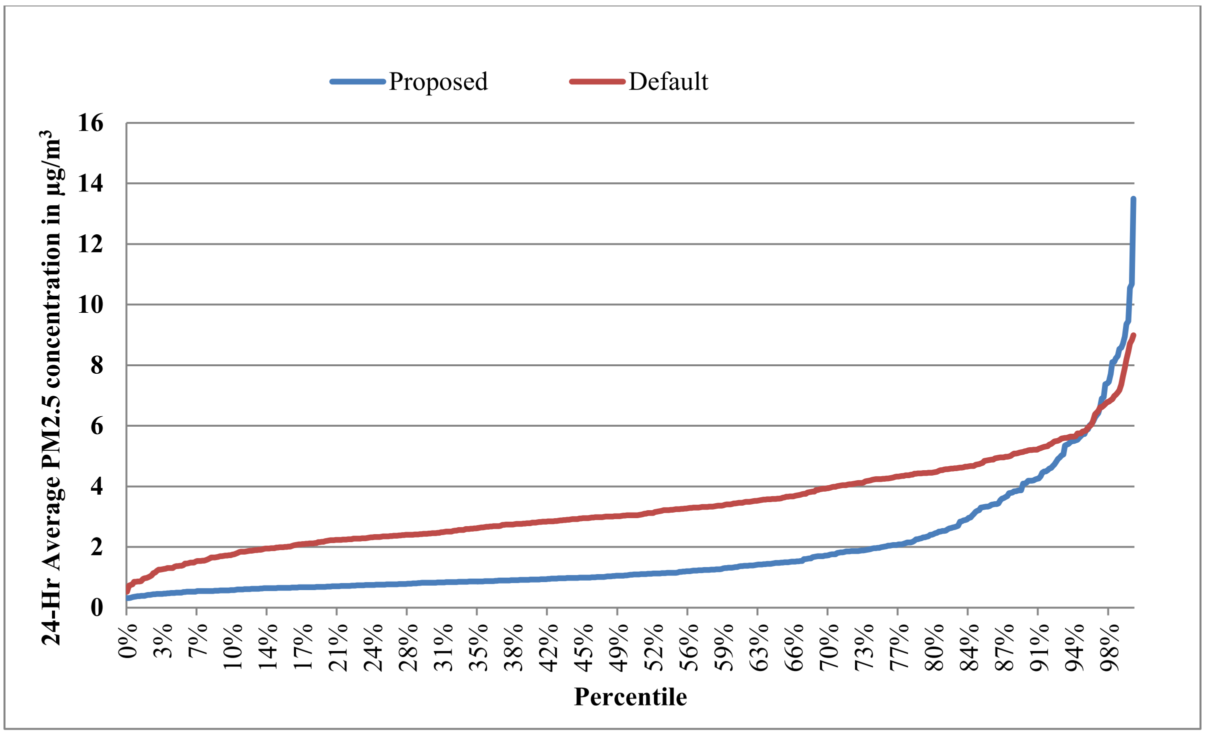

Figure 7-9: Percentile Plot of 24-hour Average PM2.5 values for the Month of July 2010 in the Entire Study Domain

For such a big domain, the concentration prediction is haphazard in nature compared to the proposed model. In the case of one-hour maximum concentration prediction, again the proposed model predicts higher concentration, and the exponential region is smaller compared to linear portion (Fig. 7-10). The curve fitting is also uniform for the proposed model. However, the default prediction in part of the curve is linear, which suggests the concentrations are uniform in the region

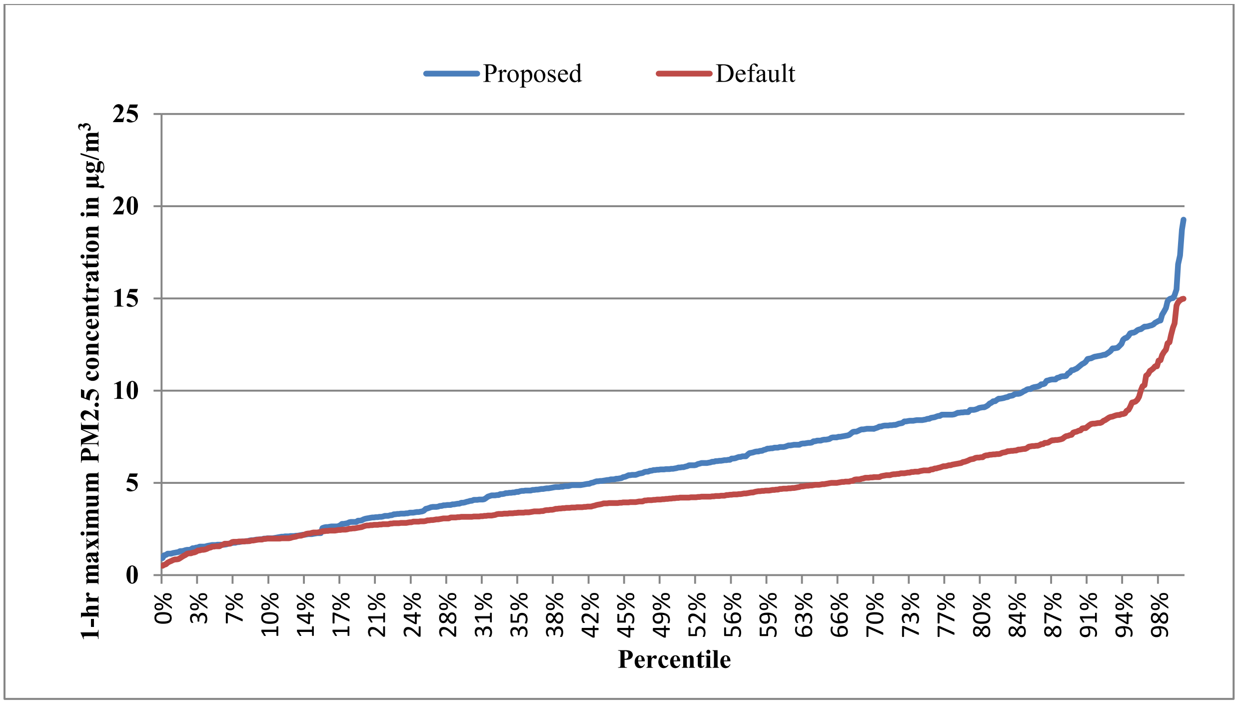

Figure 7-10: Percentile Plot of One-Hour Maximum PM2.5 Values for the Month of July 2010 for Entire Domain

For visual representation of concentration, we have created “Isopleths.” Isopleths are the lines that connect equal values, and in our case equal concentrations. We have plotted these lines on map and used color coding based on their intensity. The Isopleths created for using one-hr and 24-hr concentrations were developed using default and proposed methodologies. The one–hour Isopleth clearly indicated that the highest concentration predicted using the proposed models is higher than the values predicted using the default methodology (Fig. 7-12 and Fig. 7-14). The highest concentration was observed at the lowest elevation point south of the Ohio River, and the size is very small. The default map contour lines indicated the dispersion was more uniform and considerable east of I-71. However, the spatial distribution emission inventory indicated there were fewer emissions from I-71 as it is predominantly used by Light Duty and Heavy Duty Gas Vehicles, such as cars and pickup trucks.

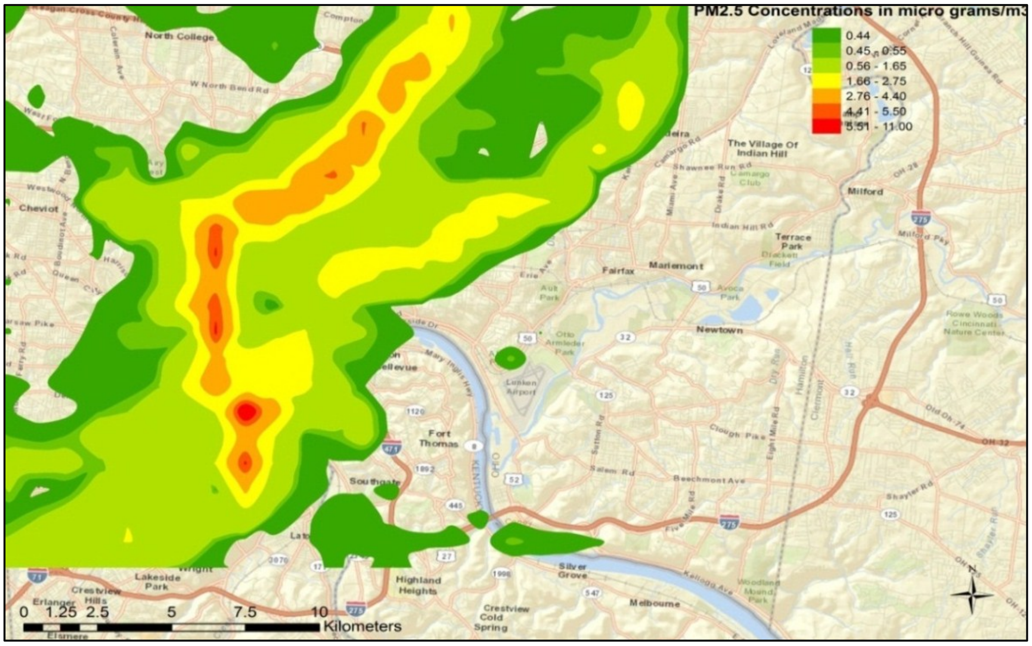

Figure 7-11: Estimated Dispersion of Truck Exhausted PM2.5 (24-hr average) in the City of Cincinnati using Proposed Methodology

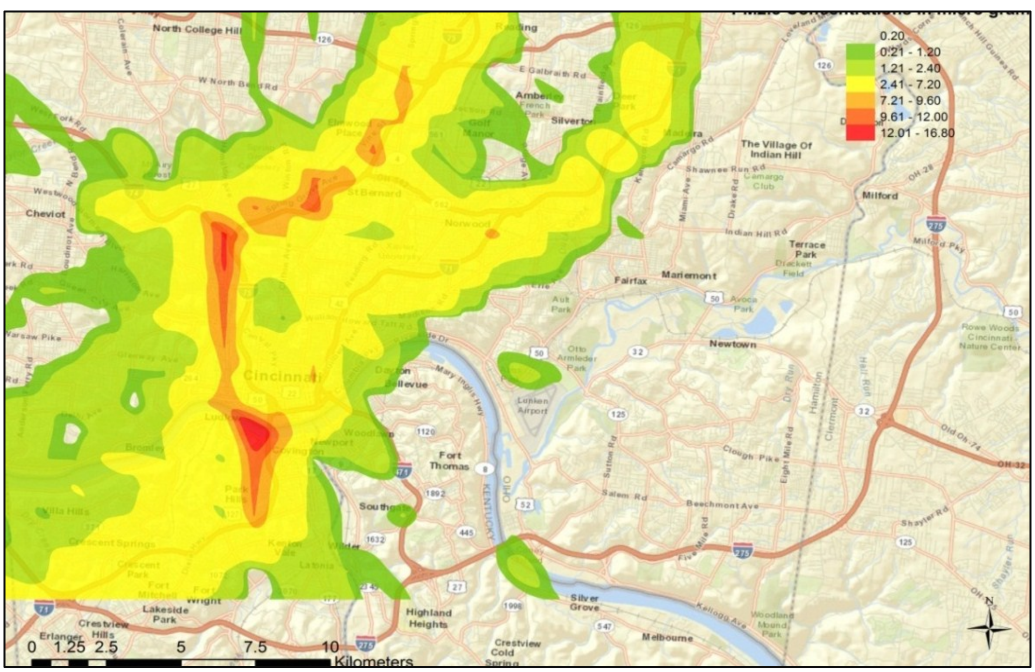

Figure 7-12: Estimated Dispersion of Truck Exhausted PM2.5 (max. 1-hr values in a Day) in the City of Cincinnati using Proposed Methodology

Similarly we have also plotted Isopleths for 24-hr average $PM_{2.5}$ concentrations (Fig 7-11 and Fig. 7-13). These contours also lead to the same conclusion as the analysis. Since the values are averaged for a 24-hour period, the values are actually smaller. The proposed methodology predicted more concentrated areas near I-75 and I-71 merge area south of downtown Cincinnati. Historically, the above-mentioned locations observe lot of $PM_{2.5}$ pollutions, which hints that the proposed methodology may predict reliable concentrations. As a next step, we have compared these predicted values with observed concentration at discrete locations.

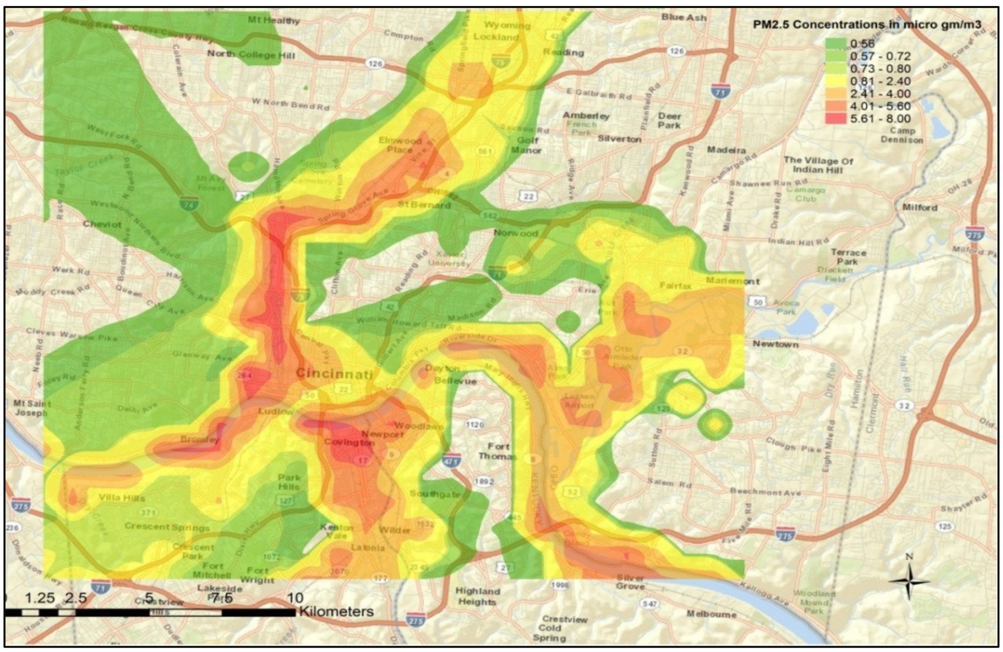

Figure 7-13: Estimated Dispersion of Truck Exhausted PM2.5 (24-hour average) in the City of Cincinnati using Default Methodologies

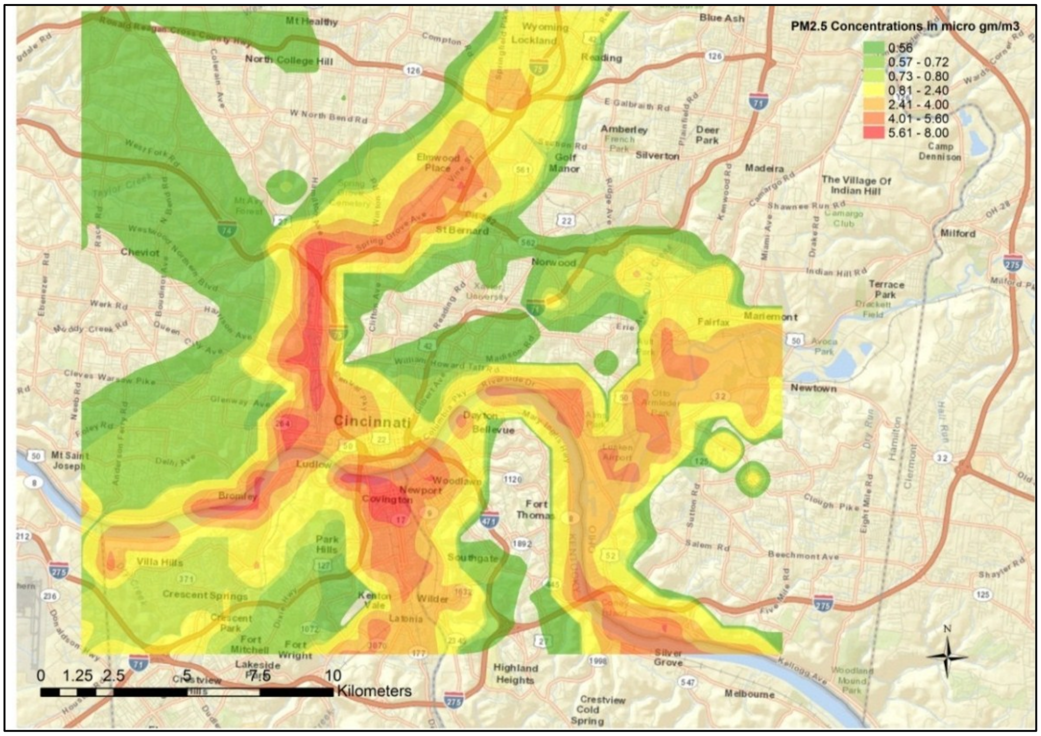

Figure 7-14: Estimated Dispersion of Truck Exhausted PM2.5 (max. 1-hr values in a day) in the City of Cincinnati using Default Methodologies

7.8 Comparison with $PM_{2.5}$ Concentrations Observed at Receptors

US-EPA monitors urban air quality to judge compliance with progress made toward meeting ambient air quality standards as part of their ambient air monitoring program. The SLAMS (State and Local Air Monitoring Stations) network is part of this ambient air monitoring program and consists of ~ 4,000 monitoring stations in the United States. Their size and distribution is largely determined by the needs of State and local air pollution control agencies to meet their respective State Implementation Plan (SIP) requirements. In the OKI region, there are 20 stations as part of the SLAMS network, and 17 stations are located in Ohio as shown in Fig. 7-15. Each of these locations monitors specific pollutants and $PM_{2.5}$ is monitored at eight different locations. Monitoring station situated on the William Howard Taft Road is the only locations within our modeling domain that collect continuous $PM_{2.5}$ concentrations in the ambient air.

CODE	LOCATION
A	Hamilton
B	Middletown
C	Batavia
D	Sacred Heart School
E	Colerain
F	Verity School
G	Taft (HCDOES)
H	Ohio Bell
J	Carthage
K	Sycamore
L	Lebanon
M	Winton (Waldorf School)
N	Addyston
O	Lower Price Hill
P	Lockland
Q	Amanda
R	Yankee

Figure 7-15: Air Quality Monitoring Stations in Cincinnati area

The observed and predicted hourly concentrations are plotted as a function of time during the typical weekdays in 2010. To verify the reliability of proposed modeling methodology, we have used observations collected for four different seasons during months of January, April, July and October as shown in Fig. 7-16 to Fig. 7-19. Overall, the proposed set of models captures reasonably well the general trend in $PM_{2.5}$ throughout the weekday. The monitor records the maximum $PM_{2.5}$ concentration at a point of time, whereas the modeling could only predict maximum values for one-hour time aggregation, thus, a slight compromise of accuracy is expected. Also note that the predicted concentrations are less than a factor of four of the observations, and this is due to the fact that only on-road mobile sources are modeled in this research. The predicted values through application of default models were also used for comparison purposes. Clearly, the proposed models show the same pattern as observed values.

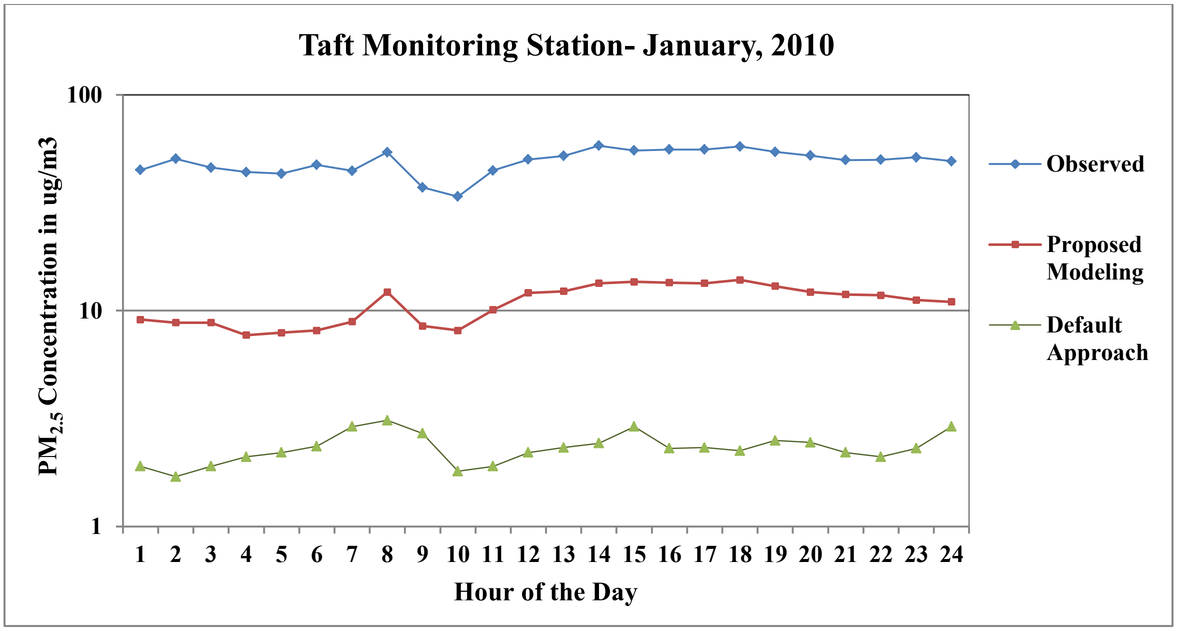

Figure 7-16: Hourly Average PM2.5 Concentrations at Taft Station, Cincinnati during a Typical Weekday in the Month of January, 2010 (Winter Season)

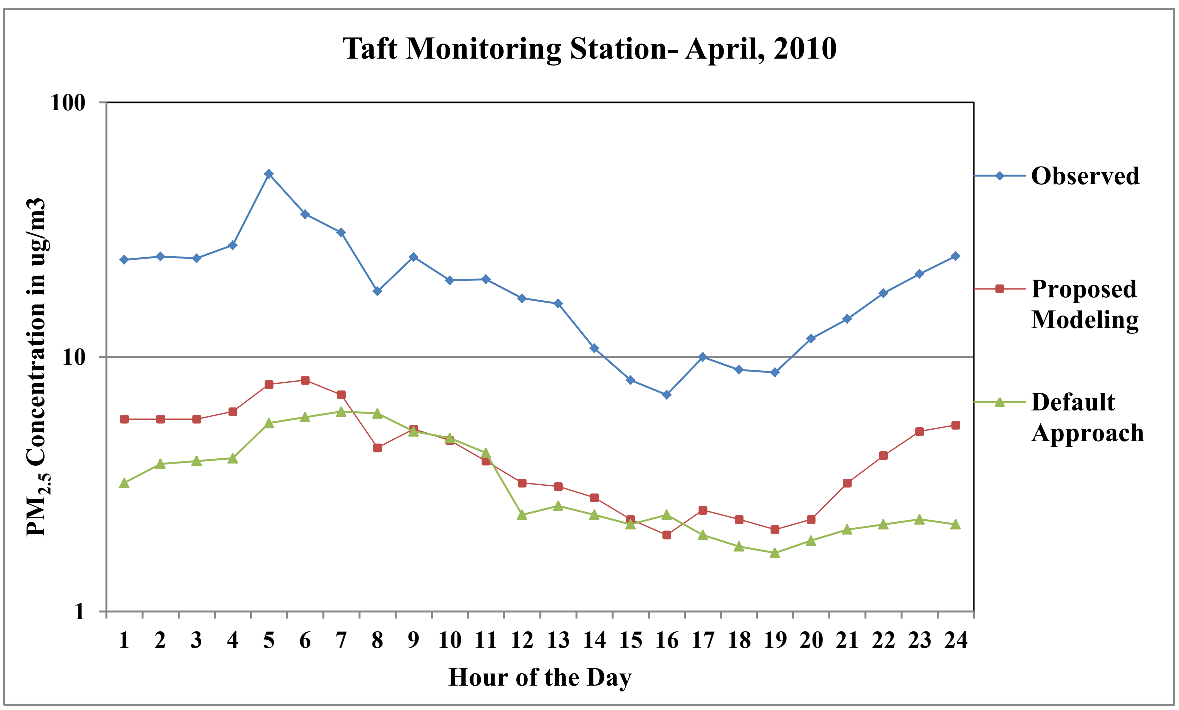

Figure 7-17: Hourly Average PM2.5 Concentrations at Taft Station, Cincinnati during a Typical Weekday in the Month of April 2010 (Spring Season)

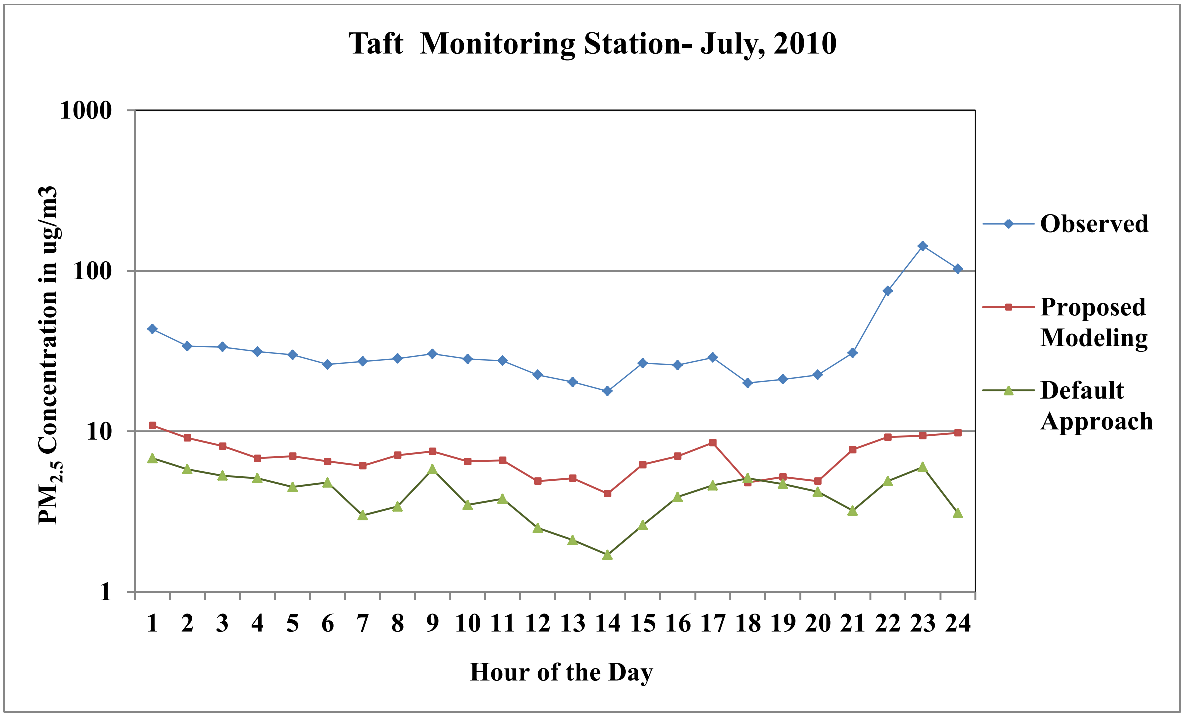

Figure 7-18: Hourly Average PM2.5 Concentrations at Taft Station, Cincinnati during a Typical Weekday in the Month of July, 2010 (Summer Season)

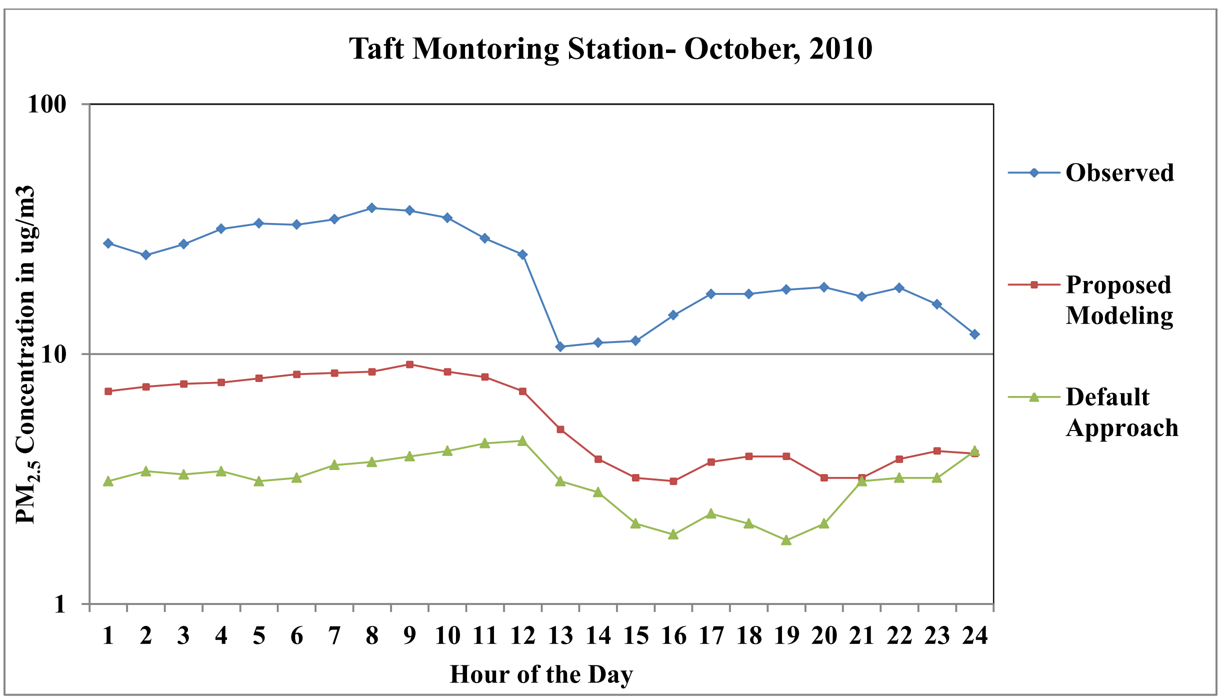

Figure 7-19: Hourly Average PM2.5 Concentrations at Taft Station, Cincinnati during a Typical Weekday in the Month of July, 2010 (Fall Season)

It seems the default predictions are homogenous throughout the day and the aberrations can be attributed to meteorological impact. Overall, the predictions during all seasons are much lower for default models, and we suspect this may be due to the inability to estimate the ground-truth truck activity and related emission estimation.

We have also estimated the Spearman's rank correlation[6] between observed and predicted $PM_{2.5}$ concentrations using default and proposed methodologies for four different seasonal weekdays. As shown in the Table 7-1, the correlation coefficients are ranging between 0.85 and 0.93 for all weekday hourly observations. On the other hand, the values predicted using default methodology has somewhat weak correlation with observed values and the Spearman's rank correlation coefficient values range between 0.34 and 0.71. In this research, we sequentially applied a series of models to identify the contributions of trucks to urban air pollution, using proposed truck activity models. The results are also presented in Table 7.1. Based on the proposed method with advanced truck models and directly using link-level emissions data, trucks contribute on average 22 percent of $PM_{2.5}$ emissions in urban areas based on monitoring station information. However, based on the default method of growth factor based truck model and disaggregating county emissions using VMT surrogates, trucks contribute 9 percent of urban $PM_{2.5}$ emissions.

---

[6] Spearman's correlation coefficient is a statistical measure of the strength of a monotonic relationship between paired data. It is constrained between values -1 and +1, and this statistic uses ranked data. Closer the value to 1 means stronger the monotonic relationship.

$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$. where d_i is difference between ranks of each observation and n is sample size

Table 7-1: Estimated $PM_{2.5}$ Contribution from Heavy Duty Trucks at Taft Monitoring Station using Different Methodologies

Method	Average Estimated Value	Average Observed Value	Spearman's Correlation	Contribution	Month
Default	2.32 $\mu g/m^3$	49.3 $\mu g/m^3$	0.34	4.7%	January
Proposed	10.9 $\mu g/m^3$		0.92	22.1%	
Default	3.36 $\mu g/m^3$	20 $\mu g/m^3$	0.71	16.8%	April
Proposed	4.37 $\mu g/m^3$		0.93	21.8%	
Default	4.18 $\mu g/m^3$	37.4 $\mu g/m^3$	0.53	11.2%	July
Proposed	7.04 $\mu g/m^3$		0.93	18.8%	
Default	3.15 $\mu g/m^3$	23.3 $\mu g/m^3$	0.55	13.5%	October
Proposed	5.86 $\mu g/m^3$		0.85	25.1%	

The contribution of diesel trucks to urban atmospheric fine particulate pollution seems to be varying by season if default methodology is applied. However, the HPMS data for the region seems shown very slight deviation in weekday truck activity among different months of the year. The default estimation of truck emissions comprises 46 percent of total on-road mobile source $PM_{2.5}$ emissions. It also seems to be underestimating their impact on urban air quality. The proposed model estimated that 81 percent of urban mobile source related $PM_{2.5}$ pollution is caused by truck activity. These contribution results indicate that the impact of truck activity on $PM_{2.5}$ pollution is very significant, and they can be very influential in air quality regulatory analysis. The analysis also provides evidence that supports that emissions calculated on the traditional truck-travel-demand modeling process tend to be underestimated compared to advanced regression-based truck models and they are not sufficiently accurate for air quality research. Further, estimating link-level emission inventory is particularly useful for such air quality modeling work.

Chapter 8: Conclusions

8.1 Contributions of this Research

The contribution of the research will be reflective of the following aspects:

1) *Comprehensive modeling approach-* Development of the comprehensive methodology for high resolution heavy duty truck related $PM_{2.5}$ air pollution prediction is useful for modelers who are preparing mobile source emission budgets, transportation conformity and who are modeling community health impact due to $PM_{2.5}$ pollution. Application of present research provides them improved inputs for emission and dispersion models.
2) *Transferability and ease of application-* The proposed methods can be applicable to any other region in the United States as all of the data used in the current research is obtained from local metropolitan planning organization and US-EPA. The truck activity and hourly distribution models have been developed using limited data i.e. .traffic counts from only three percent of the total links in the region. Most of the metropolitan planning organizations have traffic counts of at least five percent of links in the modeling domain for regional planning and travel demand model validation purposes. Even though proposed truck modeling methodology needs time series data, missing values can be imputed.
3) *Spatially detailed activity-* The spatial regression and optimization model proposed in this research can predict link level truck activity for different types of heavy duty trucks. This kind of detailed activity inputs are useful to predict more accurate truck emissions and they actually take into account the socio-economic and roadway improvement changes into account.

4) *Temporally detailed activity-* The hourly distribution fractions obtained through application of the advanced multinomial probit modeling methodology can provide modelers more reliable temporal truck activity. Improved diurnal truck activity provides us better emission inventories.

5) *Application of advanced statistical modeling techniques-* Empirical data used in this research (or any other similar research) consists of spatial attributes and collected over consecutive years. So, it is expected that the individual samples have inherited location and time-wise serial correlation. Even though the interaction among covariates is complex, the specification of both the models used in activity modeling (i.e. Spatial Regression for truck volumes and Multinomial Probit Model for hourly factors) is relatively straight forward. Moreover, the estimation procedure of these models not only nullify the errors inherent with location and time-wise serial correlation in the data but also have very significant goodness-of-fit measures when compared to traditional OLS regression models.

6) *Better congestion based speeds-*In this research, we have used HCM highway assignment procedure for the link speed prediction and it requires hourly trip data to take care of peak hour congestion during the day. So, the hourly speeds estimated should have taken care of the congestion.

7) *Better validation of the models-* The model outputs have been extensively validated against latest traffic count data. Both the link level volume validation and hourly comparison analysis of the output from proposed models provided us high fidelity in proposed methodology.

8) *Improved emission rates and emission inventory*- Since the activity inputs used in this case study are a lot different from the inputs currently used by the local air quality modelers the emission rates are also different. The significant contribution of this study is that the emission inventory (daily and annual) estimated using the input data prepared/disaggregated using proposed models is higher than the emission quantities estimated using default aggregated input. This improvement is very important for regional air quality planning agencies since it would affect their emission inventory budgets revision for future years.

9) *Bottom-Up approach*- The proposed model provide an opportunity to prepare gridded temporal $PM_{2.5}$ emission inventories for air quality modeling as we can estimate link level hourly truck emissions. This bottom-up approach can predict much reliable emission inventories for the dispersion and photochemical models, thus much effective than top down approach of using surrogates for allocating county level emission totals to the grids. Further, the link level emissions also useful to aggregate the emissions at a finer grid size like 1km X 1km, which would difficult if top-down approach is used.

10) *More ground-truth prediction of hot-spots*-The default approach and relative data predicted dispersion over large area, thus unable to predict hot spots in the region. On the other hand the dispersion predicted using the data from proposed models predicted distinctive hotspots which are useful in community health impact modeling

11) *Realistic estimation of contribution of heavy duty truck emissions to urban air quality*- Unlike any other previous mobile source $PM_{2.5}$ air quality studies, this methodology predicted the contribution of heavy-duty diesel (HDD) and heavy heavy-duty diesel (HHDD) trucks contribution to the urban air quality using limited truck counts as this

improvement is very useful for decision makers and air quality modelers. Previously, modelers could not estimate actual contribution of truck activity to urban air quality independently and this information would be very useful in designing different travel demand management strategies to improve air quality.

Bibliography

Anselin, L. *Spatial Regression Analysis in R.* A Workbook for Students, Urbana-Champaign Urbana, IL 61801: Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics, University of Illinois, 2005.

Anselin, L. "Under the Hood. Issues in the Specification and interpretation of Spatial Regression Models." *Agricultural Economics* 27, no. 3 (2002).

Anselin, L., Ibnu S., and Youngihn K. ""GeoDa: An introduction to spatial data analysis." ." *Geographical analysis* (University of Illinois, Urbana- Champaign) 38, no. 1 (2005): 5-22.

Bachman, W. W., Sarasua, S. H., and Guensler, R. "Modeling Regional Mobile Source Emissions in a Geographic Information System Framework." *Transportation Research: Part C*, 2000: 205-229.

Bai, S., Chiu, Y. E., and Niemeier D. A. "A comparative analysis of using trip-based versus link-based traffic data for regional mobile source emissions estimation."." *Atmospheric Environment* 41, no. 35 (2007): 7512-7523.

Baltagi, Budi Long Liu. "Testing For Random Effects and Spatial Lag Dependence in Panel Data Models." (Statistics and Probability Letters) 78, no. December 2008 (2008).

Bhat, C.R., Conoor, S., and Nerella, S. *GIS-Integrated Traffic Input Models for Mobile-Source Emissions Analysis.* Technical Report, Austin, TX 78763-5080: Texas Department of Transportation, 2003.

Bigazzi, A.Y., and Figliozzi, M.A. "Impacts of freeway traffic conditions on in-vehicle exposure to ultrafine particulate matter." *Atmospheric Environment* 60 (2012): 495-503.

Boilé M., and Golias, M.M. "Freight demand statistical modeling: A classification and review." *46th Annual Forum of the Transportation Research.* Washington, DC: TRF Foundation, 2005.

Boile, M. P. *Estimation of Truck Volumes and Flows.* Technical Report, Washington D.C.: Federal Highway Administration, 2004.

Boilé, M., Golias, M.M., and Ivey, S. "A Bayesian hierarchical network for truck demand modeling." *A Transportation Research Board SHRP 2 symposium.* Washington, DC: Transportation Research Board of the National Academies, 2010.

Citi Labs Inc. *Cube Analyst User Guide.* Tallahasse, FL: Citi Labs Inc,, 2010.

Cohen, J., Cook, R., Bailey, C.R. and Carr, E. "Relationship between motor vehicle emissions of hazardous pollutants, roadway proximity, and ambient concentrations in Portland, Oregon." *Environmental Modelling & Software* 20, no. 1 (2005): 7-12.

Dallmann, T. R., and Harley, R.A. "Evaluation of mobile source emission trends in the United States." *Journal of Geophysical Research* 115, no. D14 (2010): 1-12.

Dresser, C. D. *A Midwest Regional Inventory of Heavy-Duty Diesel Vehicle Emissions.* MS Thesis, Madison: University of Wisconsin, 2008.

Elhorst, J.P. "Spatial Panel Data Models." In *Handbook of Applied Spatial Analysis*, by Getis A Fischer MM, 78-102. NewYork: Springer, 2010.

ESRI Inc.,. *Spatial Statistics.* July 11, 2010. http://forums.arcgis.com/forums/110-Spatial-Statistics (accessed April 17, 2011).

Federal Highway Administration. "Assessing the Effects of Freight Movement on Air Quality at the National and Regional level." *Federal Highway Administration, Office of Natural and Human Environment* 07/06/2011. http://www.fhwa.dot.gov/environment/air_quality/-publications/effects_of_freight_move-ment/ (accessed 06 29, 2012).

Federal Highway Adminstration. *Guidelines for Travel Demand Model Validation.* Referene Manual, Washington D.C.: Federal Highway Adminstration, 2005.

Federal Highway Adminstration. *Guidelines for Travel Demand Model Validation.* Technical Report, Washington D.C.: Federal Highway Adminstration, 2005.

Fraser, M.P., Cass, G. R., and Bernd R.T. "Particulate organic compounds emitted from motor vehicle." *Atmospheric Environment*, 1999: 2715-2724.

Frazier, C., and Kockelman, K. "Spatial Econometric Models for Panel Data: Incorporating Spatial and Temporal Data." *Transportation Research Record: Journal of the Transportation Research Board* (Transportation Research Record) 1902, no. 80-90 (January 2005): 80-90.

Frey, C.H., Rouphail, N.M., and Zhai, H. "Link-Based Emission Factors for Heavy-Duty Diesel Trucks Based on Real-World Data." *Transportation Research Record: Journal of the Transportation Research Board* (Transportation Research Board of the National Academies) 2058, no. 1 (November 2008): 23-32.

Gardeazabal, J. "Vote Shares in Spanish General Elections as a Fractional Response to The Economy and Conflict." *Economics of Security* Economics of Security Working Paper, no. 33 (2010): 1-25.

Gauderman, W. James. "The Effect of Air Pollution on Lung Development from 10 to 18 Years of Age." (N Engl J Med) 2004, no. 351 (2004).

Gertler, A. W., Gillies, J. A., and Pierson,W. R. "An assessment of the mobile source contribution to PM10 and PM2. 5 in the United States." *Water, Air, & Soil Pollution* 123, no. 1 (2000): 203-214.

Hallmark, S. L., and Lamptey, S. "Evaluation of Different Methods to Calculate Heavy-Truck VMT." Washington, DC: U.S Department of Transportation, December 2004.

Health Effects Institute. *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects.* Technical Report, Boston, MA: Health Effects Institute, 2010.

Kanaroglou, P.S., and Buliung, R.N. "Estimating the contribution of commercial vehicle movement to mobile emissions in urban areas." *Transportation Research: Part E*, 2008: 260-276.

Kapoor, M., Kelejian, H.H., and Prucha, I.R. "Panel Data Models with Spatially Correlated Error Components." *Journal of Econometrics* 140, no. 1 (2007): 97-130.

Krzyanowski, M. *Health effects of transport-related air pollution: summary for policy makers.* Technical Report, Copenhagen, Denmark: World Health Organization, 2005.

LeSage, J. P. "Course on Spatial Econometrics." *Faculty of Economics, University of Coimbra, Portugal.* March 03, 2006. www4.fe.uc.pt/spatial/doc/lecture1.pdf (accessed May 11, 2011).

Lindhjem, C.E. et al. "Use of MOVES2010 in Link Level On-Road Vehicle Emissions Modeling Using CONCEPT-MV." *19th Annual International Emission Inventory Conference* . San Antonio, TX: US-EPA, 2010. 1-16.

Liu, F., Kaiser,R.G., Zekkos, M., and Allison, C. "Growth Forecasting of Vehicle Miles of Travel at County and Statewide Levels." *Transportation Research Record: Journal of the Transportation Research Board* (Transportation Research Record: Journal of the Transportation Research Board,Transportation Research Board) 1957, no. 1 (January 2006): 56-65.

Malcolm, C., Younglove, T., Barth, M., and Davis, N. "Mobile Source Emissions: Analysis of Spatial Variability in Vehicle Activity Patterns and Vehicle Fleet Distributions." *Proceedings of Annual Transportation Research Board Meeting.* Washington D.C.: Transportation Research Board, 2003.

Markakis, K., U. Im, A. Unal, D. Melas, O. Yenigun, and S. Incecik. "Compilation of a GIS based high spatially and temporally resolved emission inventory for the greater Istanbul area." *Atmospheric Pollution Research* 3, no. 1 (2012): 112-125.

Mittal N., Golias M.M., Boilé M., Spacovic L., and Ozbay K. "Estimating truck volumes on state highways-A statistical approach." *84 th Annual Meeting of Transportation Research Board.* Washington, DC: Transportation Research Board, 2005. 9-13.

Nair, H.S., and Bhat, C.R.,. *VMT-Mix Modeling for Mobile Source Emissions Forecasting: Formulation and Empirical Application.* Technical Report, Austin, TX 78705-2650: Texas Department of Transportation, 2000.

OKI Regional Council of Governments. *OKI Travel Demand Model Methodology and Validation Report.* Technical Report, Cincinnati, OH 45202: OKI Regional Council of Governments, 2009.

Pallia, D. et al. "DNA adducts and PM10 exposure in traffic-exposed workers and urban residents from the EPIC-Florence City Study." *Science of the Total Environment*, 2008: 105-112.

Papke, L. E., and Wooldridge, J. L. "Panel data methods for fractional response variables with an application to test pass rates." *Journal of Econometrics* (Elsevier) 145 (2008): 121-133.

Perugu, H. C. "A Rational Procedure to Estimate Hourly Traffic Distribution Factors for Post Model Applications." *Proceedings of TRB AppCon Meeting 2009.* Houston,TX: Transportation Research Board, 2009.

Perugu, H., Wei, H., and Rohne, A. "Modeling Roadway Link $PM_{2.5}$ Emissions with Accurate Truck Activity Estimate for Regional-Level Transportation Conformity Analysis." *ransportation Research Record: Journal of the Transportation Research Board* (Transportation Research Board of National Acadamies of Science) 2270, no. 2012 (October 2012): 87-95.

Prucha, I. R. "Ingmar R. Prucha's Research Programs." *University of Maryland,Department of Economics.* University of Maryland. May 05, 2011. http://econweb.umd.edu/~prucha/Research_Prog3.htm (accessed June 13, 2011).

Sivakumar, A., and Bhat, C.R. "Fractional Split-Distribution Model for Statewide Commodity-Flow Analysis." *Transportation Research Record* (Transportation Research Board of the National Academies) Transportation Research Record: Journal of the Transportation Research Board, no. Volume 1790 / 2002 (2002): 80-88.

Stata Corp.,. *Stata 11 help for xtgee.* StataCorp LP. March 17, 2011. http://www.stata.com/help.cgi?xtgee (accessed May 17, 2011).

US- EPA. *Air Quality Criteria for Particulate Matter.* Technical Report, Washington, DC: U.S. Environmental Protection Agency, Oct 2004.

US-EPA. *Integrated Science Assessment for Particulate Matter.* Technical Report, Research Triangle: National Center for Environmental Assessment-RTP Division, 2009.

US-EPA. *Mobile Source Emissions - Past, Present, and Future.* Policy Document, Research Triangle, NC: US-EPA, 2012.

US-EPA. *MOVES User Guide.* User Guide, Research Triangle Park, North Carolina 27711: United State Environmental Protection Agency, 2010.

US-EPA. *Particulate Matter.* Technical Report, Research Triangle, NC: US-EPA, 2011.

US-EPA. *$PM_{2.5}$ NAAQS Implementation.* Policy Document, Research Triangle, NC: US-EPA, 2012.

US-EPA. *Policy & Guidance.* Policy Document, Research Traingle, NC: US-EPA, 2009.

Wang, H., Jang, K., and Chan, C-Y. "Analysis and Multi-Level Modeling of Truck Freight Demand." *A Transportation Research Board SHRP 2 symposium.* Washington, DC: Transportation Research Board of the National Academies, 2010.

Wang, X., and Kockelman, K.M. "Forecasting Network Data: Spatial Interpolation of Traffic Counts using Texas data." (Transportation Research Board of National Academies) 2105 (2009).

Yoon, S., Rodgers, M.O., Pearson, J.R., and Guensler, R.L. "Engine and Weight Characteristics of Heavy Heavy-Duty Diesel Vehicles and Improved On-Road Mobile Source Emissions Inventories: Engine Model Year and Horsepower and Vehicle Weight." (Transportation Research Board of the National Academies) 1880 (2004).

Appendix I: Publication Accomplishments

1. Perugu, H., Wei, H., and Yao, Z. (2013). "An Improved Methodology for Modeling Truck Contribution to Regional Air Quality." *Proceedings for the 14th TRB National Transportation Planning Applications Conference*, Columbus, Ohio, May 5-9, 2013.
2. Perugu, H., Wei, H. and Rohne, A. (2012). "Modeling Roadway Link $PM_{2.5}$ Emissions with Accurate Truck Activity Estimate for Regional-Level Transportation Conformity Analysis." *Transportation Research Record: Journal of the Transportation Research Board, Vol. 2270 / 2012:87-95*.
3. Perugu, H., Wei, H. and Rohne, A. (2012). "Accurate Truck Activity Estimate for Roadway Link $PM_{2.5}$ Emissions." *ASCE Proceedings of 12th COTA International Conference of Transportation Professionals (CICTP 2012)*, Beijing, China. August 3-6, 2012.
4. Perugu, H., and Wei, H. (2011). "Development of an Integrated Model to Estimate Link Level Truck Emissions." *Proceedings of Futura 2011-Annual International Users Conference*, Palm Springs, California, October 29- November 4, 2011 (This paper is also selected as the 1st prizewinner of the Cube Student Challenge Competition 2011).
5. Wei, H., and Perugu, H.(2009). "Oversaturation Inherence and Traffic Diversion Effect at Urban Intersections through Simulation" *Journal of Transportation Systems Engineering and Information Technology, Vol.9, Issue(4):72-82*
6. Wei, H., Perugu, and Ai, Q., (2009). Integrated Arterial Signal and Access Control Strategy under Non-Notice Emergency Evacuation: Modeling and Simulation-Based

Analysis. *ASCE Proceedings of 9th International Conference of Chinese Transportation Professionals (ICCTP),* Harbin, China. August 5-9, 2009.

7. Wei, H., Konda, P., Yang, X.K., and Perugu, H. (2007). "Model for Quantitatively Estimating the Efficacy of Freeway-Arterial Traffic Diversion Strategy for Incident/Emergency Response: Methodology and Case Study," *Compendium of Papers CD-ROM for 86th Transportation Research Board Annual Meeting*, Washington, D.C., January 21-25, 2007.

Appendix II: Sample Cube Voyager Script for Application of Truck Volume Prediction Models

```
RUN PGM=HIGHWAY MSG='Calculate Refuse Truck Volumes'
FILEI TURNPENI = "C:\TruckEmissionProject\Hamilton.PEN"
FILEI LOOKUPI[1] = "C:\TruckEmissionProject\BPRCURVE.CSV"
FILEO PRINTO[1] = "C:\TruckEmissionProject\TEHWY00A.PRN"
FILEI NETI = "C:\TruckEmissionProject\TEHWY00A.NET"
FILEI MATI[1] = "C:\TruckEmissionProject\HAMILTON_RT.MAT"
FILEO NETO = "C:\TruckEmissionProject\LOADED_HAM_RT.NET"
LOOKUP NAME=BPRFUNC,
      LOOKUP[1]=1, RESULT=2, ; LINK CLASS 1
      LOOKUP[2]=1, RESULT=3, ; LINK CLASS 2
      LOOKUP[3]=1, RESULT=4, ; LINK CLASS 3
      LOOKUP[4]=1, RESULT=5, ; LINK CLASS 4
      LOOKUP[5]=1, RESULT=6, ; LINK CLASS 5
      INTERPOLATE=Y,lookupi=1
Damping = 0.5
PARAMETERS    COMBINE=EQUI    GAP=0.005    MAXITERS    =    10
time_cost  =  0.75         ;Weight  for  travel  time  as  impedance
distance_cost = 0.25    ;Weight for travel distance as impedance
PROCESS PHASE=LINKREAD   ; Calculate link travel time
      t0 = LI.TIME
      LW.Impedance=t0*time_cost+LI.DISTANCE*distance_cost
    ; Calculate the weighted impedance for path finding
    ; Calculate AM period roadway capacity
      C = LI.CAPACITY   ; 53% of traffic in AM period occurs in
the highest hour
    ; Set link class
ENDPROCESS
```

```
PROCESS PHASE=ILOOP
      PATHLOAD  PATH=LW.Impedance, PENI= 1,VOL[1]=MI.1.1
       IF(i=zones)
     LINKLOOP
          LW.PrevImp = LW.Impedance
     ENDLINKLOOP
  ENDIF
ENDPROCESS
PROCESS PHASE=ADJUST
function
{
 tc[1]=t0/BPRFUNC(1,V/C)    ; congested time function for Link Class 1
   tc[2]=t0/BPRFUNC(2,V/C)   ; congested time function for Link Class 2
   tc[3]=t0/BPRFUNC(3,V/C)   ; congested time function for Link Class 3
   tc[4]=t0/BPRFUNC(4,V/C)   ; congested time function for Link Class 4
   tc[5]=t0/BPRFUNC(5,V/C)   ; congested time function for Link Class 5
}
LW.Impedance=(time*time_cost+LI.DISTANCE*distance_cost)*(1-Damping) + LW.PrevImp*Damping
ENDPROCESS
ENDRUN
```