ABSTRACT

Dissertation Title:	MULTIMODAL TRAVEL BEHAVIOR ANALYSIS AND MONITORING AT METROPOLITAN LEVEL USING PUBLIC DOMAIN DATA
	Bo Peng, Doctor of Philosophy, 2019
Dissertation directed by:	Professor Lei Zhang, Department of Civil Engineering

Travel behavior data enable the understanding of why, how, and when people travel, and play a critical role in travel trend monitoring, transportation planning, and policy decision support. Conventional travel behavior data collection methods have been the primary source of travel behavior information for transportation agencies. However, the relatively high cost of traditional travel surveys often prohibits frequent survey cycles (currently once every 5-10 years). With decision makers increasingly requesting recent and up-to-date information on multimodal travel trends, establishing a sustainable and timely travel monitoring program based on available data sources from the public domain is in order. This dissertation developed advanced data processing, expansion, fusion and analysis methods and

integrated such methods with existing public domain data into a comprehensive model that allows transportation agencies to track monthly multimodal travel behavior trends, e.g., mode share, number of trips, and trip frequency, at the metropolitan level.

Advanced data analytical methods are developed to overcome significant challenges for tracking monthly travel behavior trends of different modes. The proposed methods are tailored to address different challenges for different modes and are flexible enough to accommodate heterogeneous spatial and temporary resolutions and updating schedules of different data sources.

Based on the number of trips by modes estimated using the proposed methods, the monthly trend in mode share can be timely estimated and continuously monitored over time for the first time in the literature using public domain data only.

The dissertation has demonstrated that it is feasible to develop a comprehensive model for multimodal travel trend monitoring and analysis by integrating a wide range of traffic and travel behavior data sets of multiple travel modes. Based on findings, it can be concluded that the proposed public domain databases and data processing, expansion, fusion and analysis methods can provide a reliable way to monitor the month-to-month multimodal travel demand at the metropolitan level across the U.S.

MULTIMODAL TRAVEL BEHAVIOR ANALYSIS AND MONITORING AT METROPOLITAN LEVEL USING PUBLIC DOMAIN DATA

by

Bo Peng

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee in charge:
Professor Lei Zheng, Chair
Professor Benjamin Kedem
Professor Ali Haghani
Professor Paul Schonfeld
Professor Qingbin Cui

© Copyright by
Bo Peng
2019

TABLE OF CONTENTS

List of Figures and Tables

Chapter I Introduction

1.1 Motivations and Research Objectives

Travel behavior data enable the understanding of why, how, and when people travel, and play a critical role in travel trend monitoring, transportation planning, and policy decision support. Departments of Transportation (DOTs) at both Federal and State levels have strategically invested in travel behavior information gathering. The National Household Travel Survey (NHTS) provides detailed information on trips in a given time period taken by a representative sample of households nationwide. Survey data such as the NHTS and regional/metropolitan travel surveys have been the primary source of travel behavior information for transportation agencies. The relatively high cost of traditional travel surveys often prohibits frequent survey cycles. Even for a large metropolitan area, comprehensive household travel surveys may be conducted once every 5~10 years or even longer. With decision makers increasingly requesting recent and up-to-date information on travel trends, establishing a sustainable and timely travel monitoring program based on available data sources from the public domain is in order.

Recent transportation policies also emphasize multimodal solutions and data driven approaches. Moving Ahead for Progress in the 21st Century Act (MAP-21) requires the establishment of a performance and outcome-based program at national, state, and metropolitan planning organizations (MPO) levels. To track performance measures timely and apply performance-driven approaches in practice, decision makers desire multimodal travel behavior information in frequent time intervals. For instance, Maryland Department of Transportation (MDOT) has introduced multimodal accessibility and non-auto mode share as performance measures for land development and transportation investment projects. Virginia House Bill 2 (HB2) calls for a performance-based prioritization process for statewide project selection. Transportation agencies are also interested in travel trend changes upon major new project openings and unusual incidents (e.g., adverse weather and disaster, extended infrastructure closure due to maintenance projects, *etc.*). These emerging information needs require more frequent estimation of multimodal travel trends and the associated transportation system performance of finer temporal resolution. Despite unprecedented emphasis on multimodalism, most transportation agencies currently do not have established data sources or tools for monitoring monthly or annual mode shares at the metropolitan level continuously.

Although there are many potential data sources for estimating monthly multi-modal travel trend, public domain data is the most preferable because of its low cost, open accessibility, relative stability, and transparency in data collection and processing methods. In the literature, there is no study that has comprehensively reviewed the public domain data of travel behavior from various sources and evaluated the feasibility of integrating these data to monitor month-by-month travel trends of multiple modes. To address this gap theoretically and practically, the main objective of the theoretical portion of this research is to construct a data fusion framework to provide monthly multimodal trend analysis at metropolitan level by only using public domain data sets which are accessible to everyone.

Chapter II Literature Review

2.1 Driving Mode

Transportation agencies have used many measures to gauge the travel demand for driving in a metropolitan area. For example, Vehicle Miles Traveled (VMT) has been used worldwide by transportation and planning agencies for various purposes. VMT is closely linked to urban/rural mobility, highway safety, fuel consumption, economic level, and environmental quality. While most transportation agencies only publish annual data, VMT estimation in higher resolutions, for instance, in different seasons of the year, weeks of the month, days of the week, and even hours of the day, can be vitally helpful for understanding the detailed travel demand patterns of a particular metropolitan study areas (Wang, 2011). The detailed information of travel demand is very valuable for decision-making in practice and can help the agencies better plan the transportation infrastructure. For instance, the accident occurrence rate is found to be correlated to passenger car VMT and truck VMT (Jovanis, 1986). With detailed VMT estimates, resources can be better allocated to the critical locations and the critical time-of-day in order to enhance traffic safety. Furthermore, high-resolution VMT estimates can also play an important role in estimating other transportation-related factors, such as environmental impacts (emissions such as PM 2.5 are highly correlated to VMT), land use impacts (VMT per capita is strongly and positively associated with population density (Cervero R, 2010)), etc. Despite all these imperative needs for accurate VMT estimates, the disaggregated and detailed VMT data or a rigorous estimation process based on existing data sources are not available in practice. In order to fill this research gap, it is of great necessities to develop a comprehensive data analysis method that takes the advantage of all the available public-domain data and estimates VMT in sufficiently fine-grained resolution.

There are two existing methods to estimate VMT: traffic count-based method and non-traffic count-based method (FHWA, 2010). The non-traffic count methods usually use other data sources such as population, number of licensed drivers, fuel sales, and number of registered vehicles (Liu, F, 2006). These data usually come from travel surveys and highway statistics. One recognized limitation of non-traffic count methods is that the data is expensive to collect; as a result, rough and/or out-of-date data from previous data collection efforts are often used, which may lead to questionable results (Kumapley, R, 1996). Traffic count-based methods are developed based on the assumption that the VMT can be estimated by the traffic volume data on some representative sections of roadways using rigorous statistical methods. Procedures developed by FHWA for "factoring" short-duration traffic counts can produce accurate estimates of annual average daily traffic (AADT) through a similar method of statistical analysis, AADT estimates for a set of road sections can be used to produce unbiased estimates of total vehicle miles traveled (VMT) for systems of roads. Rentziou, Gkritza and Souleyrettethe (2012) developed simultaneous equation models for predicting VMT on different levels of functional classes and examined how different technological solutions and changes in fuel prices could affect passenger VMT. Once AADT data was collected via traffic monitoring systems, VMT could be calculated by multiplying the volume to the length of the roadway segment and then scale up to the system using a set of weighting factors. The accuracy of this approach is highly related to the quality of the traffic volume data.

In the literature, most VMT studies were at the State level. To estimate mode share for a metropolitan area, we need to develop accurate VMT estimates for a metropolitan area and translate VMT into the number of vehicular trips. The latter is also related to the average trip length and vehicle occupancy in a metropolitan area, which has not been sufficiently discussed in the literature. To fill the aforementioned research gap, this dissertation proposes a data analytical method that will first accurately estimate high-resolution VMT in a metropolitan area and then estimate month-by-month vehicular trips based on VMT estimates using public domain data sources. A national level traffic count data (i.e. HPMS data) and an automatic traffic recorders (ATR) raw data will be used in this analysis.

2.2 Transit Mode

Transit ridership data has been analyzed at different levels to evaluate the performance of transit service providers, and to improve their efficiency. However, its application goes beyond operation optimization. To improve the transportation system, researchers also investigate travel behavior of transit riders and try to capture the factors that could help to boost the demand of transit by linking transit ridership data to other data sources. There have been many studies in the literature that address the travel behavior issues of transit riders.

Some studies investigated a wide range of factors that may affect transit ridership. For example, Taylor et al. (2009) conducted a cross-sectional analysis of transit use in 265 US urbanized areas using two-stage simultaneous equation regression models. National Transit Database (NTD) annual ridership data was used in the study. The authors concluded that ridership fluctuation crossed different regions was mostly explained by factors that out of control of transit service providers (e.g regional geography, metropolitan economy, population characteristics, and auto/highway system characteristics). They also confirmed that service frequency and fare levels did affect transit ridership. Thompson and Brown (2006) further extended the framework and considered more factors such as the geography and the building environment of the city using data from NTD. Taylor and Fink (2003) reached similar conclusions after a comprehensive review of existing literature.

Other studies focused on a more specific system and influential factors. Wang and Skinner (1984) investigated fare and gasoline price changes on monthly transit ridership using data of seven urban areas provided by American Public Transportation Association. Klan and Liu (1999) analyzed transit ridership in

Houston (all bus) and San Diego (bus and light rail) using annual data and concluded that the large ridership increases in both areas were due to service increases and fare reductions, and metropolitan employment and population growth. Hickey (2005) analyzed the impact of transit fare increases on ridership and revenue using monthly ridership data of New York City transit and found a lower than expected price elasticity. Sharaby and Shiftan (2012) studied the impact of shifting from a distance-based fare structure to a zone-based fare structure using fare box data. Chen et al. (2011) found both gas price and transit fare had significant impact on transit ridership using time series analysis and data collected from New Jersey Transit. A similar study was also done by Doi and Allen (1986).

Many researchers also considered factors beyond price. For example, Cervero and Landis (1997) analyzed the interaction between the Bay Area Rapid Transit system and land use patterns in the region. Ryan and Frank (2009) studied the correlation between walking environment and transit ridership, using data from San Diego's Metropolitan Transit Systems. Arana et al. (2014) analyzed the correlation between transit ridership and weather conditions using smartcard ridership data collected in Spain. Singhal et al. (2014) conducted a similar study using data collected from New York City Transit and further analyzed the impact of station characteristics such as weather protection, accessibility, proximity and the connecting bus services on transit ridership under different weather conditions. Kashfi and Bunker (2015) also conducted a similar study using data from New Zealand. Litman (2008) investigated the value of transit service quality through a comprehensive literature review and a comparable study with cases of driving.

Compared to the studies using actual transit ridership data, a lot more studies relied on stated-preference survey data and discrete choice modeling framework (e.g. Goodwin 1992, Ben-Akiva and Morikawa 2002, Cervero 2002, Ewing and Cervero 2010, Kim et al. 2007, Polydoropoulou and Ben-Akiva 2001, Frank et al. 2008). However, we must be cautious about generalizing findings from discrete choice analysis based on a small sample. For example, Pickrell (1989) compared the predicted and actual ridership of ten major capital improvement projects in nine urban areas during 1971-1987 and indicated significant gaps. Many of the ridership predictions were based on estimated choice models and elasticities.

The literature review shows that transit services are an important component of the modern transportation network in a metropolitan area and transit ridership is an important trend indicator of travel demand that transportation agencies closely monitor. However, most monitoring programs and analyses were based on one particular system, while many transit operators may coexist in the same metropolitan area. And none of the previous studies tried to monitor the trend in mode split in a metropolitan area over time, which could have important policy implications for transportation agencies. Based on the literature review and practice scanning, NTD is the only database that provides month-to-month ridership statistics for a majority of transit operators across the U.S. Therefore, it offers a great data source to address the needs identified in the introductory section and will be analyzed in detail in following sections.

2.3 For-hire Mode

The for-hire mode is a critical component of the multi-modal transportation system for metropolitan areas. It could serve either as the first-mile and last-mile solutions for other modes such as transit, or a standard-alone mode competing with other modes. Traditionally, taxi and escort service providers dominated the for-hire market. However, emerging mobility-on-demand start-ups such as Uber and Lyft are growing in popularity with urban travelers (Cramer and Krueger, 2016). Recognizing the potential impacts of service providers such as Uber and Lyft, several studies have been conducted to analyze the impact of these emerging ride-sharing companies on traditional for-hire service providers. Because of the relatively small market share of for-hire modes when compared to other modes, they attracted relatively fewer attention in empirical travel demand analysis. There are even fewer empirical studies on emerging ride hailing services such as Uber and Lyft because of data availability.

Among the few that exist in the literature, Correa et al. (2017) explored the spatiotemporal patterns of the demand for Uber and taxi at a Neighborhood Tabulation Area (NTA) level. Especially with the ArcGIS tools, the authored examined the spatiotemporal trip patterns of those two modes in 195 NTAs in NYC. They aimed to explore the changes in demand over time and space and factors that induced such changes. They also investigated the relation between the Uber and taxi demands. Three types of demand forecasting models (linear, spatial error, and spatial lag models) were developed and the spatial lag model outperformed the best. The models were based on transit accessibility, socio-economic and transportation-related factors and both spatial and temporal variations were considered. This paper is the first one literature that provides quantitative empirical analysis on for-hire mode ridership using Uber trip data. In addition, the demand forecast models considered the spatial dependence of Uber and taxi demands on accessibility to transit services. Results from the demand forecasting models indicated that the areas with lower transit access time, higher length of roadways, lower vehicle ownership, higher income and more job opportunities tended to generate higher Uber and taxi trips. From the empirical analysis, the authors showed a high correlation between taxi and Uber pick-ups, while Uber demands were more evenly distributed throughout the city and had a longer peak spread compared to taxi demand.

Cramer and Krueger (2016) investigated the efficiency of UberX versus taxis with the comparison of capacity utilization rates in five cities: Boston, Los Angeles, New York, San Francisco, and Seattle. The capacity utilization rates were defined as the ratio of 'the number of hours with a passenger in a car' to 'drivers' work hours in a given day.' The capacity utilization rate for taxi was estimated using the detailed service logs. However, the raw data for Uber was not directly accessible and the capacity utilization rates in the five cities for UberX were directly reported by Uber Research Staff. The latter illustrated the challenges of working with proprietary data, which motivated the project to rely on public domain data only. Authors pointed out four possible reasons for higher capacity utilization rate of UberX drivers, including 1) Uber's efficient driver-passenger matching technology; 2) larger number of available Uber drivers than taxi drivers in most cities due to flexible pick-up locations allowing more potential customers; 3) inefficient taxi licensing regulations only allowing pick-ups within the designated areas; and 4) Uber's flexible labor supply model and surge pricing. The capacity utilization rate indicated that if fares were linear without fixed costs, Uber could charge 28 percent less than taxis to have the Uber drivers earn the same amount of hourly revenue as the taxi drivers. The different utilization rates also suggested that taxi drivers drove more miles without passengers that could lead to more congestion and fuel consumptions.

Hall et al. (2015) conducted a comprehensive analysis on Uber drivers. The study compared the characteristics of Uber drivers and those of other workers based on the temporal changes in aggregated and anonymized historical driving data, schedules and earnings of Uber drivers from 2012 to 2014 as well as several survey results. According to the Beneson Strategy Group (BSG) survey, many Uber drivers were satisfied with the flexibility to choose their working time. Due to the flexibility of working days and hours, Uber attracted people who wanted to serve passengers without commitment of fixed working schedules. Uber drivers are more similar to the general workforce in age and education levels than to taxi drivers or chauffeurs. The authors suggested the reasons could be due to the higher unemployment rate during the study period, lower entry barriers for drivers than traditional for-hire modes, and the flexibility of working schedules. In addition, they mentioned that Uber drivers made similar or higher earnings compared to taxi drivers.

Wallsten (2015) studied the competitive effects of ridesharing in the taxi industry by testing hypothesis that the growth in ridesharing had led to a decrease in consumer complaints on taxis. He also introduced an index to measure the growing popularity of ride sharing. Trip datasets from NYC TLC, records of taxi complaints from NY and Chicago, and information from Google Trends on the popularity of Uber were employed for the analysis. The study showed that the taxi complaints were decreasing after Uber services entered, which might imply better taxi services. The Google Trend index also showed that Uber quickly gained its popularity after entering the market.

These studies in the literature showed some interesting trends in the for-hire mode market, especially after services such as Uber and Lyft entered the market. However, none of those studies developed a systematic method to continuously monitor the month-by-month travel demand for various for-hire modes, which is of great interest for understanding multimodal travel behavior for both policy makers and researchers. The proprietary nature of ridership data from companies such as Uber and Lyft posed a big challenge. This study will explore to what extent such challenges can be addressed using public domain data and what future efforts are needed to fully solve the problem.

2.4 Non-motorized Mode

Non-motorized travel demand data, such as the number of bicycles or pedestrians, are usually collected in selected locations during certain time periods (usually morning peak or evening peak). To derive the number of bicycle and pedestrian trips month by month from data collected during short time windows and a limited number of locations, the data analysis method has to take account of the spatial dependence among the locations, as well as the temporal correlation of the observations from the same location. Despite the growing interests of promoting non-motorized cities and the investments of agencies in non-motorized data collection, data entries of non-motorized traffic counts are very sparse and are deficient for time series analysis at the metropolitan level. In the literature, there have been some studies on the spatial and temporal analysis of the non-motorized travel demand. Some of them focused on developing regression models based on data for a specific location. For example, Phung and Rose (2007) used automatic count data in Melbourne to study bike path usage at an hourly level and a monthly level. Schneider et al. (2009) developed an OLS regression model to estimate the pedestrian intersection crossing volumes. Griswold et al. (2011) established a log linear ordinary least squares (OLS) regression model using 2-hour bicycle counts from 81 intersections. They drew several conclusions on the relationship between the bicycle volumes and the land use characteristics of the surrounding areas. Lewin

(2011) analyzed temporal patterns of available bicycle counts at weekly, monthly and seasonal levels using five-year continuous detector data from two permanent bicycle counting locations (with four counting stations). A regression model was estimated to evaluate the impact of weather conditions on bicycle volume. Hankey et al. (2012) collected non-directional count data from 259 locations on weekdays of September from 2007-2010. They compared the goodness of fit between a linear regression model with negative binomial model. Strauss and Miranda-Moreno (2013) evaluated the impact of demographics, built environment, bicycle facilities, road and transit network characteristics, and weather on bicycle volumes using data from both automatic counting stations and manually collected 8-hour bicycle counts.

Other studies targeted a larger area. Hudson et al. (2010) evaluated the data collecting methods for non-motorized trips. Casello et al. (2011) worked on the bicycle trip forecasting by conducting a two-step survey among 100 cyclists, including a stated preference survey and a GPS trip survey. They estimated the correlations between the bicycle trip rate and land use density. Nordback et al. (2013) tried to estimate the annual average daily bicycle and pedestrian trips (AADBP) utilizing both the short-term and the continuous count data from the count program of Colorado Department of Transportation (CDOT). They compared the hourly pattern of AADBP with the annual average daily traffic (AADT) of different motor vehicle types in order to identify the feasibility of applying the adjustment factors from motorized traffic to non-motorized modes. They also explored multiple statistical methods to consider the impact of the hourly weather conditions. Nordback and Sellinger (2014) outlined a sample-based method to calculate Bicycle and Pedestrian Miles Traveled (BMT and PMT) for the state of Washington. They derived the seasonal, daily and hourly adjustment factors from the continuous count data and further applied them to the short-term counts collected in the sample locations, each of which came from a unique roadway functional group. Gosse and Clarens (2014) leveraged biannual 2-h counts of approximately 15 locations and three total months of observations from a tube counter moved among three locations. They developed the temporal factoring method for sparse bicycle counts through Markov Chain Monte Carlo (MCMC) sampling process and introduced a novel spatial factoring method to expand bicycle usage estimates to all network edges.

The literature review showed that bicycle and pedestrian counts are usually sparse in both time and space for a metropolitan area. Most existing studies only focused on monitoring travel trend of non-motorized modes at one particular location. Data availability is the common challenge for extending these methods to a large metropolitan area. Statistical methods may help to address the challenges of scale estimates at a few locations to a larger metropolitan area. However, these methods need to be tested in more empirical studies.

2.5 Summary

The comprehensive literature review and practice scan showed that there have been studies on each of the fours modes that tried to develop method to monitor or estimate travel demand. However, many studies were case specific and could not support timely and continuous monitoring of multi-modal travel demands in a metropolitan area, which is very important for transportation agencies to understand emerging trend in travel patterns and make informed decisions accordingly. Data availability is the common challenge across all modes, although the data is particularly sparse for for-hire modes and non-motorized modes. Moreover, none of the existing studies in literature investigated the month-to-month trend in mode share for a metropolitan area by integrating methods for each mode into a coherent framework and developed a practice ready methodology for transportation agencies using only public domain data. This study intends to fill this research gap.

Chapter III Vehicle Miles Traveled Disaggregation

3.1 Vehicle Miles Traveled Review

Vehicle miles traveled (VMT), which has been used worldwide by governments and planning agencies, is closely linked to urban/rural mobility, highway safety, fuel consumption, economic level and environmental quality. While most transportation agencies only publish annual data, VMT estimation in higher resolution; in different seasons of the year, weeks of the month, days of the week, and even hours of the day, VMT can be vitally helpful to understand the detailed travel demand patterns of a nation, different states, or even smaller local areas. This information can be fully utilized in the decision-making process and help agencies better plan transportation infrastructures. For example, the accident occurrence rate is found to be correlated to passenger car VMT and truck VMT. With detailed VMT estimates, resources can be better allocated to the critical locations and the critical time-of-day to enhance traffic safety. Furthermore, high-resolution VMT estimates can also play an important role in estimating other transportation-related factors, such as environmental impacts (emissions such as PM 2.5 are highly correlated to VMT) and land use impacts (VMT per capita is strongly and positively associated with population density). Despite the imperative need for accurate VMT estimates, disaggregated and detailed VMT data or estimates have not attracted enough research attention. In order to fill this gap, a comprehensive data analysis method must be developed that takes advantage of available public-domain data, and that estimates VMT in sufficiently fine-grained resolution.

With the intention of filling the research gap, this study puts forward a data analysis method that accurately estimates high-resolution VMT by utilizing existing data sources. Specifically, HPMS data and automatic traffic recorders (ATR) data are employed in this analysis. The proposed method is widely applicable. Temporally, VMT statistics in different scales such as months, days and hours are estimated. Geometrically, VMTs on different roadway functional classes are estimated separately. Finally, this study differentiates VMT statistics by vehicle types. An accurate determination of passenger VMT and truck VMT is important, since trucks carry around 53% of the total freight in the U.S. While the estimation of VMT plays an important role in different aspects of urban development, it is usually limited by missing data, which can be efficiently handled by the proposed approach. The similarity between states and functional classes is quantified. The data gap is filled with statistics from other states weighed by the similarity index.

3.2 Methodology for Computing TEMPORAL VMT Adjustment Factors by vehicle types

The VMT estimation algorithm is computed based on the ATR raw data and HPMS data. These two data types play different roles in the proposed methodology. The ATR data were mainly used in the adjustment factor computing, while the HPMS data were regarded as the references of functional class classification. In addition, truck daily trend for each available function class was calculated based on the TMAS (Traffic Monitoring Analysis System) data. Data pre-processing procedures need to be done beforehand because the abnormal data will interrupt the program. The proposed method will be divided into two parts, including factors estimation and addressing data gap issue. The whole procedure and selected methods of each adjustment factor estimation will be described in the following subsections. To provide a more direct image for the method, the computational flow chart is visualized and presented in Figure 1.

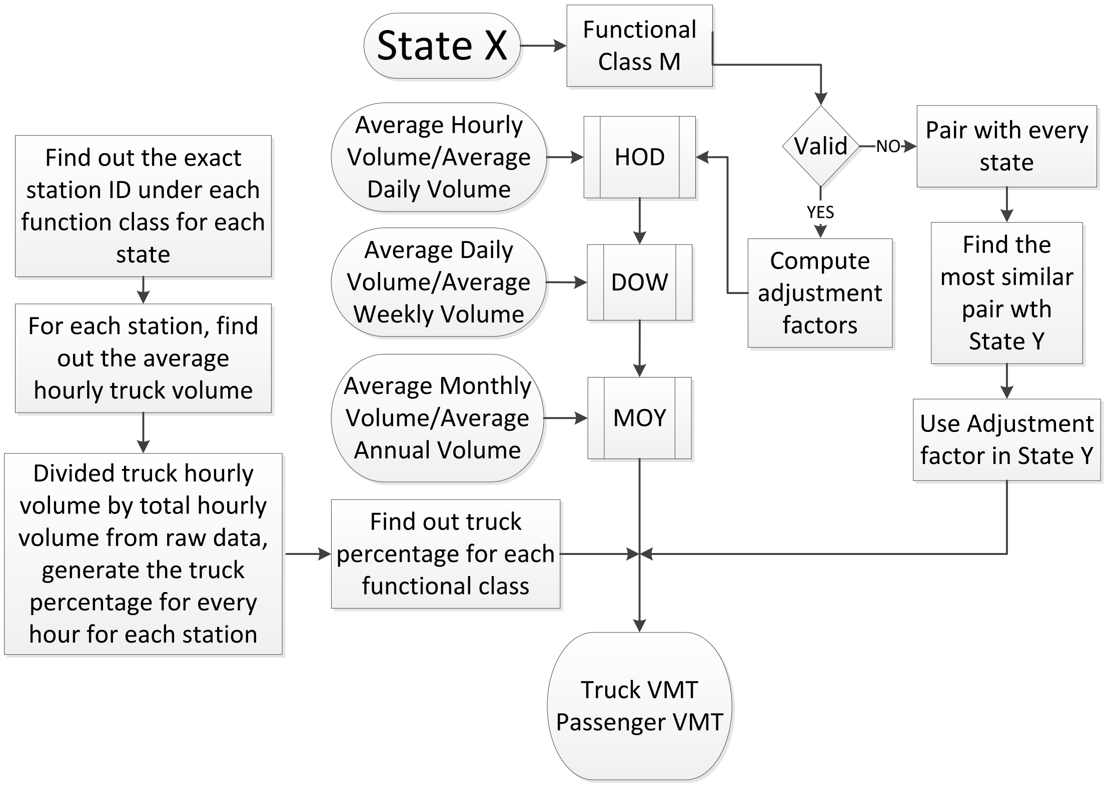


Figure 1 Flow chart for temporal VMT adjustment factors by vehicle types

3.3 Computational Algorithm

The illustrative explanation about whole procedure is explained below. The data pre-processing is required to ensure the validity of inputs. In estimating time-of-day factors, if the data from a specific counting station exhibit unreasonable trends (e.g., abnormally high hourly volumes) for any hour of a particular day, all data from that counting station from that day will be removed from the analysis. It was also decided to use those months with at least fifteen days of valid data to further ensure the quality of the data in analysis.

If there are valid counting stations in a State for a particular functional class of roads (e.g., rural interstate), the final hour-of-day, day-of-week, and month-of-year adjustment factors for that functional class in that particular state will be estimated by taking the volume-weighted average of the adjustment factor values of individual counting stations.

3.3.1 Hour of Day (HOD) Adjustment

HOD factors are used to describe the traffic volume trend within a single day. The HOD factor is calculated using the following steps. First, find out what and how many stations provide the data for a particular function class. Next, compute the HOD adjustment factors by dividing the hourly volume by the daily volume for each station.

Hour-of-Day factors for a particular station,

$$S_{HOD} = \frac{X_i}{x_i} / \sum_{i=i}^{24} \frac{X_i}{x_i} , \qquad (1)$$

where X_i denotes the total volume for hour i and xi represents the total number of days provides the data for hour i. After computing the HOD factor for all stations, the volume weighted average for all the stations with valid data for a particular function class was calculated; eventually, the results will provide the HOD factors for all the function classes for a particular state.

3.3.2 Day of Week (DOW) Adjustment

DOW is used to describe the traffic volume trend within a week. DOW factor is calculated by the following steps: First, find out the correlation between a particular function class and its stations for a particular state. For each station, calculate the average daily volume and average weekly volume and generate the DOW factor by dividing the daily volume by the average weekly volume for each station.

DOW factors for a particular station are calculated using the following equation,

$$Q_{DOW} = \frac{A_j}{a_j} / \sum_{j=1}^{7} \frac{A_j}{a_j} , \qquad (2)$$

where A_j denotes the total traffic volume for Day j and aj is the total number of days recorded for day j. Take the weighted average for all the stations with valid data for a particular function class and obtain DOW factors for all the function class for a particular state.

3.3.3 Month of Year (MOY) Adjustment

MOY is used to describe the traffic volume trend within a week. MOY factor is calculated using average monthly volume and annual volume. However, it is of great necessity to point out that unlike HOD and DOW factors, MOY factor has a minimum requirement for a valid month in that the data within that month should cover over fourteen days. After extracting the invalid month, MOY factor is computed by dividing the volume of the valid month by the annual volume for each station. Month of Year factors for a particular station are calculated using the following equation,

$$M_{\text{MOY}} = \frac{D_{\text{k}}}{d} / \sum_{k=1}^{12} \frac{D_k}{d} , \tag{3}$$

where D_k is the summation volume for a valid month and d represents the total number of days in that month. Finally, MOY factors for all function classes for a particular state can be obtained by taking the weighted average for all the stations with valid data for a particular function class.

Data gaps exist when, for example there may be no counting stations for a functional class in a state at all. A Similarity Test is adopted to describe the similarity between two different states for addressing the data gap and quality issue. It uses $\chi 2$ statistical tests on daily and monthly traffic trends as fundamental theory to define the level of similarity within a state pair. The outcome of the similarity test can be implemented into the adjustment factors substitution between different states and consequently, solve the data issue. The similarity test will be implemented through the following steps:

1. Identify the function classes for all states and Washington, D.C. with valid adjustment HOD factors.
2. Pair every single state with each other.
3. Conduct $\chi 2$ statistical tests with a state pair among all the available function classes with valid results.
4. Select the most similar state(s) to address the data gap issue.

This calculation estimates the vehicle type adjustment factors initially based on HPMS truck percentage records by functional class and by state, and then updates the vehicle type adjustment factors based on vehicle classification count data in the TMAS datasets. If there are data coverage gaps (e.g., truck percentages may not be available or a particular functional class in a state), statistical estimation methods similar to those described in previous will be applied to fill the data gaps.

3.4 Implementation of State of Maryland

In order to illustrate the application of the proposed method, the state of Maryland was analyzed. Detailed results in Maryland are presented below. In addition, the cross-state similarity analysis results are also presented using Maryland as the study area.

Figures 2 and 3 show the hourly traffic trends in Maryland for all functional classes in rural and urban settings, respectively. Each line is plotted by the percentage of traffic volume during each hour. Several features in the hourly trend figures concur with traffic demand logic. First, the results show the expected AM and PM peak periods. Except for rural local streets, other rural roadways have a rather flat AM peak travel volume. However, the AM peak-hour traffic trend in urban areas is more evident. This phenomenon could be caused by the fact that work trips and school trips are more likely to concentrate in urbanized areas and these trips are usually time sensitive; work/school schedules are usually not flexible. Therefore, the trip generation rate is greatly reduced after AM peak. I also find a smaller peak around noon using the urban HOD statistics, mainly for urban minor arterials, collectors and local streets. This trend could be caused by lunch trips localized in urban areas. Other significant findings include: 1) rural local streets and minor arterials carry more traffic during the mid-day period; 2) rural collectors (major and minor) serve a significantly higher proportion of the PM peak travel demand; VMT on urban interstates and major arterials are significantly higher in the early AM Peak (5:00 a.m. - 7:00 a.m.). These findings can be extremely valuable for understanding time-of-day travel demand patterns on different roadway classes and therefore will be useful in supporting decision-making processes.

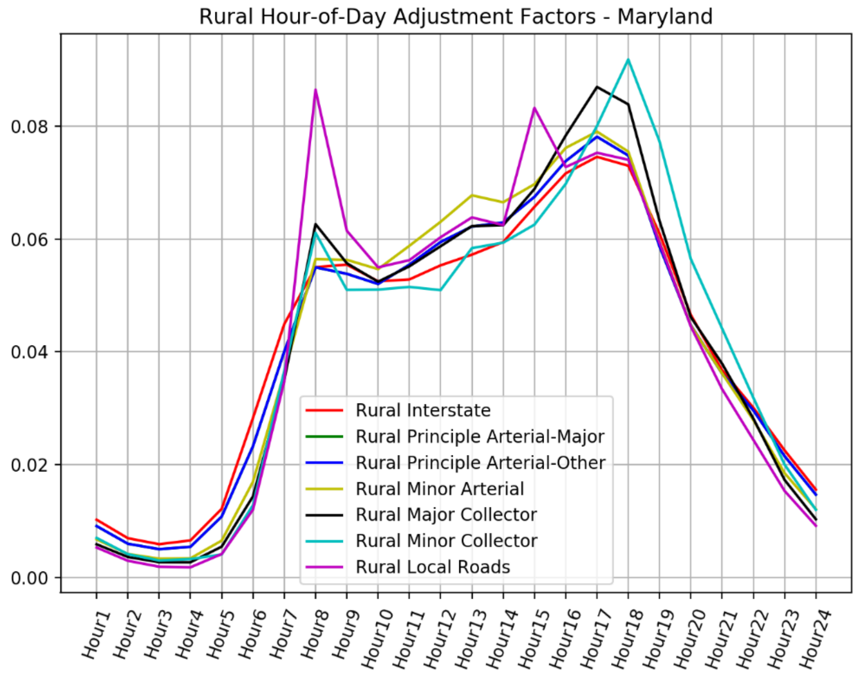


Figure 2 Rural hour-of-day adjustment factor – Maryland

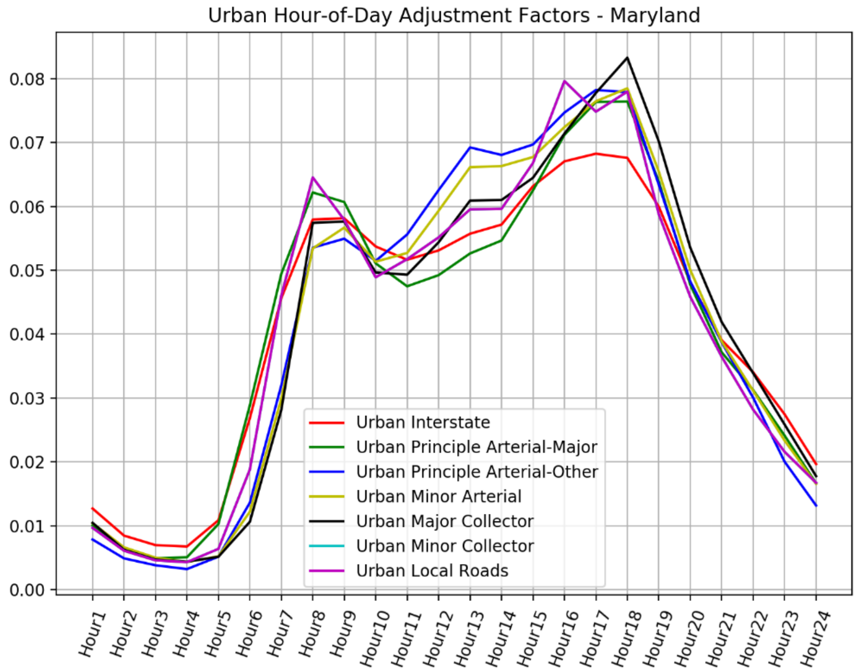


Figure 3 Urban hour-of-day adjustment factor – Maryland

Figure 4 illustrates hourly truck percentage within one day. The most significant finding is that the truck percentage of rural interstate exhibits an extremely large percentage of traffic volume during nighttime (between 10:00 p.m. and 6:00 a.m., roughly) and the peak truck VMT occurs at around 4:00 a.m. This is mainly due to two reasons: first, the majority of truck drivers will select off-peak periods as their departure time in order to avoid congestion; second, based on the result of Hour-of-Day adjustment factors shown in Fig. 2, the total VMT at 4:00 a.m. has the lowest percentage, which means that the truck percentage change can be significantly amplified, even though the increase of trucks is not conspicuous.

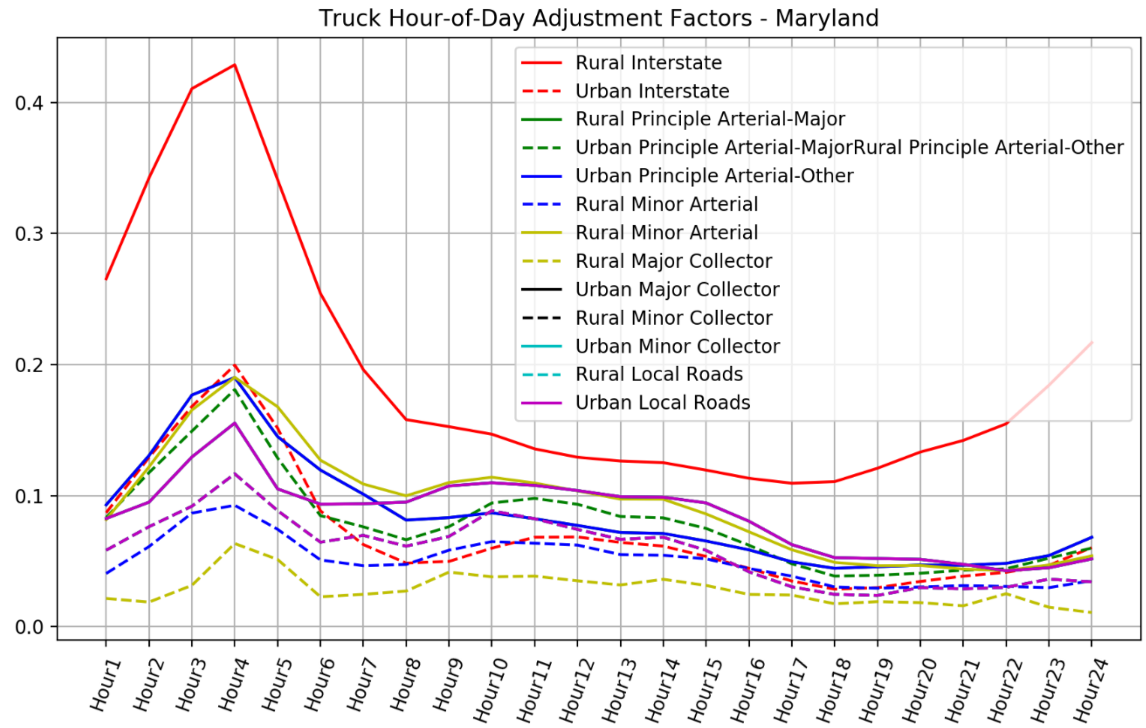


Figure 4 Truck hourly percentage for different function classes – Maryland.

Figures 5 and 6 present the day of week traffic trend in Maryland for rural and urban roadways. Based on the daily trend analysis, the results agree with common daily traffic demand trends. First, the highest volumes occur on Fridays in both rural and urban settings. This observation is due to the mixture of trips, including work, recreation, long-distance trip and short-distance trip, occurring on Friday. Moreover, the difference in the weekday and weekend volumes is more pronounced in urban settings. This observation may be explained by the higher average weekday volumes in urban areas related to commuting trips. Again, if we zoom into different roadway classes, findings can specifically help identify critical roadways on different days of the week. For instance, rural interstates serve significantly higher traffic on Sundays. Most returning legs of long-distance trips occur on Sundays, which could contribute to this observation. In addition, higher VMTs are registered on major urban arterials and local streets on weekdays.

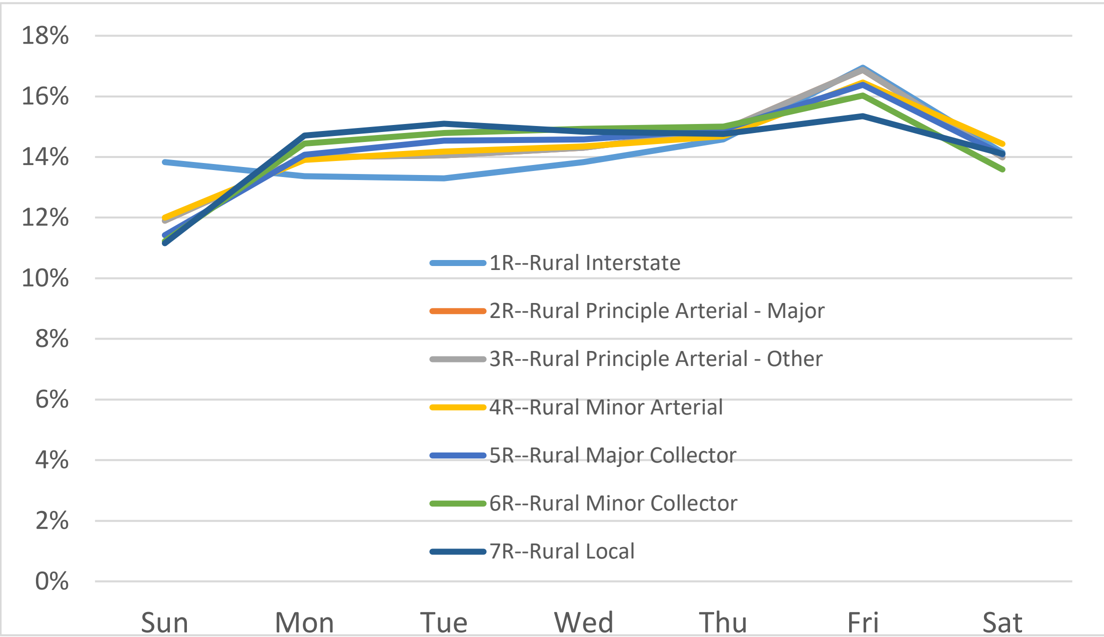


Figure 5 Rural day-of-week adjustment factor – Maryland

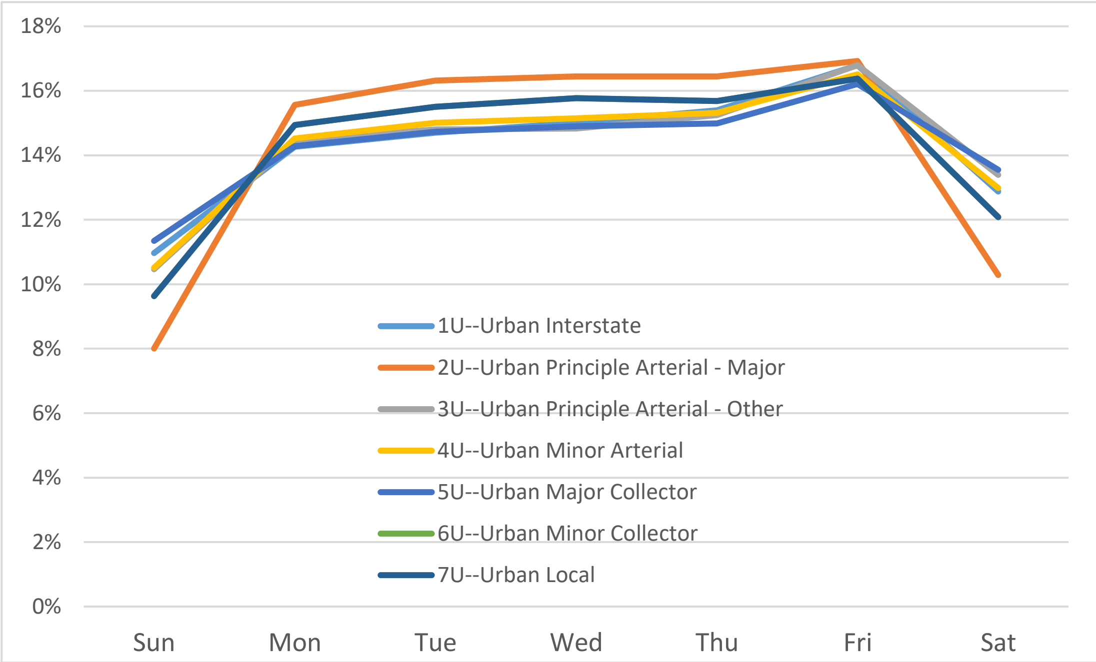


Figure 6 Urban day-of-week adjustment factors – Maryland

Figures 7 and 8 show the monthly traffic trend in Maryland for rural and urban roadways individually. The rural traffic volume has a significant increase from May to September. It is of great importance to point out that all the months mentioned display the same characteristic – the average temperature in these months are above the annual average temperature. These months also synchronize with the summer vacation of most schools. This observation will result in higher trip generation rate from May to September because people are more willing to participate in outdoor activities. Meanwhile, the monthly level traffic trend presented in urban areas is less fluctuated than rural area. Similar to the analysis above, this difference between urban area and rural area could be explained by their land use characteristics. Urban human activity and trip purpose are less likely to be influenced by the seasonal factors or weather condition. Therefore, the monthly level traffic volume in urban area is more uniformly distributed.

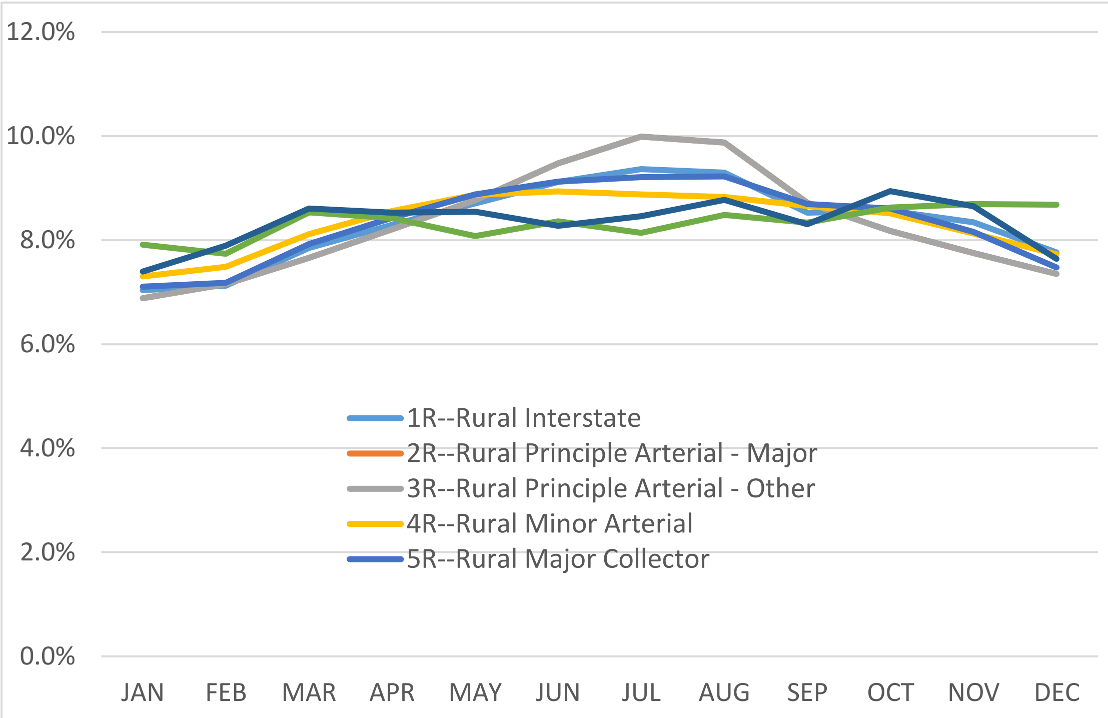


Figure 7 Rural month-of-year adjustment factor – Maryland

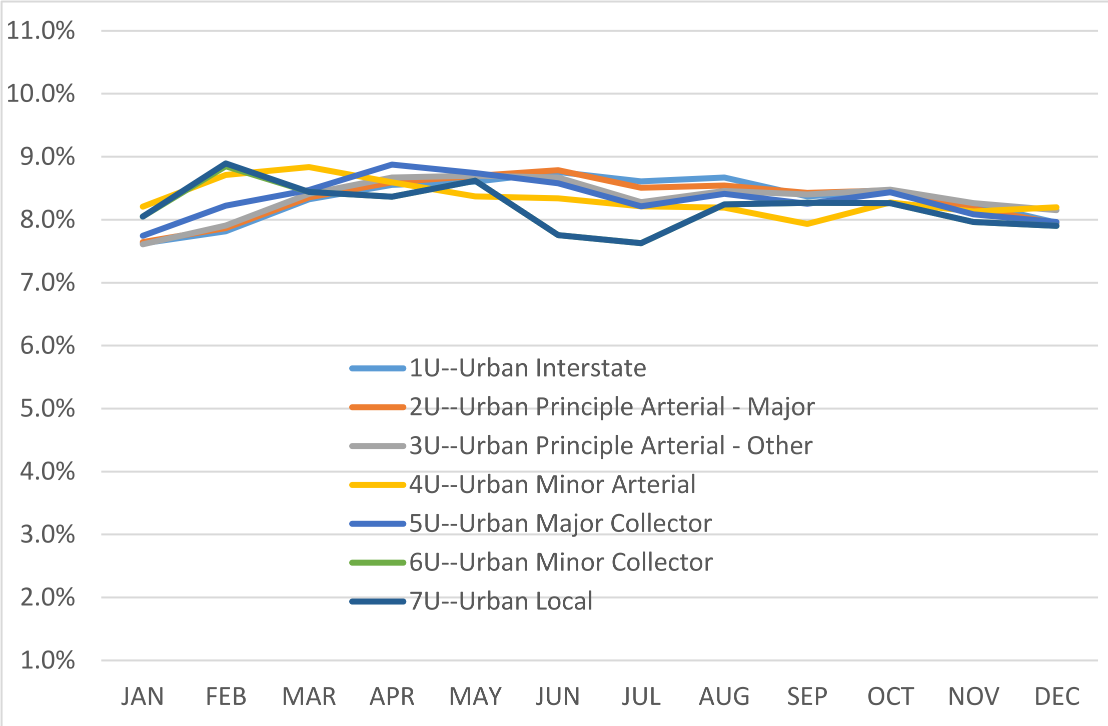


Figure 8 Urban month-of-year adjustment factor – Maryland

Functional classification is the process by which streets and highways are grouped into classes—or systems—according to several factors that contribute to the overall importance of a given roadway to a region or area. All streets and highways are grouped into one of seven classes, depending on the characteristic of the roadway and the degree of land access that they allow. Each group class has two different levels of roads, which means there are 14 types of functional classes in the FHWA monitoring system. However, the figure above shows that Washington, D.C. provides only three of them. This data gap and quality issue happens frequently. Similarity test will be implemented to address this problem. Based on $\chi 2$

statistical tests, functional class A will be compared with every state to obtain the most similar state pair.

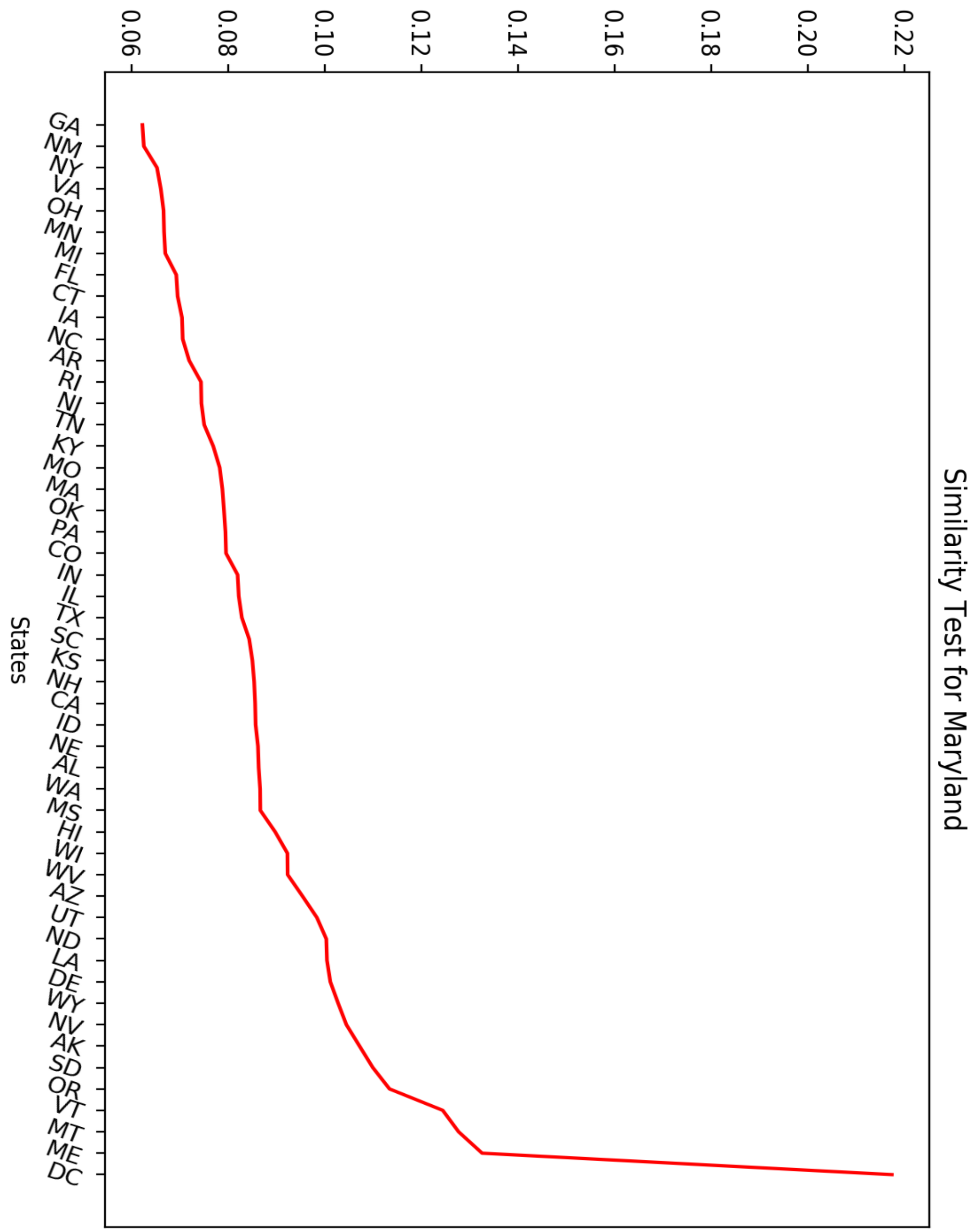


Figure 9 Similarity test of comparing Maryland with other states

Figure 9 shows the result of the similarity test. Maryland is selected as an example. Maryland is compared with the other 49 states and Washington, D.C., and the state of Georgia is most similar states compared to Maryland. The gap filling procedure is hence based on the result of the similarity test. Values developed for functional class in the state of Georgia will also be used for the same functional class in Maryland if the data for that functional class is missing.

After required preparation tasks have been done, the VMT can be estimated for any state at any time by different vehicle type by means of the following procedures.

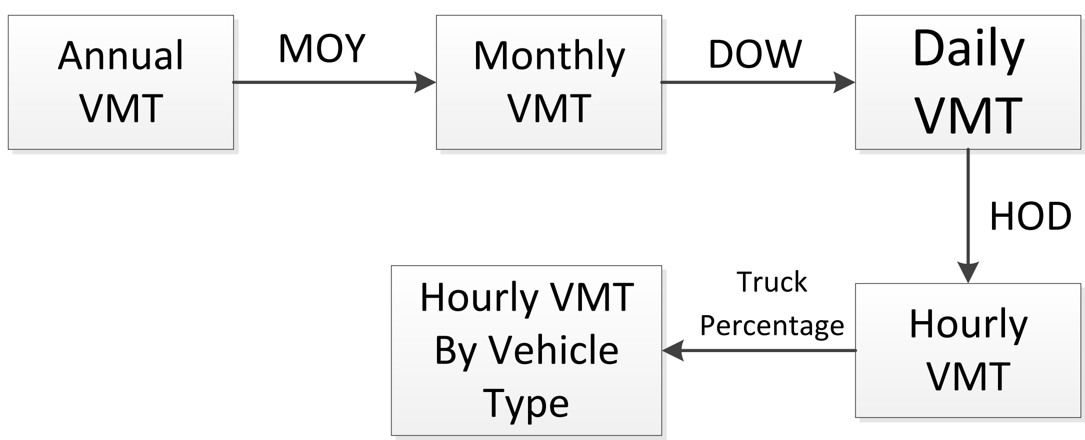


Figure 10 VMT estimation procedures

3.5 Demonstration and Validation of Final VMT product

Having shown the temporal and vehicle type adjustment factors and associated data visualizations, the VMT disaggregation can be completed by following the procedure summarized in Figure 10. Since ground truth data on validating the disaggregated vehicle miles traveled is not available, the proposed method is validated by number of trips estimation. Number of driving trips is represented by the quotient of vehicle miles traveled and average trip length.

Results	Number of Trips / Day	Number of Trips / Month
AirSage Data	24,145,753	696.67 million
Proposed Method	22,769,000	683.07 million

Table 1 Number of trips validation

In table 1, number of trips from proposed method is derived by dividing disaggregated vehicle miles traveled by average trip length from regional travel survey. It is further compared with AirSage data. AirSage data provides the weekday daily OD table for July in Maryland. It is noticeable that the results from proposed method are very close to what provided by AirSage even though the comparison shows discrepancy at some level. This fact can be explained by the fact that OD table in AirSage data involved all modes while the proposed method mainly targets at driving trips, which further enhance the credibility of vehicle miles traveled disaggregation method.

Beyond that, the TVT report, published by the Federal Highway Administration, is selected as another validation source to ensure the reliability of monthly trend derived by the proposed method.

Traffic Volume Trends is a monthly report based on hourly traffic count data reported by the States. These data are collected at approximately 5,000 continuous traffic counting locations nationwide and are used to estimate the percent change in traffic for the current month compared with the same month in the previous year. Monthly vehicle miles traveled by different states will be reported as well, which can be used to calculate the monthly trend and considered as a validation source. In this study, New York, Seattle, Washington D.C. metropolitan statistical areas are selected to demonstrate the validation results.

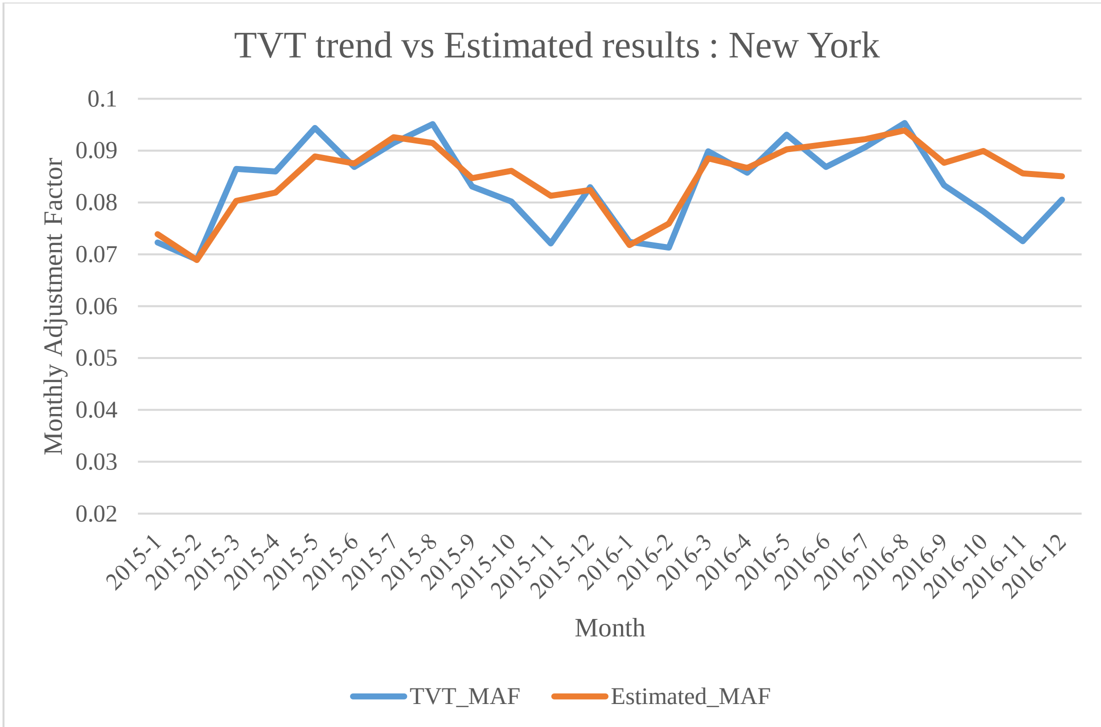


Figure 11 Compare estimated monthly adjustment factor (MAF) with TVT report in NY

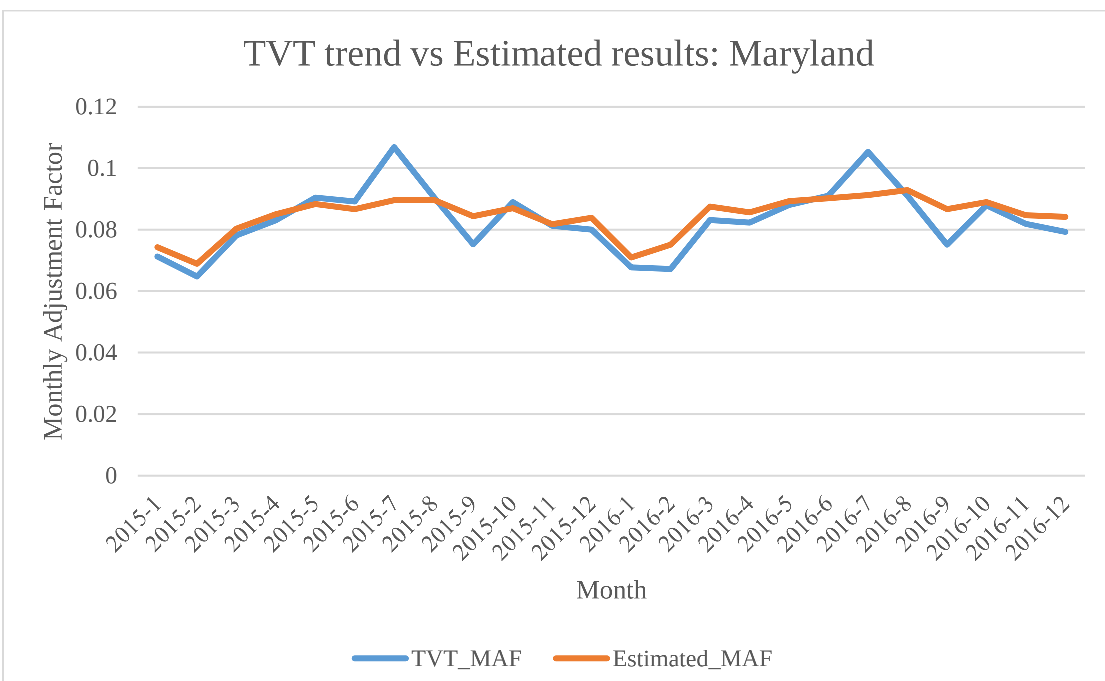


Figure 12 Compare estimated monthly adjustment factor (MAF) with TVT report in MD

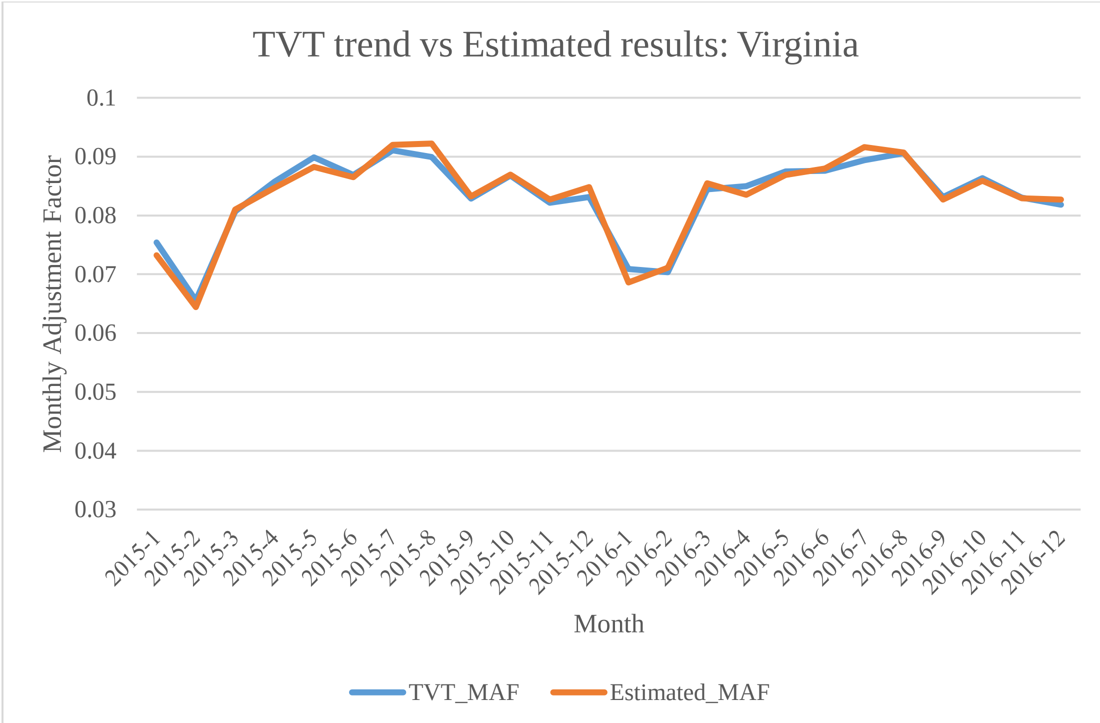


Figure 13 Compare estimated monthly adjustment factor (MAF) with TVT report in VA

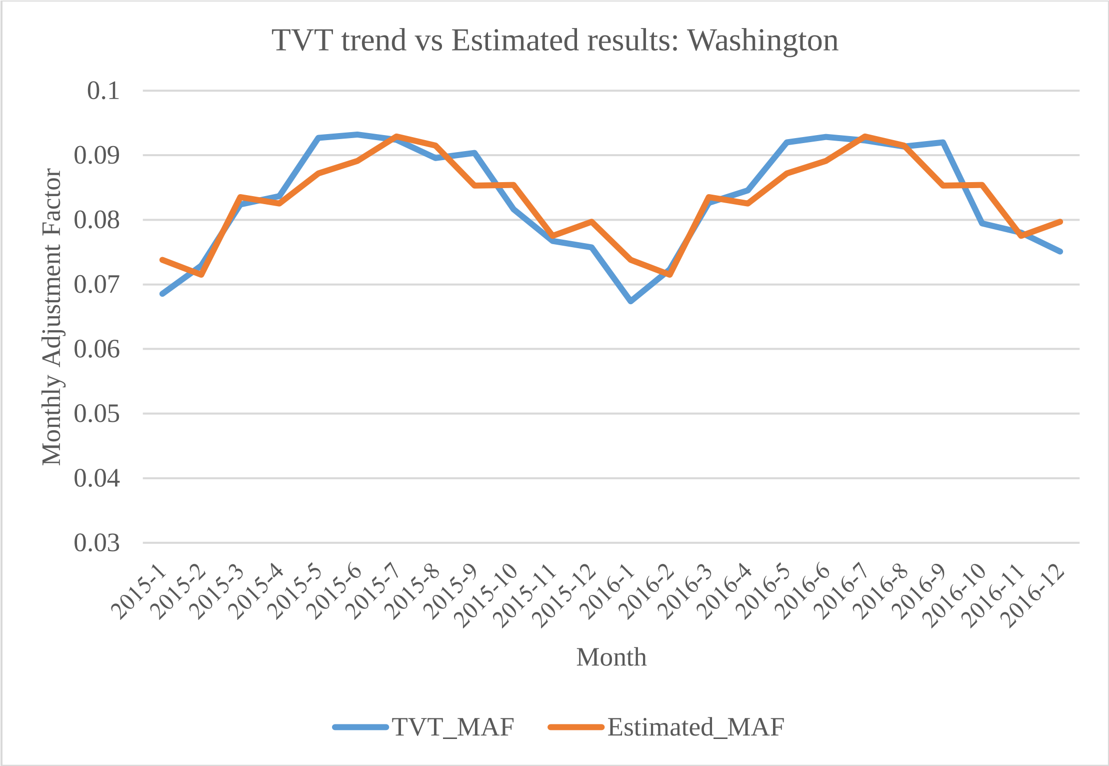


Figure 14 Compare estimated monthly adjustment factor (MAF) with TVT report in WA

Figure 11-14 present results of comparison between estimated results and TVT reports from 2015 to 2016 in NY, MD, VA and WA with orange line representing the monthly adjustment factors derived in this study and blue line showing the trend reported by FHWA. It is observable from the plots that results obtained using proposed method in this study show similar trend comparing with TVT report for most of the time, which prove the reliability and credibility of VMT disaggregation method proposed. However, what I can notice is the trend discrepancy in state of Maryland on July 2015 and 2016. Comparing with other study areas, the portion of vehicle miles traveled in Maryland at July significantly higher than other three states.

This could be possibly explained by the unexpected change on number of reported counting stations without knowing the weighting and processing algorithm on the raw data. In this case, some additional research works will be required to address this issue.

3.6 Summary

This section proposed a systematic method for VMT estimation by computing adjustment factors in different time intervals based on ATR raw data and HPMS data. The proposed method can be implemented on the analysis of traffic trends for a particular functional class and vehicle types for a relatively small-time interval. In other words, this section illustrates a proposed method that implements a statistical modeling approach to split total VMT estimates by vehicle type and time interval.

The proposed method is the first of its kind in estimating temporally high-resolution VMT statistics by comprehensively utilizing publicly available data sources from federal and state agencies. Moreover, the method is comprehensive and transformable in the way that different functional classes and vehicle types are identified and differentiated. The results can be extremely useful in understanding travel demand patterns and helping agencies' decision-making processes. Finally, the approach deals with the issue of missing data. This ensures the robustness of the proposed approach. As I demonstrated in the numerical example for the state of Maryland, the proposed approach produces reasonable and fine-grained VMT estimates for Maryland and is ready to be transferred to other states as well.

This can be applied to further research on the correlation between traffic volume and accident rate for all function classes. This section also successfully addresses the data gap and quality issue, which occurs by the coverage of a nation-wide permanent count station. It adopted a statistical method into the computing procedure aimed at establishing a linkage within a state pair, even though they are not similar in geographical condition. The credibility of this statistical method can be guaranteed, because the idea of 'similar state,' mentioned in this paper, is directly generated by the comparison of daily traffic patterns. Consequently, the outcomes of the similar pair are more representative than the traditional understanding of similar state, because this method focuses more on the similarity of traffic trends rather than the other indexes.

Chapter IV Theoretical Framework of Multimodal Travel Trend Analysis

4.1 Number of vehicular trips estimation

Figure 15 provides an overview of the process and steps to estimate the month-to-month number of trips for driving mode in a metropolitan area based on the HPMS and TMAS. The process also needs inputs from the latest Regional Household Travel Survey data. The TMAS data were mainly used to develop the monthly adjustment factor, while the HPMS data were adopted to calculate the annual vehicle miles traveled. Data pre-processing is needed to identify and fix the abnormal data, which will interrupt the program.

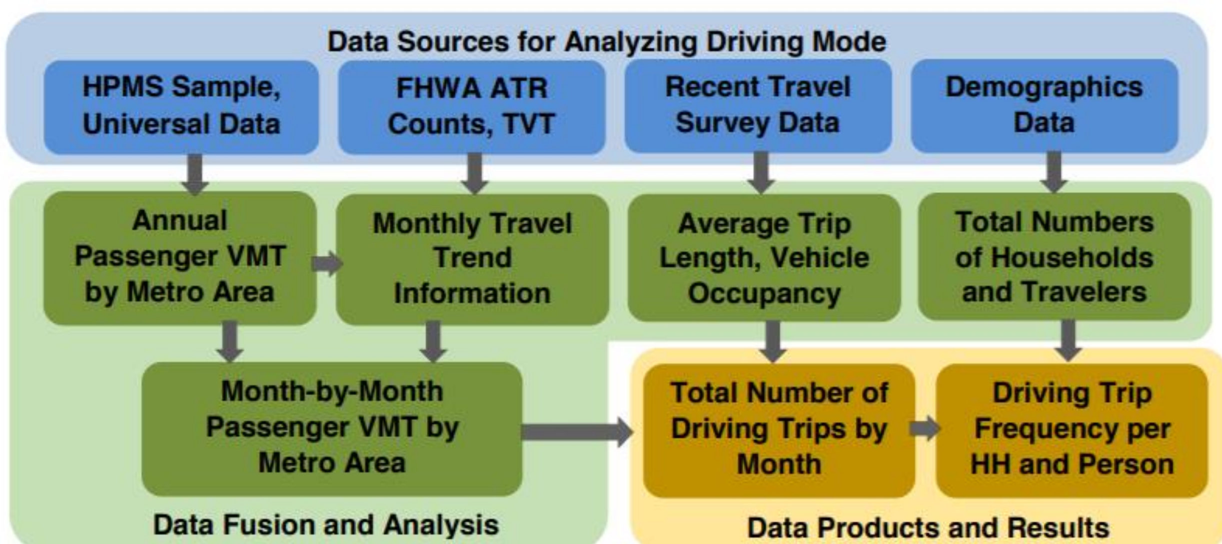


Figure 15 Computational flow chart of driving trips estimation

4.1.1 Adjustment factor

The VMT estimation algorithm has been discussed in chapter III, which is computed based on the ATR raw data and HPMS data. Other than to understand the vehicular travel trend at state level, this section mainly focuses on understanding multimodal travel trend at metropolitan level. The theoretical framework proposed in chapter III is still applicable in this section with some modification.

4.1.2 Number of Trips Estimation

The objective is to estimate the number of driving trips each month in a metropolitan area. It is critical to define the boundary of the metropolitan area before running the proposed algorithm. A comprehensive list of the counties located in each metropolitan area should be developed based on the latest MSA definition.

After the study area is defined, the VMT will be computed based on the states which are included in the study area. For example, the Washington D.C. metropolitan area includes counties from four states: Maryland, Virginia, Washington D.C. and West Virginia. VMT will be computed for these states using HPMS data. The VMT for each road section could be obtained by multiplying the section length, AADT, and number of days per year. The total VMT for a county can be calculated by aggregating VMT for all roads within a county. Since the study area has been defined, the VMT will be calculated by counties in the study area.

However, the HPMS does not provide good data for local road VMT because of reporting exemptions for roads of such functional classes. To address this data gap, it is assumed that the proportion of local road VMT of each county compare with whole state, should be consistent to the proportion of all road VMT compare with the state. For example, the percentage of local road VMT of 5 counties in Maryland compared with the local road VMT of State of Maryland will be equal to the percentage of its overall VMT within the State of Maryland. This assumption will enable us to extract the local road VMT from Highway Statistics Report and consequently improve the reliability and credibility of proposed method. After adding local road VMT, the VMT will be disaggregated into monthly level.

The HPMS program only provides AADT estimates for each road segment. In order to estimate the monthly trend of traffic volumes, which will later be used to estimate the number of vehicular trips, we need to allocate annual traffic volumes into each month. The distribution of monthly volume over a year may vary by geographic locations and by road functional classes. Therefore, this step is introduced to estimate the Month-Of-Year (MOY) factor for each road functional class. This step has been discussed in Section 4.1. Because we have only limited number of ATRs, only one set of factors is estimated for each State. Should more ATRs become available in the future, this method will be further developed by adopting a finer spatial resolution.

Once the annual VMT is obtained, the proposed method allows us to dynamically estimate the monthly vehicular trips by using monthly adjustment factor obtained from the previous step. Since the monthly adjustment represents the traffic volume percentage for a particular state, the monthly VMT can be computed by the product of monthly adjustment fact with annual VMT. In addition, this dissertation also assumes that the travel trend for a particular county is consistent with the travel trend in that state. Consequently, the annual level VMT can be disaggregated into monthly and county level. The annual VMT is disaggregated into target month VMT by using the monthly adjustment factors of target month.

The monthly VMT for any metropolitan statistics area can be estimated by following the above steps. However, planners and policy makers may be interested in the number of vehicular trips instead of VMT. The former is directly tied to the trend in mode shift over time. Therefore, this final step will be to compute the number of vehicular trips. Based on the average trip length and vehicular occupancy estimated from the most recent Household Travel Survey Data collected in the same metropolitan area.

4.2 Transit Mode

The following figure summarizes the methodology developed in this project for estimating the month-to-month transit ridership in a particular metropolitan area. It addresses three major challenges for estimating monthly transit ridership for a metropolitan area using NTD and other public domain data: 1) obtaining a comprehensive list of transit operators for each metropolitan area; 2) developing

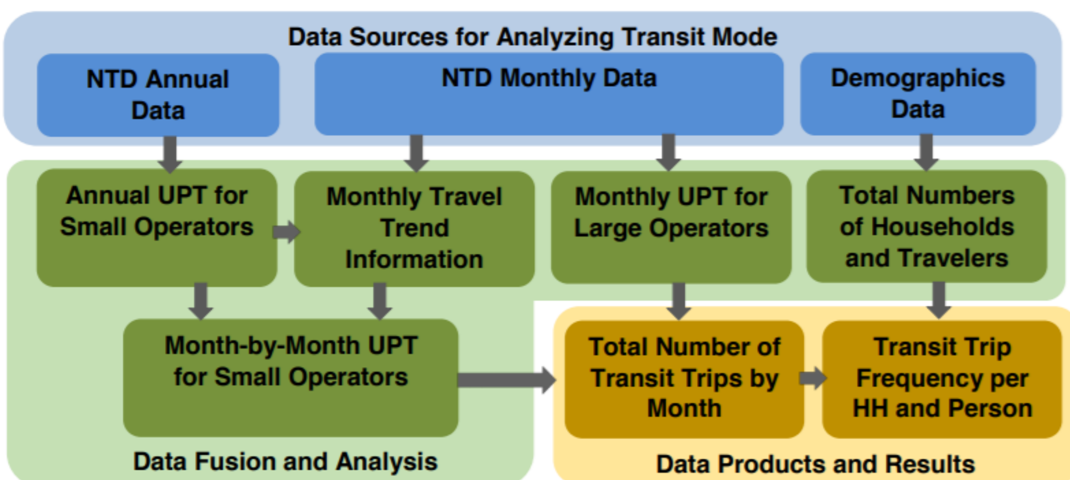


Figure 16 Flow chart for estimating the month-to-month transit ridership

methods to split ridership of large transit operators whose service network covers multiple metropolitan areas; 3) developing models to split annual ridership data reported by small operators into month-by-month trips.

4.2.1 Developing a List of Transit Operators

The first step is to develop a comprehensive and up-to-date list of transit operators that provide services in a metropolitan area. NTD assigns each transit operator to an urbanized area (UZA Name). However, the UZA Name is not a reliable identifier for identifying all transit operators within a metropolitan area. For example, although most transit operators in the Washington D.C. metropolitan areas are labeled as "Washington, DC-VA-MD" in NTD, this identifier does not cover operators such as Maryland Transit Administration (MTA), and Fredericksburg Regional Transit and etc., all of which have significant services operating within the Washington D.C. metropolitan areas. Some agencies such as MTA serve more than one metropolitan areas and are labeled only for one of them. Others such as Fredericksburg Regional Transit may originally not part of a metropolitan area and becomes so later because of the expansion of the metropolitan area. Therefore, it is important to develop a comprehensive and up-to-date list of transit operators for each metropolitan area.

To develop such a list, it is required to work with the Metropolitan Planning Organizations (MPOs) to make sure that all transit operators considered in their planning models are included in the list. This study also checked the website of all the counties and cities within the metropolitan area and make sure none of the operators supported by individual counties or cities (usually smaller ones) are missed.

This list can pull the month-by-month UPTs for full reports and annual UPTs for small systems reporters from the NTD. However, in order to calculate the month-by-month transit ridership for a metropolitan area, two problems have to be addressed: splitting the ridership for those operators whose network covers multiple network and estimating month-by-month ridership of small systems who only report annual data.

4.2.2 Geo-analysis for Splitting the Ridership of Cross-Border Service Providers

The first problem is related to the fact that some large operators may provide services in multiple metropolitan areas. If a trip either starts from or ends in the targeted metropolitan area, it should be included in the monthly total. However, if a trip neither starts from nor ends in the targeted metropolitan area, it should not be included. The NTD only reports the monthly total UPTs and does not provide a natural way to differentiate trips that fall in different metropolitan areas served by the same service provider. Therefore, additional efforts are needed to split the UPTs for operators that serve more than one metropolitan areas.

In addition, one agency may provide multiple services. NTD defined 13 modes: Motorbus (MB), Heavy Rail (HR), Light Rail (LR), Demand Response (DR), Commuter Bus (CB), Commuter Rail (CR), Light Rail (LR), Heavy Rail (HR), Street Car (SR), Ferry Boat (FB), Inclined Plane Vehicle (IP), trolleybus (TB); and Vanpool (VP). It also includes two types of services: Direct Operated (DO) and Purchased Transportation (PT). A large transit operator may provide multiple services: some of them cross the border of a metropolitan area, while others do not. For example, Maryland Transit Administration operates MARC Trains, commuter buses, and local buses. While MARC Trains and commuter buses serve both the Washington D.C. and the Baltimore metropolitan areas, the local buses only serve the Baltimore metropolitan area. Therefore, in order to accurately split ridership of cross-border service providers, a lot of efforts are required to understand the service networks of each type of services and their geographic locations in comparison with the boundary of each metropolitan areas.

General Transit Feed Specification (GTFS) Data Exchange (Czebotar 2016) website provides a data portal for transit networks across the world. It currently includes more than 13,000 files, but the updates were discontinued in 2016. In this study, all available transit networks in ArcGIS from GTFS Data Exchange and complement the dataset are collected with latest updates from open data efforts initiated by local governments such as Open Data DC (DC 2016). The transit networks were compared with the MSA boundary in ArcGIS. If the transit network falls completely within the Washington D.C. area, its ridership will be included in the monthly total for this area. However, if only part of the network falls within a metropolitan area, the trips that neither start from nor end in any transit stops locating in the targeted metropolitan area should not be included in the monthly total.

4.2.3 Analysis of Commute Rail Ridership

In the ideal case, if both the geolocations of transit stop and complete transit Origin-Destination (OD) demand matrices are available, we can easily decide which trips should be included in the total for a metropolitan area by comparing the most plausible route between each OD pair and the MSA boundary. However, in many cases, either one or both pieces of information are missing. Based on different level of information availability, new methods need to be developed using plausible assumptions.

It is unclear where passengers are going after boarding at one particular station. Without additional OD information, I assume that passengers boarding at one station are equally likely to go to any other stations along the same line. Following these assumptions, the proportion of a specific train ridership to be included in a metropolitan area should be:

$$p_i = \frac{N_i(N_i - 1) - J_i(J_i - 1)}{N_i(N_i - 1)}$$

Where p_i is the proportion of ridership of line i that should be included in the ridership total of the current metropolitan area; N_i is the total number of stations of line i; J_i is the number of stations of line i that falls within the metropolitan boundary.

4.2.4 Analysis of Commute Bus Ridership

Compared to the commute rail system, commute bus routes are more fluid. Therefore, most agencies did not keep complete records of bus stops in ArcGIS. Therefore, it is infeasible to determine the split of commute bus ridership between different metropolitan areas through OD analysis similar to that for the commute rail system.

Most commute bus riders use the system for commuting to the metropolitan area that particular line serves and are not likely to get off the bus right after boarding in the suburban area. Therefore, I assume all riders of a particular route would go to the metropolitan area that route is designed to serve. If by-line ridership information is not available, the percentage of transit ridership that belongs to the metropolitan area of interest can be approximated by:

$$p = \frac{K}{N}$$

Where K is the number of commute bus lines that serve the metropolitan area of interest, while N is the total number of commute bus routes operated by a transit operator. However, this method could be further improved if month-by-month commute bus ridership by route becomes available. The updated percentage number is:

$$p_m = \frac{\sum_k q_{km}}{\sum_n q_{nm}}$$

Where p_m is the proportion of commute bus ridership of transit administration that should be included in the ridership total of a metropolitan area in month m, and q_{km} is the monthly ridership of bus route k in month m.

4.2.5 Monthly Ridership of Small Operators

As mentioned in the introductory section, small operators only report the annual total ridership to NTD. Without additional information about the month-by-month ridership patterns of these small operators, it is reasonable to assume that the trend is consistent with large operators in the same geographic area.

4.3 For-hire Mode

4.3.1 Computation algorithm

As discussed in previous chapter, ridership data for some for-hire modes are not available in certain metropolitan area. This sub-section will discuss methods developed in this study to fill those data gaps. Figure 17 provides an overview of the process to estimate the month-to-month number of trips for for-hire modes in a metropolitan area based on available data set.

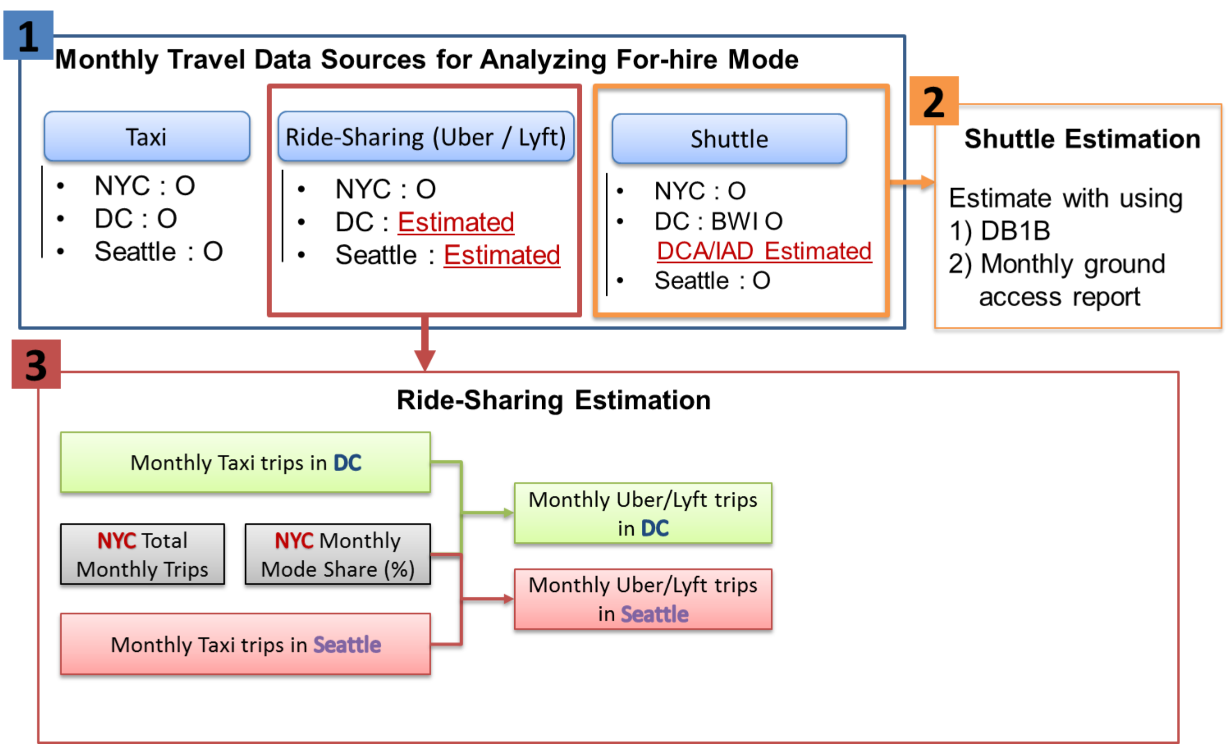


Figure 17 Monthly for-fire trip estimation

4.3.2 Estimating ride-hailing trips

NYC represents a rare case where monthly ridership of ride-hailing services such as Uber and Lyft become part of public domain information and was made available through FOIA. The same data is not available for other metropolitan area. However, data in NYC would allow us to estimate the monthly trips of ride-hailing modes in DC and Seattle if it is assumed the same trend in market share between ride-hailing services and taxi industry exists in all three metropolitan areas. The ride-hailing estimation procedure is under the assumptions that 1) market shares (%) of the ride-hailing modes are consistent for metropolitan areas within the same month; and 2) the service launching times of Uber/Lyft in each city do not affect the market share; The NYC market shares are listed in Table 2 and Figure 18 displays the column chart of NYC market shares, which are calculated from the equation below.

$$Market\ share_m = \frac{T_{i.m}}{\sum_i T_{i,m}}$$

Where $T_{i,m}$ $is\ the\ number\ of\ trips\ for\ mode\ i\ in\ month\ of\ m$

$i\ is\ for\ hire\ mode\ (taxi, Uber, Lyft)$

$m = month\ of\ year\ 2015\ (1, 2, \ldots, 12)$

Taxis in NYC include Yellow and Green taxis. It can be noted that the market shares of taxis are constantly decreasing from 90.4 to 77.3 percent as time goes by whereas Uber and Lyft shares are increasing.

Month	**Taxis**	**Uber**	**Lyft**	**Total**
January	90.4	9.6	-	100.0
February	88.9	11.1	-	100.0
March	89.1	10.9	-	100.0
April	88.6	10.9	0.4	100.0
May	86.8	12.7	0.5	100.0
June	85.3	14.1	0.6	100.0
July	82.2	16.7	1.1	100.0
August	80.8	17.4	1.8	100.0
September	79.8	18.1	2.1	100.0
October	79.8	18.3	1.9	100.0
November	79.1	18.7	2.2	100.0
December	77.3	19.8	2.8	100.0

Table 2 2015 monthly market shares of for-hire modes in NYC

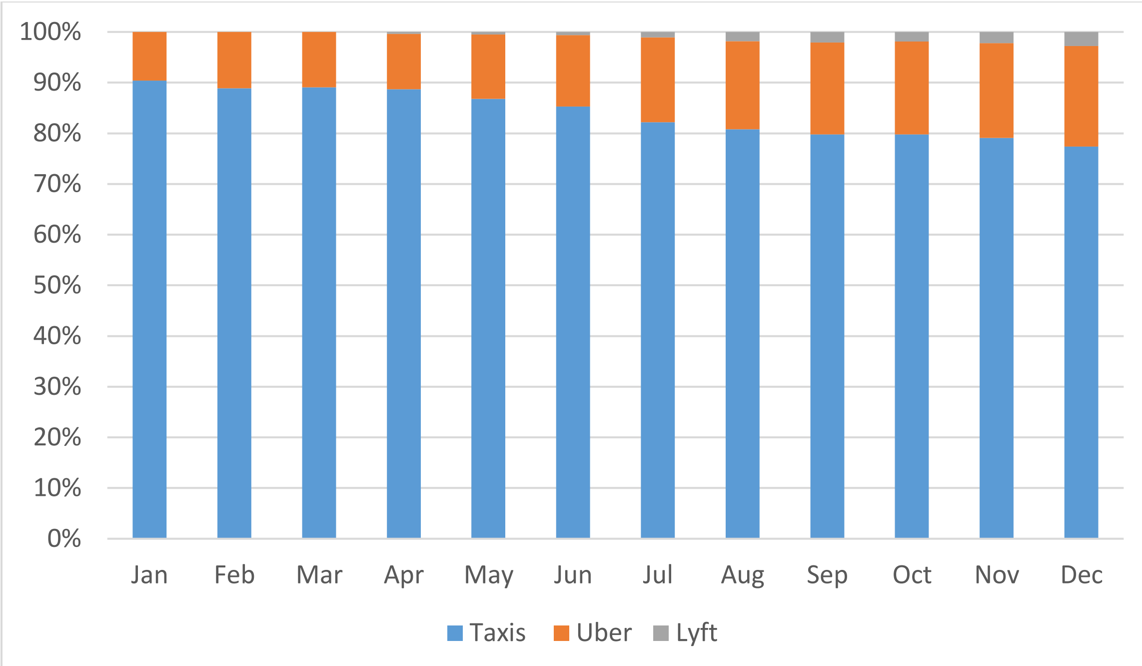


Figure 18 Column chart of NYC market share in 2015

With the NYC market shares calculated, the number of dynamic ride-hailing trips for each month can be obtained by assuming that the for-hire market shares are consistent across the three cities. Since taxi trips are known most metropolitan areas, Uber and Lyft trips can be obtained by following the NYC markets share portions.

4.3.3 Estimating shuttle trips

DB1B data from USDOT is collected for estimating shuttle trips at metropolitan level.

There are three types of data from DB1B as shown in Figure 19: 1) coupon, which describes each flight trips including layover OD; 2) market, which indicates one-way flight trips and only counts final OD; and 3) ticket, which contains round trip information.

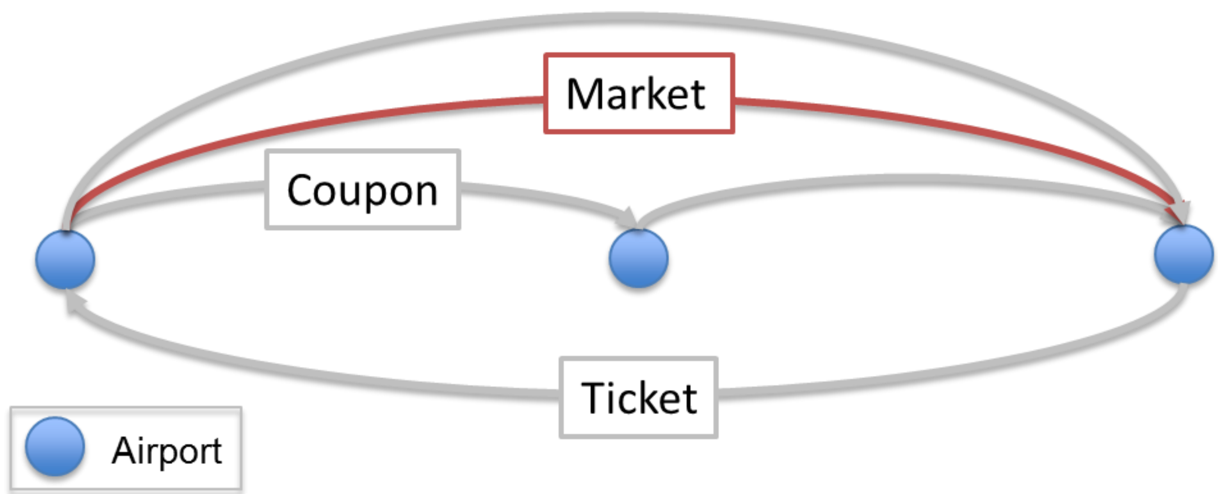


Figure 19 Description on three different types of DB1B data

Since we are interested to know the total number of passengers who are leaving from and arriving at an airport, the 'market' type data are adopted. The assumptions for estimating shuttle trips are 1) mode shares for the airport shuttles are consistent among within a metropolitan area; and 2) every month within each quarter contains the same number of passengers.

The estimation process for shuttle trips uses the following steps. First, collect the original DB1B data for major airports in a metropolitan area. Second, the estimated total number of passengers per quarter are populated by multiplying 10 for each quarter since the DB1B samples 10% of all air passengers. Finally, the total number of passengers per month are estimated from the quarterly total passengers dividing by number of months.

4.4 Non-motorized Mode

4.4.1 Overall Framework

To address the data challenges for the non-motorized mode, a method consisting of two modules is proposed in Figure 16. Unlike other travel modes, there is no source reporting the total number of non-motorized trips whatever the geographic area is. Therefore, Module 1 leverages the 1-year/5-year estimates from American Community Survey (ACS) and the regional household travel survey to estimate the daily non-motorized trip total. ACS produces many tables reflecting different aspects of the population, including the means of transportation to work. The geographies can be defined by users as small as a county. In the data products, the number of people commuting by biking or walking (biking commuter or walking commuter) is estimated for an average weekday.

To further estimate the number of biking/walking trips, the specialized trip rate for biking/walking commuters is derived from the raw data of the regional household travel survey (RHTS) in Equation (1). After that, the specialized trip rate from RHTS performs as an input to compute the average weekday biking/walking trips, together with the number of biking/walking commuters from ACS in Equation (2).

$$STR_{Biking/Walking} = \sum\nolimits_{\forall t \in BT/WT} \omega_{trip}^{RHTS} \Big/ \sum\nolimits_{\forall c \in BC/WC} \omega_{person}^{RHTS} \qquad (1)$$

$$STR_{Biking/Walking} = N_{Biking/WalkingTrips}^{RHTS} \Big/ N_{Biking/WalkingCommuters}^{RHTS} \qquad (1)$$

$$AWT_{\widehat{Biking/Walking}} = \sum\nolimits_{i} N_{Biking/WalkingCommuters}^{ACS} \times STR_{Biking/Walking}$$

(2)

Where $N_{Biking/WalkingTrips}^{RHTS}$ is the weighted number of biking/walking trips reported to RHTS, which can be linked or unlinked trips; $N_{Biking/WalkingCommuters}^{RHTS}$ is the weighted number of biking/walking commuters, who are identified from RHTS by examining whether they bicycle or walk to work; $STR_{Biking/Walking}$ is the specialized trip rate for one single mode, either biking or walking;

$N^{ACS}_{Biking/WalkingCommuters}$ is the daily number of biking/walking commuters estimated by ACS; $\widehat{AWT_{Biking/Walking}}$ is the average weekday biking/walking trip estimate.

In household travel survey, linked trip file records one complete trip from the origin to the destination while unlinked trip file specifies every segment in one trip such as transferring the travel mode. The weights given by RHTS for each respondent or even each trip is considered as well when calculating the number of biking/walking trips and commuters.

In the meantime, Module 2 establishes a Poisson Multilevel Model (PMM) to address the sparsity and scarcity of bicycle/pedestrian count data. With the weekend effect factor, one of the model outputs, the average weekday biking/walking trip estimate can be expanded to annual total estimate along with the number of weekdays and weekends in that year. The detailed model specification of PMM is described in Section 4.8.2.

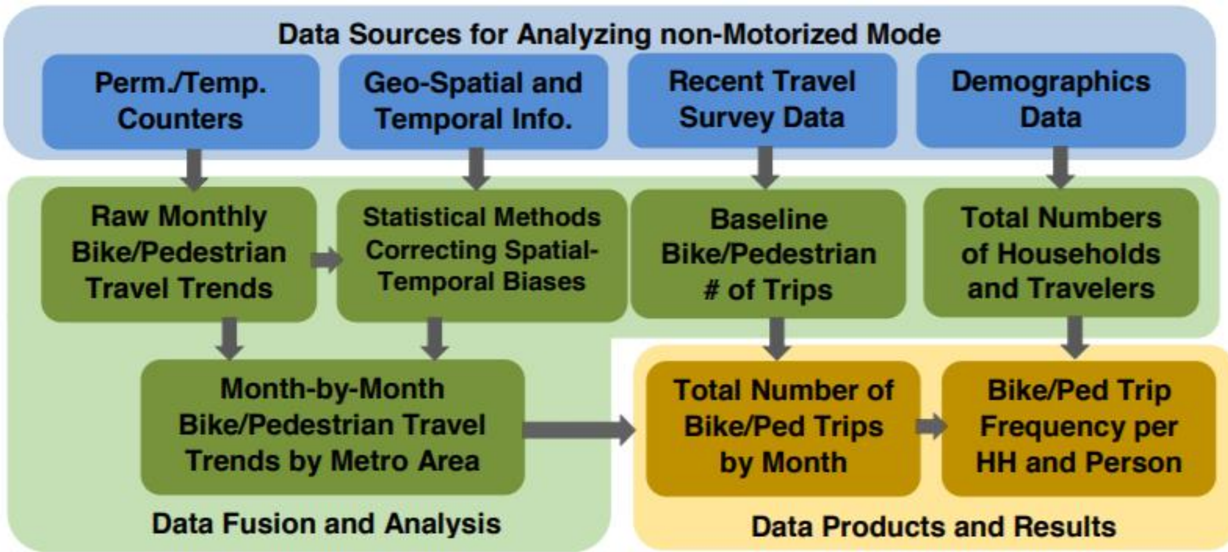


Figure 20 Framework for monthly non-motorized trip estimation

4.4.2 Poisson Multilevel Regression Model

To accommodate the sparsity of the data source and to address the distribution of the count data, a Poisson Multilevel Model (PMM) is proposed based on the characteristics of the typical non-motorized counts. The universal model specification is given as two parts: the first part is the Poisson Regression Model and the second part is the Multilevel Model with random intercepts.

$$P(y|x) = \frac{\lambda^y e^{-\lambda}}{y!} \tag{3}$$

$$\log\big(E(Y|x)\big) = \log(\lambda) = \theta x \tag{4}$$

Where y is the hourly bicycle volume, x is a set of the independent variables, and λ is the parameter of the predicted Poisson distribution, which is greater than zero and equal to the mean and variance of the random variable. θ represents the coefficients to be estimated. The multilevel model is specified as following.

Level 1: $\log(\lambda_{ij}) = \beta_{0j} + \beta_1 M_{ij} + \beta_2 Y_{ij} + \beta_3 WD_{ij} + \beta_4 W_{ij} + \cdots$ (5)

Level 2: $\beta_{0j} = \beta_0 + u_{0j}$ (6)

Where i is the index of each data entry in the dataset and j is the identification of each count station;

M_{ij} is a vector of indicators for months, where the indicator for the first available month is the reference category;

Y_{ij} is a vector of indicators for different years, where the indicator for the first available year is the reference category;

WD_{ij} is a binary variable indicating if the counting date is a weekend;

W_{ij} is a categorical variable for weather description such as "Sunny or clear", "Windy or forecast", and "Rainy or snowy", where the first weather category is the reference category;

β_{0j} is the random intercept composed of the fixed intercept β_0 and the random error term u_{0j}, whose distribution is assumed to be normal;

$\beta_1, \beta_2,$ ... are the fixed effects of the attributes to be estimated and in this case, they are row vectors, where each element is a fixed slope.

To extract the monthly trend from various types of count data, the necessary independent variable to include is the month indicator vector, M_{ij}. Other attributes like year indicators and weather indicators can be also considered based on their availability in the count dataset. For a reliable model estimation, it is desirable to have multiple records from different months in one count location.

Chapter V Case study in D.C., Seattle and New York

The proposed method described in Chapter IV has been applied to estimate the month-to-month travel demand of driving, transit, for-hire, and non-motorized modes in New York City, Seattle, and Washington D.C. metropolitan areas. With decision makers increasingly requesting recent and up-to-date information on travel trends, establishing a sustainable and timely travel monitoring program based on available data sources from the public domain is in order. The literature review showed that data availability is not merely an obstacle for estimating driving trips, but also a common challenge for all modes to develop a longitudinal travel behavior monitoring method for a metropolitan area. Based on findings from the literature review, a few promising public domain data sources have been identified to support the research method to be developed in this dissertation. This section will introduce the data set involved in this case study and discuss the results for multimodal travel trend analysis as well.

5.1 Public Domain Data

5.1.1 Driving Mode

Highway Performance Monitoring System (HPMS) data is the official Federal government source of data, especially on the condition, performance and characteristics of nation's highways. HPMS data provides specific and comprehensive data items about all highways in the U.S., including length, lane-miles, AADT and more. The HPMS function class classification system is regarded as the reference of the available function class for 50 states and Washington, D.C. TMAS stands for Travel Monitoring Analysis System. Certain data items, including station description, truck volume and other related factors, are required for the data. TMAS data is submitted by State traffic offices and is the only public domain database that could provide volumes of truck traffic for highways in the U.S. TMAS data will be used to generate truck daily trend in this study.

For the reporting of HPMS, State highway agencies have deployed a limited number of permanent automatic traffic recorders (ATR) for continuous detection and more spread short duration detectors for 48-hour traffic counts. Two publicly available datasets generated based on traffic count data collected from those detectors will be used in this study: Traffic Volume Trend (TVT) and Highway Statistics (HS).

Generally, TVT provides monthly VMT by highway functional groups for each of the states, and HS summarizes annual VMT by highway functional class at the state level.

To translate the VMT estimates derived from the HPMS data into the number of vehicular trips, the average vehicle occupancy and trip length data for a metropolitan area is needed. Based on a comprehensive literature review, Household Travel Survey is the only reliable data source for such information. In this study, the vehicle occupancy and average trip length will be extracted from the most recent updated Household Travel Survey for each metropolitan statistics area. The average trip length will enable us to transfer the monthly total vehicle miles traveled into the number of trips while the average occupancy will be used in the computation of number vehicular trips per person. Therefore, these two factors could be various from time to time. The Household Travel Survey presented in the case study will be regularly updated to ensure the accuracy and representativeness.

5.1.2 Transit Modes

The National Transit Database provides a unique and centralized data hub for most transit operators in the US. All US transit agencies who receive funding from the Urbanized Area Formula Program (5307) or Rural Formula Program (5311) are required to report a wide range of performance data to NTD. The NTD performance data will be used to apportion over $5 billion of FTA funds annually to transit agencies in urbanized areas (UTZs). Currently, about 850 transit operators in UTZs are reporting to the NTD through the Internet-based system. Two major transit ridership data, the Unlinked Passenger Trips (UPTs) and Passenger Miles of Travel

(PMTs) are reported annually, and the time series dates back to 1991. Since 2002, large transit operators are also required to report up-to-date time series of monthly UPTs. One huge advantage of a centralized NTD is the standardization of data format, collection methods and quality, which greatly facilitate data analysis and applications. NTD divides transit into 8 rail modes and 12 non-rail modes. It further differentiates directly operated services and purchased transportation services. All estimates must meet 95% confidence and 10% precision levels.

NTD requires 100% counts of UPT and PMT if such data is available and reliable (USDOT 2013). For example, 100% counts of UPTs and PMTs can be derived from the smartcard system where both swipe-in and swipe-out are required. When such data is not available, sampling method can be used. National Transit Database Sampling Manual (USDOT 2009) provided detailed description of a ready-to-use sampling method and procedure. APCs can also be used for transit ridership collection. However, the use of APCs for NTD reporting requires prior FTA approval and a validation of the APC data for UPT and PM data against a separate data sample covering a full year to ensure the data quality. Because this universal reporting requirement, and the standardized data format and quality, NTD becomes a great asset for transit ridership monitoring and analysis. In this study, NTD will be the primary source for month to month transit ridership analysis for a metropolitan area.

5.1.3 For-hire Modes

For-hire modes include taxis, ride-hailing (Uber/Lyft), and shuttles. Unlike the driving mode and transit modes, there is no established national database for for-hire modes. Moreover, some for-hire modes, such as the rider-hailing, are operated by private companies and the ridership data for those modes becomes proprietary in most cases. This study has explored the data availability in the three metropolitan areas selected for demonstration (New York City (NYC), Washington D.C. (DC), and Seattle) case by case, while recommendations on standardizing such data collection efforts will be made based on findings from this study.

There are 7 airports in the three metropolitan areas to be investigated in this study:

- New York City (NYC): John F. Kennedy International Airport (JFK), Newark Liberty International Airport (EWR), and LaGuardia Airport (LGA)

- Washington D.C.: Washington Dulles International Airport (IAD), Ronald Reagan Washington National Airport (DCA) and Baltimore Washington International Airport (BWI)
- Seattle: Seattle Tacoma International Airport (SEA)

This dissertation has also explored various ways to acquire the airport shuttle ridership data for all 7 airports. The data collection efforts for each metropolitan area is summarized below.

For NYC, yellow taxi, green taxi, and ride-hailing (Uber and Lyft) trip data are available at the New York Taxi & Limousine Commission (NYTLC) website. Airport ground access trip data are collected from the port authority of NY & NJ website. Among the collected airport ground access modes, ground transportation counter booking, and airport coach are considered as airport shuttles. The entire for-hire mode data are available for the year of 2015 except Lyft. Lyft data from January to March in 2015 are missing.

For Seattle, taxi trips of 5 operators (Far West, Green, Orange, STITA, and Yellow) from January to December 2015 are obtained. Port of Seattle provided the airport ground trip data, but ride-hailing data are not available. Among the available airport ground access modes, pre-arranged limousine, on demand limousine, courtesy vehicle, crew van, scheduled airporter, and downtown airporter are considered as shuttle modes.

For DC, taxi data are obtained via FOIA request at the District of Columbia Government's Freedom of Information Act Public Access website. One year of taxi data are available from April 2015 to March 2016. Ride-hailing data are not available from public domain. Baltimore-Washington (BWI) airport passenger ground access data that include shuttles, limousines, and taxis are obtained requesting at the Maryland Department of Transportation Maryland Aviation Administration. On the other hand, the information for monthly shuttle trips in two D.C. airports, Ronald Reagan Washington (DCA) and Washington Dulles (IAD), is currently not available.

The following table summarizes the data availability in each metropolitan area and the terms in parenthesis indicate source of data. To estimate the missing airport shuttle ridership for DCA and IAD (shown in Table 3), Airline Origin and Destination Survey (DB1B) data from United States Department of Transportation (USDOT) are collected and additional data processing method will be developed in Chapter III to estimate the missing ridership.

	Washington D.C.	New York City	Seattle
Taxi	(DC-Government FOIA)	(NYC TLC)	Valid
Airport Shuttle	BWI: (MDDOT)	JFK: (NY Port Authority)	
	DCA: Estimated*	LGA: (NY Port Authority)	SEA: (Port of Seattle)
	IAD: Estimated*	EWR: (NY Port Authority)	
Uber/Lyft	Estimated**	(NYC TLC)	Estimated**

Table 3 For-hire mode data status

** Airport shuttle trips for DCA and IAD are estimated by DB1B data and the airport shuttle trips from BWI*
*** Ride-hailing modes (Uber/Lyft) in Washington D.C. and Seattle are estimated by applying the market shares from New York City.*

5.1.4 Non-motorized Modes

Compared to the other three modes that have been discussed in this report, even few data were collected for non-motorized modes. Moreover, the data standard and collection methods vary even within the same metropolitan areas. Because of the data sparsity, it was hard to rely on count data to estimate the total number of trips for non-motorized modes. The count data must be integrated with regional travel survey data to provide reliable estimates. This section will review the existing efforts on collecting non-motorized traffic counts in the three metropolitan areas.

All the datasets considered in the report are accessible to public. American Community Survey (ACS) provides statistical data nationwide. For Washington metropolitan area, the most recent household travel survey is the 2007/2008 Transportation Planning Board (TPB) Household Travel Survey. For New York metropolitan area, the most recent one is the 2010/2011 Regional Household Travel Survey conducted by New York Metropolitan Transportation Council. For Seattle metropolitan area, the most recent one is the 2015 Seattle Household Travel Survey. There are also different bicycle and pedestrian count programs in the three case study areas while they have produced similar data products. The trip data from the bikeshare programs (Capital Bikeshare, Citi Bike, and Pronto Bikeshare) are incorporated as the validation for monthly trend.

5.1.4.1 Public Count Data in Washington Metropolitan Area

The count dataset from the District of Columbia Department of Transportation (DDOT) contains data entries from 2001 to 2015 at 71 count stations. Each observation records the date, the time duration, the facility characteristics, and the number of cyclists. Some of them also includes the number of Capital Bikeshare users and the weather information. Since the observations before 2004 is scarce and may provide bias information for the results, they are excluded before modeling. There are only very few data entries in each year comparing to the number of count stations. Figure 21 shows the number of data records in each year from 2004 to 2015, while Figure 22 shows the number of data records in each Monday when all those records are aggregated. According to the records, most data were collected in summer months, while no data was collected in December.

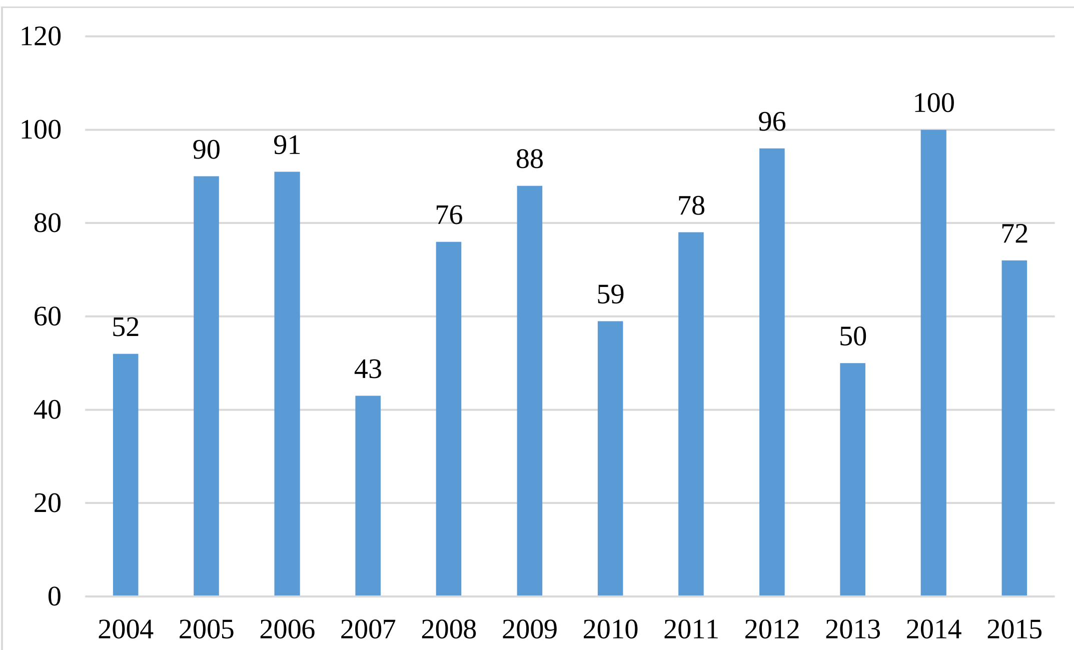


Figure 21 D.C. MSA: numbers of records in each year

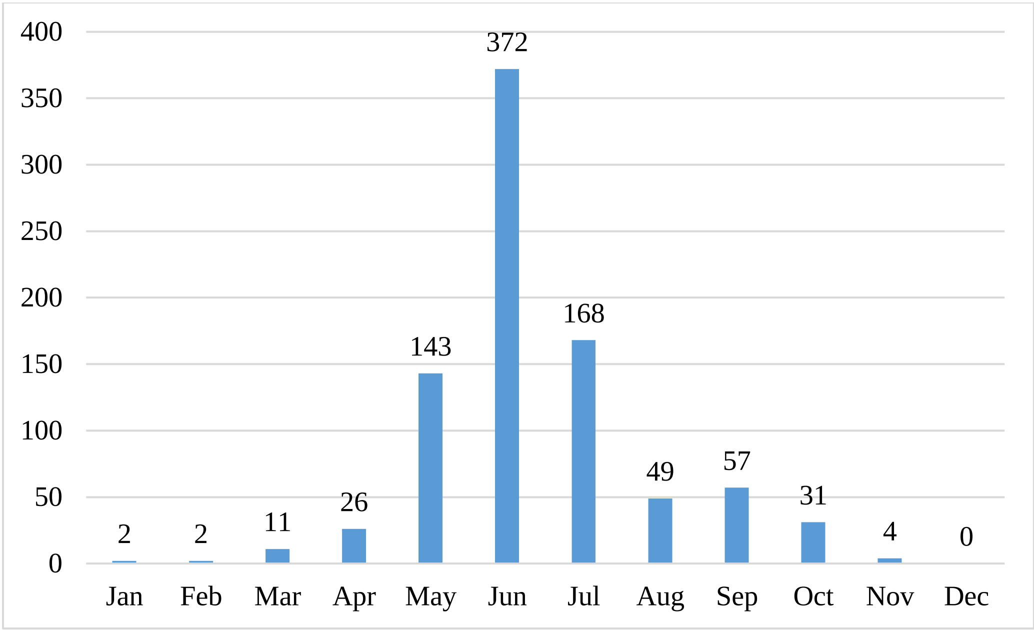


Figure 22 D.C. MSA: numbers of records in each month

5.1.4.2 Public Count Data in New York Metropolitan Area

New York City Department of Transportation (NYCDOT) conducts ten 12-hour weekday bicycle counts within one year at 22 selected locations from 2008 to 2015. Although the numbers of records in each year do not fluctuate too much, the numbers of records in each month are very unbalanced. No data were collected in either March or November. Figure 23 and 24 show the frequency of the observations over different years and in each month, respectively.

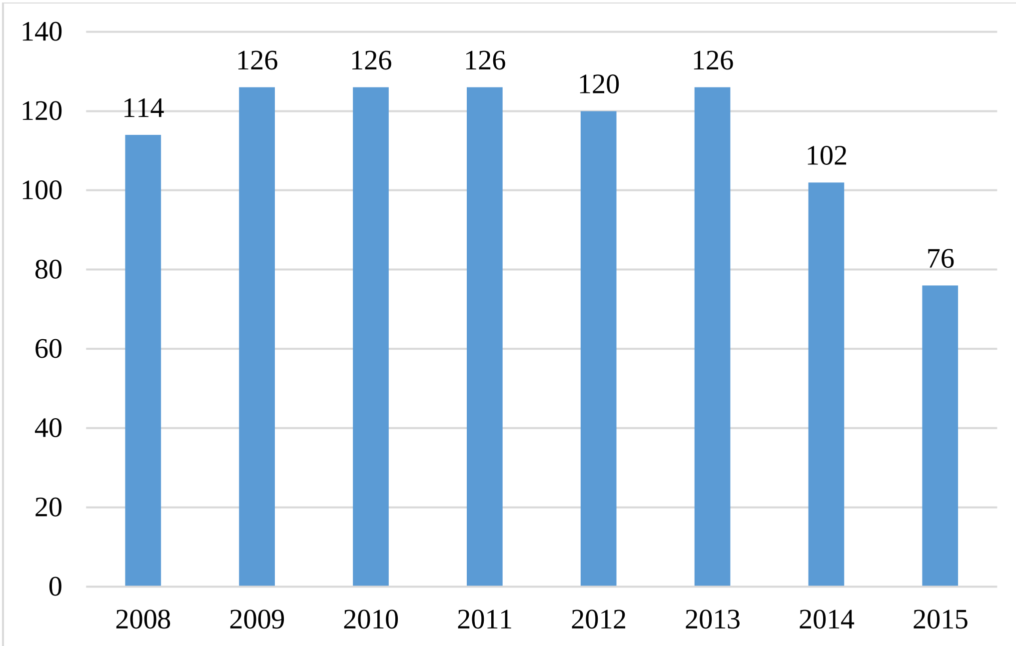


Figure 23 NYC MSA: numbers of records in each year

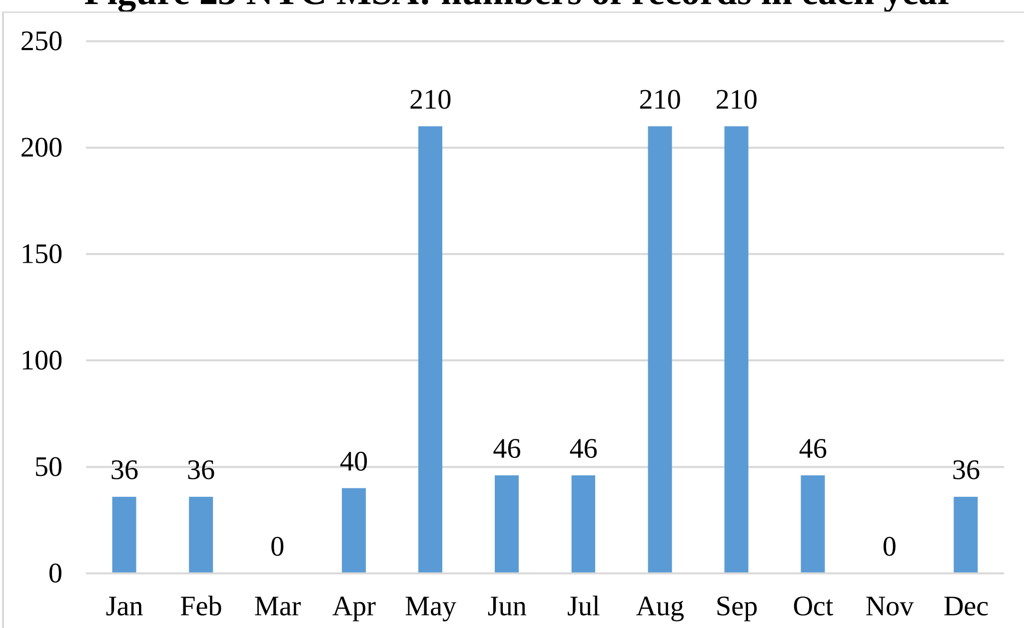


Figure 24 NYC MSA: numbers of records in each month

5.1.4.3 Public Count Data in Seattle Metropolitan Area

Seattle Department of Transportation (SDOT) installed the automatic bike counters in 10 locations across Seattle, four of which also provide pedestrian counts. Data from the continuous count stations was aggregated hourly. The occasional counter failures remain as a problem and additional data cleaning effort was required before using those data for non-motorized travel demand estimation. As shown in Figure 25 and 26, the continuous counts are recorded from 2012 to 2016 and the numbers of records are stable in each month. This stability and continuation in data shows great advantage of deploying automatic count stations for non-motorized modes compared to manual data collection efforts.

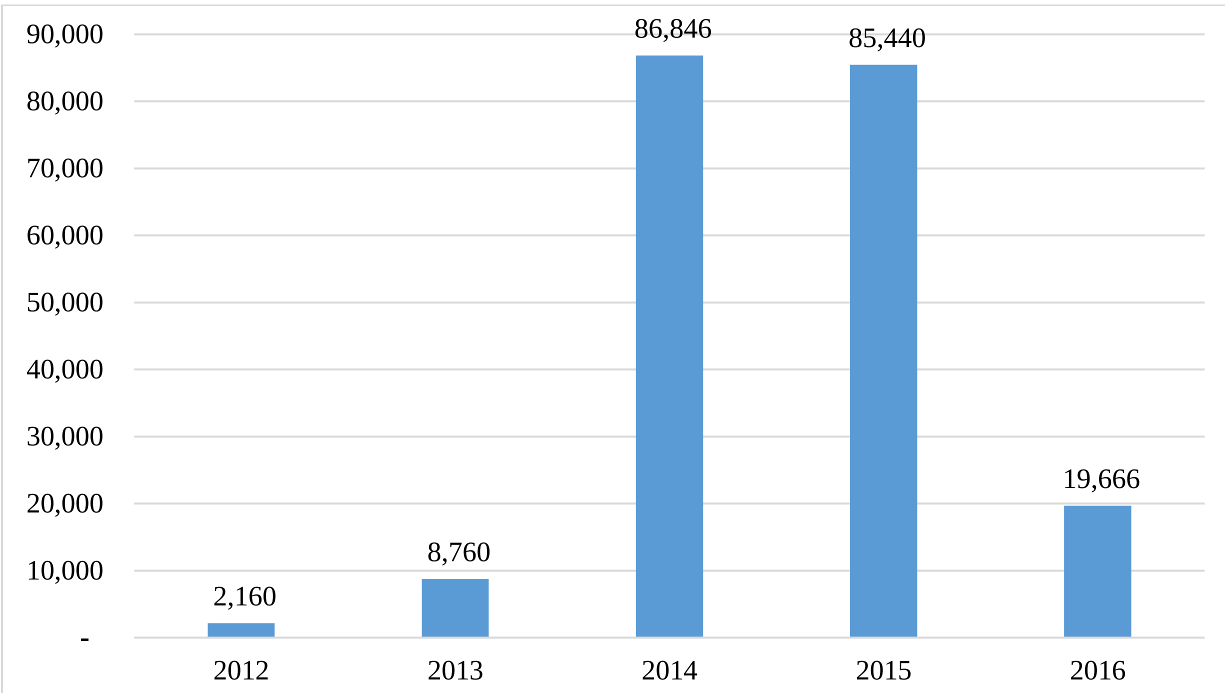


Figure 25 Seattle MSA: numbers of records in each year

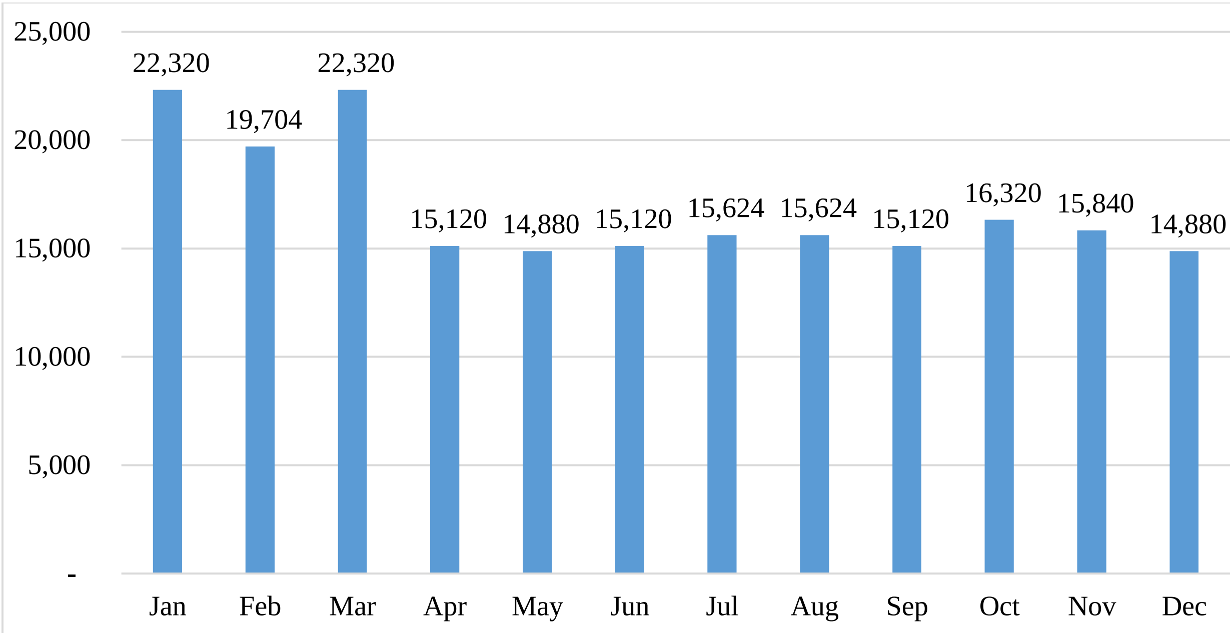


Figure 26 Seattle MSA: numbers of records in each month

For standardized modeling, all the data entries have been converted into hourly bicycle/pedestrian volume.

5.2 Emerging Data Sources

Passively Collected Location Data (PCLD) is one of the most effective data collection methods to provide longitudinal travel behavior monitoring for a large geographic area and sample size, and with a reasonable cost, which has significant potential as supplementary data input to the integrated public domain data warehouse. To take advantage of synergies between different PCLD datasets, and between the emerging and conventional data sources, a comprehensive data warehouse for PCLD datasets will be developed and integrated with additional supporting datasets. For example, INRIX trip trajectory data provides OD information and vehicle trajectories for about 1-2% of all vehicles in the entire State of Maryland for one year; AirSage data provides multimodal OD matrices in Maryland for different trip purposes, different periods of the day, and different months of the year; Social Media data such as Twitter data will enable us to extract traveler's locations over time and space for a long time; Other significant emerging potential data sources that has been archived include 2011/2012 DC-Baltimore GPS-based Travel Survey, the travel trajectories collected during the SafeTrack impact studies using the smartphone app "Travel Helper" and other GPS travel location data collected by UMD in previous research efforts in D.C. and Maryland.

5.3 CASE STUDY: D.C., Seattle and New York

The proposed method described in Chapter IV has been applied to estimate the month-to-month travel demand of driving, transit, for-hire, and non-motorized modes in New York City, Seattle, and Washington D.C. metropolitan areas using public domain data sources. This chapter summarizes the results and major findings.

5.3.1 Monthly Number of Driving Trips Estimation

The objective is to estimate the number of driving trips each month at metropolitan level. It is critical to define the boundary of the metropolitan area before running the proposed algorithm. A comprehensive list of the counties located in each metropolitan area is developed based on the latest MSA definition. Since DC MSA is selected as the template for demonstration, the list of counties in predefined study area is obtained. All the following procedures will be operated based on the counties list within this study area. After the study area is defined, the VMT will be computed based on the states which are included in the study area. For example, the Washington D.C. metropolitan area includes counties from four states: Maryland, Virginia, Washington D.C. and West Virginia. VMT will be computed for these states using HPMS data. The VMT for each road section could be obtained by multiplying the section length, AADT, and number of days per year. The total VMT for a county can be calculated by aggregating VMT for all roads within a county. Since the study area has been defined, the VMT will be calculated by counties in the study area.

2014 VMT	VMT (10^6 Miles)
DC	4039.508
Maryland	53857.052
Maryland (5 Counties)	20453.588
West Virginia	19079.221
Jefferson County	475.889

Virginia	82215.400
Virginia (17 Counties)	24220.45

Table 4 VMT computation by counties

The HPMS does not provide good data for local road VMT because of reporting exemptions for roads of such functional classes. To address this data gap, it is assumed that the proportion of local road VMT of each county compare with whole state, should be consistent to the proportion of all road VMT compare with the state. For example, the percentage of local road VMT of 5 counties in Maryland compared with the local road VMT of State of Maryland will be equal to the percentage of its overall VMT within the State of Maryland. This assumption will enable us to extract the local road VMT from Highway Statistics Report and consequently improve the reliability and credibility of proposed method. After adding local road VMT, the VMT will be disaggregated into monthly level. Table 5 presents the results of local road VMT disaggregation by states.

2014 VMT	**VMT (10^6 Miles)**	**Percentage**	**Local Rd VMT (10^6 Miles)**	**Total**
DC	4039.508	1.0000	774.000	4813.508
Maryland	53857.052	1.0000	4700.835	58557.887
5 Counties in Maryland	20453.588	0.3797	1785.261	22238.850
West Virginia	19079.221	1.0000	1329.000	20408.221
Jefferson County	475.889	0.0249	33.149	509.038

Virginia	82215.400	1.0000	8038.178	90253.579
17 Counties in Virginia	24220.45	0.2946	2368.027	26588.48

Table 5 Results of local road VMT disaggregation

5.3.2 Temporal adjustment factors

MOY is used to describe the traffic volume trend within a year. Similarly, the MOY factor is calculated using the following steps: For each station, calculate the average monthly volume and annual volume. The minimum requirement for a valid month is that the data within that month should over fourteen days. Generate the MOY factor by dividing the volume of the valid month by the annual volume for each station. Take the weighted average for all the stations with valid data for a function class. Calculate the MOY factors for a particular state by taking the weighted average for all function classes. The TMAS data will be selected to compute the monthly adjustment factors. Monthly adjustment factors present percentage of target month traffic volume versus annual historic traffic volume from the ATR data.

$$M_{\mathrm{MOY}} = \frac{D_{\mathrm{k}}}{d} / \sum_{k=1}^{12} \frac{D_k}{d} ,$$

where D_k is the summation volume for a valid month and d represents the total number of days in that month. Finally, MOY factors for all function classes for a particular state can be obtained by taking the weighted average for all the stations with valid data for a particular function class. Table 6 presents the results of MOY factors for each state in D.C. MSA

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Virginia	0.0761	0.0759	0.0827	0.0858	0.0870	0.0879	0.0860	0.0875	0.0839	0.0847	0.0824	0.0801
Maryland	0.0783	0.0779	0.0826	0.0848	0.0885	0.0853	0.0851	0.0872	0.0842	0.0859	0.0816	0.0786
West Virginia	0.0762	0.0729	0.0814	0.0809	0.0860	0.0878	0.0850	0.0847	0.0803	0.1010	0.0831	0.0807
DC	0.0867	0.0840	0.0900	0.0926	0.0862	0.0856	0.0884	0.0825	0.0688	0.0793	0.0798	0.0761

Table 6 Monthly adjustment factors (MAF)

Once the annual VMT is obtained, the proposed method allows us to dynamically estimate the monthly vehicular trips by using monthly adjustment factor obtained from the previous step. Since the monthly adjustment represents the traffic volume percentage for a particular state, the monthly VMT can be computed by the product of monthly adjustment fact with annual VMT. In addition, the research team also assume that the travel trend for a particular county is consistent with the travel trend in that state. Consequently, the annual level VMT can be disaggregated into monthly and county level. The annual VMT is disaggregated into target month VMT by using the monthly adjustment factors of target month.

However, planners and policy makers may be interested in the number of vehicular trips instead of VMT. The former is directly tied to the trend in mode shift over time. Therefore, this final step will be to compute the number of vehicular trips. Based on the average trip length and vehicular occupancy estimated from the most recent

Household Travel Survey Data collected in the same metropolitan area. Table 8 shows the number of monthly VMT, the average trip length, the average vehicle occupancy, and the number of vehicular and person trips estimated using the three inputs.

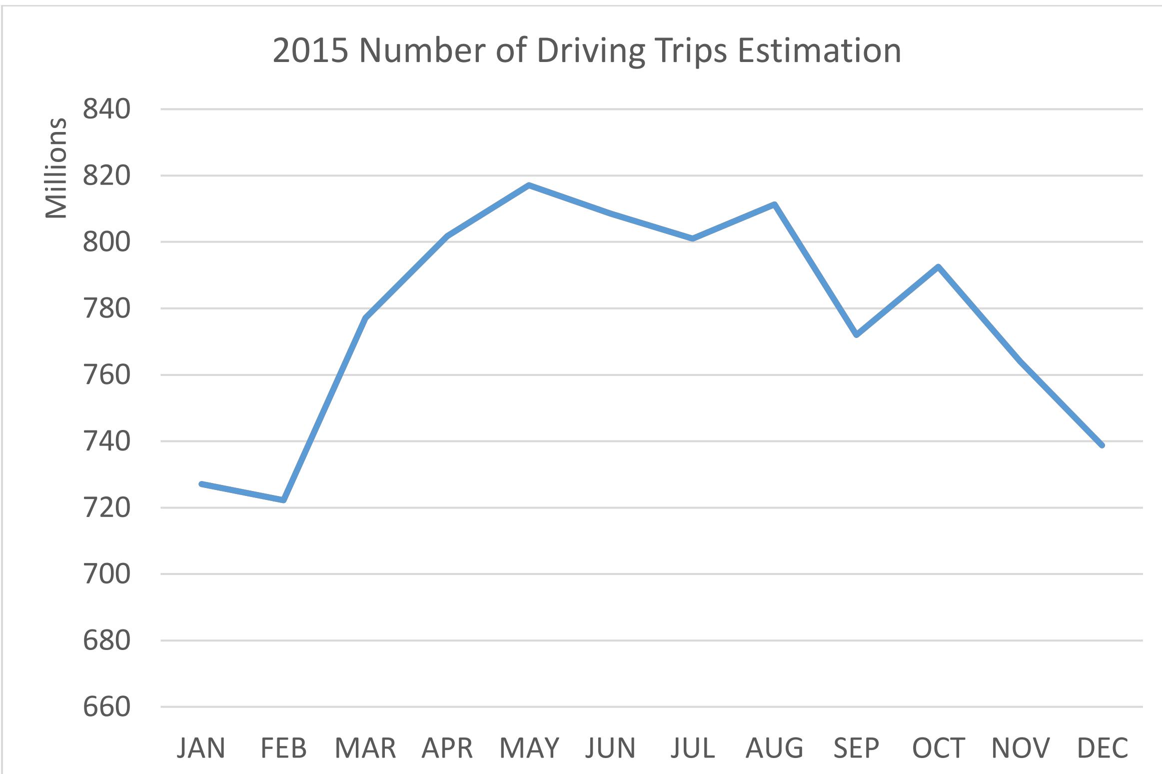


Figure 27 D.C. MSA 2015 number of driving trips estimation

Figures 27 show the monthly traffic trend in DC MSA 10 2015. The traffic volume shows a significant increase from February to August. It is consistent with the fact that the average temperature in these months presents an increasing trend. This observation will result in higher trip generation because people are more willing to participate in outdoor activities. It is noticeable that February also has a lower number of driving trips compared with other months in the winter season. This could be explained by the fact that February has two to three fewer days than other months.

5.4 Monthly Number of Transit Trips Estimation

The first step is to develop a comprehensive and up-to-date list of transit operators that provide services in a metropolitan area. Although most transit operators in the Washington D.C. metropolitan areas are labeled as “Washington, DC-VA-MD” in NTD, this identifier does not cover operators such as Maryland Transit Administration (MTA), and Fredericksburg Regional Transit and etc., all of which have significant services operating within the Washington D.C. metropolitan areas. Some agencies such as MTA serve more than one metropolitan areas and are labeled only for one of them. Others such as Fredericksburg Regional Transit may originally not part of a metropolitan area and becomes so later because of the expansion of the metropolitan area. Therefore, it is important to develop a comprehensive and up-to-date list of transit operators for each metropolitan area.

For the Washington D.C. Metropolitan area, the counties considered by the regional planning agency, Metropolitan Washington Council of Governments (MWCOG), in their regional planning model is slightly different from those included in the MSA. The COG model does not include Warren, Rappahannock, and Culpeper counties in Virginia, but included four additional counties in Maryland: Carroll, Howard, Anne

Arundel, and St. Mary’s, and the King George County in Virginia. To be consistent with the analysis of other metropolitan areas and make the method easily transferrable, I adopt the MSA definition in this study. Table 7 summarizes all the operators in the Washington D.C. metropolitan area.

5 digits NTDID	Agency	UZA Name	Reporter Type
30030	Washington Metropolitan Area Transit Authority	Washington, DC-VA-MD	Full Reporter
30051	Ride-On Montgomery County Transit	Washington, DC-VA-MD	Full Reporter
30058	City of Fairfax CUE Bus	Washington, DC-VA-MD	Full Reporter
30068	Fairfax Connector Bus System	Washington, DC-VA-MD	Full Reporter
30070	Potomac and Rappahannock Transportation Commission	Washington, DC-VA-MD	Full Reporter
30071	City of Alexandria	Washington, DC-VA-MD	Full Reporter
30073	Virginia Railway Express	Washington, DC-VA-MD	Full Reporter
30080	Arlington Transit - Arlington County	Washington, DC-VA-MD	Full Reporter
30081	Loudoun County Commuter Bus Service - Office of Transportation Services	Washington, DC-VA-MD	Full Reporter
30085	Prince George's County Transit	Washington, DC-VA-MD	Full Reporter
30103	Martz Group, National Coach Works of Virginia	Washington, DC-VA-MD	Full Reporter
30072	Transit Services of Frederick County	Frederick, MD	Full Reporter
30088	County Commissioners of Charles County, MD	Waldorf, MD	Full Reporter
30034	Maryland Transit Administration	Baltimore, MD	Full Reporter
30079	Fredericksburg Regional Transit	Fredericksburg, VA	Small Systems Reporter
30106	National Capital Region Transportation Planning Board	Washington, DC	Small Systems Reporter

Table 7 Transit operators in the Washington D.C. metropolitan area

Using this list, the research team can pull the month-by-month UPTs for full reports and annual UPTs for small systems reporters from the NTD. However, in order to calculate the month-by-month transit ridership for a metropolitan area, two problems have to be addressed: splitting the ridership for those operators whose network covers multiple network and estimating month-by-month ridership of small systems who only report annual data.

For MARC Train system, MTA only collects the byline ridership stead of complete OD. Therefore, it is unclear where passengers are going after boarding at one particular station. Without additional OD information, I assume that passengers boarding at one station are equally likely to go to any other stations along the same line. MARC Train system includes three independent lines: Penn Line, Camden Line and Brunswick Line with only one shared station: the Washington Union Station, a terminal station in downtown Washington D.C. Therefore, transfers between different lines are not possible.

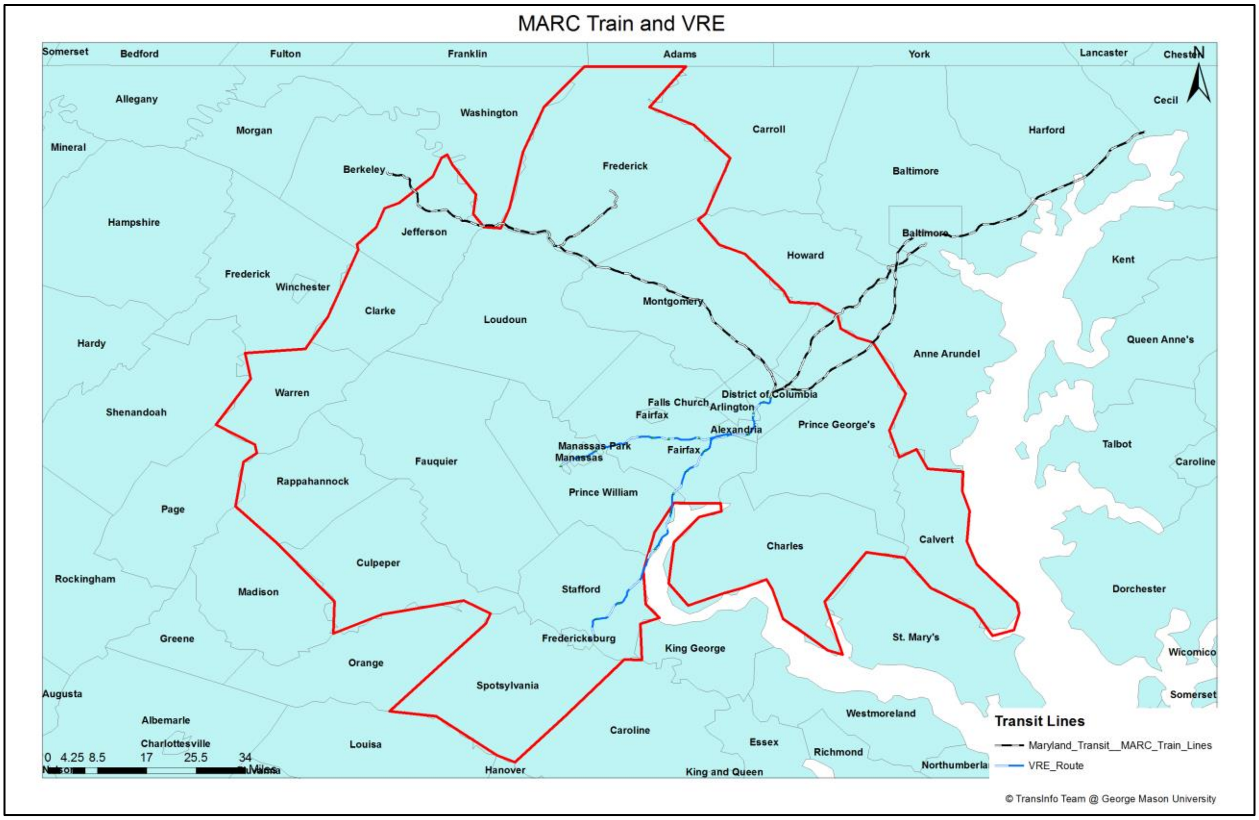


Figure 28 MARC Train and Virginia Rail Express (VRE) network

As mentioned above, it is assumed that the proportion of MARC Train ridership to be included in the Washington D.C. metropolitan area should be:

$$p_i = \frac{N_i(N_i - 1) - J_i(J_i - 1)}{N_i(N_i - 1)}$$

Where p_i is the proportion of ridership of line i that should be included in the ridership total of the current metropolitan area; $N_i$is the total number of stations of line i; J_i is the number of stations of line i that falls within the metropolitan boundary.

Among the three MARC train lines, 53.85% of the Penn Line riders should be included in the total for D.C., while the percentage number of the Camden Line is 77.27%, and that of the Brunswick Line is 100%. MTA also provided line-by-line annual ridership. Using the per-line annual ridership as the weight, the percentage of MARC train riders that go to D.C. MSA is 64.54%. Should additional information be collected by MTA, the proposal method can be further enhanced. For example, the percentage number can be refined month-by-month if monthly total were collected on a continuous basis. Riders from some particular stations may be more likely to go to one metropolitan area than the other. Surveys on their destinations could also help to refine the proposed method.

5.4.1 Analysis of Commute Bus Ridership

Figure 28 shows all commute bus routes MTA operates. Based on the name of the bus routes, we can clearly tell the metropolitan area each route serves and it is unlikely that riders would use these routes for local trips. By assuming that all riders of a particular route would go to the metropolitan area that route is designed to serve, the percentage of transit ridership that belongs to the metropolitan area of interest can be approximated by:

$$p = \frac{K}{N}$$

Where K is the number of commute bus lines that serve the metropolitan area of interest, while N is the total number of commute bus routes operated by a transit operator. According to Figure 14, among the 35 commute bus routes MTA operates, 30 routes serve the Washington D.C. area and 5 serve the Baltimore area. Therefore, the percentage of D.C. commuters served by MTA is 85.71%. However, this method could be further improved if month-by-month commute bus ridership by route becomes available. The updated percentage number is:

$$p_m = \frac{\sum_k q_{km}}{\sum_n q_{nm}}$$

Where p_m is the proportion of commute bus ridership of MTA that should be included in the ridership total of the DC metropolitan area in month m, and q_{km} is the monthly ridership of bus route k in month m. Figure 30 compared the results based on the monthly by-line commute bus ridership data (blue) and on bus route information only (red). The monthly fluctuation in ridership split ratio is not significant, but the method based on bus route information only does slightly under-estimate that ratio.

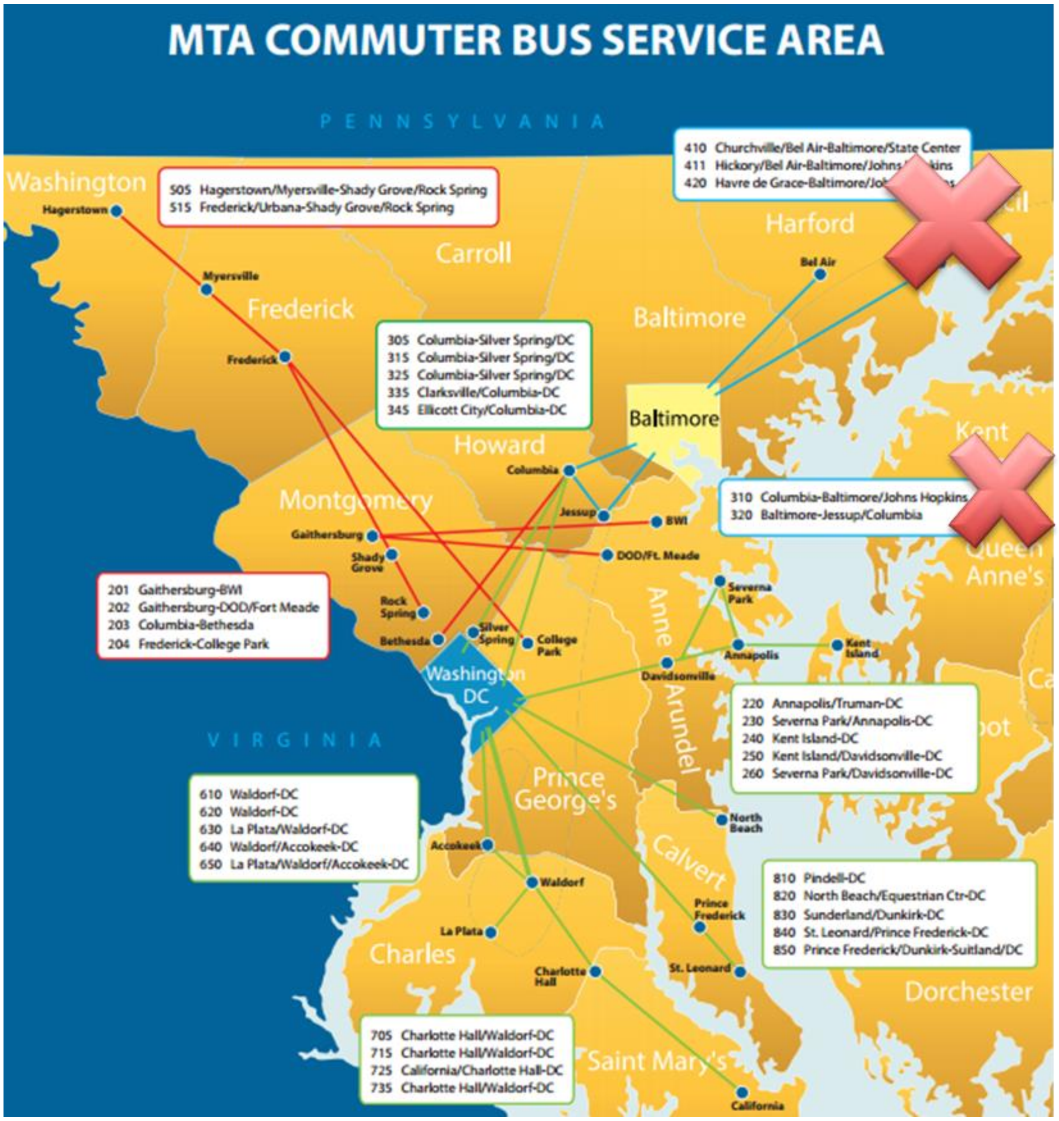


Figure 29 Commute bus line operated by Maryland Transit Administration

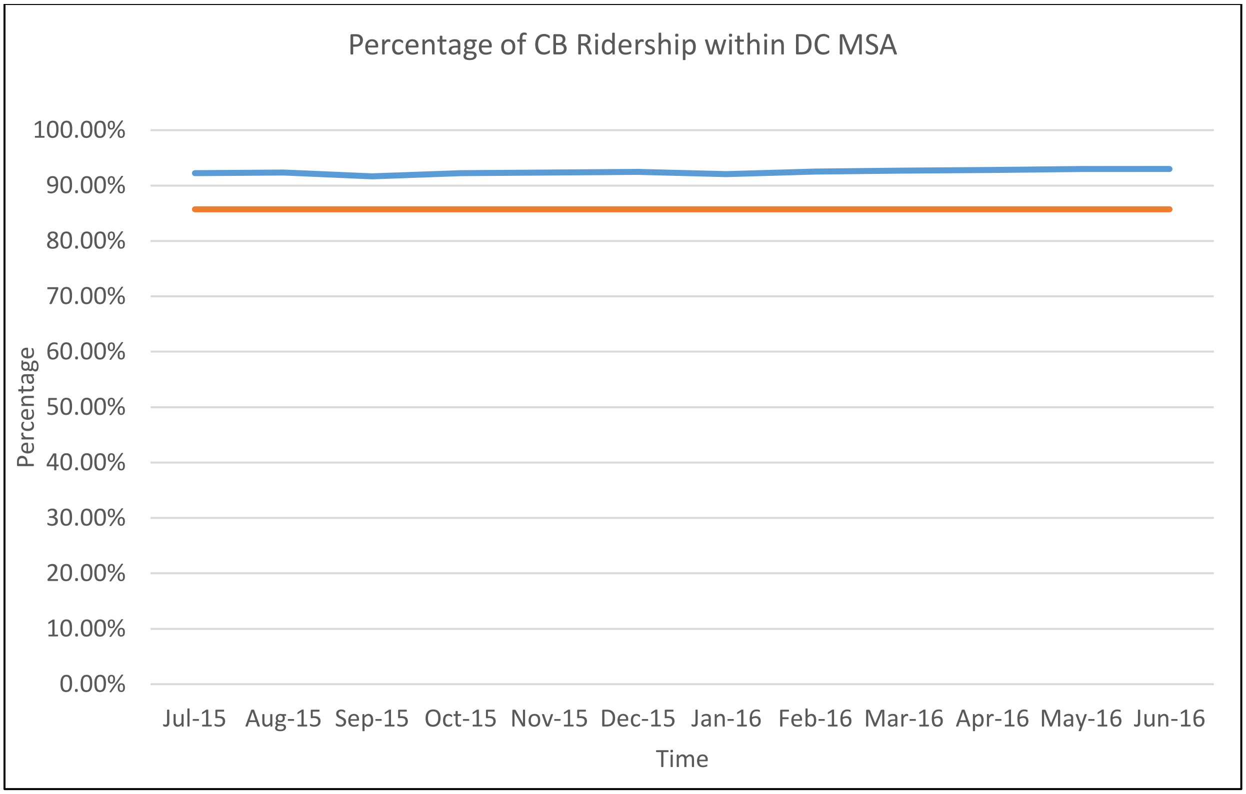


Figure 30 Monthly MTA Commute Bus ridership that belongs to D.C. metropolitan area based on by-line ridership data (blue) or simply number of routes (red)

5.4.2 Monthly Ridership of Small Operators

As mentioned in the introductory section, small operators only report the annual total ridership to NTD. Without additional information about the month-by-month ridership patterns of these small operators, I assume that the trend is consistent with large operators in the same geographic area. Figure 16 shows the percentage of monthly ridership compared to annual total based on data of large operators in the Washington D.C. metropolitan area. This monthly trend is then used to convert the annual ridership reported by small operators into month-by-month ridership.

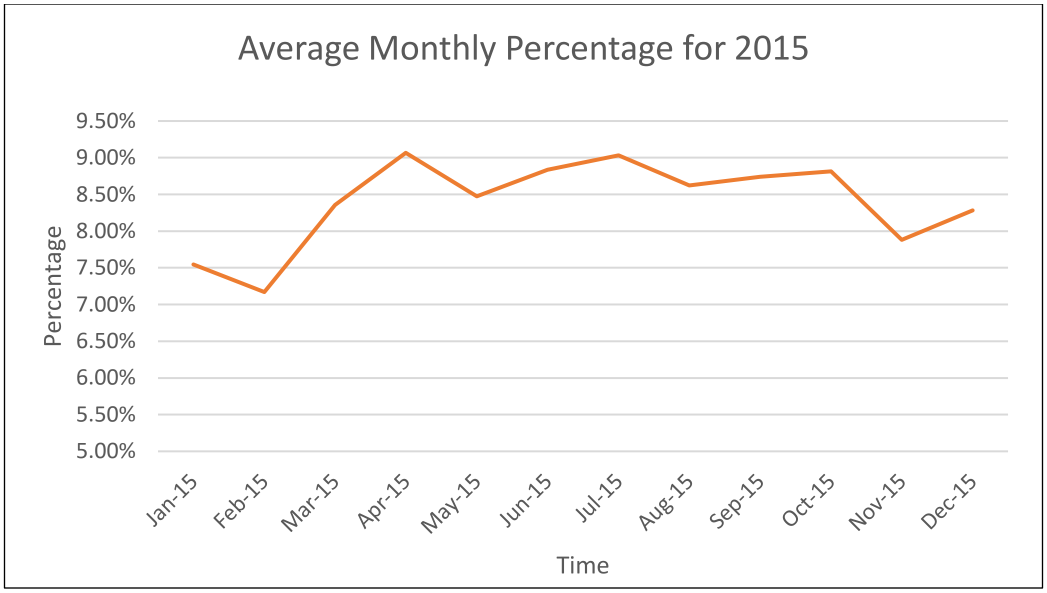


Figure 31 Monthly trend for small transit operators estimated using month-by-month ridership of large operators in the Washington D.C. metropolitan area

After addressing the issues of cross-border trips of large operators and the monthly trend of small operators, we can estimate the month-by-month transit ridership within a metropolitan area using UPT data from NTD and the method developed in this project. Figure 32 shows the estimated monthly UPT for the Washington D.C. metropolitan area (blue curve) from for the period of 2010-2015.

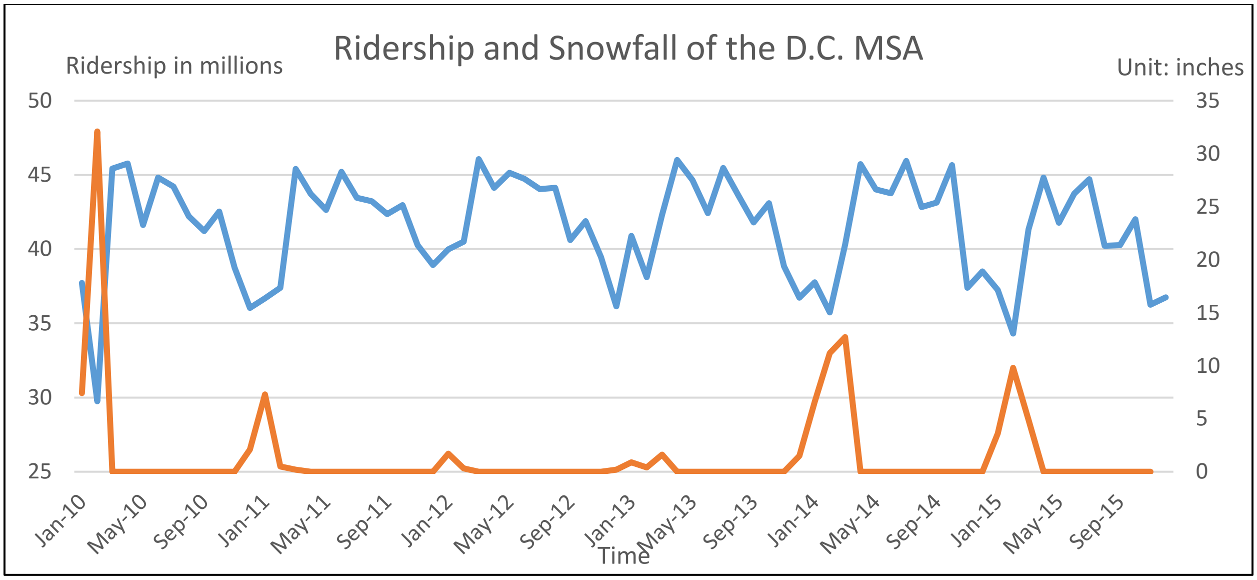


Figure 32 Month-by-month transit trips (blue) in the Washington D.C. area from 2010 to 2015, in comparison to show falls (orange) in inches during the same period

According to literature, transit ridership is sensitive to weather conditions, especially severe weather such as heavy snow falls. Therefore, I also plotted the accumulative snow falls in inches for each month on the same graph with line in color orange. Significant snow falls are correlated with unusually ridership drops, which enhances our confidence about the estimated results.

To further test the quality of ridership report in NTD, I compared the reported UPT (green line in Figure 33) with the annual revenue (blue line) for the Washington Metro system. The revenue data comes directly from the fare collection system and is usually much more accurate. When there is no change in fare structure, the curves for UPT and the total revenue should be in parallel. When a fare hike happened, revenue should increase faster than the UPT. These trends are reflected in Figure 33.

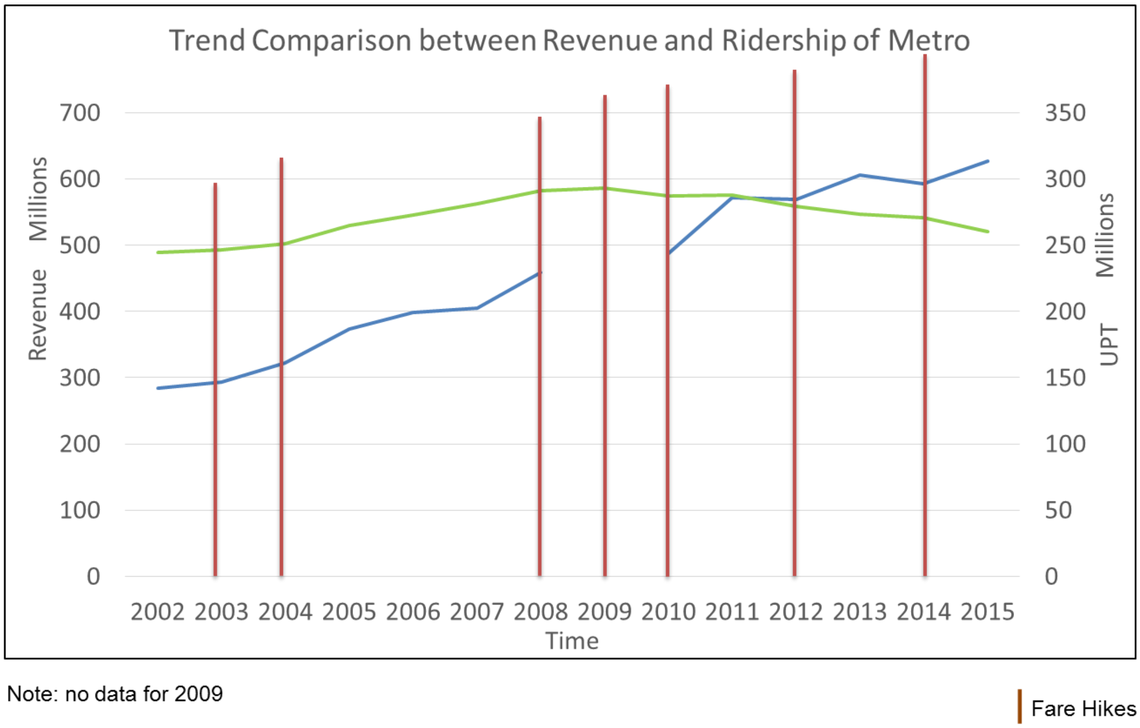
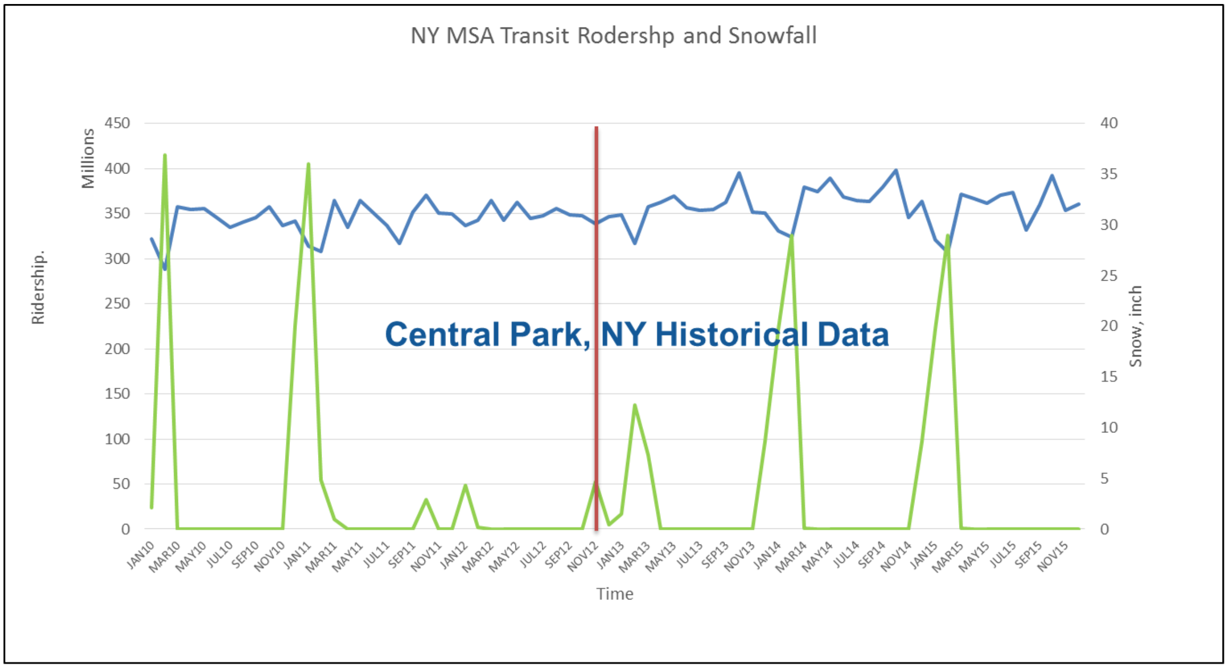


Figure 33 Comparison of Washington Metro ridership in UPT (green) and the annual revenue (blue)

Applying the same method using data collected from NTD, I also estimated the month-by-month transit trips in UPT for the New York City (Figure 34) and the Seattle metropolitan areas (Figure 35).



Data source: National Weather Service, Central Park Station

Hurricane Sandy

Figure 34 Month-by-month transit trips (blue) in the New York City metropolitan area from 2010 to 2015, in comparison to show falls (green) in inches during the same period

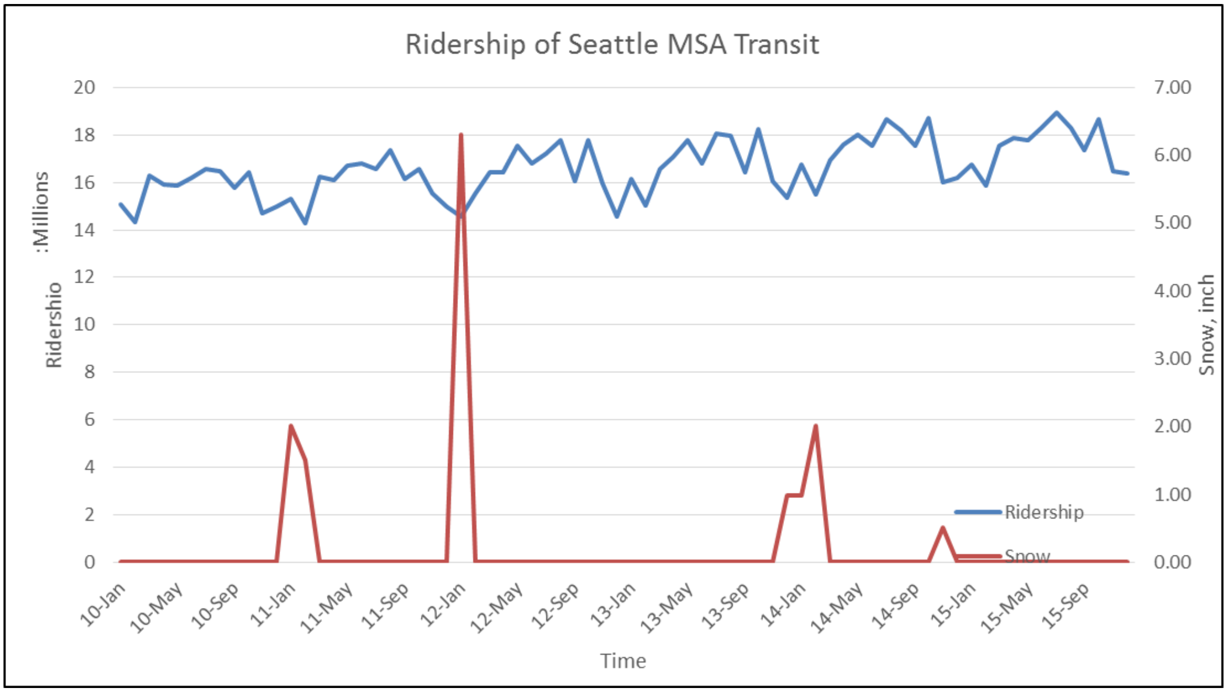


Figure 35 Month-by-month transit trips (blue) in the Seattle metropolitan area from 2010 to 2015, in comparison to show falls (red) in inches during the same period

5.5 For-hire Mode

Applying the two methods proposed for ride-hailing and shuttle in the Chapter IV, the 2015 monthly trip totals are obtained, and the results of each city are shown in Figure 36, Figure 37, and Figure 38. The method is based several assumptions that may be further validated or rejected if new data sources become available.

In Figure 36, 2015 monthly trips for the entire for-hire modes in NYC are plotted. Monthly trips of for-hire modes are based on the ground truth data for NYC due to better data availability than other metropolitan areas. Figure 37 and Figure 38 showed the for-hire monthly trips in DC and Seattle, respectively. Ride-hailing trips in both DC and Seattle and two airport shuttle trips in DC are estimated using methods described in Chapter III. It is based on an assumption that the trips in the same month are consistent regardless of year and the number of taxi trip data are necessary for any cities to estimate ride-hailing trips by NYC market share. In addition, Lyft trips from January to March are missing for all cities since Lyft trip data in NYC became available as of April 2015. The number of taxi trips from January to April in 2015 are replaced by trips from January to April in 2016 in DC area due to data availability. Since taxi data is of public domain and will be released under FOIA once it becomes available, this problem will be fixed in the near future.

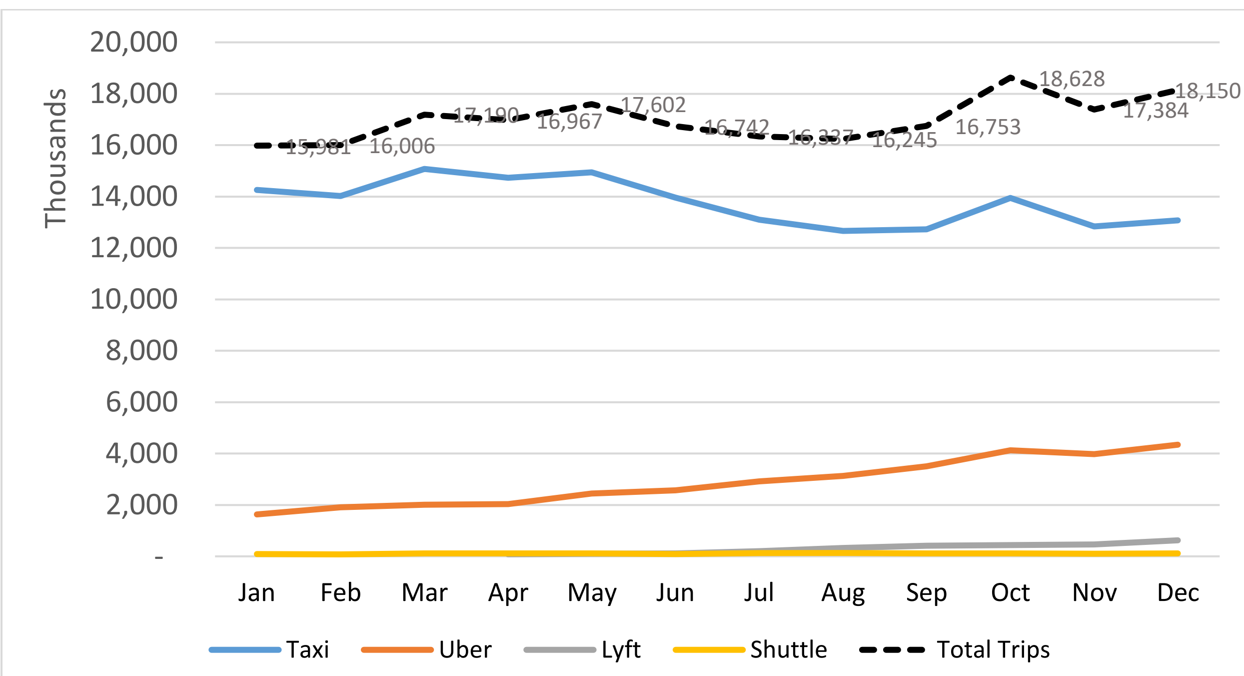


Figure 36 2015 monthly trip totals in New York City

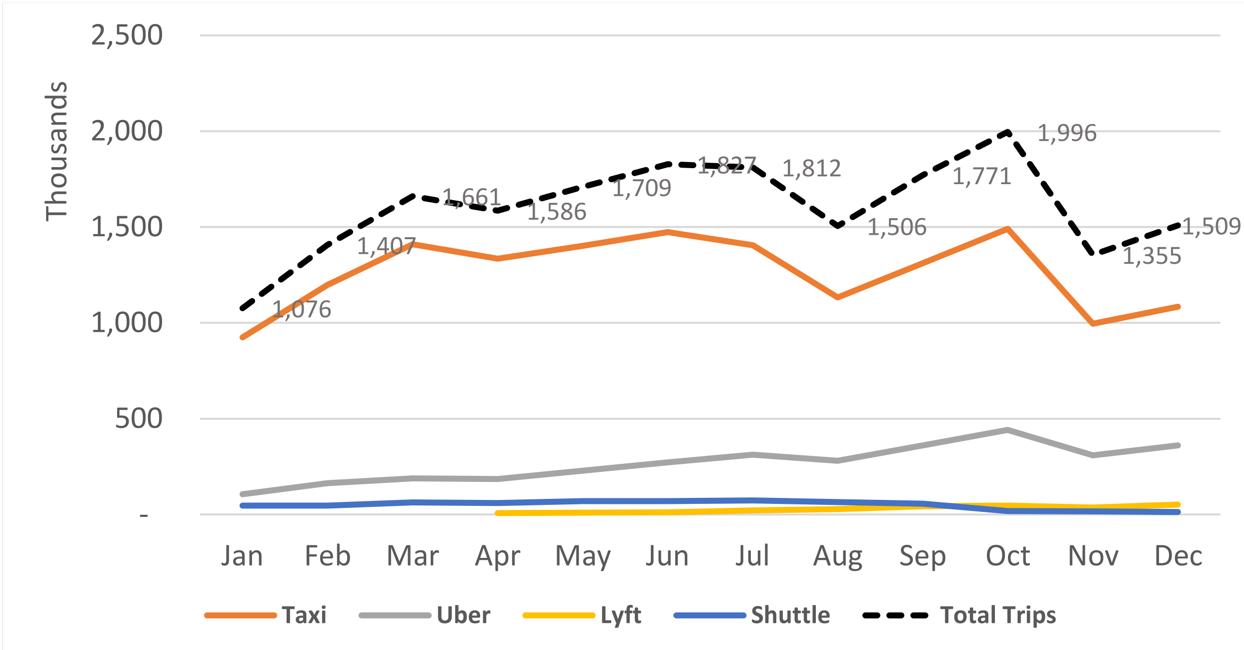


Figure 37 2015 monthly trip totals in Washington DC

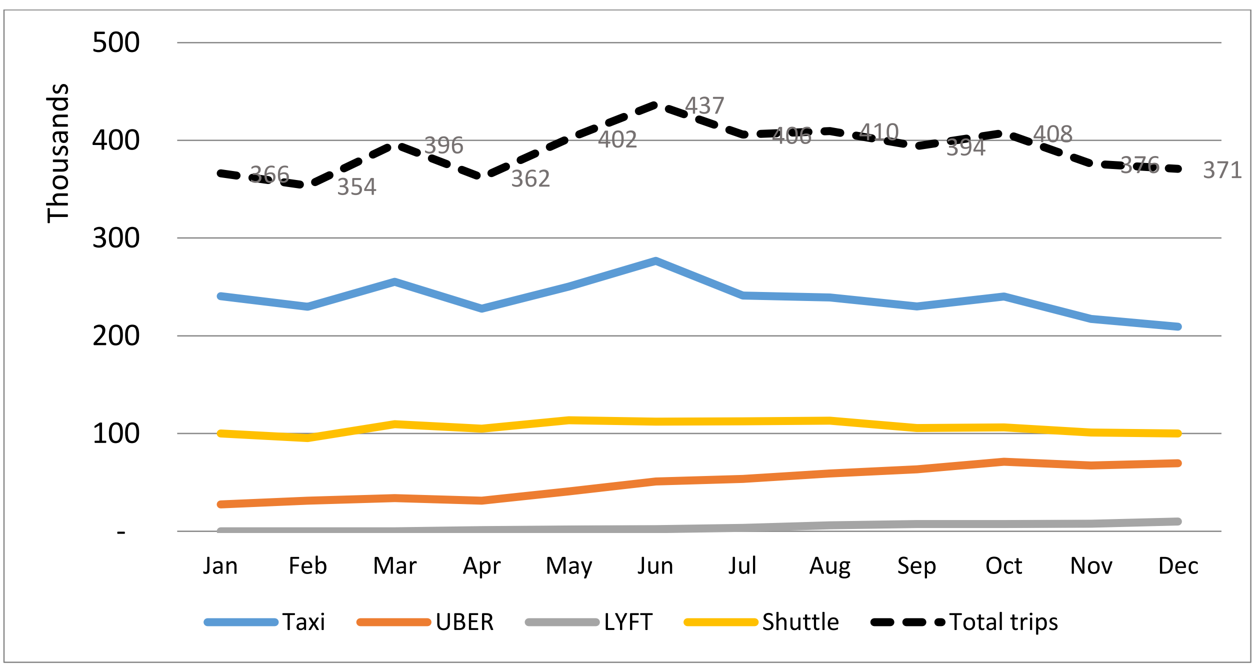


Figure 38 2015 monthly trip totals in Seattle

5.6 Non-motorized Mode

This sub-section presents the estimates of monthly non-motorized trips in the Washington D.C., NYC, and Seattle metropolitan areas.

5.6.1 Washington Metropolitan Area

The annual total estimate for Washington metropolitan area in 2015 is given in Table 8.

2015		Biking	Walking
State	County	Estimate	Estimate
District of Columbia	District of Columbia	14,718	50,165
Maryland	Frederick	291	2,692
	Montgomery	3,228	11,590
	Prince George's	564	10,636
	Calvert*	0	438
	Charles	100	726
Virginia	Alexandria	1,121	3,908
	Arlington	2,818	7,323
	Clarke*	26	251
	Culpeper*	0	391
	Fairfax	2,080	13,269
	Fairfax City*	74	557
	Falls Church*	99	221

	Fauquier*	0	553
	Loudoun	466	2,956
	Manassas*	40	326
	Manassas Park*	120	74
	Prince William	187	3,139
	Rappahannock*	0	98
	Spotsylvania	0	686
	Stafford	0	534
	Fredericksburg*	93	721
	Warren*	13	424
West Virginia	Jefferson*	17	744
Daily Commuting Total		26,055	112,422
Commuting Ratio		30%	5%
Annual Total		**63,400,500**	**1,641,361,200**

Table 8 Annual total of non-motorized trips in Washington metropolitan area
*: The results of these counties come from the 2011-2015 ACS 5-year estimates.

(Data source: American Community Survey and 2007/2008 TPB Household Travel Survey)

The monthly trends from different sources have been compared in Figure 39, which includes the estimation results from PMM with or without the effects of weather controlled and the monthly trends of the total bikeshare trips and those conducted by registered users. In Figure 39, there was an obvious decrease of bikeshare trips in July, probably due to a facility maintenance that affected the supply. It shows that the model including weather conditions provides more reasonable results.

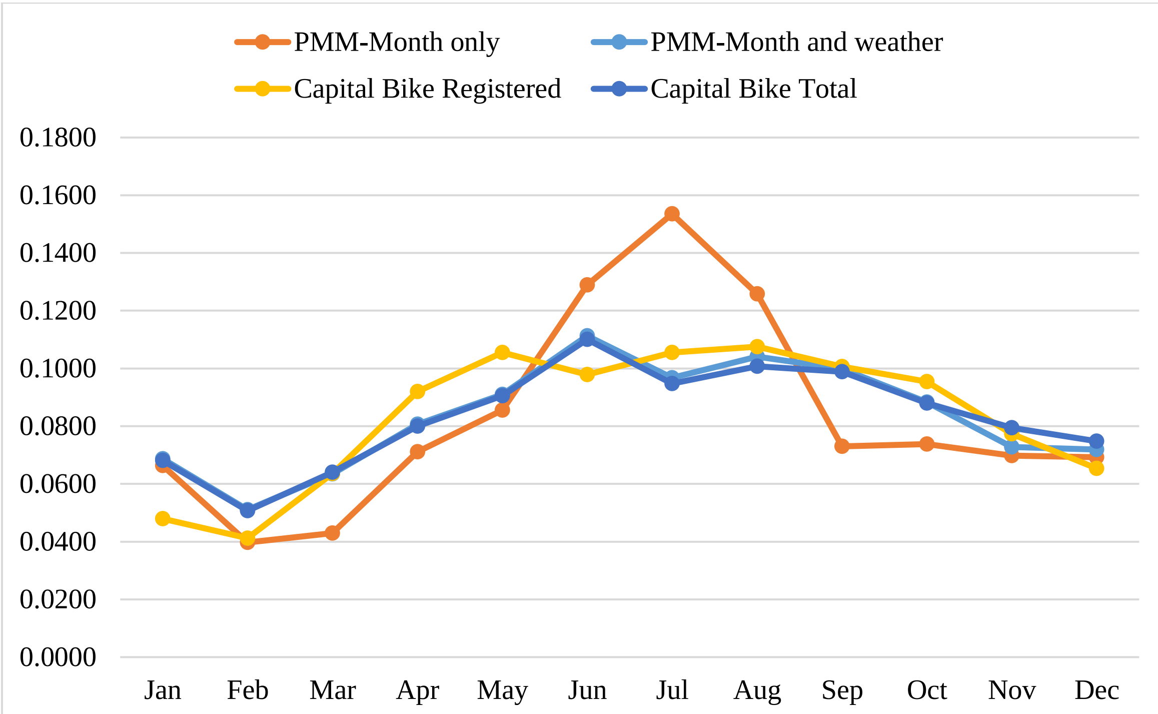


Figure 39 Comparison of the bicycle trip monthly trends

The monthly trip estimate is further derived by combining the annual total estimate and the monthly trend estimate. The bicycle trip estimate is shown in Figure 40 and the walking trip estimate is shown in Figure 41. Since the data quality of the pedestrian count data cannot meet the model requirements, the monthly trend of walking trips is assumed to be the same as bicycle trips. The procedure can be significantly improved should more pedestrian data become available in any metropolitan areas.

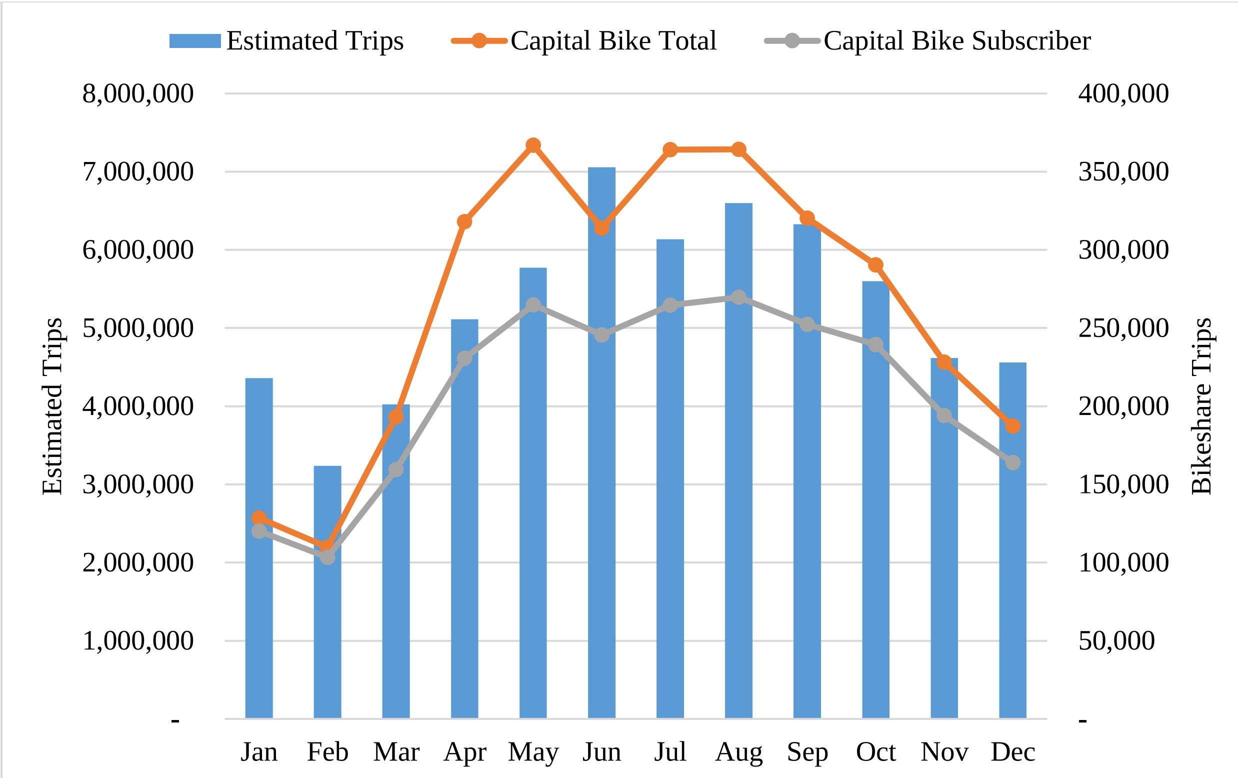


Figure 40 2015 monthly bicycle trips in Washington metropolitan area

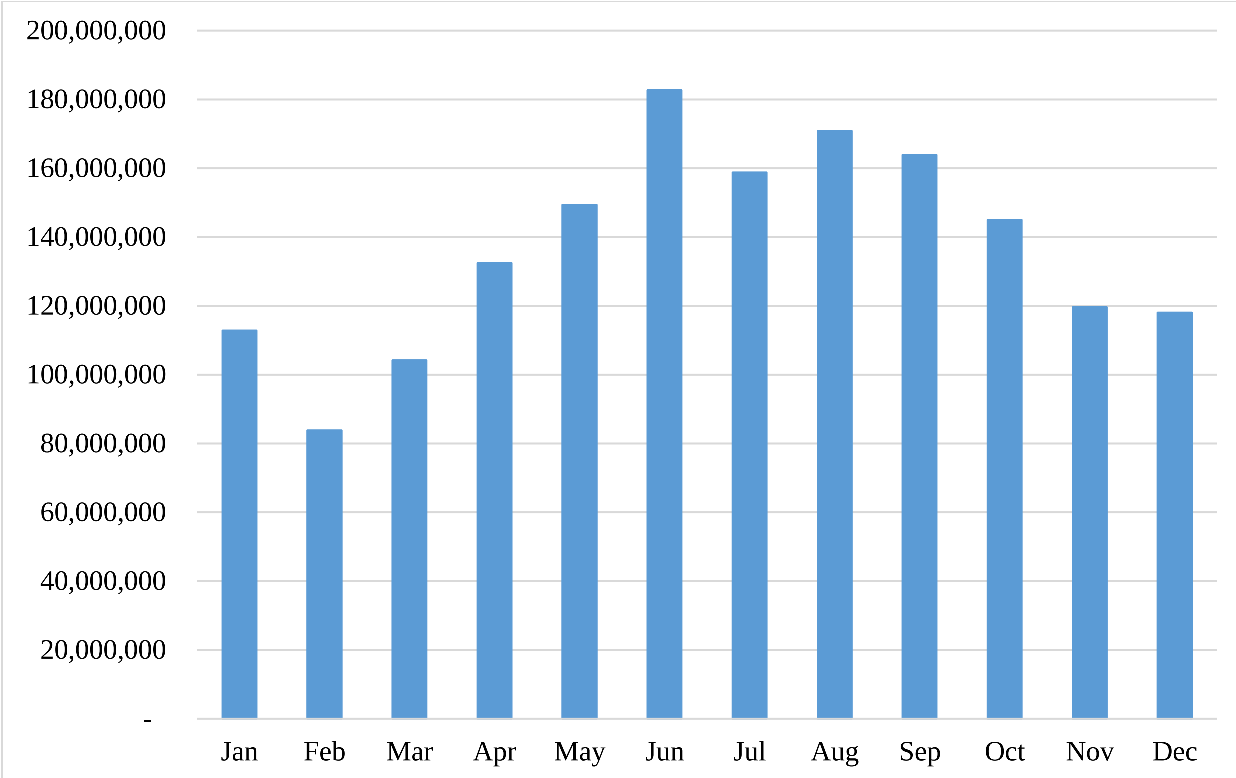


Figure 41 2015 monthly walking trips in Washington metropolitan area

5.6.2 New York Metropolitan Area

The annual total estimate for New York metropolitan area in 2015 is given in Table 9.

2015		Biking	Walking
State	County	Estimate	Estimate
New York City	NYC	46,057**	403,211
New York	Nassau	1,829	14,359
	Suffolk	1,840	10,896
New Jersey	Bergen	763	13,468
	Essex	371	16,144
	Hudson	1,843	29,386
	Hunterdon	278	1,543
	Morris	415	4,512
	Passaic	806	7,392
	Somerset	590	3,684
	Sussex	69	1,145
	Union	949	8,130
	Warren*	262	1,553
	Middlesex	1,066	16,149
	Mercer	926	6,516
	Monmouth	1,719	6,253
	Ocean	1,186	5,120
Connecticut	Fairfield	641	11,629
	New Haven	2,012	15,677

	Litchfield	44	2,425
Pennsylvania	Monroe	0	1,434
	Pike*	3	812
	Carbon*	76	859
	Lehigh	292	4,881
	Northampton	300	3,566
	Warren*	76	823
Daily Commuting Total		64,413	591,567
Commuting Ratio		18.40%	5.29%
Annual Total		255,551,576	8,163,400,945

Table 9 The annual total of non-motorized trips in New York Metropolitan Area

*: The results of these counties come from the 2011-2015 ACS 5-year estimates.

**: The average daily commuters in 2015 is estimated to be 45,000 by NYC DOT.

(Data source: American Community Survey and 2010/2011 Regional Household Travel Survey)

The monthly trends from different sources in NYC have been compared in Figure 42, which includes the estimation results from PMM and the monthly trends of the total bikeshare trips and those conducted by subscribers. For the estimated monthly trend, there is a slight decrease in June and July, which may also be the consequence of summer vacation. The bikeshare trips considered in the graph only counts those originated or destined in the existing stations by January to exclude the influence of the bikeshare facility expansion.

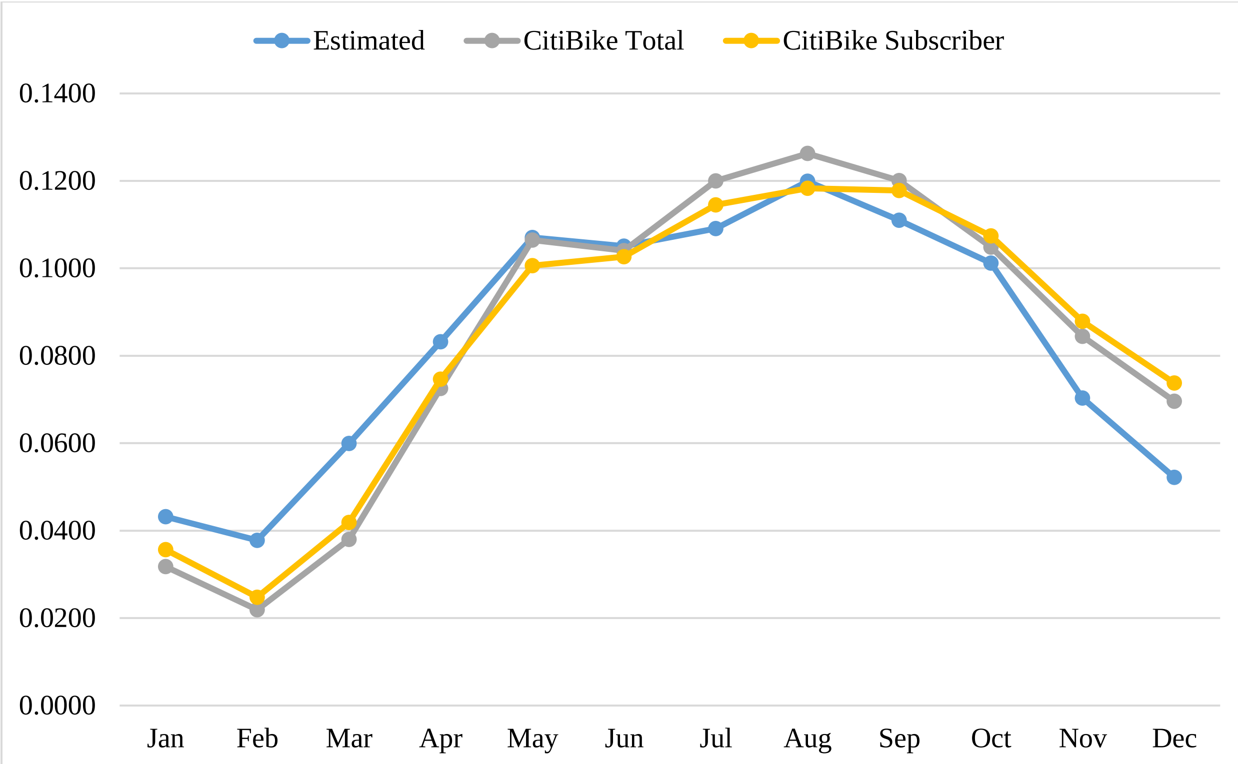


Figure 42 Comparison of bicycle trip monthly trends

The bicycle trip estimate for New York metropolitan area is shown in Figure 43 and the walking trip estimate is shown in Figure 44. Since NYCDOT only conducts a bi-annual pedestrian count program, the data is not qualified to derive a reliable monthly trend. So, the monthly trend of walking trips is also assumed to the same as the bicycle trips.

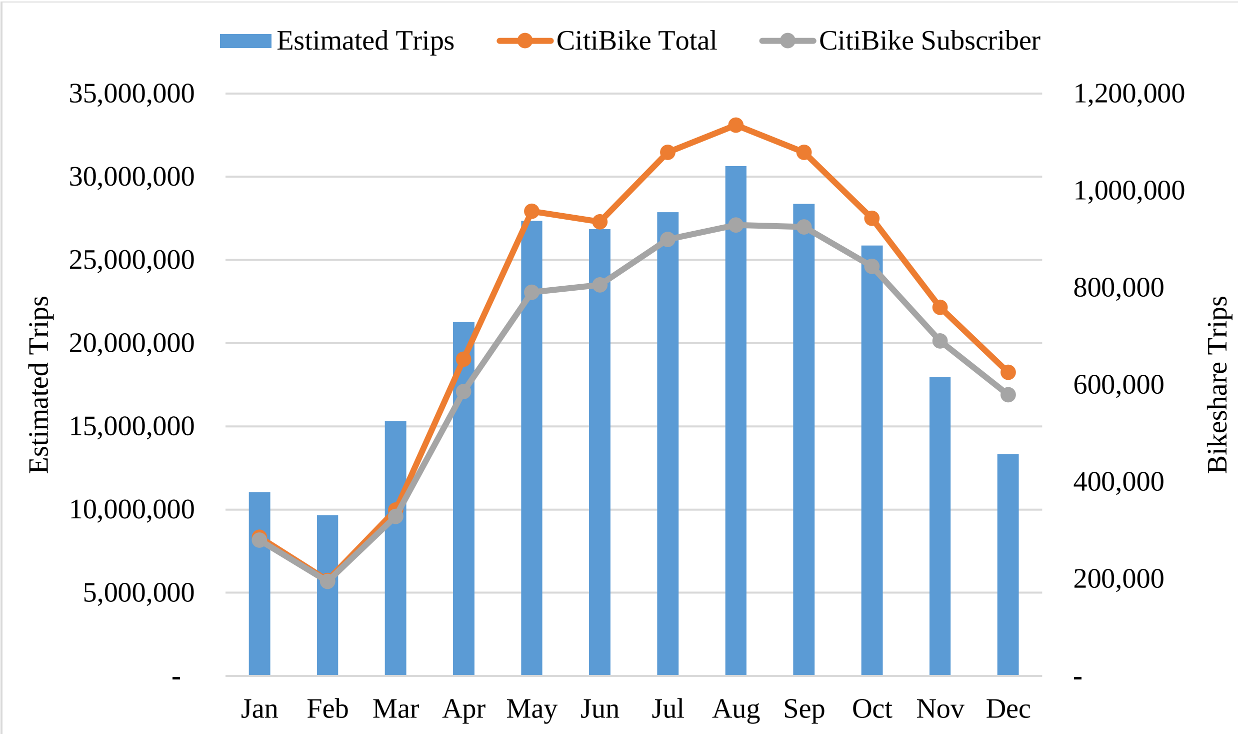


Figure 43 2015 monthly bicycle trips in New York Metropolitan Area

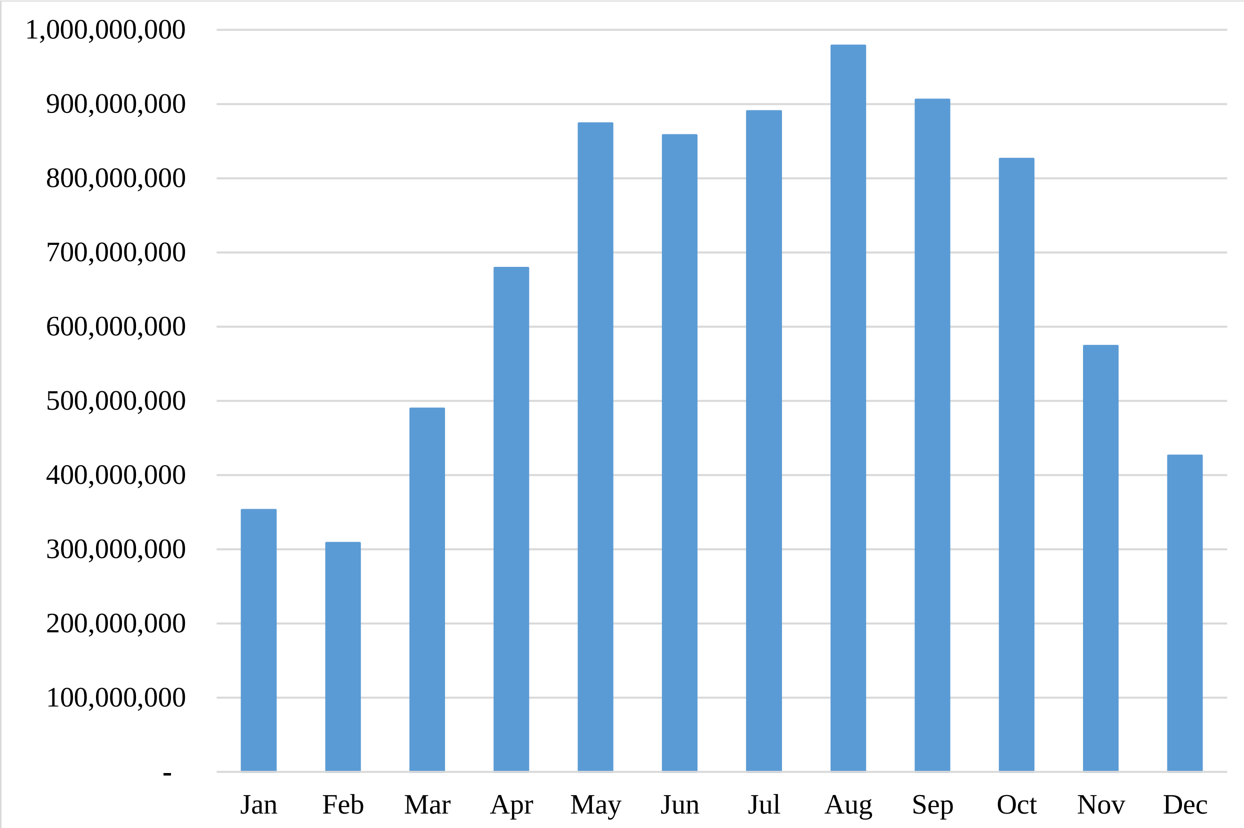


Figure 44 2015 monthly walking trips in New York metropolitan area

5.6.3 Seattle Metropolitan Area

The annual total estimate for Seattle metropolitan area in 2015 is given in table 10.

2015		Bicycle	Walked
State	County	Estimate	Estimate
Washington	Seattle	16,251	43,665
	King	19,730	57,374
	Snohomish	2,209	9,643
	Pierce	1,645	7,759
Daily Commuting Total		39,835	118,441
Commuting Ratio		29.68%	10.73%

Annual Total	97,976,920	805,796,179

Table 10 Annual total of non-motorized trips in Seattle metropolitan area
(Data source: American Community Survey and 2015 Seattle Household Travel Survey)

The monthly trends from different sources have been compared in Figure 45, which includes the estimation results from PMM with or without the effect of weekday indicator and the monthly trends of the total bikeshare trips. For the estimated monthly trend, there is a slight decrease in June because of summer vacation. Compared to the estimated monthly trend, the bikeshare trips indicate more trips in July while less in the winter, which may imply that a large portion of the bikeshare system users comes from tourists.

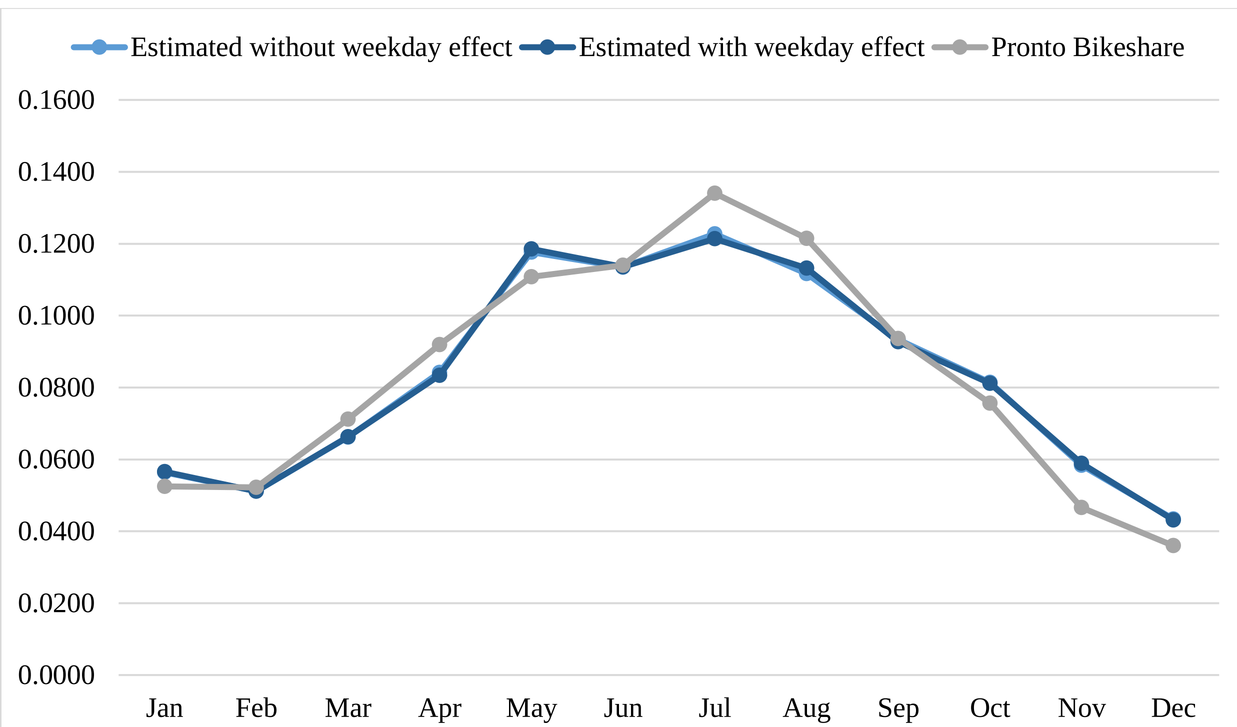


Figure 45 Comparison of the bicycle trip monthly trends

The bicycle trip estimate for Seattle metropolitan area is shown in Figure 46 and the walking trip estimate is shown in Figure 47.

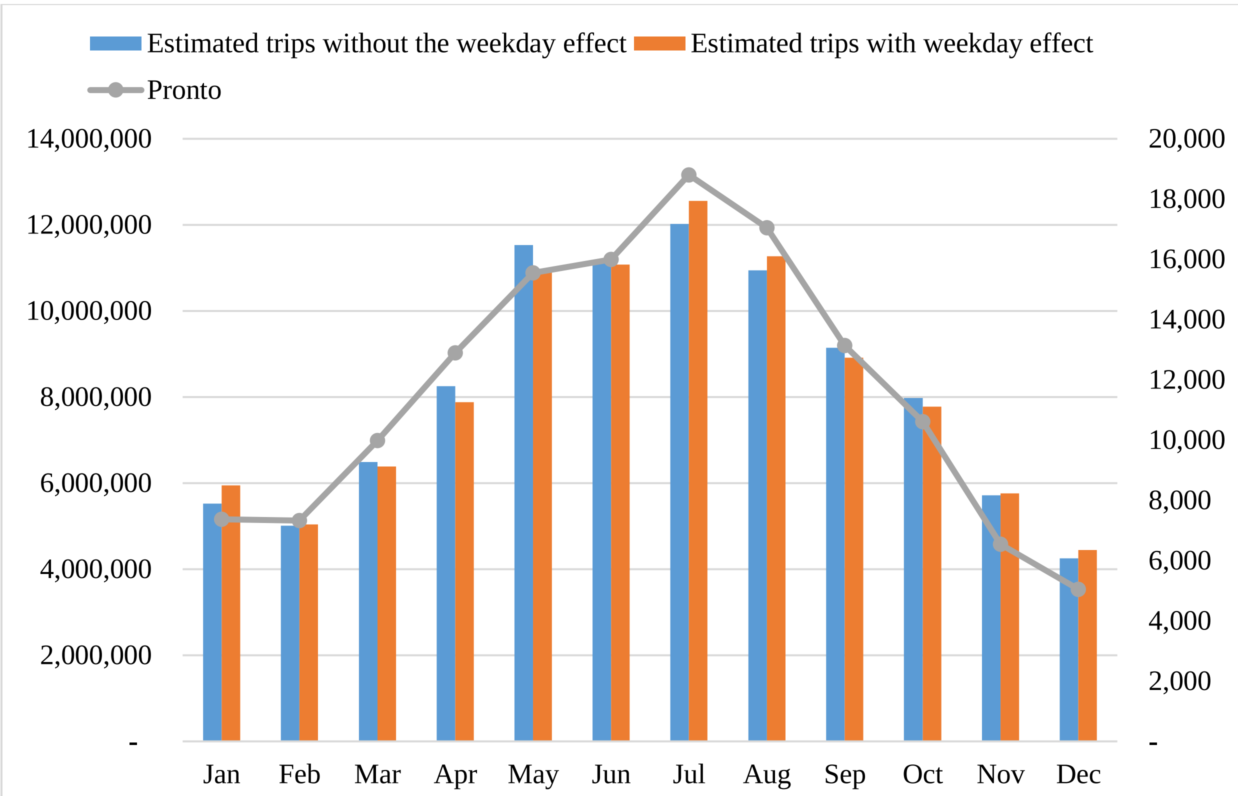


Figure 46 2015 monthly bicycle trips in Seattle metropolitan area

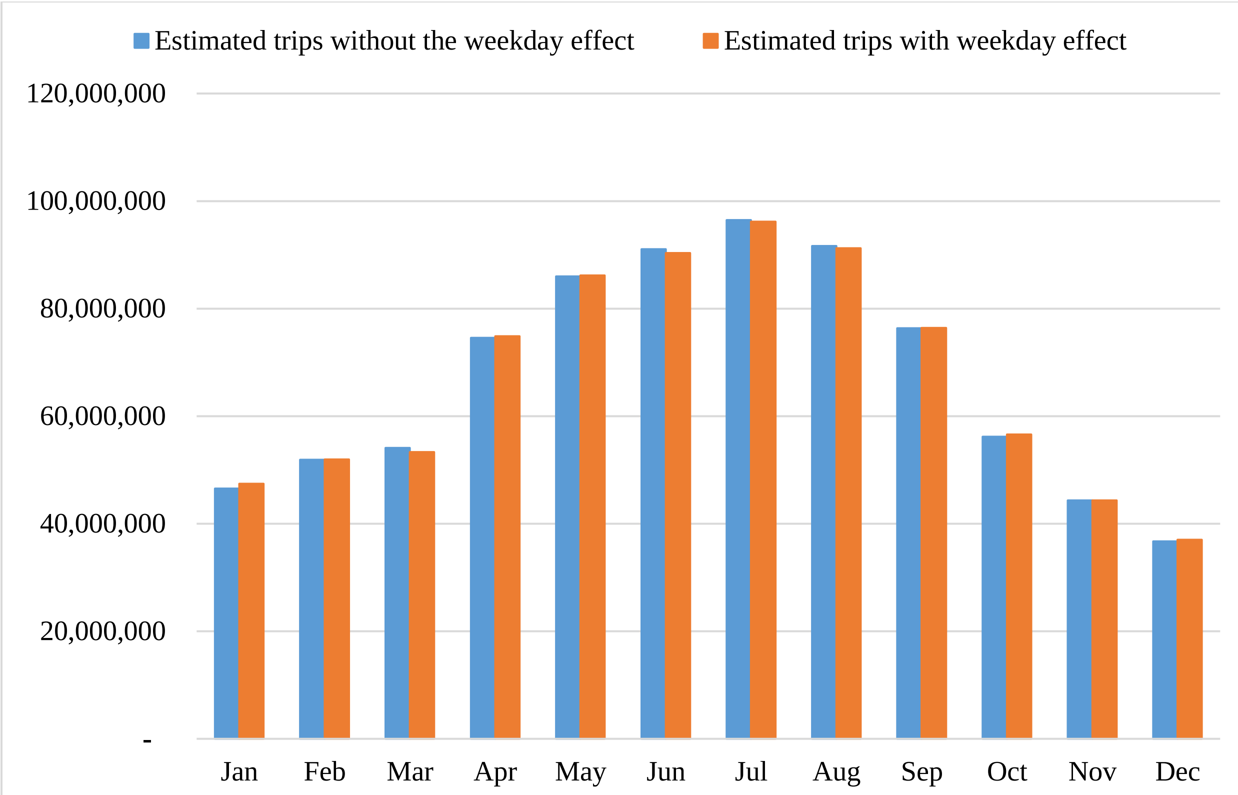


Figure 47 2015 monthly walking trips in Seattle metropolitan area

5.7 Validation

It is of great significance to validate the results derived from the proposed methods. The shortage of validation source is the most challenging issue. To address that, many potential data sources has been explored and detailed validation plan are proposed in this section by modes based on available validation sources. For driving mode, as what I have discussed in previous sections, only limited private sector data sets provide number of driving trips for every state. TVT report is considered as a major validation source to control the quality of monthly adjustment factor. The validation results are presented in section 3.5. For non-motorized mode, since the monthly adjustment factor for biking and walking trips were estimated based on statistic regression model, 10-fold cross validation is adopted to test the reliability and robustness. In addition to that, negative binomial regression model with same variable set is selected to evaluate the performance of existing model, using D.C. MSA as demonstration. Results will be presented and discussed in this section.

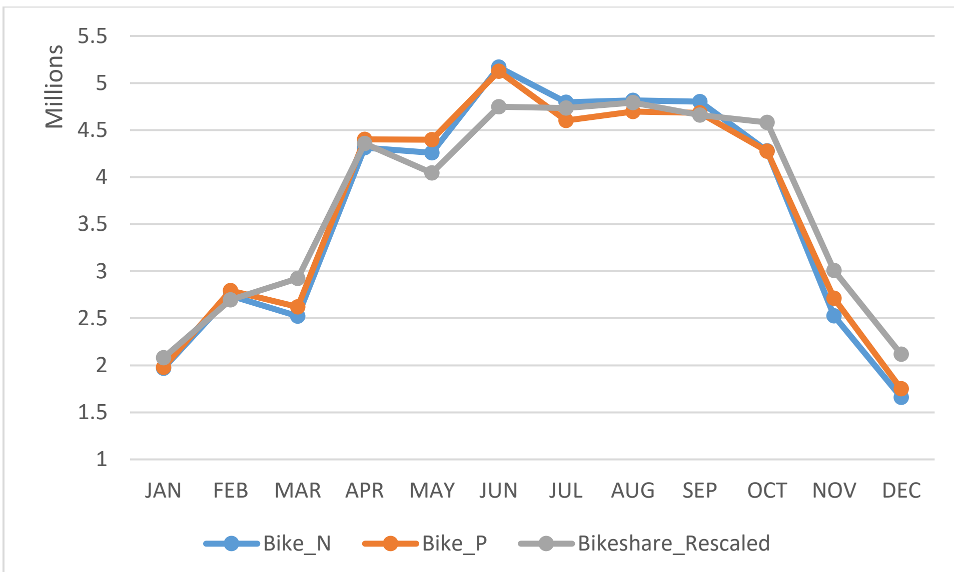


Figure 48 2017 monthly factor in D.C. Metropolitan area derived from regression models and bike-sharing data

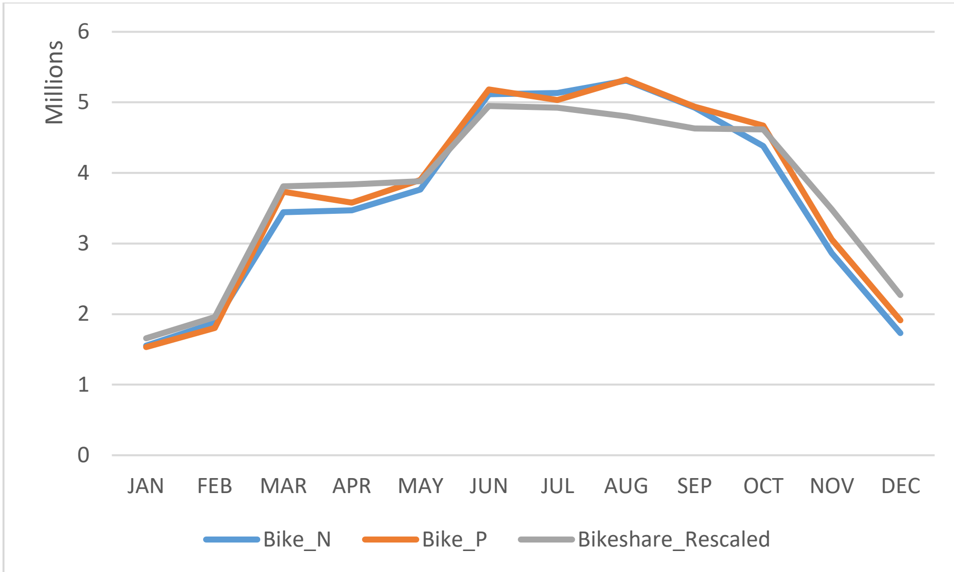


Figure 49 2016 monthly factor in D.C. Metropolitan area derived from regression models and bike-sharing data

Figure 48 and 49 present monthly trends computed using Negative binomial multilevel regression (blue line), Poisson multilevel regression model (orange line) and bike sharing data (grey line) with 2016 and 2017 bicycle count data respectively. From the plots I can qualitatively conclude that these two models have comparable performance in terms of computing monthly trend on biking trips. To further examine the reliability and robustness of these model, 10-fold cross validations on each regression model by travel mode are conducted. The validation results for are plotted in the following figures.

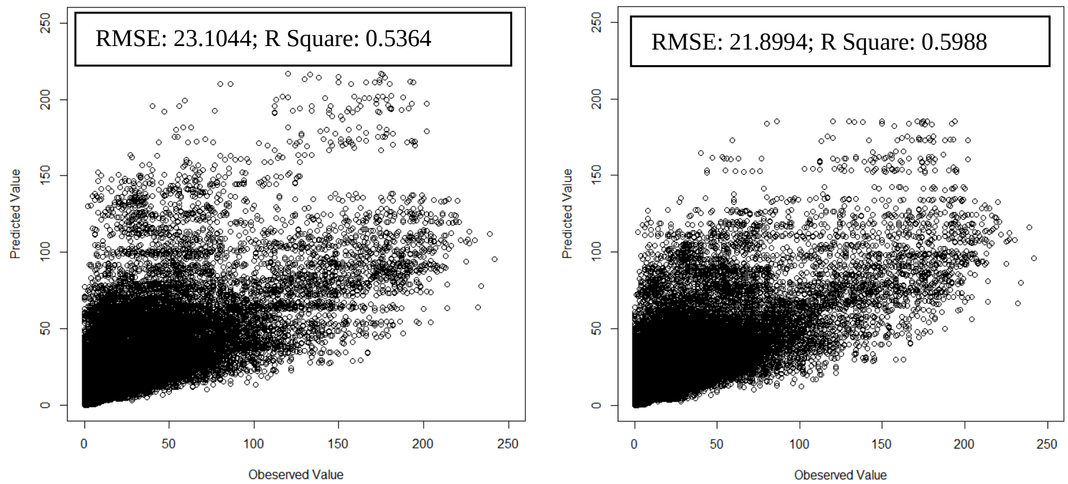


Figure 50 10-folds validation on (left) Negative binomial model and Poisson model (right) using 2017 bicycle count data

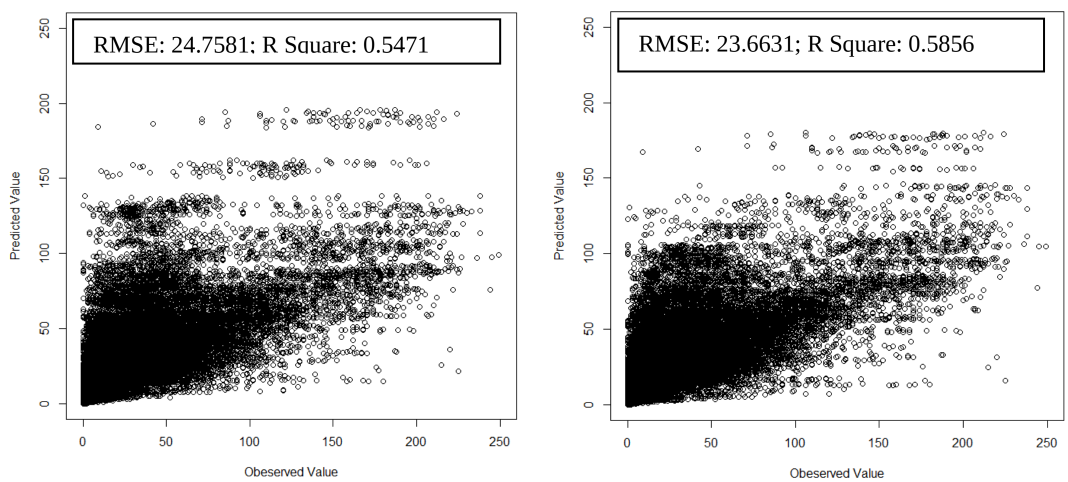


Figure 51 10-folds validation on (left) Negative binomial model and Poisson model (right) using 2016 bicycle count data

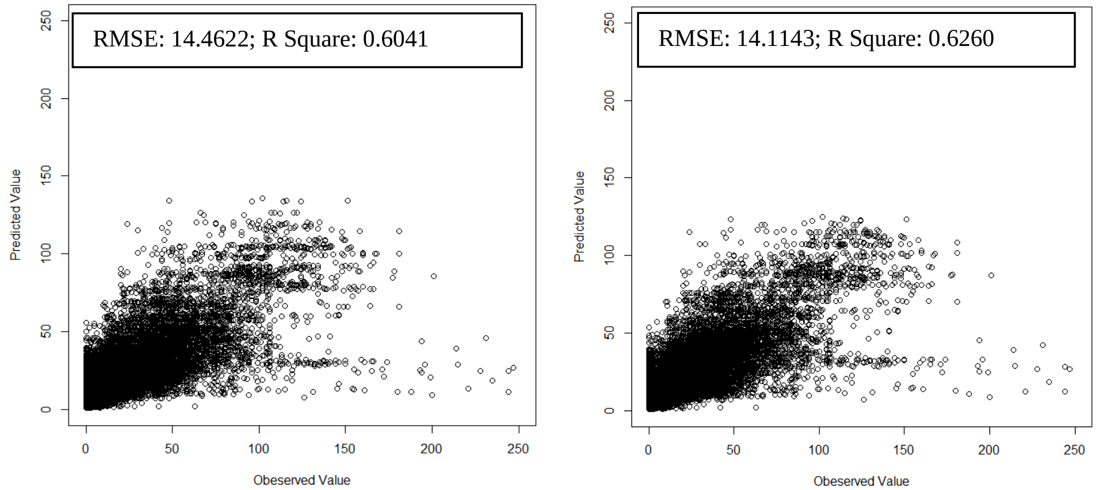


Figure 52 10-folds validation on (left) Negative binomial model and Poisson model (right) using 2017 pedestrian count data

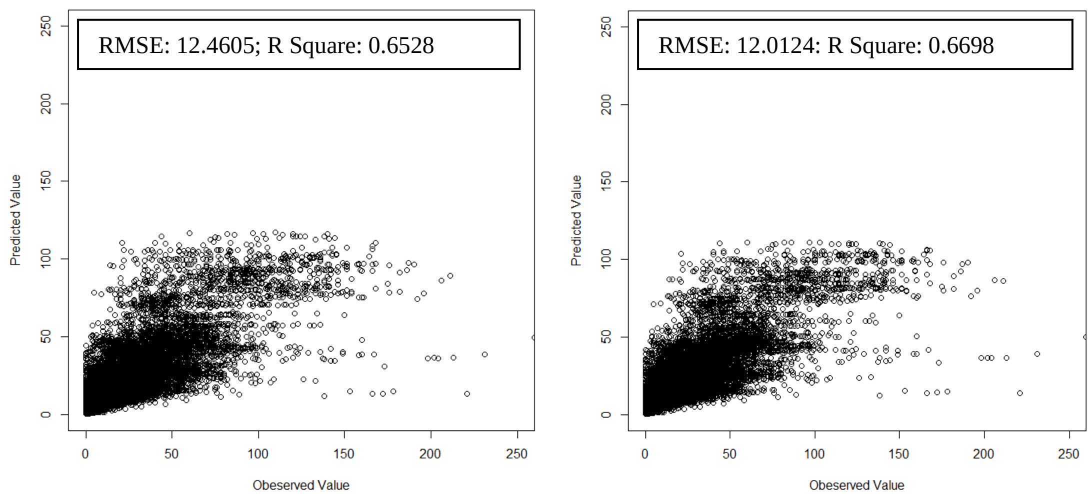


Figure 53 10-folds validation on (left) Negative binomial model and Poisson model (right) using 2016 pedestrian count data

The summarized validation results are visualized in figure 50-53. The x-axis stands for the observed value and y-axis represents value predicted using proposed model. Projected results are valid and reliable from the proposed Multilevel Poisson regression model and Multilevel Negative Binomial model based on the plot. By looking at root mean square error and R square, Poisson regression performs slightly better comparing with Negative binomial regression. The basic model only considers the monthly factor and weekend effect and time period effect. Nevertheless, to further enhance the estimation accuracy, the basic model could take more factors into consideration, such as weather effect, land use characteristics and infrastructure attributes. Another observation can be obtained by comparing bicycle and pedestrian estimation results. Despite of what model is used, walking trips estimation are more accurate than bicycle trips estimation. This could be a consequence of data quality difference. Bicycle count data involves large number counting stations yet with limited data records. Meanwhile, the variation of bicycle count data is greater than the pedestrian count data. There are two possible way of handling this problem. The first one, as discussed above, is to adapt more factors in the model; the second one is only considering the counting station that reports both bicycle and pedestrian trips. However, this could lead to insufficient input to the model because of data downsize.

To test the robustness of the proposed method, RMSE of 10-folds cross validation are compared with the results obtained from full data trained and test. The summarized results are presented in Table 11. Results show that 10% of input data downsize will only cause -2.1% impact at most on the performance, which further proves the robustness of the proposed models.

	2016				2017			
	Biking		Walking		Biking		Walking	
RMSE	NB	Poisson	NB	Poisson	NB	Poisson	NB	Poisson
90% Trained 10% Test	24.7581	23.6631	12.4605	12.0124	23.1044	21.8994	14.4622	14.1143
100% Trained 100%Test	24.7175	23.6161	12.4472	11.7852	23.0505	21.8453	14.1670	13.8739
Performance Impact	-0.164%	-0.199%	-0.107%	-1.928%	-0.234%	-0.248%	-2.083%	-1.732%

Table 11 Results of performance impact of data downsize

To comprehensively understand the model for better application in the future, comparisons between observed and predicted value are conducted at monthly level in different counties by modes. Results are exhibited as follow.

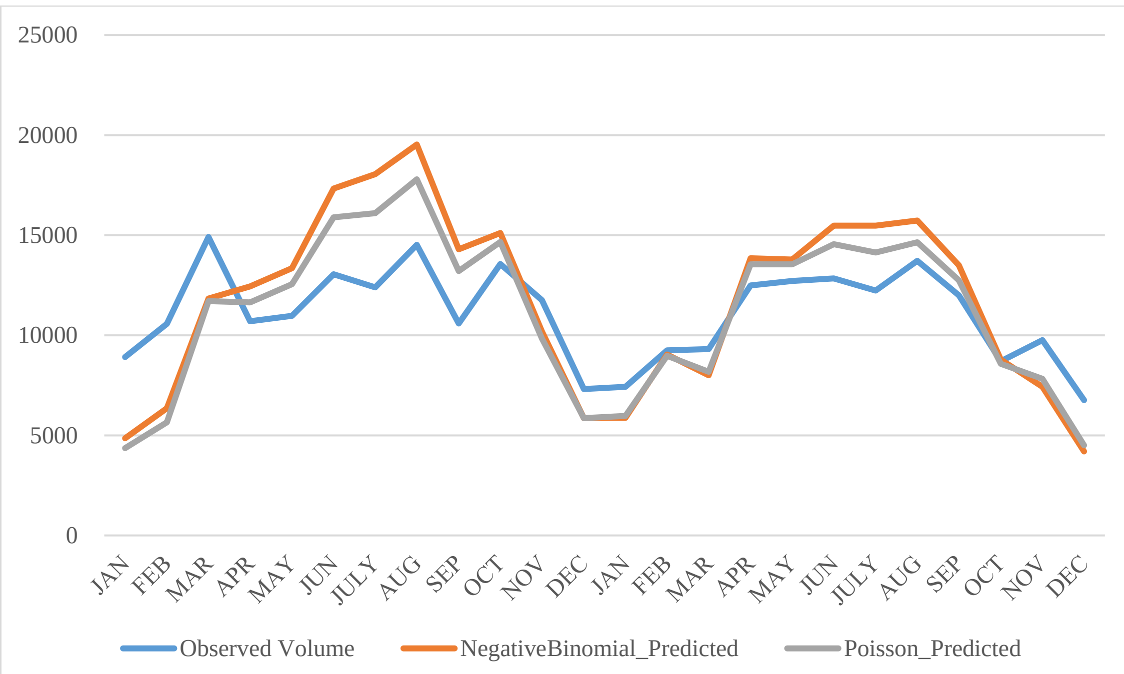


Figure 54 Biking: compare estimated results with observed value in DC from 2016-2017

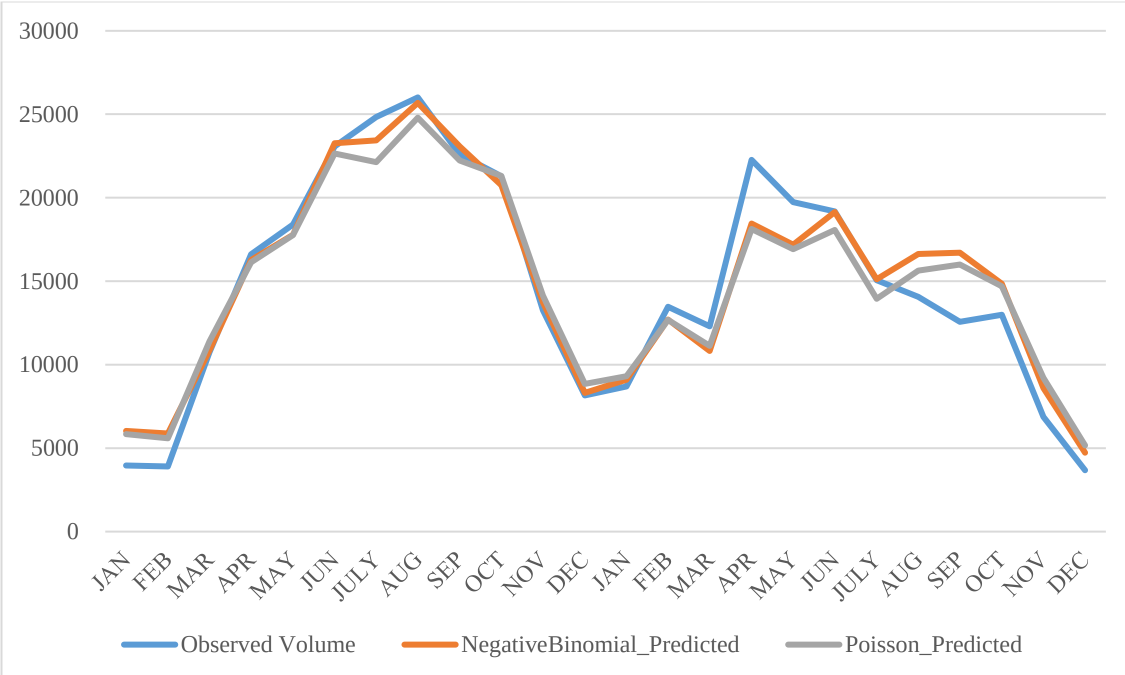


Figure 55 Biking: compare estimated results with observed value in Alexandria from 2016-2017

Figure 56 Biking: compare estimated results with observed value in Arlington from 2016-2017

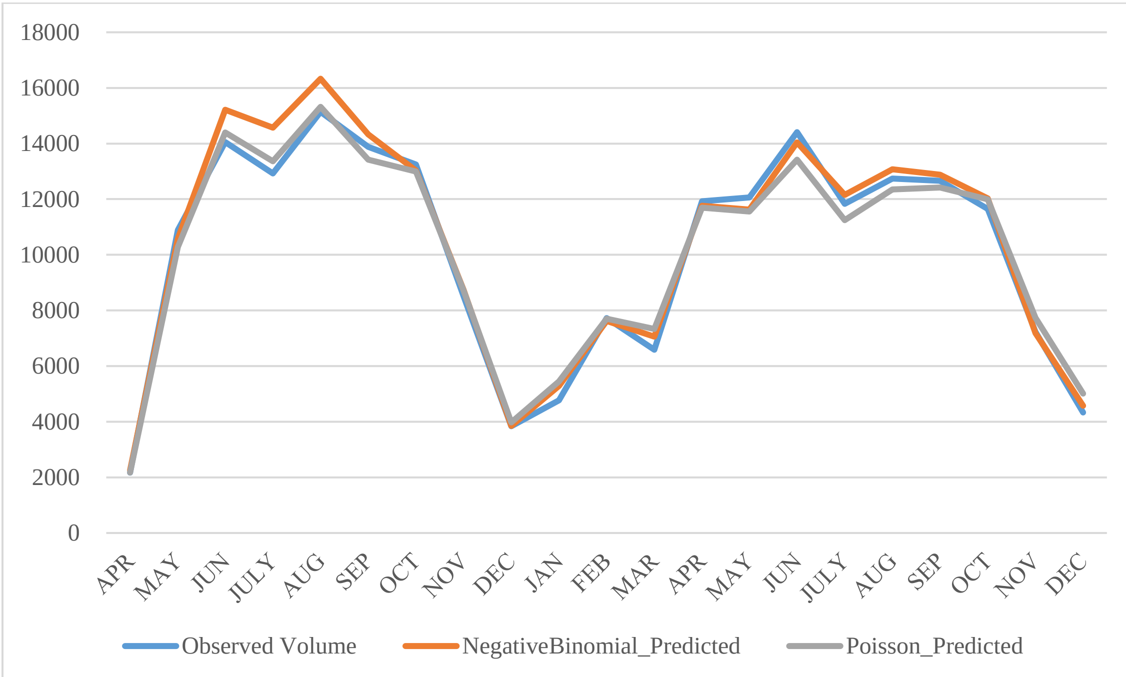


Figure 57 Biking: compare estimated results with observed value in Montgomery from 2016-2017

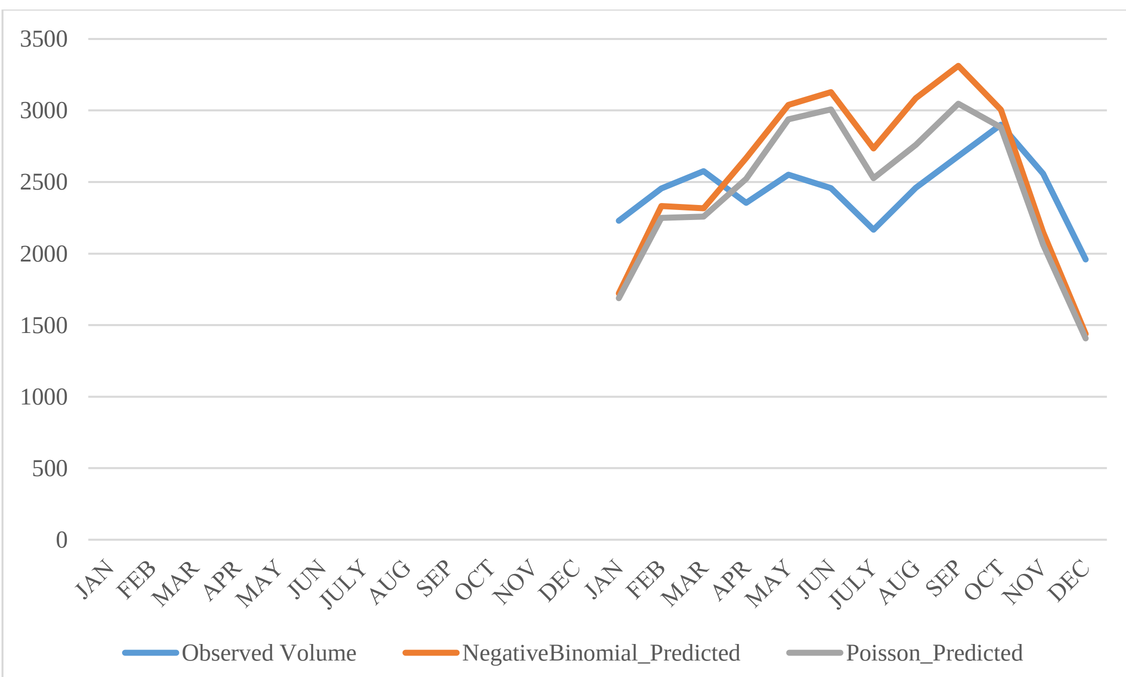


Figure 58 Biking: compare estimated results with observed value in Montgomery from 2016-2017

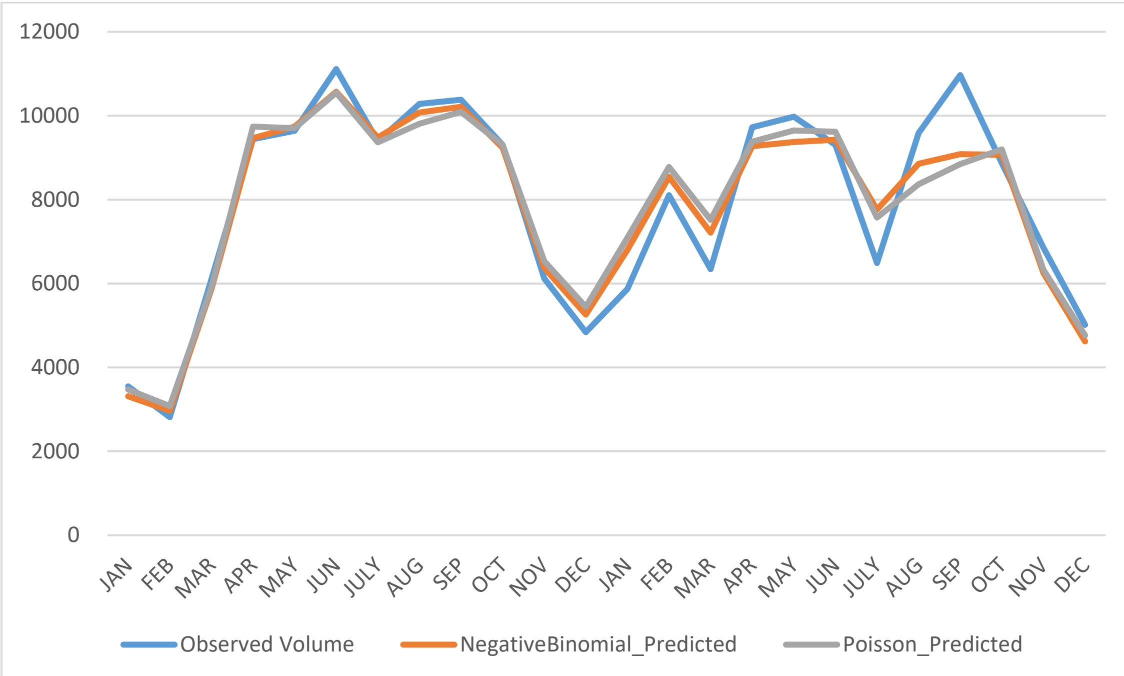


Figure 59 Pedestrian: compare estimated results with observed value in Alexandria from 2016-2017

Figure 60 Pedestrian: compare estimated results with observed value in Alexandria from 2016-2017

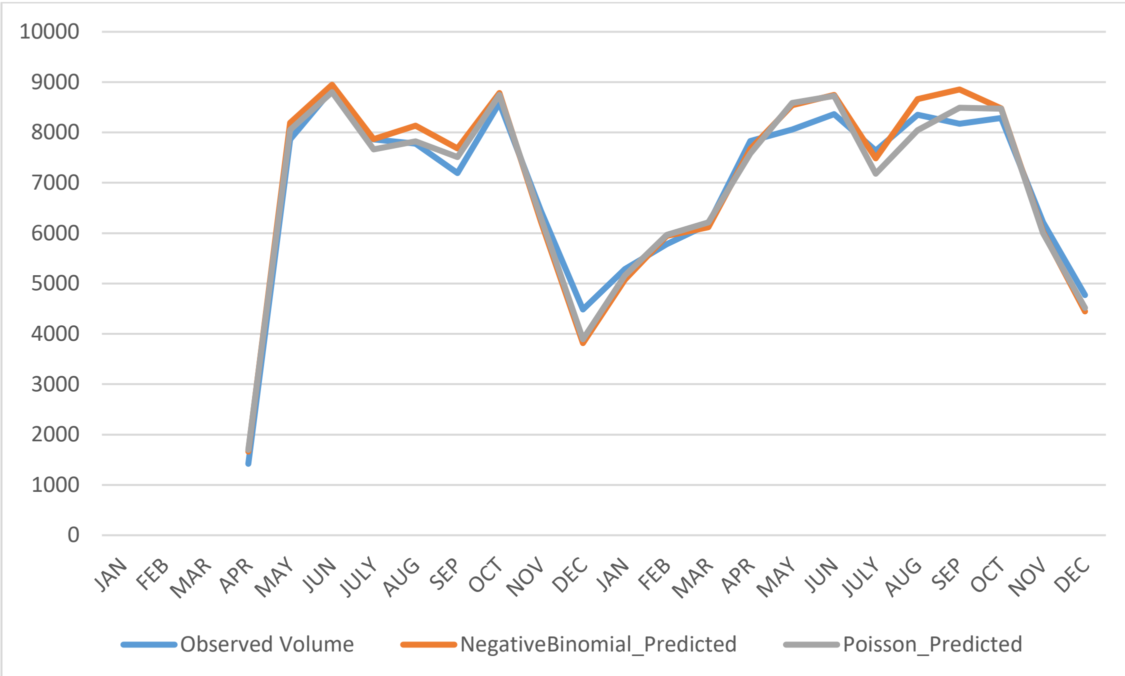


Figure 61 Pedestrian: compare estimated results with observed value in Montgomery from 2016-2017

Figure 54-61 show the results of cross validation of estimated trip volume from different model at monthly level by different counties from 2016 to 2017. In each figure, blue line displays the monthly volume derived from input data; Orange line represents the results estimated by Negative binomial multilevel regression model while grey line shows the results computed by Poisson multilevel regression model. In D.C. MSA, all the automatic counting stations are located at D.C., Arlington county, Montgomery county and Alexandria county. In figure 56, plot in D.C. only shows 2017 results due to insufficient data input. These plots visualize the cross-validation results from which some interesting insights about model performance can be derived. From the plot it is noticeable that the estimated trends from two proposed regression model have strong resemblance with the real trend, except for the results shown in D.C. This discrepancy can be a sequence of the fact that raw data with large variation are provided from only two counting stations in D.C. In this case, some model adjustments and data pre-processing are required when dealing with such dataset. Overall, I can conclude that both regression models are more suitable for application on aggregated level than individual level.

Chapter VI Mode Share Analysis

Trend in mode share plays a critical role in transportation planning and policy making. In additional to travel trend for a particular mode, the mode shift can provide a holistic vision to understand the interaction between all modes for a metropolitan area. Hence, it is important to dynamically monitor the mode shift. However, most travel surveys only report mode share at a particular time point and could not reveal the trends frequent enough to support policy debates due to the data limitation. This project develops a method to monitor the mode shift at the metropolitan level month to month using public domain data by integrating methods developed for individual modes. This section presents the mode share analysis in three metropolitan areas respectively. To evaluate the reliability and credibility of the proposed method, the estimates are compared with results reported by regional travel survey.

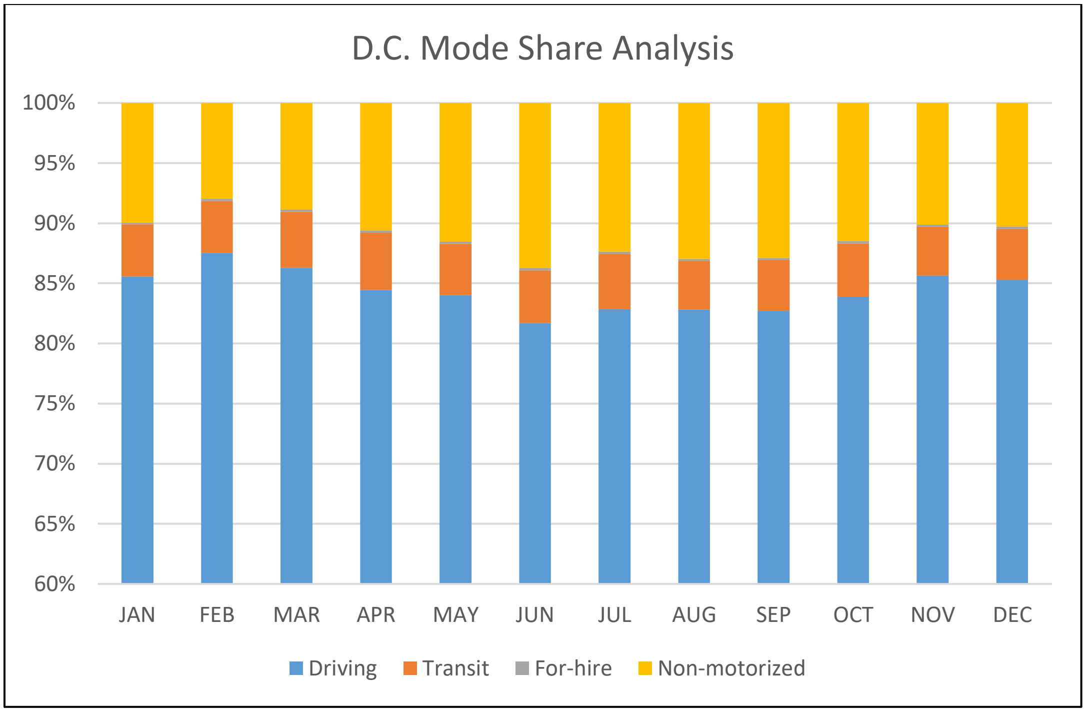


Figure 62 Mode share analysis in D.C. MSA

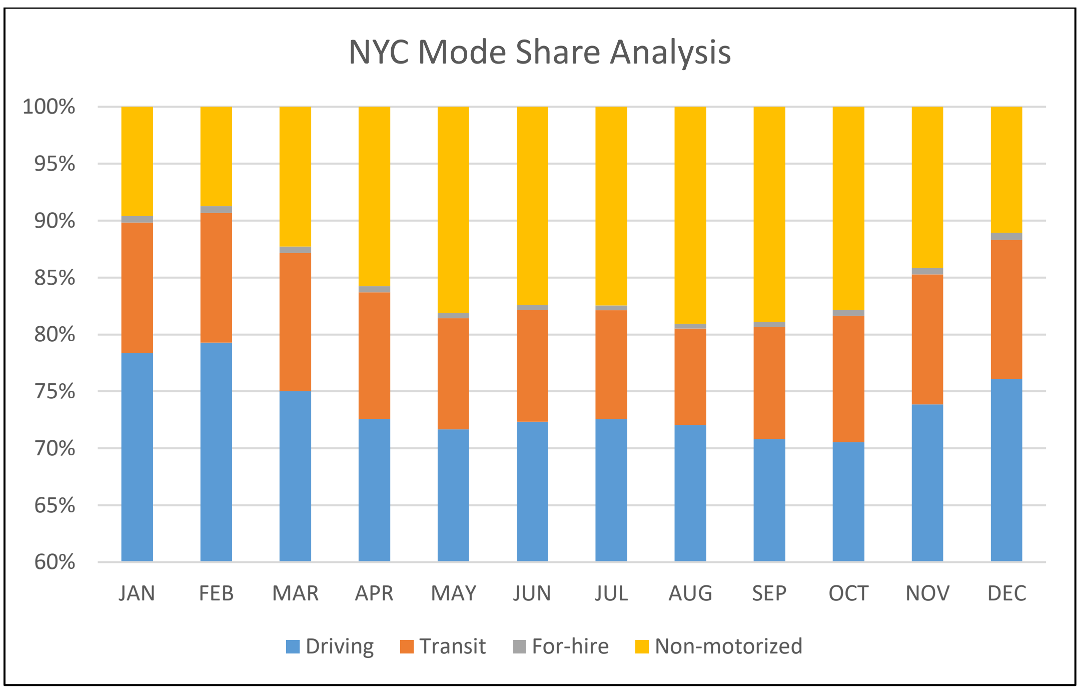


Figure 63 Mode share analysis in NYC MSA

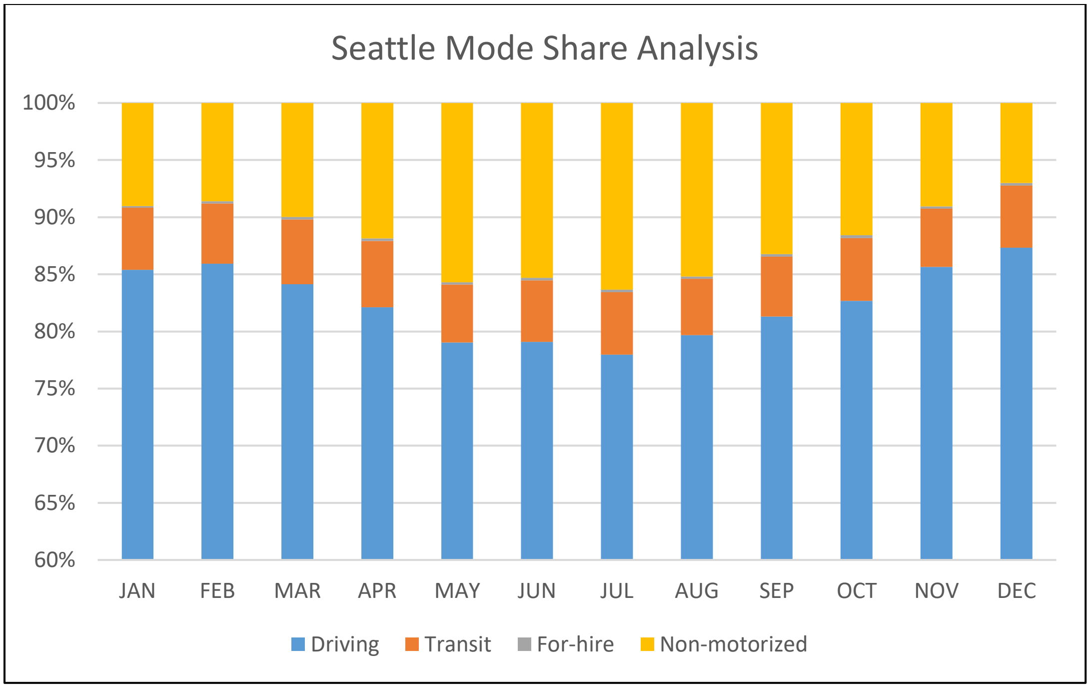


Figure 64 Mode share analysis in Seattle MSA

Figure 62 presents the mode share results for D.C. metropolitan statistical area in 2015. The y-axis range is set from 60% to 100% to improve the readability of the graph. Each bar in the figure represents the mode share for each mode by month. Blue bar represents the mode share for driving mode while orange bar in middle stands for the mode share of transit mode. The mode share of for-hire mode is presented in color gray and the yellow bar is the non-motorized mode share. Driving is still the major travel mode in all three metropolitan areas in the case study, which is presented by blue bar in the plots. However, the mode share of transit differs by cities. The transit share in NYC was about 10 percent overall (in Figure 64) because of the well-developed transit network and denser development while D.C. and

Seattle have on average 7 percent for transit (Figure 62 and 64). The non-motorized mode, including walking and biking, is similar with mode share of transit in three cities. NYC has the highest proportion of walking and biking trips compare with D.C. and Seattle. At winter season, the mode share of driving reaches the peak with the value of 86% in February and it decreases to its lowest point in summer season. This could be explained by the fact that people are less willing to use transit or non-motorized modes in severe weather condition. To enhance the credibility and reliability of the method, the mode share results are validated by comparing the mode share results from the published travel survey report. The comparison is summarized in the table 12, 13 and 14.

Percentage	**Linked**	**Unlinked**	**Survey Results**	**Linked**	**Unlinked**
Driving	84.29%	76.11%	Driving	85.80%	79.07%
Transit	4.37%	3.94%	Transit	5.28%	5.56%
For-hire	0.17%	0.16%	For-hire	0.57%	0.65%
Non-motorized	11.17%	19.79%	Non-motorized	8.35%	14.72%

Table 12 Comparison of D.C. mode share with regional travel survey report

Percentage	**Linked**	**Unlinked**	**Survey Results**	**Linked**	**Unlinked**
Driving	73.33%	61.29%	Driving	68.94%	54.38%
Transit	10.55%	8.82%	Transit	11.73%	12.88%

For-hire	0.50%	0.42%	For-hire	0.98%	0.88%
Non-motorized	14.78%	29.48%	Non-motorized	18.35%	31.86%

Table 13 Comparison of NYC mode share with regional travel survey report

Percentage	**Linked**	**Unlinked**	**Survey Results**	**Linked**	**Unlinked**
Driving	82.31%	80.31%	Driving	80.53%	78.45%
Transit	5.35%	5.22%	Transit	5.62%	5.82%
For-hire	0.21%	0.21%	For-hire	0.54%	0.54%
Non-motorized	12.13%	14.26%	Non-motorized	13.31%	15.20%

Table 14 Comparison of Seattle mode share with regional travel survey report

The estimates from the proposed method are compared with travel survey results at the annual level, which is the finest temporal resolution I can find for results based on travel surveys. In terms of the mode share, they seem to be mostly the same in three cities despite of linked and unlinked trip. Nevertheless, the mode share of for-hire mode reported in travel survey is nearly two times of the estimates from proposed method. This is caused by the fact that the definition of for-hire modes in travel survey is more liberal than what is defined in our method. There are some private carriers of for-hire modes in the regional travel survey reports whose data are not accessible to the public and are thus not included in our proposed method. In future study, additional emerging data sources are needed to complement the public domain data and to eliminate this discrepancy.

Chapter VII Conclusion and Future Work

7.1 Conclusion

This study proposed a comprehensive analytical package for multimodal travel trend monitoring and analysis by integrating a wide range of traffic and travel behavior data sets of multiple modes that are accessible to the public. This study also successfully addresses the data gap and quality issues which are common for many data sources and all modes. The proposed methods have been implemented on the analysis of traffic trends for a particular metropolitan area across all modes for a relatively small-time interval and proved to be effective as long as the target area meets three requirements: being a larger area than county, conducting a basic travel survey and owning a typical count dataset.

The major contribution of this study is three-fold: 1) the proposed method is the first of its kind in estimating multimodal passenger travel behavior across all modes; 2) the proposed analytical package provides monthly measurements on the number of trips and mode share at metropolitan level; and 3) the approach integrates various type of data. The results can be extremely useful in understanding multimodal travel patterns and helping agencies' decision-making processes. The integration of multiple data sources ensures the robustness of the proposed approach and fully utilize the performance of each data set in travel trend monitoring. As I demonstrated in the numerical example for the three metropolitan areas in the case study, the proposed approach produces reasonable and fine-grained multimodal travel trend analysis continuously and is ready to be transferred to other metropolitan areas as well.

This study demonstrates that it is feasible to develop accurate, monthly, multimodal travel trend statistics using public-domain data based on advanced data analytics and modeling methods. To ensure the reliability of the method, some important parameters should be updated routinely. For instance, the average trip length and vehicle occupancy which is enrolled as key parameter in the calculation of number of vehicular trips needs to be updated regularly from the regional travel survey.

However, there are some data gaps that need to be improved in the future. For the total trip estimates, ACS releases data annually but behind schedule and the local household travel survey updates almost every ten years. In addition, both are traditional surveys, the sample of which is limited in size. As the data evolving in recent years, passively collected data such as GPS device data and cell phone location data, can be incorporated into the travel behavior studies to generate better estimates. To further improve the method, it is also necessary to explore additional data sources and build comprehensive data warehouse to both fill current data gaps and further improve monthly travel trend estimates. Since the proposed is mainly focusing on examining the feasibility of using public domain on estimating monthly trips across all modes, the method could be further enhanced by calibration and validation on the proposed assumptions. Another improvement can be made in the count data. It is recommended to fix count locations for non-motorized data collection in different months and to include the weather description at the same time. The locations should be selected to most represent the various biking or walking environments in that city. Although it is not necessary to count every location every month, it is preferable to have balanced observations in the twelve months. In the multilevel model, each group (location) is suggested to have at least 30 observations for a promising result. Based on the recommendation, local agencies can plan the count program considering resources. The above efforts will lead to a more organized count dataset and thus definitely enhance the monthly trend estimates.

Overall, it is recommended that data sources and methods can be implemented to start tracking multimodal travel trends in metropolitan areas across the U.S. month by month. The methods and data sources require periodic updates. This may be further expanded to more cities. Meanwhile, perusing additional data sources and associated methods is required to both fill current data gaps and further improve monthly travel trend estimates.

7.2 Future Work

The conclusion drawn from this proposal indicates that it is feasible to develop accurate monthly and multimodal travel trend statistics using only public-domain data based on the advanced data analytics and modeling methods. Meanwhile, the proposed methodology can still be enhanced by additional exploratory studies on data collection methods and data fusion algorithms in order to extract more details about people's travel behaviors for MSAs. Such details can include origin/destination, mode, trip purpose, trip length/duration, etc. Though they can be obtained through the traditional regional household travel survey, neither the sample size nor the survey cost is considered satisfactory. Hence, the following research topic are proposed as the scope of future works, which would explore the emerging data sources as a parallel and supplement to the existing techniques.

1. Correct sample biases in passively collected travel behavior data.
2. Develop methods to integrate emerging data sources with conventional household travel survey data that would help to improve the accuracy and

timeliness of travel monitoring data, while reducing the required sample size and cost of conventional travel surveys.

3. Validate the proposed methods by testing the sensitivity to data requirements for emerging data sources and the sample size of the conventional household survey data.

REFERENCES

Arana, P., Cabezudo, S., & Peñalba, M. (2014). Influence of weather conditions on transit ridership: A statistical study using data from Smartcards. Transportation research part A: policy and practice, 59, 1-12.

Benekohal, R. F., & Girianna, M. (2002). Evaluation of Methodology for Determining Truck Vehicle Miles Traveled in Illinois (No. ITRC FR 99-2,).

Ben-Akiva, M., & Morikawa, T. (2002). Comparing ridership attraction of rail and bus. Transport Policy, 9(2), 107-116.

Cervero, R., Ferrell, C., & Murphy, S. (2002). Transit-oriented development and joint development in the United States: A literature review. TCRP research results digest, (52).

Cervero R, Murakami J, 2010, "Effects of built environments on vehicle miles traveled: evidence from 370 US urbanized areas" Environment and Planning A 42(2) 400 – 418

Casello, J., Nour, A., Rewa, K., & Hill, J. (2011). An analysis of stated preference and GPS data for bicycle travel forecasting. In 90th Annual Meeting of the Transportation Research Board.

Chen, C. F., & Chao, W. H. (2011). Habitual or reasoned? Using the theory of planned behavior, technology acceptance model, and habit to examine switching intentions toward public transit. Transportation Research Part F: Traffic Psychology and Behaviour, 14(2), 128-137.

Cervero, R., & Landis, J. (1997). Twenty years of the Bay Area Rapid Transit system: Land use and development impacts. Transportation Research Part A: Policy and Practice, 31(4), 309-333.

Cramer, J., & Krueger, A. B. (2016). Disruptive change in the taxi business: The case of Uber. The American Economic Review, 106(5), 177-182.

Correa, D., Xie, K., & Ozbay, K. (2017). Exploring the Taxi and Uber Demand in New York City: An Empirical Analysis and Spatial Modeling (No. 17-05660).

Cramer, J., & Krueger, A. B. (2016). Disruptive change in the taxi business: The case of Uber. The American Economic Review, 106(5), 177-182.

Doi, M., & Allen, W. B. (1986). A time series analysis of monthly ridership for an urban rail rapid transit line. Transportation, 13(3), 257-269.

Ewing, R., & Cervero, R. (2010). Travel and the built environment: a meta-analysis. Journal of the American planning association, 76(3), 265-294.

Federal Highway Administration (2005). Highway Performance Monitoring System Field Manual. FHWA, U.S. Department of Transportation.

Frank, L., Bradley, M., Kavage, S., Chapman, J., & Lawton, T. K. (2008). Urban form, travel time, and cost relationships with tour complexity and mode choice. Transportation, 35(1), 37-54

Griswold, J., Medury, A., & Schneider, R. (2011). Pilot models for estimating bicycle intersection volumes. Transportation Research Record: Journal of the Transportation Research Board, (2247), 1-7.

Goodwin, P. B. (1992). A review of new demand elasticities with special reference to short and long run effects of price changes. Journal of transport economics and policy, 155-169.

Gosse, C., & Clarens, A. (2014). Estimating spatially and temporally continuous bicycle volumes by using sparse data. Transportation Research Record: Journal of the Transportation Research Board, (2443), 115-122.

Hall, J. V., & Krueger, A. B. (2016). An analysis of the labor market for Uber's driver-partners in the United States (No. w22843). National Bureau of Economic Research.

Hickey, R. (2005). Impact of transit fare increase on ridership and revenue: metropolitan transportation authority, New York city. Transportation Research Record: Journal of the Transportation Research Board, (1927), 239-248.

Hankey, S., Lindsey, G., Wang, X., Borah, J., Hoff, K., Utecht, B., & Xu, Z. (2012). Estimating use of non-motorized infrastructure: Models of bicycle and pedestrian traffic in Minneapolis, MN. Landscape and Urban Planning, 107(3), 307-316.

Strauss, J., & Miranda-Moreno, L. (2013). Spatial modeling of bicycle activity at signalized intersections Jillian Strauss, Luis F Miranda-M.

Hudson, J. G., J. C. Duthie, K. Yatinkumar, K. A. Larsen, and J. L. Meyer. 2012. Using smartphones to collect bicycle travel data in Texas final report. College Station, TX: Texas A&M Transportation Institute.

Jovanis P P, Chang H L. Modeling the relationship of accidents to miles traveled [J]. Transportation Research Record, 1986, 1068: 42-51.

Kumapley, R., & Fricker, J. (1996). Review of methods for estimating vehicle miles traveled. Transportation Research Record: Journal of the Transportation Research Board, (1551), 59-66.

Kuzmyak, J. R. (1981). Usefulness of existing travel Data Sets for Improved VMT Forecasting. Transportation Research Record, (815).

Kashfi, S. A., Bunker, J. M., & Yigitcanlar, T. (2015). Effects of transit quality of service characteristics on daily bus ridership. Transportation Research Record: Journal of the Transportation Research Board, (2535), 55-64.

Kim, S., Ulfarsson, G. F., & Hennessy, J. T. (2007). Analysis of light rail rider travel behavior: impacts of individual, built environment, and crime characteristics on transit access. Transportation Research Part A: Policy and Practice, 41(6), 511-522.

Liu, F., & Kaiser, R. G. (2006). Forecasting Truck VMT Growth at the County and Statewide Levels. In National Urban Freight Conference.

Litman, T. (2008). Valuing transit service quality improvements. Journal of Public transportation, 11(2), 3.

Lewin, A. (2011). Temporal and weather impacts on bicycle volumes. In Transportation Research Board 90th Annual Meeting (No. 11-2536).

Nordback, K., Marshall, W., Janson, B., & Stolz, E. (2013). Estimating annual average daily bicyclists: Error and accuracy. Transportation Research Record: Journal of the Transportation Research Board, (2339), 90-97.

Nordback, K., & Sellinger, M. (2014). Methods for estimating bicycling and walking in Washington state (No. WA-RD 828.1).

Polydoropoulou, A., & Ben-Akiva, M. (2001). Combined revealed and stated preference nested logit access and mode choice model for multiple mass transit technologies. Transportation Research Record: Journal of the Transportation Research Board, (1771), 38-45.

Pickrell, D. H. (1989). Urban rail transit projects: forecast versus actual ridership and costs. Final Report (No. PB-90-148693/XAB). Transportation Systems Center, Cambridge, MA (USA).

Phung, J., & Rose, G. (2007, September). Temporal variations in usage of Melbourne's bike paths. In Proceedings of 30th Australasian Transport Research Forum, Melbourne.

Rentziou, A., Gkritza, K., & Souleyrette, R. R. (2012). VMT, energy consumption, and GHG emissions forecasting for passenger transportation. Transportation Research Part A: Policy and Practice, 46(3), 487-500.

Ryan, S., & Frank, L. F. (2009). Pedestrian environments and transit ridership. Journal of Public Transportation, 12(1), 3.

Sharaby, N., & Shiftan, Y. (2012). The impact of fare integration on travel behavior and transit ridership. Transport Policy, 21, 63-70.

Singhal, A., Kamga, C., & Yazici, A. (2014). Impact of weather on urban transit ridership. Transportation research part A: policy and practice, 69, 379-391.

Schneider, R., Arnold, L., & Ragland, D. (2009). Pilot model for estimating pedestrian intersection crossing volumes. Transportation Research Record: Journal of the Transportation Research Board, (2140), 13-26.

Taylor, B. D., Miller, D., Iseki, H., & Fink, C. (2009). Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. Transportation Research Part A: Policy and Practice, 43(1), 60-77.

Thompson, G., & Brown, J. (2006). Explaining variation in transit ridership in US metropolitan areas between 1990 and 2000: multivariate analysis. Transportation Research Record: Journal of the Transportation Research Board, (1986), 172-181

Taylor, B. D., & Fink, C. N. (2003). The factors influencing transit ridership: A review and analysis of the ridership literature. University of California Transportation Center.

Usdot, U. S. (2013). Department of Transportation, 2012. National Transportation Statistics. Research and Innovative Technology Administration. Bureau of Transportation Statistics. Washington, DC

Wang, G. H., & Skinner, D. (1984). The impact of fare and gasoline price changes on monthly transit ridership: empirical evidence from seven US transit authorities. Transportation Research Part B: Methodological, 18(1), 29-41.

Wang, S. G., Bai, L., & Bao, Y. (2011, August). Modeling and Algorithms of VMT and AADT Estimation for Community Area Traffic Networks. In IFAC World Congress 2011.

Wallsten, S. (2015). The competitive effects of the sharing economy: how is Uber changing taxis? Technology Policy Institute, 22.