Project ID: NTC2014-SU-R-21

U.S. NATIONAL AND INTER-REGIONAL TRAVEL DEMAND ANALYSIS: PERSON-LEVEL MICROSIMULATION MODEL AND APPLICATION TO HIGH-SPEED RAIL DEMAND FORECASTING

Final Report

by

Lei Zhang
University of Maryland, College Park

Yijing Lu
University of Maryland, College Park

for

National Transportation Center at Maryland (NTC@Maryland)
1124 Glenn Martin Hall
University of Maryland
College Park, MD 20742

January 2015

ACKNOWLEDGEMENTS

This project is financially supported by the National Transportation Center (NTC) @ Maryland. The authors are solely responsible for all statements in this report.

DISCLAIMER

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXCUTIVE SUMMARY

The objective of this proposed research project is to develop a prototype microsimulation-based national and inter-regional passenger travel demand model for High Speed Rail demand forecasting and other national-level travel analysis. The proposed research represents the first attempt to develop a microsimulation-based national long-distance travel demand for high speed rail and national travel analysis. All major behavioral dimensions of long-distance travel will be considered, except for route choice and network loading that require significant new network data collection/coding efforts and cannot be achieved with the limited budgeted of this project. Compare to the traditional four-step approach, microsimulation-based techniques offer several advantages: (1) It is easier to consider tours, multi-day and multi-stop trips, and intermodal access/egress transfers that are important for long-distance travel modeling; (2) Households and persons are the basic units of analysis, which enables detailed behavioral representations and interactions; and (3) It provides a rich framework in which travel is analyzed as a multi-day, monthly, quarterly, or yearly pattern of behavior, derived from activity participation. There are also significant differences between long-distance trips considered in the proposed microsimulation-based model and trips on a daily/weekly basis represented in metropolitan/state-level tour/activity-based models developed in previous research. For instance, it is often the case that households first choose travel modes for long-distance vacation trips based on travel budget before selecting destinations. Categorization of trip purposes is also different for long-distance trips. Cost of travel for long-distance trips is not just travel disutility, but also includes lodging, food, etc., and the same with the total travel time for long-distance which usually covers not only in-vehicle travel time but also the ingress/egress time, transfer time, and lodge time. The much lower frequency of long-distance travel may also imply a different decision-making process. This research is exploratory in nature, and it is hoped that the final product, the prototype microsimulation-based model, will be able to predict high speed rail travel demand among various OD pairs at the national level.

1.0 INTRODUCTION

The increasing interest of national transportation policies from strategic infrastructure investment to infrastructure operation and management with regard to efficiency, sustainability, and safety has attracted researchers and decision makers to call for advanced and policy-sensitive tools for analysis *(Lundqvist & Mattsson, 2001)*. The increase of national travel also requires the analysis tools beyond the urban and regional level. The highway infrastructure investment, the high-speed rail and the airport development all depend on national travel markets. To ensure that the infrastructure meets the demand growth it is imperative to model and analyze the passenger travel behavior at the national level *(FHWA, 2013)*.

Americans travel a lot including inter-city, interregional and international travel. According to the 1995 American Travel Survey on the long-distance travel of persons in the U.S. (Bureau of Transportation Statistics, BTS), the U.S households made over one billion national-level long-distance trips and 41 million international trips *(Zhang et al., 2012)*. National long-distance trips in the U.S can be of various purposes including business, leisure, personal business, family or friend visit and so on. All of the long-distance activities could constitute the economic and recreational opportunities that would benefit both the person and the area where the long distance activities occur. Thus, it is essential for the U.S from the economic and social perspectives to have the capability to support high level personal long-distance travel, which requires that we have sufficient data and accurate analysis tools to be able to understand the long-distance travel behavior and forecast the travel patterns in the future. Without the analysis tools we could risk making inefficient and costly investments in our transportation infrastructure and management.

The needs for analyzing transportation capital expenditure decisions at the national level in the 1970s led to two U.S. National Transportation Studies (NTS) in 1972 and 1974 respectively *(Weiner, 1976)*. These early national travel studies inventoried existing and planned U.S. transportation systems; and estimated future travel demand, system costs, performance, and broader impacts under alternative funding scenarios. With the completion of major investments on the Interstate Highway System, the development of national-level long-distance passenger travel analysis tools in the U.S. has been stagnant since the 1970s, though there have been continual academic interests in improving the theory and methods for multimodal intercity passenger travel demand analysis with a focus on mode choice *(Lundqvist & Mattsson, 2001;Koppelman & Sethi, 2005;Bhat, 1995;Winston, 1985;Mannering, 1983)*. The lack of a capable long-distance passenger travel analysis tool in the U.S. is in sharp contrast with important emerging needs for analyzing various national transportation policies related to long-distance passenger travel. The Obama administration allocated $8 billion in the 2009 stimulus funds for high-speed passenger rail, hoping that the supertrains would operate throughout the American landscape as they do in Europe and Asia *(Billitteri, 2013)*. The U.S federal and state planners are prompted to provide the high speed rail services through selected major corridors. However, lacking a capable long-distance passenger travel analysis tool in the U.S. has hindered the decision makers' and politicians' ability to systematically design and quantitatively evaluate the high speed rail. Under this circumstance, it will be desirable to quantitatively forecast the high speed rail demand, systematically design and evaluate the operational effectiveness of the investment. Besides the high speed rail, there are also other national transportation investment strategies in need of the long-distance travel analysis tool to conduct quantitative analysis and evaluation, such as reconstructing and expanding the capacity of the Interstate Highway System, and building the next-generation air transportation system. In addition to these multimodal capacity investment needs for long-distance passenger travel, there are also urgent needs to assess a variety of operational and management strategies at the national level, which could significantly improve transportation efficiency and productivity, support and stimulate economic growth, and produce positive social and environmental impacts. Examples include: (1) congestion pricing on the Interstate and National Highway System; (2) Congestion management at airports; (3) Separation of passenger vehicles and heavy trucks on highway facilities; (4) National transportation financing options such as fuel tax increase and mileage fees; and (5) Substitution between long-distance travel and teleconference/telecommuting.

In addition to enabling national-level infrastructure investment and operational analysis, a long-distance passenger travel demand model for the U.S. also has the following important benefits: (1) Analyze the impact of socio-demographic, economic, and transportation infrastructure changes on long-distance travel demand. (2) Anticipate the influence of energy (e.g. fuel price) and environmental factors (e.g. climate change and related regulations) on long-distance passenger travel. (3) Improve the long-distance passenger travel module in statewide and even some metropolitan travel demand models and provide an authoritative tool for multi-state transportation corridor analysis. After the Intermodal Surface Transportation Efficiency Act (ISTEA) in 1991 was enacted, a lot of state departments of transportation started to develop statewide travel demand models and use them as critical analysis tools in addressing legislative requirements in statewide planning. However, the statewide models are weak in external trips which are usually generated with information from federal and neighboring states instead of available socioeconomic data (National Travel Demand Forecasting Model Phase I Final Scope). A national long-distance travel demand model therefore can provide external trips for statewide models in base-year and future-year. Meanwhile, it can also generate the travel demand for multistate corridors based on available datasets as well as standard and rigorous procedures which can minimize the duplication and efforts of the statewide models. (4) Support large-scale evacuation planning and operations due to natural disasters or targeted attacks; and (5) Enable micro-level analysis of the spread of pandemic diseases resulting from long-distance travel.

National long-distance passenger travel demand analysis has been an understudied area in transportation planning. As the nation and various states engage in funding transportation infrastructure improvements (interstate highway tolling/expansion, high speed rail, next-generation passenger air transportation system relying more on smaller airports and aircrafts) to meet future long-distance passenger travel demand, it is pressing to develop effective and practical modeling methods for long-distance passenger travel analysis.

This project represents the first attempt to develop a prototype activity-based national long-distance travel demand model for national travel analysis. All major behavioral dimensions of long-distance travel will be considered. There are also significant differences between long-distance trips considered in the proposed activity-based model and daily/weekly trips in metropolitan/state-level tour/activity-based models developed in previous research. For instance, the long distance trips usually take days or weeks and may involve car, airplane, train, bus, or multiple modes of the four. It is often the case that households firstly choose travel time for long-distance vacation trips based on time and money budget before selecting destinations and mode.

Categorization of trip purposes is also different for long-distance trips. Cost of travel for long-distance trips is not just travel disutility, but also includes lodging, food, and etc. The same applies to the total travel time for long-distance which usually covers not only in-vehicle travel time but also the ingress/egress time, transfer time, and lodge time. The much lower frequency of long-distance travel may also imply a different decision-making process.

2.0 LITERATURE REVIEW

During the middle of 20th century, the U.S started its travel demand modeling at urban and metropolitan level. Since the passage of the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991, an increasing number of states began to develop their statewide travel demand modelling in order to meet policy and legislative needs *(Zhang et al., 2010)*. At the national level, there are three notable studies on travel demand modelling. Baik et al. *(2008)* developed a travel demand model at the national level which can predict the annual county-to-county personal travel for commercial airline, air taxi, and automobile in the U.S at 1-year interval through 2030. The transportation systems analysis model (TSAM) adopted the four-step travel demand model process which includes trip generation, trip distribution, mode choice and network assignment. An intercity trip in the system is defined as the one with one-way route distance larger than 100 miles, excluding commute travel. Different from the traditional network assignment in four-step travel demand model, the network assignment in TSAM is developed for the commercial airline and air taxi. The TSAM outputs annual county-to-county person round-trips by travel mode, trip purpose (business and non-business) and household income group. The network assignment then outputs the annual flights between all the commercial and air taxi airports. Cambridge Systematics *(2008)* conducted a study in order to provide specifications for national travel demand forecasting model development. They proposed a four-step demand model structure which focuses on providing travel information for statewide models. Data sources including network and zone system, demographic and employment data, freight data and travel behavior data are assessed prior to preparing the input data. Nostrand *et al. (2013)* presented a research on national long distance travel demand modeling only for leisure purpose. An annual vacation destination and time choice model is developed, using Multiple Discrete-Continuous Extreme Value (MDCEV) structure *(Bhat, 2005)*, to predict the destinations that a household would visit during a year and the time allocated for each of the vacation destinations. The model mainly relied on the 1995 American Travel Survey data for analysis, and a total of 210 zones were divided for the whole nation. The output of the model can be used to construct a national-level OD table for leisure travel.

Meanwhile, in Europe, a lot of attention has been paid to national travel demand modeling during the last two decades. The National Model System (NMS) in Netherlands was developed in 1986 and is being updated since then *(HCG & TOI, 1990; Gunn, 2001)*. It has served as a "prototype" disaggregate model in Europe and was built based on behaviorally oriented tour-based method *(Gunn, 2001)*. The model system consists of a series of connected choice models including license holding and car ownership models, tour frequency, tour mode and destination choice, and time-of-day. Total 345 zones were divided in the NMS system and 1302 sub-zones were used in the sub-module of mode and destination choice models. The NMS is sensitive to a variety of socio-economic, land use, transportation systems, and policy factors. Applications were observed in the rail demand prediction for railway options, impact of raising fuel prices, effect of the introduction of motorway signaling, high speed trains demand analysis, and etc *(Hofman, 2001)*. The Norwegian national model system followed the structure of the Dutch NMS, but with the objective of emissions (CO_2 and NO_x) prediction at the national level *(HCG & TOI, 1990)*. Therefore, no detailed link-loadings on a national network were needed *(Gunn,* *2001)*. Sweden started its national model development in the beginning of 1980s, and has improved it to the current new version of SAMPERS which belongs to the mainstream trip-based four-step model *(Zhang et al., 2012)*. The new model covers the trips in Sweden and to neighboring countries in detail and trips to and from other parts of Europe in a coarser way *(Widlert, 2001),* which results in three different model systems: regional models, domestic long distance models and international models. All models in the three systems, such as car ownership, trip frequency, destination choice, mode choice, and departure time choice, adopted discrete choice logit model, except for an ordinary least squares trip frequency model for foreigners' travel to Sweden *(Beser & Algers, 2001).* The Danish PETRA model was developed as an activity-based method to travel demand analysis *(Fosgerau, 2002)*. In the model, a person's daily travel is represented in terms of chains of tours instead of separate trips or tours. In order to reduce the complexity, the observed chains were transformed to a simplified version of chain types. The model system first deals with the cohort effects on car ownership and license holding *(Lundqvist & Mattsson, 2001)*. Then a mode and destination choice model (nested logit model) was estimated for each tour in a chain. Finally, the choice of chain type is modeled, and the accessibility measured by the logsum from the mode and destination model is incorporated in the chain type choice model. Since there is no network assignment module in PETRA, no congestion analysis is considered. A variety of applications have been observed with the model in analyzing the effects of different policy measures on people's travel behavior. From the perspective of the geography and the travelling population size, many of the European national travel modeling efforts are closely equal to statewide model studies in the U.S *(Zhang et al., 2012)*. The pan-European models include the countries forming the European Union, which to some extent are close to the national model study in the U.S. The pan-European travel demand model developments started with the estimation of multimodal OD matrices without behavior framework in the MYSTIC *(Peter Davidson Consultancy, 2000)* project in the early 1990s and then the DATELINE project of 2004 *(Brog et al., 2004; Davidson & Clarke, 2004)*. Then aggregate methods and joint aggregate-disaggregate methods were employed in the 201-zone NUTS2 STREAM model *(Williams, 2001)* and the 1275-zone NUTS3 STEMM model *(Gaudry, 2001)* respectively. Finally, the most recent pan-European travel demand model is the disaggregate TRANS-TOOLS model integrating European transportation and economic models *(Zhang et al., 2012; Burgess et al. 2006)*.

In addition to the applications of the national travel demand model in the transportation field, we also found that it is used in other areas. Epstein, Parker, and et.al *(Epstein et al., 2008)* developed an agent-based microsimulation model for intercity travel in a research on spatial-temporal epidemic dynamics. The model simulates individual's travel decisions on trip frequency and destination choice based on a zip-code-level OD system. Since travel demand analysis is not a focus of the study, no mode choice or assignment models were employed. Even so, this study demonstrated the benefit of long distance model in other fields in addition to the transportation area.

3.0 DATA

The data employed in the dissertation contains two parts: one for model estimation and the other for the base year model simulation. The main component data source for model estimation is the travel survey data, besides the transportation OD skim data and economic/demographic information of traffic analysis zones (TAZs) which are also the input in the model simulation part. The other input data for model simulation is the population data which contains individual and household characteristics such as age, gender, employment status, household type, household income, etc.

3.1 TRAVEL SURVEY DATA

The 1995 American Travel Survey is the primary source of national travel survey data used in the research for model estimation. It is a long distance nationwide travel survey of the United States, and was conducted by the Bureau of Transportation Statistics (BTS) between April 1995 and March 1996. The 1995 ATS data collected detailed long distance travel (>100 miles) information and demographic information from more than 80,000 random selected households from the 50 states and the District of Columbia, and each household was interviewed every three months during the survey period. "One of the ATS's main objectives is to comply with two requirements of the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA): (1) to provide information on the number of people carried in intermodal transportation by relevant classification, and (2) to provide information on patterns of movement of people carried in intermodal transportation by relevant classification in terms of origin and destination." *(Hwang & Rollow, 2000)*. Undeniably, the 1995 ATS is a little old, but it is the most recent dataset in long distance travel over the course of one year in the U.S. Although the household travel survey data, 2001 National Household Travel Survey (NHTS), has a long-distance travel component, it is a relatively small sample and has less long distance information than the 1995 ATS.

The 1995 ATS provides comprehensive information. To be specific, the gathered demographic information in the survey includes the characteristics of both the household (family type, household income, household size and etc.) and the household members (gender, age, employment status, race and etc.). The detailed long distance travel information contains the origin and destination of the trip, stops along the way from and to destination, side trips originating at the destination, the means of transportation, the reason of the trip, the lodging type, the number of nights spent away from home, travel party size, and etc. All the location information was recorded at regional, state, and metropolitan level. Such detailed information cannot be found in other U.S. national travel surveys and makes the 1995 ATS a useful source of long distance travel survey data.

The 1995 ATS database consists of four data sets: household trips, household characteristics, personal trips, and personal characteristics, of which the latter three datasets were utilized in this research. The personal trip data included 556,026 trip records for both domestic and abroad long distance trips, and a total of 45374 long distance trips were made by people older than 18 years old and dozens of travel modes were recorded for the entire travel (inbound trip and outbound trip). In the dissertation, only adult domestic long distance trip will be covered. Based on the characteristics of the detailed travel modes, 18 classes of the travel modes were aggregated into 6 including car, air, bus, train, recreational vehicle (RV) and others. From the sample, the percentage of the long distance tour travel modes can be plotted (Figure 1) where Others refer to the combinations of multiple travel modes and other modes such as motorcycle and bicycle. It can be observed that almost 80% of the long distance activities were made by car for the entire tour, and the second most popular travel mode is the air travel which accounts for 17%. From the 11 activity types in 1995 ATS, three main activity types (business, personal business, pleasure) were aggregated based on the activity similarities, to reduce the model computation complexity (Table 1). Figure 2 illustrates the percentage of the aggregated three trip purposes. 60% of the long distance trips are made for pleasure, and 24% and 16% of the long distance activities are business and personal business related respectively.

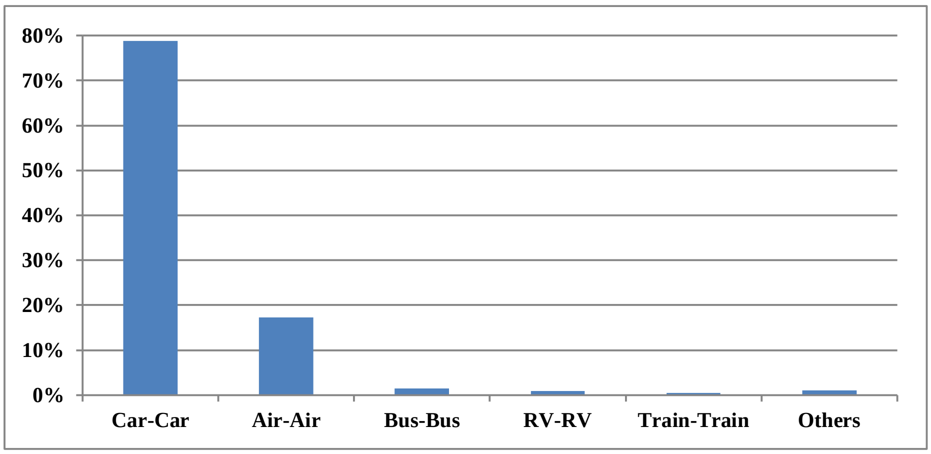


Figure 1: Percentages of Long Distance Tour Travel Modes

Table 1: Decoding Reported Trip Purposes

Reported Trip Purpose	Decoded Trip Purpose
Business	Business
Combined Business/Pleasure (B/P)	Business
Convention, Conference, or Seminar	Business
School-related activity	Personal Business
Visit relatives or friends	Pleasure
Rest or relaxation	Pleasure
Sightseeing, or to visit a historic/scenic attraction	Pleasure
Outdoor recreation	Pleasure
Entertainment	Pleasure
Shopping	Pleasure
Personal, family or medical	Personal Business
Others	Deleted

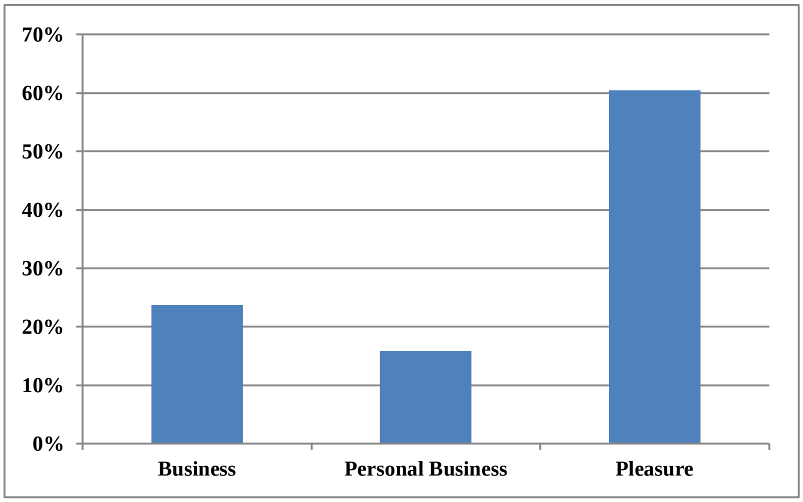


Figure 2: Percentages of Long Distance Trip Purposes

3.2 TRANSPORTATION OD SKIM AND ECONOMIC/DEMOGRAPHIC DATA

According to the percentages of the travel modes and the difficulty of accessing the bus and RV data, only three travel modes (car-car, air-air, and train-train) were considered for the entire tour of the long distance activities. The traffic analysis zone in our national travel demand model system is Metropolitan Statistical Area (MSA) or non-MSA that is the remaining area of a state not belonging to an MSA. Although the non-MSA usually has larger area than MSA and it will be desirable to divide the non-MSA into smaller zones, the finest geographic resolution in the ATS data is MSA and non-MSA. In the sample data of 1995 ATS, there are total 208 zones (161 MSAs and 47 non-MSAs). The OD skim data or level of service variables primarily refer to the travel time and costs between each origin and destination pair via different travel modes, and it can be observed as the functions of distance between TAZs. Usually an MSA or a non-MSA is made up of more than one county, so the distance between two TAZs can be estimated by averaging the distance between all the county pairs in each zone (Equation 1).

$$D_{ij} = \frac{\sum_{m,n} d_{mn}}{m \times n} \tag{1}$$

where *i*, *j* refer to the zone of the MSA or non-MSA; *m*, *n* indicate the number of the county in zone *i* and zone *j*, respectively; d_{mn} is the distance between county *m* in zone *i* and county *n* in zone *j*. The Census Bureau provides the geographic information of each county in the U.S., which assists us to estimate the great circle distance[1] between each zone pair.

Auto travel times and costs were derived as a series of functions with the information of the great circle distance of a zone pair, the average driving speed, the vehicle's characteristics, and etc.

[1] The great circle distance: the shortest distance between two points on the surface of a sphere.
Distance= 6371*acos(sin(latitue1)*sin(latitue2) + cos(latitue1)*cos(latitue2)*cos(longitude2-longitude1)), where 1 and 2 refer to the point 1 and point 2 on the surface of a sphere.

The cost of the vehicle usually contains the fuel cost, the insurance, the maintenance and the tire costs, and among them only the fuel cost is out-of-pocket expense for the trip, while the other costs are paid separately after the trip. Therefore, only fuel cost is considered for the vehicle cost during the long distance travel. Several assumptions are made in order to estimate the auto travel time and costs: 1) the average auto speed is 65 miles/hour; 2) the auto travel time/cost consists of two parts spent on driving and lodging; 3) people on business travel will stop for an overnight stay every 9 hours, while people taking personal business and pleasure travel will stop every 13 hours; 4) the auto average fuel efficiency is 19.7 mpg *(Grush, 1998)* and the average retail fuel price is $1.48/gallon in the U.S. in 1995 which was obtained from Energy Information Administration; 5) the average lodge cost per person night for business travel from the low-income to the high-income are $70, $90, and $110 respectively, while the lodge cost for personal business and pleasure are $30, $50, and $70 respectively; 6) the travel party size is 1 person for business travel, and 2 persons for personal business and pleasure trip, which helps to estimate the vehicle cost for each person.

Air fare and the number of layover were collected from the Airline Origin and Destination Survey (DB1B) provided by the Bureau of Transportation Statistics, Research and Innovation Technology Administration (RITA). DB1B is a 10% sample of airline tickets from reporting carriers; therefore, in order to obtain a sample size as large as possible the DB1B data from 1994 to 1996 was employed. The air travel time is made up of access/egress time (time spent traveling to the airport and from the airport to the final destination), air fly time, and transfer wait time between flights. Air fly time is estimated with the obtained great circle distance and the average flight speed which is assumed as 500 mph (*Boeing, 2011*). The average total access and egress time is assumed as 2 hours for all the air travel, and the average wait time per transfer is set as 1.5 hours. The number of layover is obtained based on the airport groups in the DB1B data, which list the airport codes of all the airports in the flight itinerary. Since the airport code is of three characters and the airports in the airport group variable are separated by colons, the number of layovers can be obtained according to the number of the characters in the variable. By multiplying the average wait time per transfer and the number of layovers, we can acquire the total transfer wait time in the itinerary between TAZs. Air fare was taken from the DB1B data eliminating the first/business class fare to reduce the travel cost variance and to be in line with the fact that the majority of the travelers choose the economic class for their air travel. A MSA or non-MSA may have more than one airport from the geographic perspective, therefore, the air fare and time between TAZs should be the average value between all the airport pairs in the corresponding zones.

It is hard to get the train fare and time data in 1995 or neighboring years. Therefore, we collected the train fare and travel time in August, 2013 from Amtrak, a national railroad passenger corporation, as a proxy. The Amtrak website provides people an access to look up station-to-station timetable and ticket information. The Amtrak train ticket has several classes including saver, value, flexible, and premium. Generally, saver class ticket charges the least, flexible or premium class ticket charges the most, and the price of the value class ticket is at the middle level of all the classes' fares. While collecting the train fare for our study, we chose the value class price and then converted the fare to 1995 dollar according to the Consumer Price Index (CPI). The travel time from Amtrak contains both the train travel time and the transfer waiting time from origin station to destination station. In economically developed regions, a TAZ may have multiple rail stations. Under this circumstance, the TAZ-to-TAZ train fare and time is achieved by aggregating the fare and time between all station pairs from the two zones.

Data for the zones' attractiveness indexes in this research mainly include the total population, the number of employment by industry sector, and the number of households. These economic and demographic data for each MSA and non-MSA was obtained from the Complete Economic and Demographic Data Source (CEDDS) by Woods & Poole Economics. This database offered historical, current and projected socioeconomic indicators (e.g. population, employment, households, etc.) for all the regions, states, statistical areas and counties in the U.S.

3.3 PUBLIC USE MICRODATA SAMPLE DATA

The input data for the base year model is the 2000 Census Public Use Microdata Sample data (PUMS) that contains the full range of population and housing information collected in Census 2000. The Microdata includes individual records representing 5% sample of the people and the housing units in the U.S., which is over 14 million people and over 5 million housing units (*Census Bureau, 2008*). Detailed person and household information were stored in person record and housing unit record in one PUMS file for each state. Each record had a unique identifier linking the people to the proper housing unit record. The specific geographic unit was defined as Public Use Microdata Areas (PUMAs) for the 5% sample data, and each PUMA contained a minimum population threshold of 100,000. The housing unit record in PUMS file contains detailed household information such as home ownership, real estate taxes, number of vehicles, number of persons in the household, household type, household unit weight, presence and age of own children, PUMA code, state code, household income in 1999 (changes of income), MSAPMSA code, and etc. The person record includes person's information like age, gender, race, marital status, education attainment, school enrollment, employment status, means of transportation to work, travel time to work, class of worker, income in 1999 by type, person weight and etc. The weights in PUMS file for each person and housing unit can be used to expand the sample to the relevant total. Since each state is comprised of one or more PUMAs and some large metropolitan areas may be divided into several PUMAs, a PUMA could contain parts of multiple TAZs. Geographic equivalency between PUMA and MSA/non-MSA needs to be identified prior to the data being employed. All the persons or housing units located in a PUMA containing mixed TAZs should be allocated to each TAZ according to the population percentages of the TAZs in the specific PUMA.

4.0 METHODOLOGY

The activity-based national travel demand model we developed can predict the long distance passenger trips made by auto, air, and train in the U.S. in one year period. It can serve as a forecasting tool of long distance travel in the U.S. The model system has root in econometric model developments including discrete choice model and duration model. These models are employed to guarantee the maximum behavior realism and model sensitivity to regional and national projects and policies. The model is implemented in a micro-simulation framework which simulates the long distance travel for each adult in the U.S. Since the finesse spatial resolution in 1995 ATS data is metropolitan statistical area, we adopted the MSA and Non-MSA as our traffic analysis zone system which includes 378 zones in total.

The model system consists of three tiers, see Figure 3: 1) the yearly long distance activity pattern level which predicts the number of different types of activities a person will choose during one year (Figure 4); 2) the tour level model system (Figure 5) which contains choices of tour destination, time of year, tour duration, and tour mode; 3) the stop level model system including the number, the purpose, and the location of each intermediate stop made during the inbound and outbound legs of the tour (Figure 6).

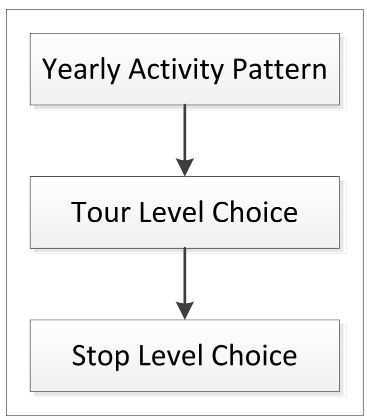


Figure 3: Activity-Based Long Distance Travel Demand Model System

4.1 ACTIVITY PATTERN LEVEL MODEL

The demand for long distance activities and travel can be considered as a choice among all the possible bundles of activities and travel annually. Dissimilar to regular urban-level activities organization, people choose their long distance activities in one year with few interactions due to much less frequent long distance travel. As shown in Figure 4, the yearly long distance activity schedule can be presented as a set of different long distance activities per year. From the 12 activity types in 1995 ATS, three main activity types (business, personal business, pleasure) were aggregated based on the activity similarities, to reduce the model computation complexity. So the yearly long distance activity pattern can be presented as {B-x, PB-y, P-z}, where B, PB, and

P stand for the activity type business, personal business and pleasure, respectively, and x, y, z are integers (x, y, z $\geq$ 0) referring to the number of the corresponding activities during one year. In the prototype model, the Multiple Classification Analysis (MCA) method is employed to estimate the long distance trip rates by activity type.

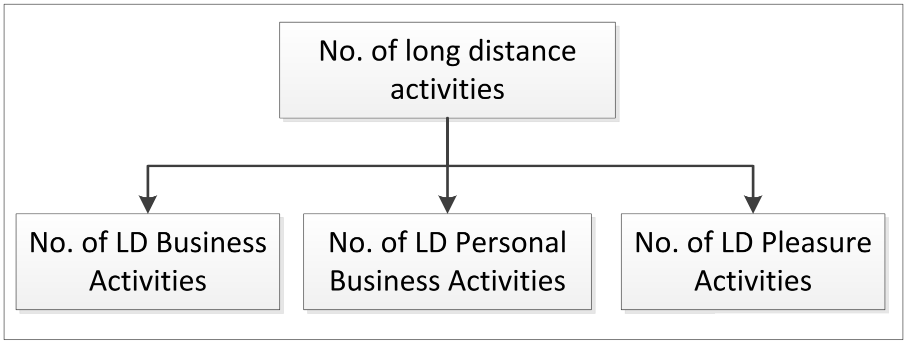


Figure 4: Yearly Long Distance Activity Pattern Level

4.2 TOUR LEVEL MODEL STRUCTURE

Each long distance activity schedule has a primary tour, and may have zero or more intermediate stops during the legs of the tour. In our model system, the tours that occur based on the long distance primary destination are ignored due to the data limitation. The tour level model system defines the characteristics of the primary tour of each long distance activity such as the tour destination, time of year, tour duration and tour mode. When we develop and estimate each model component at the tour level, it is assumed that the outcomes of the upper-level model, the household, person characteristics and mobility attributes are already known. So the solid arrow in the figures indicates that the output of the upper level can be used as an explanatory variable at the lower level, while the dash arrow means that the expected utility of the lower-level models can affect the choices at the upper level. When people decide to make a long distance activity, they usually have different priority considerations and decision procedures for different activity types. In our study, we made a set of assumptions about people's decision making process at the tour level. For example, the long distance pleasure activity (a discretionary activity) requires people to consider their time availability prior to other decisions. When they have a period of time (days, weeks or months) for pleasure, they will decide when to spent it, where to go and how to go sequentially. In contrast, people taking long distance business and personal business activities usually give priority to the decision of the activity location and time (including the time of year and duration), followed by the tour mode choice. Therefore, Figure 5 shows two different tour level structures are proposed for business/personal business and pleasure. According to the direction of the dash lines in the figure, both time of year models and destination models should include the expected utility variable from the mode choice model (mode choice logsum). Due to the fact that the finest temporal resolution in ATS is quarter, our proposed model system will function at a resolution of three-month or one-quarter. The three-month increments begin in January and end in December, thus four quarters in total. In the ATS data, few records are observed that depart from and arrive at home across quarters, which results in only four alternatives {(Q1, Q1), (Q2, Q2), (Q3, Q3), (Q4, Q4)} for each person when he/she decides what time of the year to travel and what time of the year to get back. Three travel modes are modeled at the tour level, i.e. {(car, car), (air, air), (train, train)}, and no combination of different travel modes is considered due to the small sample size in the ATS data (approximately 1%). Different from the urban- or metropolitan-level travel demand model systems, the duration in long distance trip is measured in days away from the origin, which leads to a discrete choice.

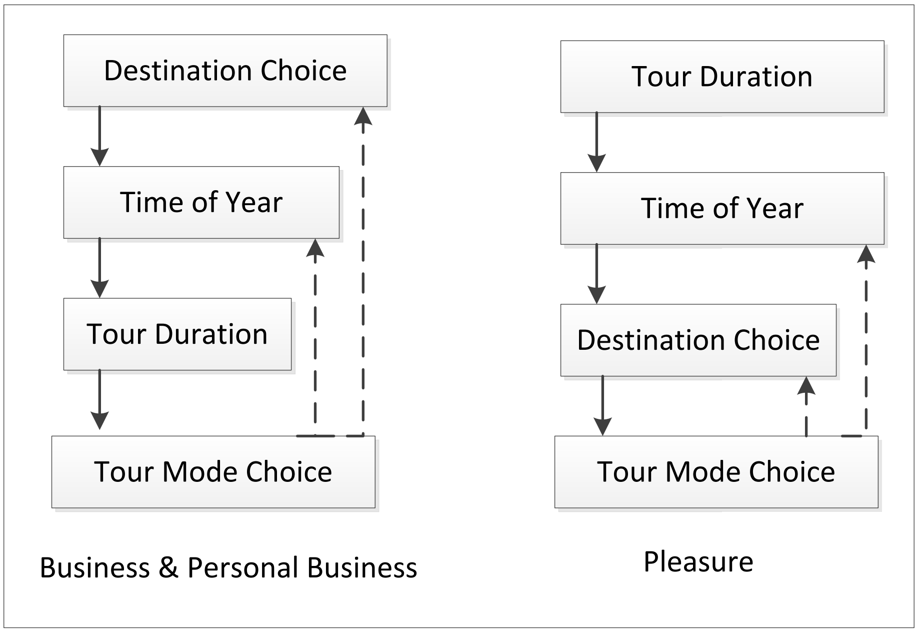


Figure 5: Tour Level Procedure and Model Components

4.3 STOP LEVEL MODEL STRUCTURE

After people have made decisions about their travel to the main destination, they will make plans for their trips on the way to and from the destination. It is assumed that people have the same logic to determine their stops or trips during the tour legs regardless of the main activity types. Consequently, the same model structure at the stop level will be applied to all the three tour-level activity types (business, pleasure, personal business) (Figure 6). The stop level structure predicts the information of the intermediate stops people would make during their inbound/outbound legs of the long distance tour.

The stop frequency model at the higher level determines the number of intermediate stops people will have on the way from/to the tour destination. In each direction, a maximum number of 4 stops can be made which results in a maximum of 5 trips on each tour leg. Once the number of stops on each tour leg is obtained, the purpose for each stop needs to be determined (the middle-level model component), and the stop purpose category follows the same tour-level activity-type which are business, personal business and pleasure. At the low tier of the stop-level structure, the location for each stop will be predicted with the similar method employed in the primary destination choice at the tour level. Since we assume that people only take one of the three modes (air, car, train) and no transfer among different modes for the entire tour, the travel mode for each trip on each half leg will share the same one predicted at the tour level. Different from the tour-level primary destination choice, the impedance of the travel to an intermediated stop in the stop location choice model should measure the additional impedance between the tour origin or stop origin and the tour primary destination if it is an outbound trip. For example, the level of service (LOS) variables for the first stop on the way to the tour primary destination are based on the additional impedance between the tour origin and the tour destination, and the LOS for the second stop is based on the additional impedance between the first stop and the tour destination. The same method works with the stops of the inbound direction but in an opposite way, as the anchor point is the tour origin instead of tour primary destination.

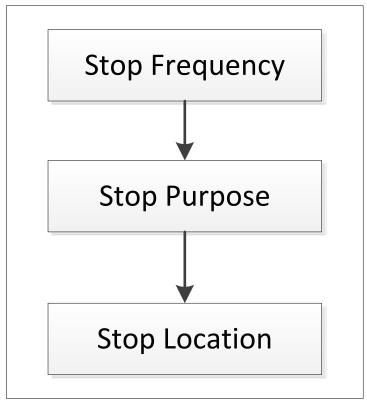


Figure 6: Stop Level Procedure and Model Components

The long distance activity-based model system has a series of analytic tools to ensure maximum travel behavior realism and model sensitivity. All the model components at the tour level and the stop level adopt the discrete choice forms (multinomial Logit model) except for the tour duration component which employs the hazard duration methodology. A micro-simulation-based framework, which simulates each person's long distance travel decision features, is developed based on the model system. The Monte Carlo simulation method is used in the framework to achieve an unbiased selection of alternatives in the case of predicting each decision.

4.4 MODELS ESTIMATION AND VALIDATION

The proposed national travel demand model system has a dozen of model components to be estimated and due to the space limitation, only several of the model estimation results are presented here. Table 2 illustrates the estimation results for the tour mode choice. And all the variables are significant at the 95% confidence level. The coefficients estimation implies that the high-income people have the largest value of time (VOT) ($45.52/hour) when they take the long distance business travel. And in general, people taking long distance pleasure trip or personal business trip have smaller VOT than taking business trip, except for the low-income people who have a smaller VOT ($9.25/hour) in business trip than in pleasure trip ($10.5/hour) and personal business trip ($11.54/hour). This can be explained by the fact that people going on a business trip can get reimbursement for their travel and accommodation expense, and the low-income people's jobs are usually low-paid. Table 3 presents the estimation results for time of year choice under the long distance personal business trip. The fourth quarter is set as the base alternative. Coefficients in bold and italic format are significant at the 95% confidence level. The mode choice logsum variable infers that people significantly tend to take their long distance personal business trip in the quarter with larger accessibility in terms of mode choice logsum. People in higher-income households are more likely to travel in the second quarter for their personal business. If an individual has a job, he or she prefers travelling in the third quarter to take care of his/her personal business.

Table 2: Tour Mode Choice Model Estimation Results

Variables	Business	Pleasure	Personal Business
Total Travel Cost(Low Income)	-0.006	-0.008	-0.008
Total Travel Cost(Medium Income)	-0.004	-0.008	-0.007
Total Travel Cost(High Income)	-0.001	-0.006	-0.006
Total Travel Time	-0.057	-0.081	-0.090
Rho-Square	0.318	0.510	0.498

** Low income level: household income < =$30,000;*
Medium income level: $30,000 < household income <= $70,000
High income level: $70,000 < household income

Table 3: Time of Year Choice for Personal Business Trip

	(Q1,Q1)	(Q2,Q2)	(Q3,Q3)	(Q4,Q4)
Mode choice logsum	***0.01***			
Household Income	***1.0e-05***	***5.64e-06***	-6.65e-07	0.000
Employed	***0.174***	***0.322***	***0.327***	0.000
Male	-0.083	***-0.21***	-0.110	0.000
Couple with Child	-0.01	0.068	***-0.249***	0.000
Age	***0.011***	***0.016***	***0.006***	0.000
Constant	***0.568***	0.224	***0.524***	0.000

After the model estimation and before the implementation in the micro-simulation framework, a with-in-sample model validation has been performed for each model component at both the tour level and the stop level. Due to the space limitation, two of the models' validation results are presented: one for destination choice under the business trip (Figure 7) and one for the mode choice under the pleasure trip (Figure 8). Figure 7 shows the validation result for only 206 zones that exist in the dataset for the business trip destination model estimation. The figure depicts that most of the destinations can be predicted with a small error except for a few ones with high prediction errors due to the relatively small sample size. Figure 8 presents the aggregate share of the observed tour mode choice and the predicted one. Results show that the car and train travel are over-predicted by a small percentage, while air travel is under-predicted slightly.

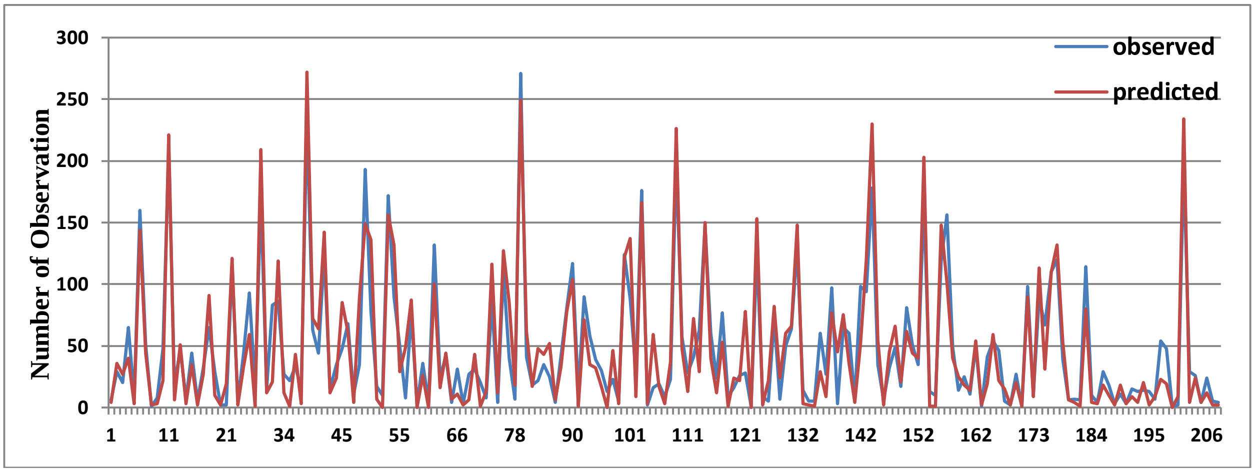


Figure 7: Tour Destination Choice Model Validation for Business Trip

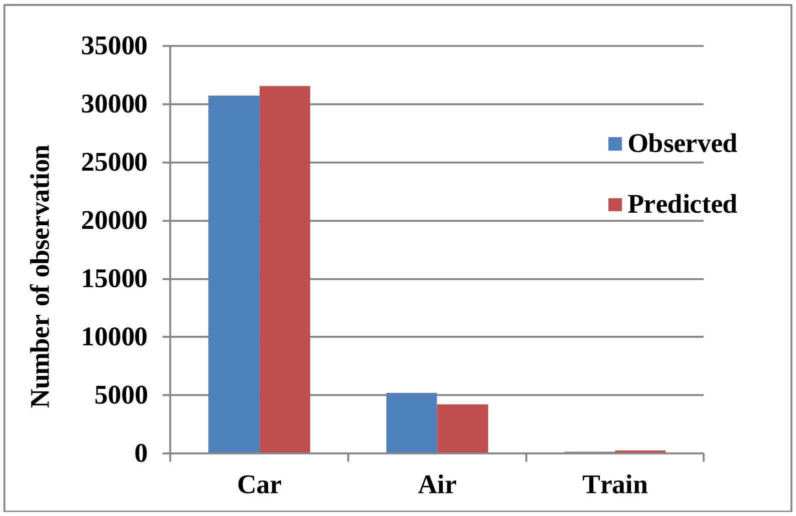


Figure 8: Mode Choice Model Validation for Pleasure Trip

5.0 PRELIMINARY BASE YEAR OD ESTIMATIONS

The year of 2000 was chosen as the base year, since it is close to 1995 in which year we got the long distance travel survey data for model estimation. From 1995 to 2000, we assume that people have small change in their long distance travel behavior. Meanwhile, the 2000 Census PUMS data contains the full range of population and housing information collected in Census 2000. The base year transportation OD skim data and economic/demographic data are also collected and processed using the same methods discussed in Chapter 3. The micro-simulation tool is developed using Java. Given all the input data, our developed micro-simulator could output the long distance activity patterns of each person in the U.S. In order to reduce the simulation complexity and time, the 5% PUMS data is used as the input and then the output trips will be expanded to be consistent with the national population according to the person weight provided by PUMS data. Based on the simulation output of the first two levels (yearly long distance activity pattern level and tour level), 12 national OD tables (4 quarters * 3 travel modes) are generated including all the activity types. The simulation at the stop level is still running and the results will be ready to present in our future work. Aggregating the cell trips in each of the 12 OD tables can give us the national trip distribution by travel mode in each quarter (Figure 9). The aggregate results imply that the car travel accounts for above 70% among the three travel modes, and a small portion of people choose train for their long distance travel. Meanwhile, people prefer taking their long distance trips during the first three months (from January to March) regardless of the travel mode. Summing all the trips yields over 1.3 billion trips a year in total which infers that a person would take an average of around 6 long distance trips during a year.

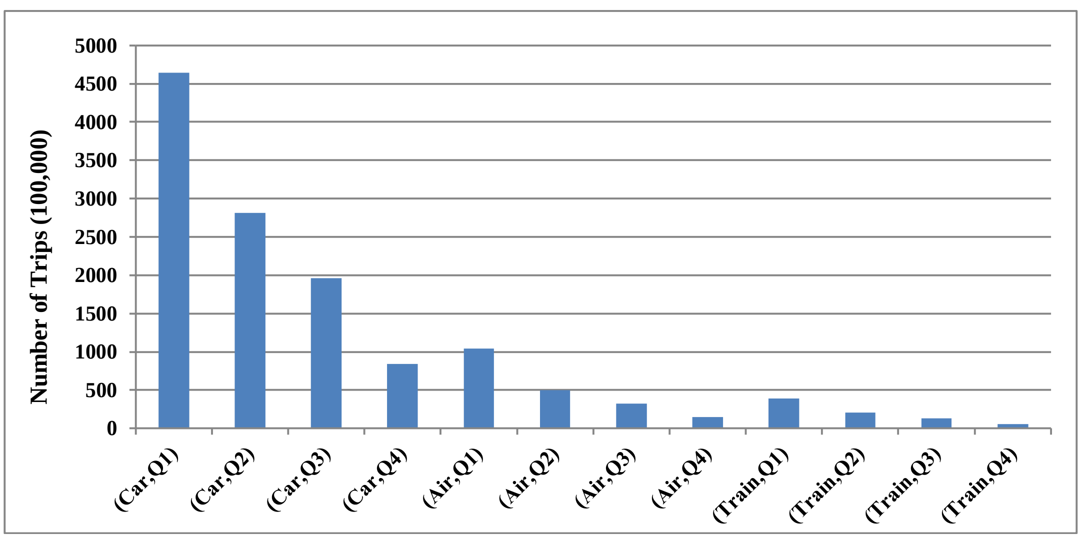
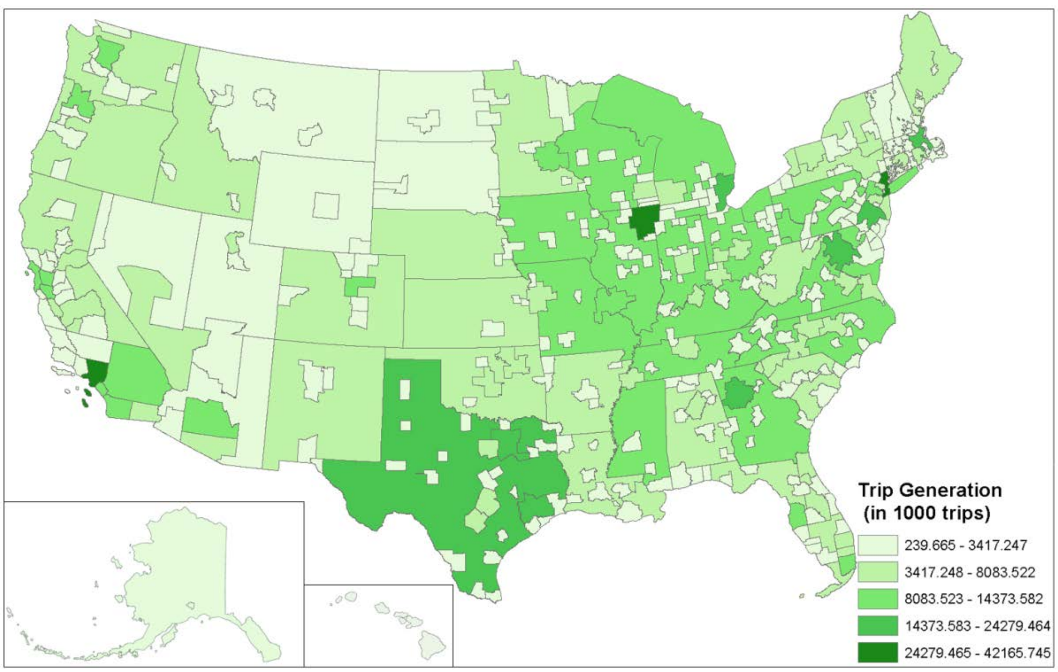


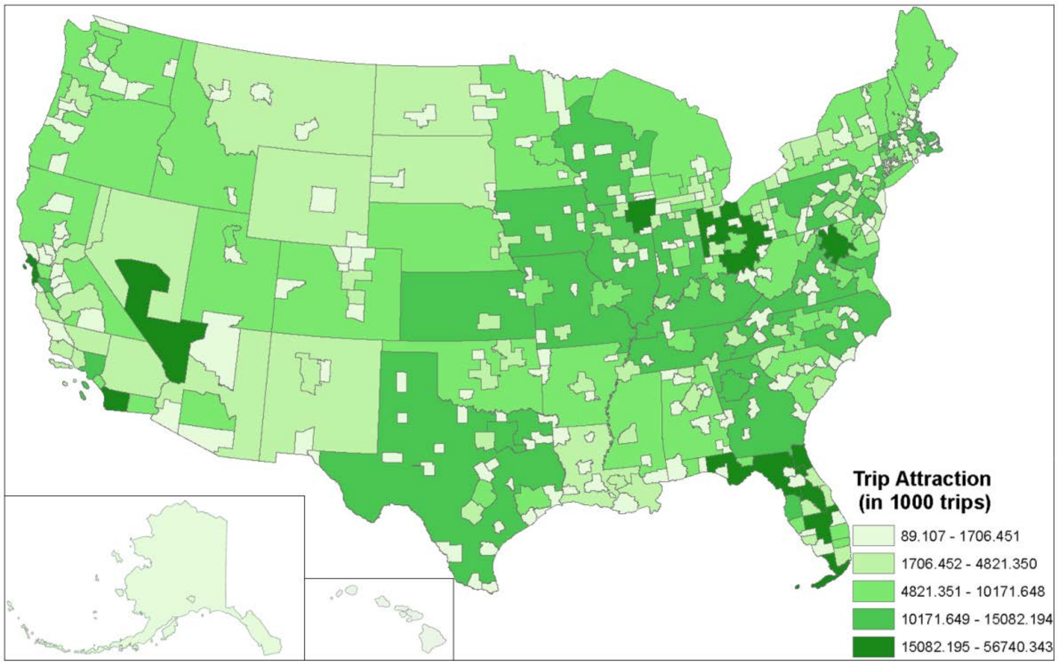
Figure 9: Trip Distribution by Travel Mode and Time of Year

Based on the OD tables, we can obtain the long distance trip generations, see Figure 10(a), and attractions, see Figure 10(b), at the MSA/Non-MSA level for the nation. It is observed from Figure 10(a) that the zones in northeast, west pacific, and Texas generally have a larger number of outgoing trips during one year, while the zones in the states of Montana, North Dakota, South

Dakota, Wyoming, Utah, and Colorado produce a relatively small number of trips which can be explained by the small size of the population and low GDP of these states. Figure 10(b) implies that the zones in the states with low trip generation also attract fewer people to travel to, in contrary, the zones in east, east north central and west of the U.S have a lot of people to visit. As an illustrative example, Figure 11 shows the trip distribution by TAZ and travel mode originating from the Washington D.C metropolitan area. Figure 11(a) infers that the car trips from D.C usually are centralized around D.C, and the farther the zone is away from DC, the fewer people will go by car. So, most of the car trips from D.C occur in the middle-east and east parts of the U.S. Since the train network is not as wide as car or air networks, the trips taken by train should be distributed in certain zones with train stations. As expected, the train trips from D.C are mainly located in east coast and centered around D.C. In general, those trips have shorter distance than the car trips and the distribution range is much smaller than that of the car trips, see Figure 11(b). The most significant advantage of air travel is that it is unquestionably the fastest mode among all the transportation means, especially when people travel in long distance. Moreover, in the U.S people can travel by air to most places in the nation due to the high density of airlines and airports. Consequently, trips by air departing from D.C are observed all over the U.S. (Figure 11(c)). A large number of air trips to non-MSA zones in Virginia and Maryland are observed due to the fact that the air fare and time between D.C and those zones are significantly small.

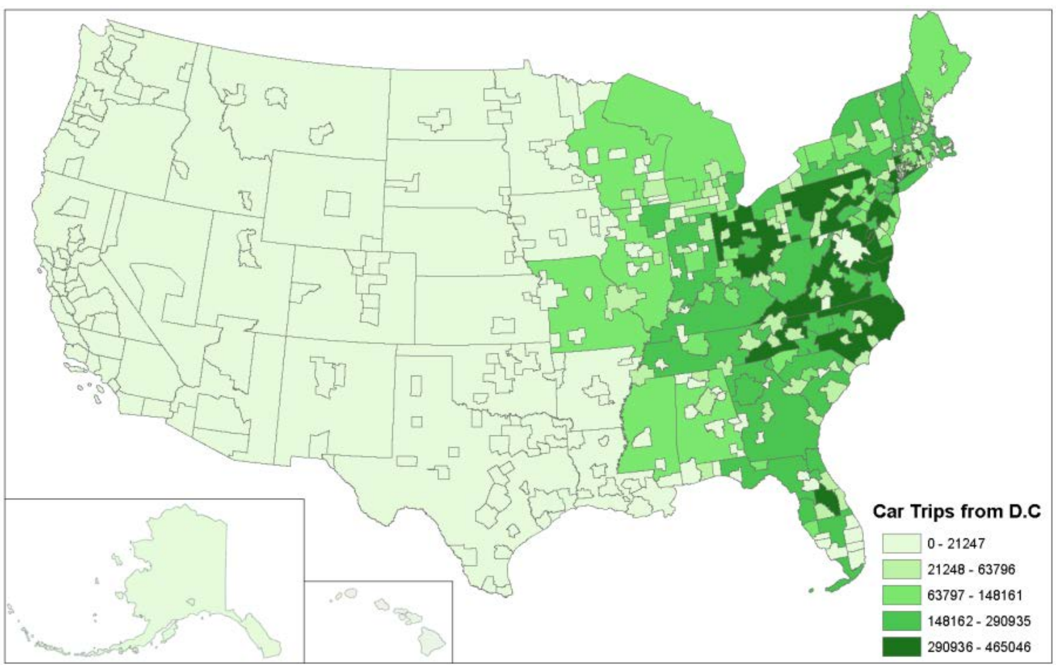


(a)

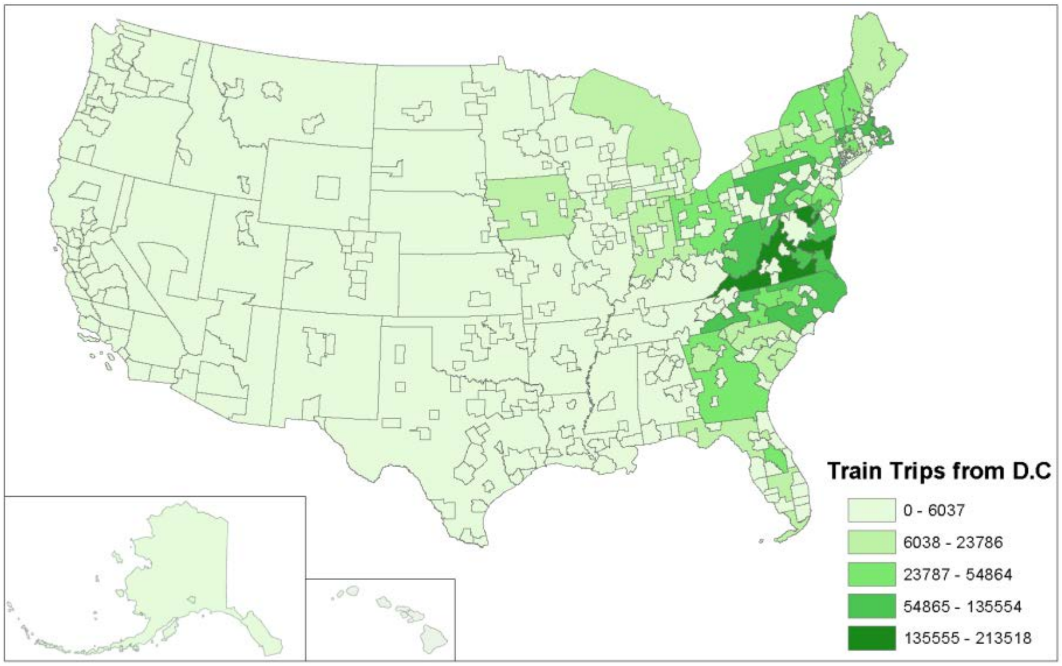


(b)

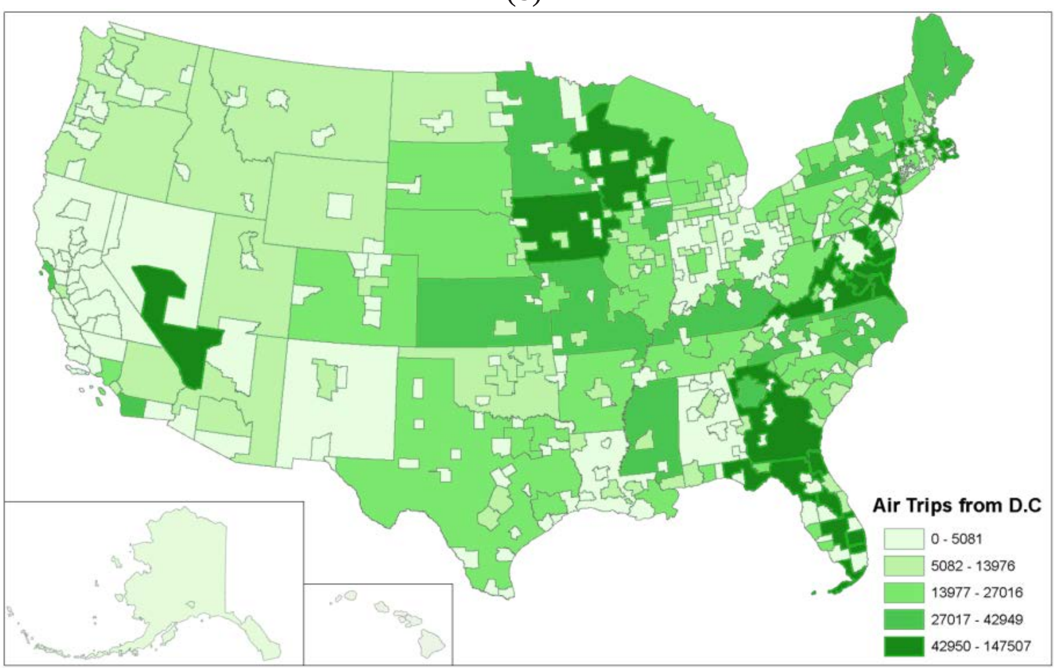
Figure 10: Trip Generation and Attraction at MSA/Non-MSA Level



(a)



(b)



(c)

Figure 11: Yearly Trip Distribution by TAZ and Mode Originating from Washington D.C

6.0 CONCLUSIONS

This research project presents a prototype national activity-based travel demand model system which can predict the long distance travel for each person in the U.S. during one year. Unlike daily travel at the urban level or metropolitan level, long distance travel has its unique features, such as low frequency of trips, long duration of activities (usually days, weeks, or even months), different set of mode alternatives, and time and money constraints. Taking these factors into account, we proposed a three-tier model structure including yearly long distance activity pattern, tour level choice and stop level choice. Furthermore at the tour level two model structures were developed for business/personal business activities and pleasure activities respectively according to the different decision-making processes. In the model system, each model component was developed using the econometric methods and estimated from the available long distance travel survey data (1995 ATS), transportation level of service or OD skim data, as well as the economic/demographic data. Then, model validation was performed for each model component prior to the model being implemented for travel analysis. A micro-simulation platform was developed to simulate and predict each person's long distance travel activities at the MSA/Non-MSA level during one year in the U.S with the 2000 Census PUMS data, the corresponding transportation level of service data and economic/demographic data. Besides the disaggregate output of each person's long distance travel, aggregate results such as OD table estimations by time of year and travel mode can also be obtained from the micro-simulation framework. Integrating the simulation results at the stop level in our future work should make the OD estimations more accurate.

This project demonstrates a more advanced academic research endeavor for national passenger travel analysis. This research aims to provide important insight and help guide federal and state to make decisions on corridor-level, region-level, and nation-level infrastructure investment, design, and management, as well as to research on long-distance passenger travel demand. The prototype model in the research exhibits the system logic and concept, statistically supports its basic structure, and provides preliminary OD estimations based on the model simulation. And further model calibration and model improvements will be conducted in our future work.

7.0 REFERENCES

Baik, H., Trani, A. A., Hinze, N., Swingle, H., Ashiabor, S., & Seshadri, A. (2008). Forecasting Model for Air Taxi, Commercial Airline, And Automobile Demand in the United States. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2052, pp9-20.

Beser, M. and Algers, S. (2001). SAMPERS – The New Swedish National Travel Demand Forecasting Tool. In Lundqvist, L., & Mattsson, L.G. (Eds.), *National Transport Models: Recent Developments and Prospects*, Springer, New York.

Bhat, C.R. (1995). A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice. *Transportation Research Part B*, Vol. 29, No. 6, pp. 471-483.

Bhat, C. (2005). A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions. *Transportation Research Part B*, Vol. 39, No. 8, pp. 679-707.

Billitteri, Thomas J. High-Speed Trains: Does the United States Need Supertrains?. [Online] library.cqpress.com. Retrieved on December 26, 2013

Boeing (Producer). (2011). 757 Commercial Transport History. Retrieved from http://www.boeing.com/history/boeing/757.html

Brog, W., Erl, E., & Schulze, B. (2004). DATELINE: Concept and Methodology. Paper presented at *the 2004 European Transport Conference*

Burgess, A., Snelder, M., Martino, A., Fiorello, D., Bröcker, J., Schneekloth, N. & Rudzi-kaite, L. (2006). TRANS-TOOLS (TOOLS for Transport forecasting ANd Scenario testing) Deliverable 1. Funded by 6th Framework RTD Programme. TNO Inro, Delft, Netherlands.

Cambridge Systematics. (2008). National travel demand forecasting model phase I final scope. *NCHRP Project* (2008): 08-36. Retrieved from http://www.camsys.com/pubs/NCHRP08-36-70.pdf

Census Bureau. (2008). Public Use Microdata Sample, 2000 Census of Population and Housing-Technical Documentation. https://www.census.gov/prod/cen2000/doc/pums.pdf

Davidson, P., & Clarke, P. (2004). Preparation of OD Matrices for DATELINE. Paper presented at *the 2004 European Transport Conference*.

Epstein, J. M., Parker, J., Cummings, D., & Hammond, R. A. (2008). Coupled Contagion Dynamics of Fear and Disease: Mathematical and Computational Explorations. *PLoS One*, Vol. 3, No. 12, p. e3955.

Federal Highway Administration. (2013). Understanding Long-Distance Traveler Behavior - Supporting a Long-Distance Passenger Travel Demand Model. FHWA-HRT-13-095.

Fosgerau, M. (2002). PETRA—An Activity-based Approach to Travel Demand Analysis. In *National Transport Models* (pp. 134-145). Springer Berlin Heidelberg.

Gaudry, M. (2001). Test of Nonlinearity, Modal Captivity And Spatial Competition Within the STEMM Multicountry Applications for Passengers. In Lundqvist, L., & Mattsson, L.G. (Eds.), *National Transport Models: Recent Developments and Prospects*, Springer, New York.

Grush, W. (1998). Usage and Vehicle Miles of Travel (VMT) per Capita. *Highway Information Quarterly*, 5(4).

Gunn, H.F. (2001a). An Overview of European National Models. In Lundqvist, L., & Mattsson, L.G. (Eds.), *National Transport Models: Recent Developments and Prospects*, Springer, New York.

HCG and TOI. (1990). A Model System to Predict Fuel Use And Emissions from Private Travel in Norway from 1985 to 2025. *Report to the Norwegian Ministry of Transport*. Hague Consulting Group.

Hofman, F. (2001). Application Areas for the Dutch National Model. In Lundqvist, L., & Mattsson, L.G. (Eds.), *National Transport Models: Recent Developments and Prospects*, Springer, New York.

Hwang, H.L. & Rollow, J. (2000). Data Processing Procedures and Methodology for Estimating Trip Distances for the 1995 American Travel Survey (ATS). http://cta.ornl.gov/cta/Publications/Reports/ORNL_TM_2000_141.pdf

Koppelman, F.S., & Sethi, V. (2005). Incorporating Variance and Covariance Heterogeneity in the Generalized Nested Logit model: An Application to Modeling Long Distance Travel Choice Behavior. *Transportation Research Part B*, Vol. 39, No. 9, pp. 825-853.

Lundqvist, L., & Mattsson, L. G. (Eds.). (2001). *National Transport Models: Recent Developments and Prospects*. Springer.

Mannering, F.L. (1983). An econometric analysis of vihicle use in multivehicle households. *Transportation Research Part A*, Vol. 17, No. 3, pp. 183-189.

Peter Davidson Consultancy. (2000). MYSTIC Toward Origin-Destination Matrices for Europe. London.

Van Nostrand, C., Sivaraman, V., & Pinjari, A. R. (2013). Analysis of Long-Distance Vacation Travel Demand in the United States: A Multiple Discrete-Continuous Choice Framework. *Transportation*, Vol. 40, No. 1, pp. 151-171.

Weiner, E. (1976). Assessing National Urban Transportation Policy Alternatives. *Transportation Research*, Vol. 10, No. 3, pp. 159-178.

Winston, C. (1985). Research on Intercity Freight and Passenger Transportation: An Economist's Perspective. *Transportation Research Part A*, Vol. 19, No. 5-6, pp. 491-494.

Widlert, S. (2001). National Models: How to Make It Happen. The Case of the Swedish National Model System: SAMPERS. In Lundqvist, L., & Mattsson, L.G. (Eds.), *National Transport Models: Recent Developments and Prospects*, Springer, New York.

Williams, I. (2001). Designing the STREAMS model of Europe. In Lundqvist, L., & Mattsson, L.G. (Eds.), *National Transport Models: Recent Developments and Prospects*, Springer, New York.

Zhang, L., Southworth, F., Xiong, C., & Sonnenberg, A. (2012). Methodological Options and Data Sources for the Development of Long-Distance Passenger Travel Demand Models: A Comprehensive Review. *Transport Reviews*, Vol. 32, No. 4, pp. 399-433.

Zhang, L., Xiong, C., & Berger, K. (2010). Multimodal Inter-Regional Origin-Destination Demand Estimation: A Review of Methodologies and Their Applicability to National-Level Travel Analysis in the US. In *the World Conference on Transport Research*, Lisbon, Portugal.