Monitoring Multi-modal Travel Demand Month-by-month through Data Fusion and Integration

Zhuo Yang[1]
Sid and Reva Dewberry Department of Civil, Environmental and Infrastructure Engineering
George Mason University
4400 University Drive, MS 6C1 Fairfax, VA 22030-4444
Tel: (703) 993-1797 Email: zyang10@gmu.edu

Bo Peng[2]
Department of Civil and Environmental Engineering at the University of Maryland, College Park
1173 Glenn Martin Hall, University of Maryland, College Park, MD, 20742
Email: bp105@terpmail.umd.edu

Yixuan Pan[2]
Department of Civil and Environmental Engineering at the University of Maryland, College Park
1173 Glenn Martin Hall, University of Maryland, College Park, MD, 20742
Email: ypan1003@umd.edu

Shanjiang Zhu, Ph.D.[1]
Sid and Reva Dewberry Department of Civil, Environmental and Infrastructure Engineering
George Mason University
4400 University Drive, MS 6C1 Fairfax, VA 22030-4444
Tel: (703) 993-1797 Email: szhu3@gmu.edu

Lei Zhang[2*]
Department of Civil and Environmental Engineering at the University of Maryland, College Park
1173 Glenn Martin Hall, University of Maryland, College Park, MD, 20742
Email: lei@umd.edu

*Corresponding author

Abstract:

Conventional travel behavior data collection methods such as the National Household Travel Survey (NHTS) 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. 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 paper develops a package of methods that are tailored to data of different quality for different modes in the public domain, and can collectively reveal month-to-month travel trends dynamically in a metropolitan area. The proposed methods will be demonstrated through case studies in three different metropolitan areas. A comparison with mode split trend based on household survey data collected in the same metropolitan area showed the effectiveness of the proposed method. Future studies will further address the data gap and reliability issue.

Key words: *Multi-modal travel demand, mode split, data fusion and integration*

INTRODUCTION

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 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 fill this gap in research and practice, this paper will develop a package of methods that are tailored to data of different quality for different modes in the public domain, and can collectively reveal month-to-month travel trends dynamically in a metropolitan area. The proposed methods will be demonstrated through case studies in three different metropolitan areas.

The next section reviews the state of the art academic studies and the state of practice on the related topics. The paper will then present the data to be used and the proposed methodology step by step. Issues with potential data quality and availability will be discussed and additional modules are developed to enhance the robustness of the proposed methods. Finally, the proposed methods will be demonstrated through case studies in three different metropolitan areas.

LITERATURE REVIEW

Different measures have been used to gauge travel demand of different modes. For example, Vehicle Miles Traveled (VMT) has been used worldwide by transportation and planning agencies for various applications related to driving, including urban/rural mobility, highway safety, fuel consumption, economic level, and environmental quality. 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 (Jovanis, 1986). 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 at a monthly level or a rigorous estimation process based on existing data sources are not available in practice.

Similar trend has also been seen for transit. Most studies focused on a few factors that may affect transit ridership of a specific transit operator, or a specific geographic area. For example, Kain 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 and Chao (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. These studies showed the fundamental importance of monitoring transit ridership in a metropolitan area (ridership of different operators in the same metropolitan area may substitute one another) month by month for investment decisions and policy, however, no studies tried to develop a robust ridership monitoring program in a metropolitan area in a monthly basis.

Most cities report ridership of taxi and escort services that they oversee. However, emerging mobility-on-demand start-ups such as Uber and Lyft are growing in popularity with urban travelers (Cramer and Krueger, 2016) and can quick overtake traditional taxi services by the number of customers served. 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 spatio-temporal patterns of the demand for Uber and taxi at a Neighborhood Tabulation Area (NTA) level using data from NYC.

Monitoring travel demand of non-motorized modes, including biking and walking, is even more challenging. Non-motorized travel demand data, such as the number of bicycles or pedestrians, are usually sparse since they are only collected in selected locations during certain time periods (usually morning peak or evening peak). Most existing studies only focused on monitoring travel trend of non-motorized modes at one particular location (e.g. Phung and Rose, 2007, Schneider et al. 2009, Griswold et al. 2011, Lewin 2011, Hankey et al. 2012, Strauss and Miranda-Moreno 2013). Robust statistical methods are need to address the challenges of sparse data and scale estimates at a few locations to a larger metropolitan area.

In short, many studies in the literature 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 paper intends to fill this research gap.

METHDOLOGY

This section discusses the method to derive monthly travel demand in a metropolitan and the data to be used by mode. With demand of all four modes (driving, transit, for-hire, and non-motorized) estimated, the monthly trend in mode split for a metropolitan area can be easily calculated and monitored continuously.

Driving

Highway Performance Monitoring System (HPMS) is a federal program that monitors the condition, performance and characteristics of nation's highways. To meet the data need for reporting to 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. An additional data program, Travel Monitoring Analysis System (TMAS), provides volumes of truck traffic for highways in the U.S. HPMS data only reports annual VMTs. To derive monthly VMT, additional analysis is needed to derive adjustment factors based on raw ATR and TMAS data to convert annual VMTs into monthly passenger vehicle VMTs in a metropolitan area. Moreover, to convert VMTs into the number of trips, two additional factors, the average trip distance by driving and vehicle occupancy in a metropolitan area, need to be estimated using the latest Household Travel Survey Data and applied. The following figure summarizes the major steps to calculate the number of driving trips monthly for a metropolitan area.

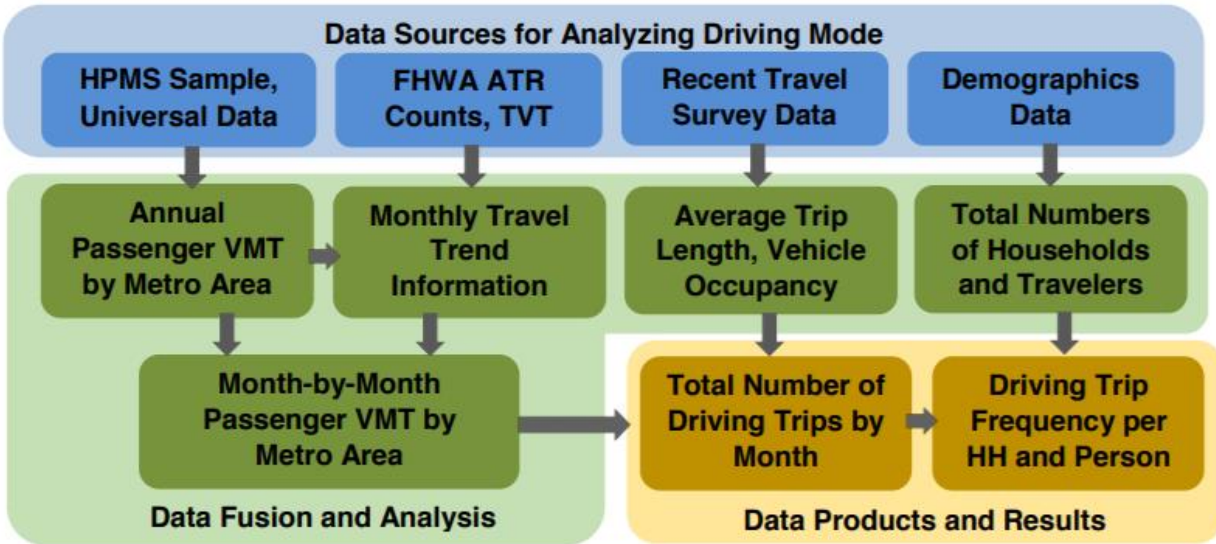

Figure1: Computational flow chart of driving trips estimation

Transit

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, including the Unlinked Passenger Trips (UPTs). Since 2002, large transit operators are required to report up-to-date time series of monthly UPTs, while small operators only report ridership data annual. Therefore, methods are needed to allocate annual ridership of small transit operators to different months. In addition, some large transit operators may serve a geographic area that covers more than one metropolitan areas. For example, the MARC train in Maryland serves both the Washington D.C. and the Baltimore metropolitan areas. And these operators usually do not report their ridership by transit line, or by metropolitan areas. 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.

This paper will demonstrate the method using the MARC Train system in Maryland (Figure 2). 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, we 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. Following these assumptions, 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.

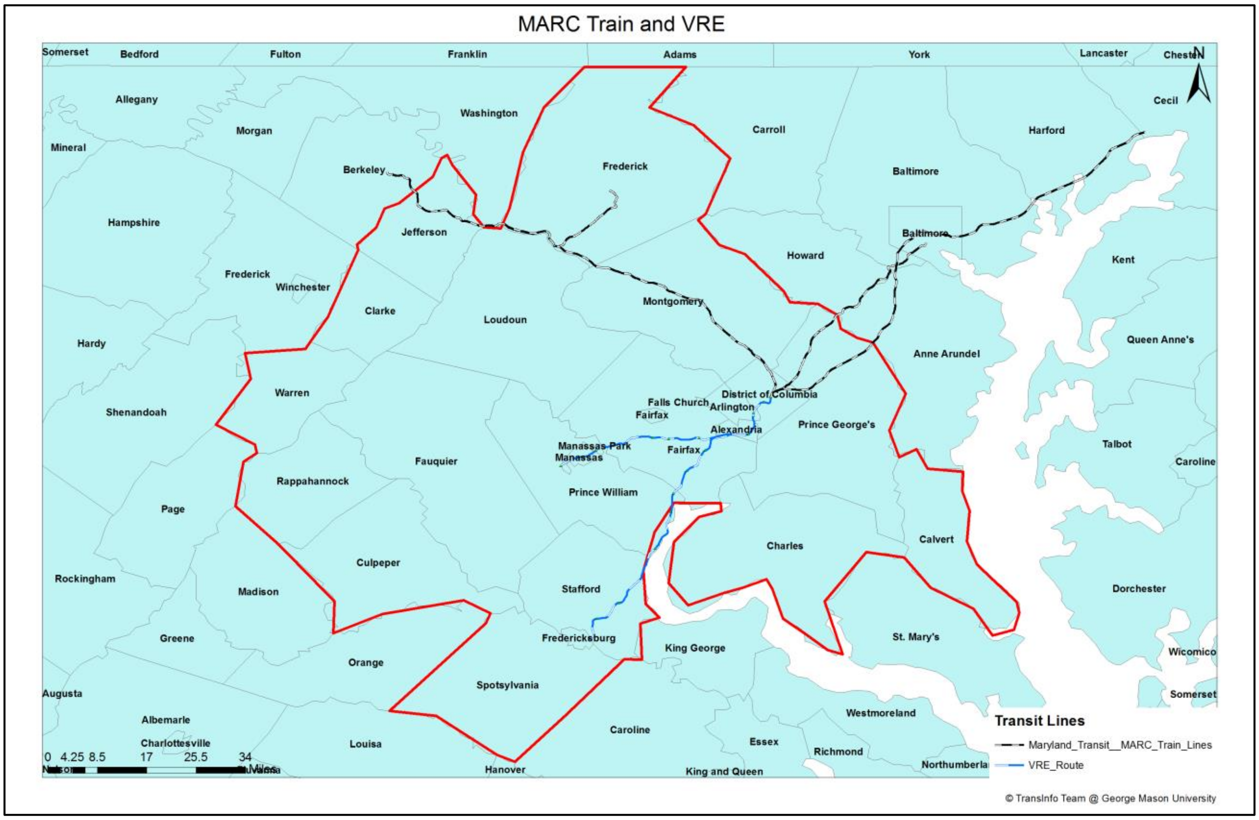

Figure 2: MARC Train and Virginia Rail Express (VRE) network

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%.

With these issues addressed, we can calculate the total number of transit trips by cross-referencing the network of service providers and the boundaries of a MSA.

For-hire modes

For-hire modes in this study include taxi, ride-hailing services such as Uber and Lyft, and airport shuttles. Ridership by taxi in a metropolitan area is usually collected by local jurisdictions and is accessible by the public. However, ridership by ride-hailing services are usually not available because it belongs to the private sector. NYC is the only exception where monthly ridership by Uber and Lyft can be made available through the Freedom Of Information Act (FOIA). Figure 3 shows the market share of taxi, Uber, and Lyft in NYC derived from monthly ridership data. Without better information sources, the number of monthly trips using for-hire modes can only be calculated by assuming the trend in market share is consistent in all major metropolitan areas.

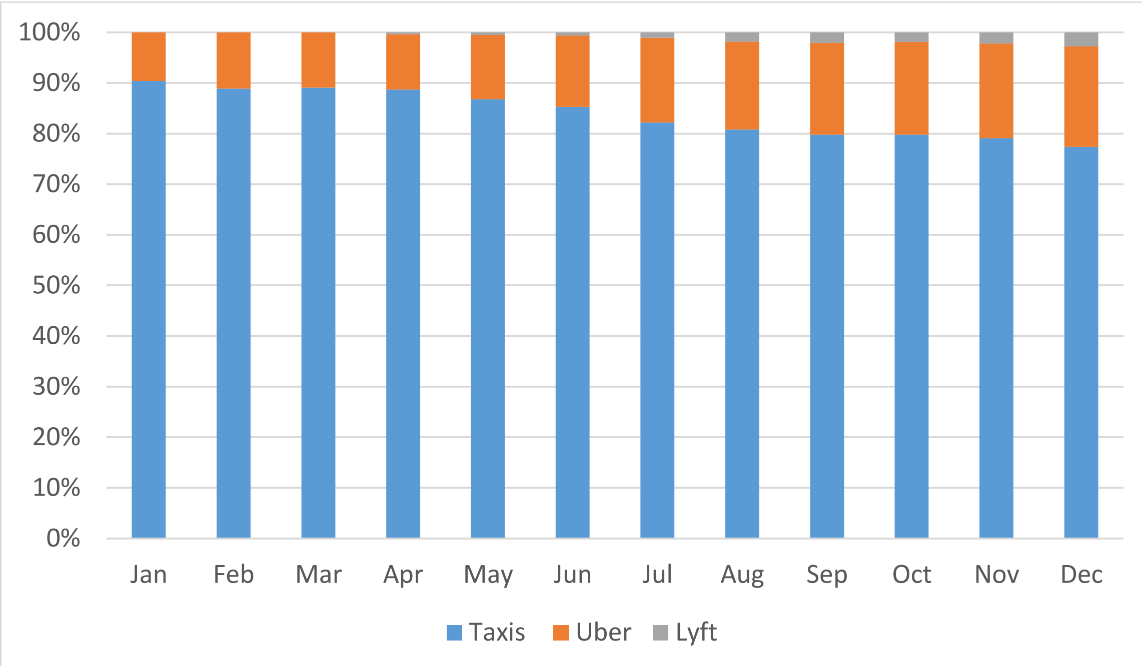

Figure 3: Column Chart of NYC Market Share in 2015

Trips by airport shuttles are usually not directly reported. To fill this data gap, the research team integrated USDOT DB1B data, which provides an estimate of the number of visitors at each airport, and ground transportation data that a few airports did report. For example, Figure 4 shows the mode split by modes of ground transportation that the BWI airport reported, while Table 1 shows the total number of passengers estimated from DB1B data for each quarter. The monthly number of airport shuttle trips at each airport in the Washington D.C. area can be estimated by assuming the same trend in ground transportation mode split applies to all three airports and the number of airport passengers can be equally split among the three months in each quarter.

Table 1: Populated Total Airport Passengers Per Quarter In Washington D.C. Airports

Quarter	BWI	DCA	IAD
1	1,563,270	1,878,220	803,230
2	2,051,040	2,386,700	1,049,600
3	2,022,900	2,283,790	1,009,750
4	1,981,180	2,355,180	950,610
Total	7,618,390	8,903,890	3,813,190

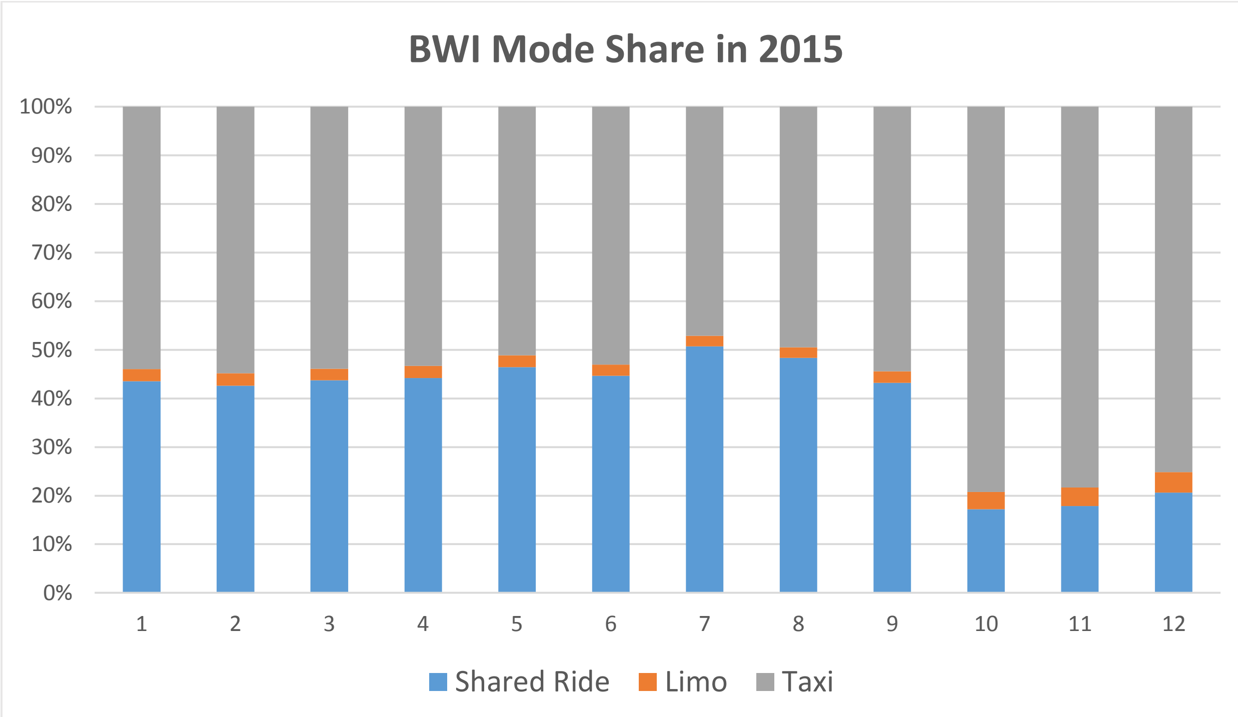

Figure 4: Monthly mode shares in Baltimore-Washington airport

The total number of trips by for-hire modes can then be estimated by adding up the trip by each mode in each month.

Non-motorized modes

The major challenge for estimating the monthly number of trips by non-motorized mode is the sparseness of counting data. As the limited number of counting locations is unlikely to be sufficient to provide the overall picture of demand for non-motorized modes, the American Community Survey (ACS) and the regional household travel survey data are used to provide an estimate of the total number of non-motorized trips in the base year when the latest survey data is available. A Poisson Multilevel Model (PMM), which is particularly efficient to address the sparsity and scarcity in count data, is developed to estimate the monthly trend of non-motorized trips using counting data collected in each month. The PMM model includes two parts that need to be jointly estimated:

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

$$\log\big(E(Y|x)\big) = \log(\lambda) = \theta x$$

Where y is the hourly bicycle volume that is assumed to follow Poisson distribution with a mean of λ. λ depends on a set of the exogenous variables x and can be estimated through a multi-level linear regression model that controls the spatial correlation. The exogenous variables include month of the year, year index, and weather information. Additional variables that affect travel demand such as population and employment density, availability of biking and pedestrian facilities, and other building environment are implicitly controlled by the location specific intercept. However, they could also be explicated treated if more data becomes available.

Mode split

With the monthly number of trips by the four modes estimated using methods tailored to each mode, the mode split trend for a metropolitan area can then be easily estimated and monitored over time.

CASE STUDIES

Figure 5-7 presents the mode share results for the D.C., NYC, and Seattle metropolitan statistical area in 2015, respectively. 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 6) because of the well-developed transit network and denser development while D.C. and Seattle have on average 7 percent for transit (Figure 5 and 7). 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 2, 3 and 4.

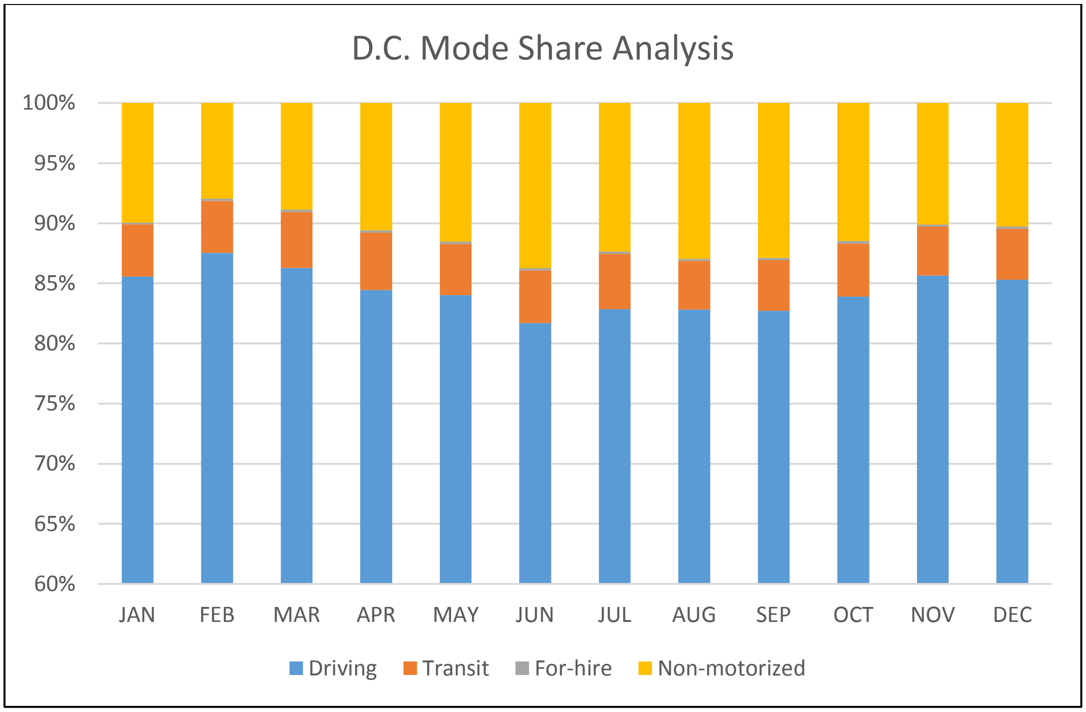

Figure 5: Mode share analysis in D.C. MSA

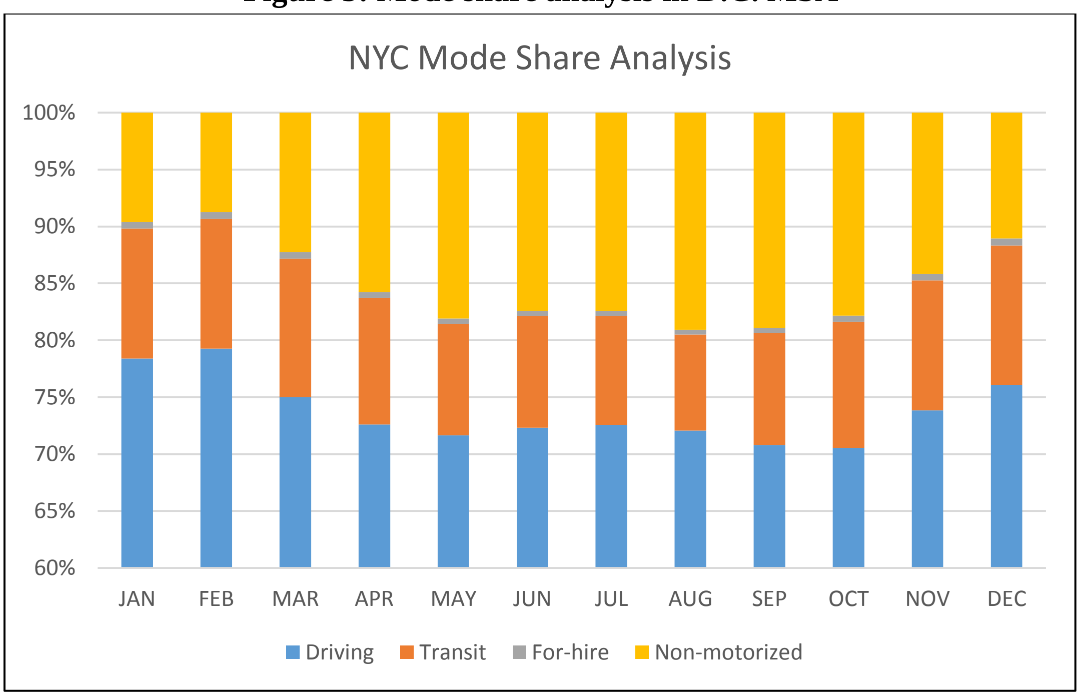

Figure 6: Mode share analysis in NYC MSA

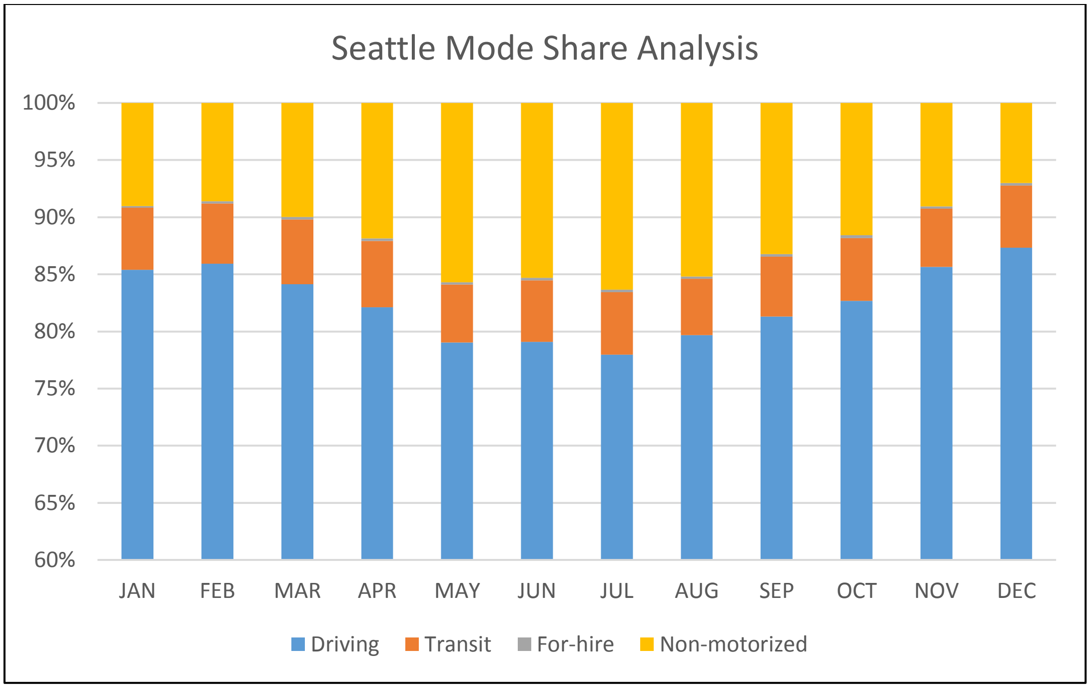

Figure 7: Mode Share Analysis in Seattle MSA

Table 2: D.C. mode share comparison with regional travel survey report

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 3: NYC mode share comparison 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 4: Seattle mode share comparison 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%

The estimates from the proposed method are compared with travel survey results at the annual level, which is the finest temporal resolution we 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 address this discrepancy.

CONCLUSIONS

This paper 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.

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 we 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.

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 of them are usually 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.

ACKNOWLEDGEMENT

This study has been supported by the Federal Highway Project: DTFH6116C00001, Advancing Travel Behavior Information Gathering through Public Domain Data. Contents of this paper do not represent the view of any government agencies and are the sole responsibility of the authors.

REFERENCES

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

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.

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).

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.

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.

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.

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

Kain, J. F., & Liu, Z. (1999). Secrets of success: assessing the large increases in transit ridership achieved by Houston and San Diego transit providers. Transportation Research Part A: Policy and Practice, 33(7), 601-624.

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

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

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.

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

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