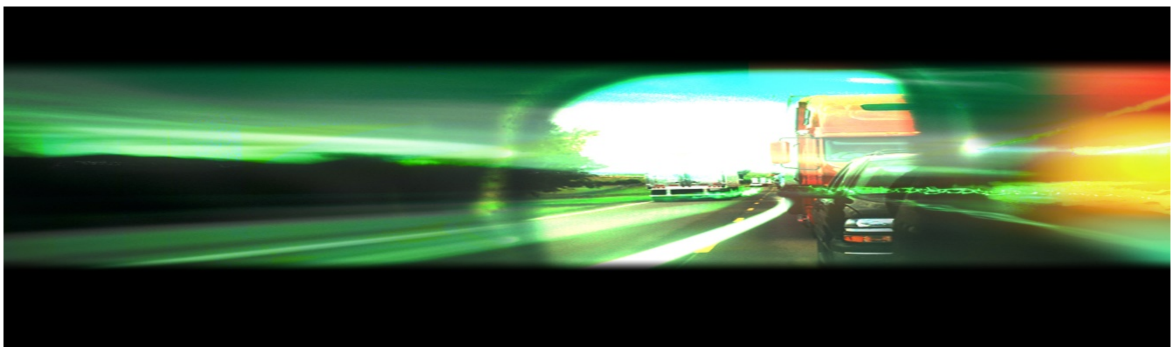
REDUCING ENERGY USE AND EMISSIONS THROUGH INNOVATIVE TECHNOLOGIES AND COMMUNITY DESIGNS

Final Report



TranLIVE

Asad Khattak & Xin Wang

May 2016

DISCLAIMER

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Reducing Energy use and Emissions through Innovative Technologies and Community designs		5. Report Date May 2016
		6. Source Organization Code
7. Author(s) Asad Khattak & Xin Wang		8. Source Organization Report No.
9. Performing Organization Name and Address Department of Civil and Environmental Engineering The University of Tennessee, Knoxville Knoxville, TN 37996-2313		10. Work Unit No. (TRAIS)
Source Organization Name and Address NIATT & TranLIVE University of Idaho 875 Perimeter Dr. MS0901 Moscow, ID 83844-0901		11. Contract or Grant No. DTRT12GUTC17
12. Sponsoring Agency Name and Address US Department of Transportation Research and Special Programs Administration 400 7th Street SW Washington, DC 20509-0001		13. Type of Report and Period Covered Final Report: 2012-2016
		14. Sponsoring Agency Code USDOT/RSPA/DIR-1
15. Supplementary Notes:		

16. Abstract: This project aimed to quantify the impacts of growth and technology strategies at the regional level by using modeling, simulation, and visualization tools, with the overall goal of enhancing livability and sustainability. A key research outcome is the application of a modeling and simulation framework that covers the role of smart growth and intelligent transportation systems on energy and emissions potentially leading to more sustainable urban development strategies and eco-friendly transportation information delivery. The knowledge generated in this study is of interest to researchers and to state and regional planning agencies who aim to reduce energy consumption and emissions. The noticeable outcomes of the project include 1) scientific understanding of a wide range of travel behaviors, e.g., from alternative fuel vehicle use to micro-level driving decisions of accelerations, braking and vehicular jerk, 2) extracting useful information from large-scale data for less volatile driving, and 3) application of tools that encourage alternative mode usage and provide eco-information to travelers. Virginia and other geographic areas in the US are used as test cases, with results documented in this technical report, academic papers, and presentations at transportation conferences. The outcomes of the project include the building blocks of planning and analysis tools that encourage the consideration, evaluation and implementation of more robust growth and eco-friendly transportation strategies.

17. Key Words Sustainable transportation, smart growth, intelligent transportation systems, eco-routing, big data	18. Distribution Statement Unrestricted; Document is available to the public through the National Technical Information Service; Springfield, VT.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 185	22. Price …

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

EXCUTIVE SUMMARY 2
INTRODUCTION 5
METHODOLOGY 6
Conceptual Framework 6
Modeling and Simulation 7
REGIONAL MACROSCOPIC DEMAND AND GROWTH ANALYSIS 8
PART 1: IS SMART GROWTH ASSOCIATED WITH REDUCTIONS IN CARBON DIOXIDE EMISSIONS? 8
PART 2: ACTIVITY SEQUENCING BEHAVIOR FROM A REGIONAL PERSPECTIVE 28
A COMPARATIVE STUDY OF TRANSIT-ORIENTED DEVELOPMENT RESIDENTS AND AUTO-ORIENTED DEVELOPMENT RESIDENTS 28
PART 3: MACROSCOPIC DEMAND MODEL 48
REGIONAL MICROSCOPIC TRANSPORTATION SYSTEMS ANALYSIS 63
PART 1: WHAT IS THE LEVEL OF VOLATILITY IN INSTANTANEOUS DRIVING DECISIONS? 63
PART 2: THE ROLE OF ALTERNATIVE FUEL VEHICLES: USING BEHAVIORAL AND SENSOR DATA TO MODEL HIERARCHIES IN TRAVEL 94
PART 3: ARE HOV/ECO-LANES A SUSTAINABLE OPTION TO REDUCING EMISSIONS IN A MEDIUM-SIZED EUROPEAN CITY? 126
APPLICATIONS AND CASE STUDIES 127
PART 1: A COMPARATIVE STUDY OF DRIVING PERFORMANCE IN METROPOLITAN REGIONS USING LARGE-SCALE VEHICLE TRAJECTORY DATA 127
PART 2: SUPPORTING CALMER INSTANTANEOUS DRIVING DECISIONS: USE OF VEHICLE TRAJECTORY DATA TO GENERATE ALERTS AND WARNINGS 157
PART 3: GENERATING EMISSIONS INFORMATION FOR ROUTE SELECTION: EXPERIMENTAL MONITORING AND ROUTES CHARACTERIZATION 180
CONCLUSIONS 181
APPENDIX-PROJECT PUBLICATIONS AND PRESENTATIONS 184

EXCUTIVE SUMMARY

Consumption of fossil fuels and the generation of local vehicle emissions (that include CO, NO_x, and VOCs) pose challenges to sustainability of cities, with global emissions (CO_2 and other tailpipe gases) becoming a key societal concern. Notably, the transportation sector is the second largest source of greenhouse gas (GHG) emissions with a national average of 28%. Reductions in GHG emissions are on the agenda of various states, e.g., Executive Order 59 in Virginia has set a goal of 30% reduction in GHG emissions by 2025. Also, managing energy use and achieving energy independence is important for the entire nation. Multiple strategies can help reduce transportation energy use and emissions, including:

- Land use/growth management, e.g., implementing smart growth strategies that result in compact and transit oriented developments vs. business-as-usual often associated with sprawl.
- Implementation of Intelligent transportation system technologies, e.g., eco-friendly technologies that include generation and distribution of information on eco-friendly routes for travelers, so drivers will choose routes that minimize fuel consumption and emissions.
- Encouraging adoption and use of alternative fuel vehicles that can directly reduce dependence on fossil fuels.

The above strategies can be implemented in different combinations. Therefore, a key research question is: which strategies or combinations of strategies can achieve the best outcomes more efficiently and effectively? To understand impacts at the behavioral and network level, a comprehensive modeling, simulation and visualization framework is needed. The framework should be able to assess the long-term impacts and implement growth and technology strategies jointly at the regional level. This project aims to quantify and understand the impacts of these strategies by applying appropriate methods to large-scale data and developing various tools, with the overall goal of enhancing sustainability, i.e., lowering energy use and emissions.

An illustrative study conducted under this project identifies behavioral and sensor data and then applies analysis techniques to extract environmentally relevant information that can be provided to travelers in the form of alerts and warnings. Over the life of this project, several research papers were written and presented at national conferences. This report synthesizes key efforts that were undertaken by the project team. Figure 1 presents a structure of the efforts that are discussed in the report.

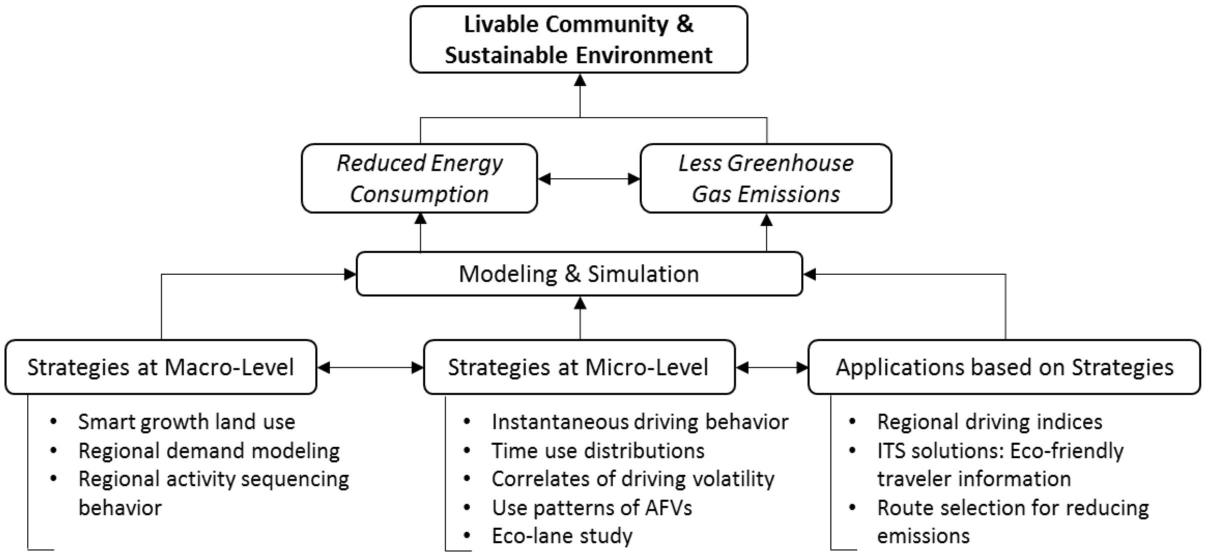


Figure 1 Report Structure.

The project addresses key goals of the TranLIVE University Transportation Center that are related to creating livable and sustainable communities. The objectives of this study are to reduce energy consumption and lower vehicular emissions by:

- Developing modeling, simulation, and visualization tools that support transportation decision making.
- Integrating data and advanced transportation applications to minimize energy and environmental impacts.

To achieve these goals, building blocks are developed and used to assess the long-term impacts of growth and technology strategies at the appropriate regional and national levels. A major effort undertaken in this project relates to understanding energy and emissions impacts of strategies at the macro- and micro-levels; and to develop applications that can achieve the objectives of this project. The core of this report consists of three components:

1) Strategies applied at the macro-level: This effort focusses on land use and smart growth strategies. First, we discuss whether smart growth land use strategy is associated with reductions in CO_2 emissions, through modeling of National Household Travel Survey (NHTS) data. Second, we obtain smart growth scenario data and use the Hampton Roads, Virginia (nearly 1.6 million individuals), macro model in TransCAD to conduct regional macroscopic analysis focusing on smart growth scenarios. Part of smart growth strategies are Transit Oriented Developments. Therefore, we further assess the impacts of TODs by comparing travel behavior of people who live in TODs with people residing in traditional communities. To develop deeper understanding, TOD travel time reliability as well as activity sequencing behaviors is compared with residents of auto-oriented developments, using behavioral data from Virginia.
2) Strategies applied at the micro-level: In this effort, travel behavior at the microscopic level is investigated in the context of developing eco-friendly technology strategies. First, instantaneous driving decisions of drivers on various journeys are analyzed. The decision to

maintain or change speed, accelerations/decelerations has implications for energy consumption and tailpipe emissions. A behavioral framework based on “driving volatility” characterized by hard braking and accelerations was developed. A large-scale GPS travel survey data (containing 51,371 trips and their associated second-by-second total 36 million seconds) from Atlanta, GA was used to study instantaneous driving decisions. Correlates of driving volatility in the Atlanta metropolitan area are quantified, showing that driving volatility varies significantly by gender, age, and trip attributes, e.g. trip lengths. Second, we highlight the role of alternative fuel vehicles (AFVs) in travel. By estimating hierarchical models, important questions about the use patterns of early AFV adopters (e.g., trip frequency and daily vehicle miles traveled) and their driving practices compared with conventional vehicles are explored. Finally, an integrated micro-scale modeling platform calibrated with real world data was developed to assess both traffic and emissions impacts of traffic management strategies. The choice of a route and smoother driving styles are shown to affect energy consumption and pollutant emissions. The trade-off between reducing CO_2/fuel consumption and local pollutants is discussed.

3) Applications and products related to strategies: Given the strategies mentioned above, potential applications and products that can help reduce energy use and emissions are discussed. First, by analyzing driving behavior data, a regional index was created to capture the differences in driving patterns across regions. This index can be used by practitioners to compare the daily performance of a region based on second-by-second vehicle trajectory data. To demonstrate the use of this index, four metropolitan areas were compared including Atlanta, Sacramento, San Francisco, and Los Angeles. Regional profiles of time use and driving performance (acceleration, deceleration and constant speed) are captured and compared. Different driving practices were found in these metropolitan areas given differences in development density, road network structures, and socio-economics. Second, we propose applications, to assist drivers with alerts and warning when they accelerate or brake hard, provided through advanced traveler information systems. In a related collaborative study conducted in Portugal, a methodology to generate information about emissions and other route characteristics for drivers faced with a choice of routes was generated. GPS equipped-vehicles were used to traverse various paths between origins and destinations to collect second-by-second trajectory data required for micro-scale emissions analysis. Selection of eco-friendly routes is discussed when information about emissions and fuel economy is provided to travelers.

The products of this research enhance the building blocks needed for research and education in sustainable systems using smart growth and intelligent transportation systems. The project has targeted a diverse body of students by involving them in technical aspects of the research. Research activities undertaken were disseminated to a broad audience through conferences, peer-reviewed publications, and mass/social-media outlets. Specifically, the project has generated many international conference presentations and several refereed papers in high-impact journals, contributing to scientific knowledge about livable and sustainable communities-see Appendix for details of project-related papers and presentations. The list of authors indicates that several individuals have contributed substantially to this project.

INTRODUCTION

According to the US Environmental Protection Agency, Greenhouse Gas (GHG) emissions from transportation have increased by about 18% since 1990. Increased travel demand along with the stagnation in fuel efficiency in vehicles has largely led to higher greenhouse gas emissions. In 2011, greenhouse gas emissions from transportation accounted for about 28% of the total U.S. emissions, making it the second largest contributor of such emissions after the Electricity sector. A large amount of greenhouse gas emissions come from combustion of petroleum-based products in internal combustion engines, commonly seen in passenger vehicles and light-duty trucks. Additionally, the use of petroleum poses energy security risks to the US. Communities have been developing ways to reduce emissions in transportation sector, including applying smart growth land use strategies, changing drivers' behavior by using intelligent transportation systems or in-car eco-driving devices, and promoting new vehicle and alternative fuel technologies. At the same time, Global Positioning Systems (GPS) and sensor technologies have become ubiquitous generating "big data" and enabling more targeted advice to users who are willing to share their travel patterns and behaviors.

To take advantage of these opportunities and examine land use, transportation, as well as energy and emissions strategies comprehensively, this study has the following specific objectives and tasks:

- Perform a comprehensive literature review to document what is known about various strategies that can potentially reduce energy consumption and emissions.
- Analyze smart/compact growth strategies using behavioral data to understand relationships of smart growth and its impacts on travel decisions as well as network performance/emissions. A deeper understanding of Transit Oriented Developments provides the basis for developing smart growth scenarios.
- Analysis of big data to understand the role of intelligent transportation systems, especially eco-friendly traveler information. This work relates to generating information that supports instantaneous driving decisions, i.e., emissions information that can help users select more eco-friendly routes.
- Investigation of alternative fuel vehicle use by early adopters based on behavioral and sensor data. This work is based on analysis of trips, vehicles, and drivers in an integrated analytical framework.
- Transforming big data coming in from sensors to create regional indices for better understanding how people drive in a region and over time, given different land use and local contextual conditions. This work further examines the role of HOV/eco-lanes as a sustainable option to reducing emissions.

METHODOLOGY

This chapter presents the methodologies used for modeling and simulation. A conceptual framework for better understanding the roles of travel behavior, especially pre-trip decisions and eco-route decisions is presented.

Conceptual Framework

Travelers' decisions impact their transportation emissions outcomes, so it is important to understand critical decision points before and during trips and access whether, how, and to what extent these decisions may affect energy use and transportation emissions. The behavioral model, as illustrated in Figure 2, conceptualizes the routine decisions by travelers that may change as a result of relevant intelligent transportation systems (eco-information, eco-signals, etc.) and compact growth strategies. For example, people may change their residential location decisions or mobility decisions in terms of the kind and number of vehicles they own; they may change pre-trip decisions such as mode shift to public transportation or walking, or cancel their trips; or they may change route decisions by diverting to alternate routes in response to incidents or do more trip chaining; and/or they may change their driving decisions that include speed limit compliance and lane selection.

The changes in behavior can result in VMT (vehicle miles traveled), gas consumption, and emissions reductions. Notably, these measures are more sensitive to pre-trip and en-route decisions, but there is a need to explore additional performance measures that can capture information from GPS and other vehicle-based sensors that have emerged recently.

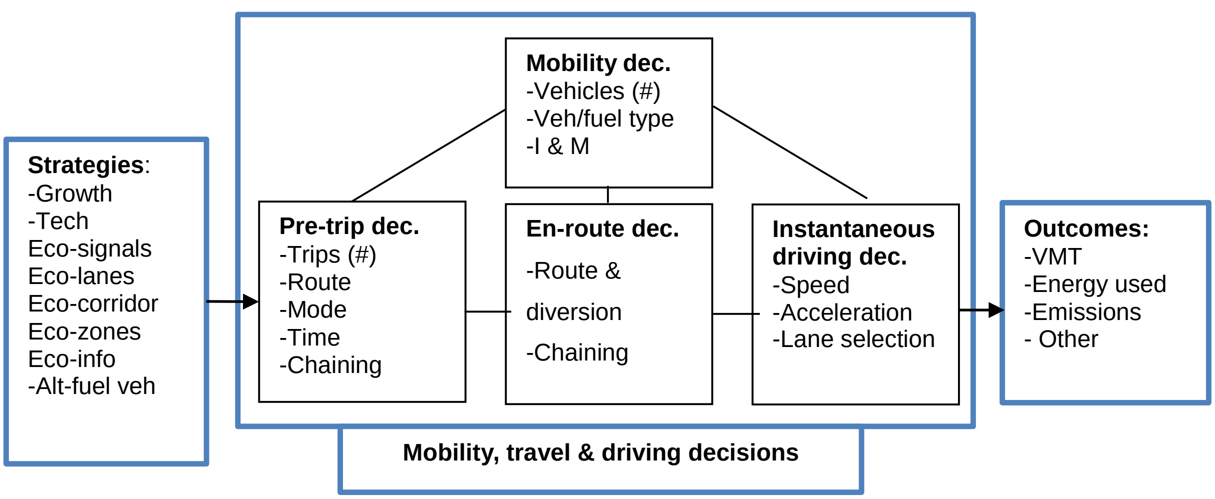


Figure 2: Eco-friendly strategies, travel behavior, and environmental outcomes.

Modeling and Simulation

A key focus of the work is to develop the building blocks for land use, transportation, and energy and emissions models. In this regard, behavioral models are a key focus. Such models are incorporated in network models, where possible to demonstrate enriched behavioral basis for network models and simulations, as shown in Figure 3 below. Macroscopic and microscopic transportation models are used as test-cases to study the impacts of intelligent transportation technologies, compact growth strategies and alternative fuel vehicles. An effort was made to connect transportation models with emissions models to obtain a clearer picture of local emissions and global emissions associated with intelligent transportation and compact growth strategies. The interoperability issues are complex and outside the scope of the work conducted under this project.

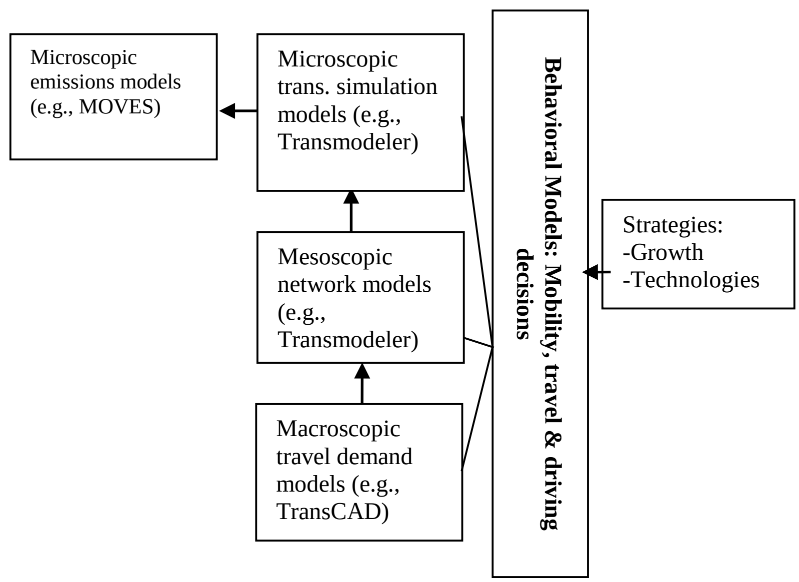


Figure 3: Integrating transportation models with emissions models.

REGIONAL MACROSCOPIC DEMAND AND GROWTH ANALYSIS

PART 1: IS SMART GROWTH ASSOCIATED WITH REDUCTIONS IN CARBON DIOXIDE EMISSIONS?

Xin Wang[1], Asad Khattak[2], and Yichi Zhang[3]

[1]Virginia Department of Transportation, [2]The University of Tennessee, Knoxville, [3]Old Dominion University

Abstract: The transportation sector is the second largest contributor to human generated carbon dioxide (CO_2) emissions. A key goal of the U.S. Department of Transportation is to implement environmentally sustainable policies that can reduce carbon emissions from transportation sources. Smart growth—characterized by compact, mixed use; greater network connectivity; and environments friendly to alternative modes—may encourage reductions in vehicle travel and emissions. A better understanding of travel behavior in conventional and smart growth communities is needed to inform policies. A behavioral data set is analyzed to determine whether smart growth developments are associated with lower CO_2 emissions. Sample selection models are estimated from a 2009 travel behavior survey of 15,213 households to capture the conditionality of emissions on the decision to drive (or not) by household members on a given day. Results indicated that 12% of responding households used alternative modes or did not travel from home; the rest of the sample traveled in an automobile and therefore contributed to CO_2 emissions. CO_2 emissions were calculated from vehicle miles traveled and the fuel efficiency of the vehicle used for specific trips taken by household members. The developed framework models whether CO2 emissions are associated with land use, socio demographics, and preferences for adopting information technology. Tailpipe CO_2 emissions are lower for households that reside in mixed land use neighborhoods with good network connections (on the order of 9%). As a long-term strategy, CO_2 emissions reductions from smart growth developments can be substantial.

Key words: Emissions, smart growth, Greenhouse Gas emissions, built environment, CO_2

1. INTRODUCTION

According to the United States Environmental Protection Agency, 1,745.5 Million Metric Tons CO_2 comes from transportation sources, accounting for 28% of such emissions. A key goal of the United States Department of Transportation is to implement environmentally sustainable policies that can reduce carbon emissions from transportation sources. The amount of carbon in the atmosphere has steadily increased over the years, and the increases are associated with the greenhouse effect, global warming, and to a lesser extent human health. Consistent with federal aims, Virginia aims at 30% reduction in CO_2 emissions by 2025. Reducing carbon emissions from the transportation sector is both important and urgent. Policy solutions can include higher corporate average fuel efficiency standards, promoting alternative fuel technologies, and smart growth-considering that land is being consumed for development at a rate almost three times faster than population growth, which has caused CO_2 emissions from vehicles to rise (Ewing and

Anderson 2008). Smart growth is characterized by higher density, mixed land use, greater network connectivity, and alternative mode friendliness. This study quantifies the associations of compact, mixed use, connected and alternative mode friendly developments on tailpipe emissions of carbon dioxide (CO_2); this is achieved by analyzing behavioral data and estimating a behavioral model. Emissions are calculated based on household level VMT (Vehicle Miles Traveled) combined with MPG (Miles per Gallon) for vehicles used for specific trips. Such detailed information (about which household vehicle is used for a specific trip) is provided by the latest National Household Travel Survey (NHTS). Relatively fine spatial resolution is used to generate local land use around residential locations, which are then used as correlates to estimate the relationships between smart growth developments and household level vehicle emissions while controlling for household demographic factors.

2. LITERATURE REVIEW

From a long-term perspective, whether smart growth developments are associated with lower CO_2 emissions was answered using simulations and empirical studies. Simulations based on link level trip assignment explore the links between VMT and emissions, using contemporary emissions models that include MOBILE6/MOVES, CMEM. Additionally, to test whether smart growth is associated with reduction of emissions, travel demand models have been used in addition to statistical techniques of estimating regressions based on behavioral data. In general, compact developments are associated with less travel and 5% to 11% lower CO_2 emissions. TOD (Transit Oriented Developments) can even reduce CO_2 emissions by 47.5% (Tiwari, Cervero et al. 2011). Reductions in emissions are equivalent to using alternative vehicle technologies (Stone Jr, Mednick et al. 2009) or changing driving behavior based on eco-driving information (Xia, Boriboonsomsin et al. 2013, Barth and Boriboonsomsin 2009).

Most of the studies concentrate on long-term CO_2 reductions, e.g., in next 30 years (Meyer, Burbank et al. 2013). Typically, smart/compact growth scenarios were compared with no-change scenarios (business-as-usual or current trend) and scenarios with different market penetration of alternative technologies (Stone Jr, Mednick et al. 2009,Rodríguez, Morton et al. 2011,Ewing, Nelson et al. 2009, Niemeier, Bai et al. 2011, Hennessy and Tynan 2012, Hankey and Marshall 2010). Most studies suggest that smart/compact growth is associated with lower vehicle miles traveled and CO_2 emissions. A study by Stone et al. (Stone Jr, Mednick et al. 2009) suggested decline of approximately 8% in CO_2 emissions as a result in more aggressive smart growth compared with business-as-usual scenario. Rodriguez et al. (Rodríguez, Morton et al. 2011) suggested compact development without penetration of alternative technologies is associated with decreases CO_2 emissions by 7.1% relative to Business-As-Usual in 2050. Niemeier et al. (Niemeier, Bai et al. 2011) suggested that total emissions (total organic gases, CO, NO_x, PM) for controlled growth scenario is 6%-10% lower than the baseline growth scenario, but the effects of future land-use growth patterns may vary among differently sized spatial areas. Hennessy & Tynan (Hennessy and Tynan 2012) concluded that CO_2 emission would be reduced by approximately 11% if aggressive growth ratio is implemented. Studies also suggested that compact growth could achieve long-term emission reductions equivalent to or higher than using alternative technologies (Stone Jr, Mednick et al. 2009, Rodríguez, Morton et al. 2011).

Multivariate statistical methods have been used to explore the associations between land use and transportation emissions based on current travel behavior data. Frank et al. (Frank, Stone et al. 2000) concluded that increases in residential density, intersection density, and mixed use are associated with significant reductions in oxides of nitrogen (NO_x), hydrocarbons (HC), and carbon monoxide (CO), while controlling for socio-demographic factors. The land use characteristics are based on relatively large spatial units—the census tract. A more recent paper by Frank et al. (Frank, Chapman et al. 2005) estimated multivariate linear regression models for daily household level vehicle HC and NOx emissions based on creating land use variables at one kilometer residential buffer level. The study concluded that the mix of high population density, street connectivity, and land use index were negatively associated with HC and NO_x in the Seattle metropolitan area. Also, the land use variables remained significant predictors of emissions even after vehicle miles of travel was included in the models; that is to say land use is not only associated with emissions through its relationship with travel demand (VMT), but also directly. A limitation of the study is the use of OLS (Ordinary Least Square) regression model for emissions or VMT, given that emissions are usually not normally distributed and more importantly, there can be a substantial number of zero observations in the sample. While log transform may relieve the first problem, it will not help the latter one. Vehicle emissions as a result of driving, is conditional on travelers decision to drive. Therefore, it is important to capture whether a household drives on a particular day and if so the amount of emissions produced. When conditionality was ignored from OLS model, the associations between land use and emissions can be underestimated. Table 1 shows a summary of relevant studies.

One benefit of simulations is that link-level vehicular flows can be extracted to calculate emissions, which provides needed details to accurately estimate emissions, e.g., speeds, acceleration, and congestion. Also, it can be used to anticipate future scenarios with innovative technology adoption such as eco-routing systems (Bandeira, Almeida et al. 2013, Guo, Huang et al. 2013), dynamic eco-driving technology (Xia, Boriboonsomsin et al. 2013) and intelligent transportation systems (Gkritza and Karlaftis 2013) based optimal emission pricing (Sharma and Mishra 2013). However, numerous assumptions must be made regarding travel behavior, VMT level, vehicles, fuel usage, regional vehicles age distribution and information technology's market penetration in the future. This brings in high levels of uncertainty, especially when the planning horizons are 20 to 40 years in the future.

Current emissions studies commonly measure regional emissions based on link level traffic assignment results (Rodríguez, Morton et al. 2011, Niemeier, Bai et al. 2011). However, the link between emissions and the households who produce them is not well understood or explained. The household decision of whether or not to drive on a particular day, and if they drive, how much emissions they generate (referred to as self-selection), is unexplored. Additionally, behavioral data regarding the vehicle used for a specific trip was generally not available. This paper attempts to fill in these gaps.

3. HYPOTHESES

The built environment is critical to how much people drive and hence the emissions that are produced. This study emphasizes the conditional link between two processes: the decision to drive or not and the extent of resulting CO_2 emissions, given that household members drive (Figure 1).

Conditionality exists because personal vehicle-based CO_2 emissions are produced if the traveler decides to drive or be driven on a particular day. If the person does not travel in a vehicle, then no emissions are produced by the personal vehicle. Therefore, a two-stage process exists in the behavioral dataset: travel by personal vehicle or not on a day, and if driving is chosen, how much emissions will be produced by driving.

Factors associated with these two decision processes are conceptually different. Travelers' decision to travel by personal vehicle or not is likely associated with household socio-demographics, and the need to participate in activities (for work, shopping or social recreational purposes). Given decision to drive, emissions produced are likely associated with local land use variables, socio-demographics, driver related factors, and living preferences. Some household socio-demographic can be associated with both processes, e.g., household income, but not all of them are likely to have a direct impact on both processes. Household has children younger than 16 years can directly impact a family's driving needs, e.g. the need to drop-off or pick-up children from school. Therefore, they are included in the binary model only. Factors that may be (theoretically) weakly associated with the self-selection issue are not used in the binary drive-or-not model, e.g., local land use variables.

Factors related with emissions may include local land use and contextual characteristics, driver related attributes, and living preference variables such as use of information technology and working from home. Residential density, land use mix, and local roadway connectivity are important indicators of smart growth residential neighborhoods. This is because higher mix of residential and commercial land use can increase the ease of satisfying activity participations needs with fewer and/or shorter trips, enhancing the work-life balance which in turn can lower VMT and decrease the transportation emissions. Driver factors can have a direct relationship with emissions. These variables include the number of drivers at home, vehicles available, frequency of automobile trips and how much the vehicles are driven. Studies have shown empirical evidence linking driving behavior with emissions, e.g., if they drive aggressively, then higher emissions are expected (Cervero and Duncan 2003). Considering data availability and the fact that the model is based on household level, the age of household head is used as explanatory variable. For instance, generally a household with an average age of 60 years can drive differently compared with a household with average age of 30 consisting of only young couples.

The discussion above leads to two specific hypotheses to be tested in this paper: 1) travelers' drive or not is associated with household socio-demographics, and need to participate in activities (for work, shopping or social recreational purposes); 2) given the decision to drive, emissions produced is correlated with local land use variables, as well as socio-demographics, driver related factors, and living preferences.

4. MODEL STRUCTURE

To capture the two-stage process in the behavioral database, i.e., drive or not and how much emissions are produced if driving, regression techniques such as Tobit model and Heckman sample selection model are appropriate. Heckman selection model is used in this study given a limitation of the Tobit model that assumes the same variables are associated with both processes. Based on the conceptual structure, a set of common as well as unique set of variables are likely associated with the two stages. The Heckman selection model separates the model into two parts: A linear regression model for emissions and a binary probit model for the decision to drive. The detailed model specification is presented below:

$$y = \boldsymbol{\beta}'\boldsymbol{x} + \varepsilon \quad \text{(linear regression model)} \tag{1}$$

$$z^* = \boldsymbol{\alpha}'\boldsymbol{v} + u \quad \text{(binary probit model)} \tag{2}$$

In the model, y is how much emissions are produced by driving a personal vehicle), which is observed only if a criterion, $z = 1$ is met (the traveler drove). The dependent variable y is related to independent variables $\boldsymbol{x}$ with the error term ε. And z^* is unobserved; it can be estimated by independent variables $\boldsymbol{v}$, with the error term, u. Here z^* is not observed and has an observed counterpart z, which is determined by:

$$z = 1 \text{ if } z^* > 0 \text{ and } z = 0 \text{ if } z^* \leq 0 \tag{3}$$

$\boldsymbol{x}$ and $\boldsymbol{v}$ are two sets of explanatory variables. In this paper, a demographic variable, household income, is common in both $\boldsymbol{x}$ and $\boldsymbol{v}$; variables related with the needs for working, and whether household has children at home are used in $\boldsymbol{v}$; variables capturing local land use, living preference, and driver factors are only included in $\boldsymbol{x}$. The model reports an index, ρ, to represent the correlation between the unobserved variables in the two equations. A statistically significant estimate for ρ indicates that modeling the two processes simultaneously is superior to modeling them separately.

5. DATA DESCRIPTION

5.1 Data Sources

The behavioral data analyzed in the paper are from the Virginia Add-on survey of NHTS conducted in 2009 (survey period covered April 2008 through May 2009). A total of 15,213 households were included in the sample after cleaning out data with unclear location information. The sample covers about 12 CBSAs (Core Based Statistical Areas) in Virginia, representing various land use types and populations. To ensure that the survey is representative of the State’s population (approximately 8 million) and their daily activity-travel patterns, geographic and demographic distribution goals for the sample were achieved. The survey relies on the willingness of households to 1) provide demographic information about the household, its members and its vehicles, and 2) have all household members record all travel-related details for a specific 24-hour period, including information about the miles traveled and the vehicle used for each trip. The database has detailed fuel economy compared with previous surveys, including MPG for the vehicles owned and gasoline price information. The database contains three different levels of data: personal data; household data, and trip data. The survey’s response rate is reasonably high at 25%.

A unique database was created by adding other data from various sources, including: 1) Publicly maintained roadway centerline shape files (Tiger files); 2) population data from Census 2010; 3) employment from the 2009 LEHD (Longitudinal Employer-Household Dynamics) database. Both the population and employment data are at the Census block level which is the finest geographic unit available for public use, providing reasonable details about residences and employment. The employment data in LEHD also provides employment information based on NAICS (North

American Industry Classification System) categories, which is very useful for grouping employment by different type which was used as a proxy for land use types.

5.2 Calculating Key Variables

CO_2 emissions

The CO_2 emissions were calculated for each household using VMT data from NHTS. A key improvement in the 2009 NHTS is that it provides information on MPG (miles per gasoline-equivalent gallon estimate) for each vehicle driven on a particular trip. Note that vehicles with alternative fuels are considered in this calculation since their MPGs are substantially different compared with regular gasoline vehicles. The amount of CO_2 created from burning one gallon of fuel depends on the amount of carbon in the fuel. EPA suggests using 8,887 grams of CO_2/gallon. The CO_2 emissions are calculated by using the equation recommended by EPA:

$$CO_2 \text{ emissions per mile} = \frac{CO_2 \text{ per gallon}}{\text{MPG}} = \frac{8887}{\text{MPG}} \text{ grams}$$

The total CO_2 emissions for each trip are equal to CO_2 emissions per mile multiplied by VMT for each trip, and then the emissions for each trip are aggregated to the household level.

Mix of land use

Entropy was used as an index to represent land use diversity. This index is used frequently to define the mixture of employment and residences (Kockelman 1997), land use balance (Brownson, Hoehner et al. 2009), and mixture of different types of land uses (Frank, Chapman et al. 2005). In this study, the entropy score is a normalized index which varies between 0 and 1, with 1 signifying maximally mixed and full balance of land uses, and 0 signifying a homogeneous used. Seven land use types are considered in the calculation of land use mix entropy index, including residential, commercial, service, office, institutional, industrial, and agricultural land uses. The employment densities for each of these land uses were exacted from Census LEHD and calculated at the Census block level. There are two types of land use entropy, "balanced" and "blended." Higher values on balanced entropy represents neighborhoods with a balanced combination of residential, commercial, service, office and institutional land uses, which can provide better access for activity participation. Blended entropy represents neighborhoods where residential land use is mixed with industrial and agricultural land use. Higher values of blended entropy will be typically associated with lower levels of accessibility (especially alternative mode connections). The entropies are calculated in the following equation:

$$\text{Entropy} = - \sum_j \frac{P_j \times \ln(P_j)}{\ln(J)}$$

Where,

P_j is the proportion of density in the j^{th} land use type;

J is the number of land use types, which is equal to 5 for balanced entropy and 3 for blended entropy.

Measures of local transportation network

Studies of spatial analysis, often use circular buffers around residences to measure land use around a residence, including buffers with a radius from 0.25 miles to 1 mile (Brownson, Hoehner et al. 2009). Note that the NHTS and its add-on data do not release the exact location of each residence. Therefore, to obtain accurate residential buffers, a geo-imputation method was applied to assign each residence a synthetic location (latitude and longitude) within their census block (Wang, Khattak et al. 2013). A 0.75-mile buffer (equivalent to a 45-60 minutes walking distance) is created for each synthetic residence (shown in Figure 2) to capture the local land use characteristics in the neighborhood. Research (Wang, Khattak et al. 2013) has shown that using a 0.75-mile buffer around a geo-imputed residence can produce local land use variables with reasonable accuracy. The following network measurements are based on this buffer and are meant to capture accessibility of the local network.

Network Connectivity

Connected Node Ratio (CNR) was selected to represent the connectivity of roadway network. It is calculated using the number of all intersections divided by the number of intersections plus cul-de-sacs. The maximum value is 1.0, which means all the intersections in a buffer were 4-way or 3-way intersections; implying that the roadway network is in a perfect grid.

The Link-Node Ratio (LNR) was selected to represent the connectivity of roadway network. Link-Node Ratio is equal to the number of roadway segments divided by the number of intersection or cul-de-sacs within a buffer. A perfect grid has a ratio of 2.5.

Roadway length within residential buffer

Note that CNR and LNR do not account for roadway length in a residential buffer. The roadway length within a buffer can be used to represent the density of roadway in a neighborhood. Also the average length per node was used to represent the separation between nodes, which is a coarse indicator of walkability. Lower roadway length per node is expected to be better for walking while longer roadway length per node is preferred for driving.

5.3 Descriptive Statistics

Table 2 presents the descriptive statistics of variables. On average, the sampled household traveled 74 miles per day, including all vehicles used. 12% of respondents reported that they did not drive on the assigned travel day. For those who drive, the average CO_2 emissions produced are about 29 kilograms (64 pounds). For those who did not drive, zero emissions are produced. Given that the emissions variable is skewed, it was log-transformed before using it as dependent variable.

For household characteristics, the average driver count per household is 1.8, and the average vehicles owned is 2.2. About 4% of households own hybrid vehicles. Analysis shows that the average CO_2 emissions of the households who own hybrid vehicles are slightly higher than others. This result is unexpected. However, further analysis confirms that hybrid vehicle owner households travel more. On average, households who own hybrid vehicles are higher income families, with more drivers at home (2.0 vs. 1.8), making more daily trips (8.5 vs. 7.9 trips per day, on average), and having higher VMT (87 vs. 73 miles per day). This suggests that consumers buying fuel efficient vehicles may have greater travel needs. Given that whether hybrid vehicles were driven has been considered in the MPG (i.e., calculation of CO_2 emissions—the dependent variable), it was not used in the model specification.

On average, 22% of sample households have children younger than 16 years old. Nearly 7 daily trips per household are made by automobile. For work related factors, the average worker per household is about 1, the home to work distance (one way derived distance calculated based on home and work address) between home and work is about 10.8 miles, if there are more than one person working in the family, the maximum distance among all working family members is reported in this measure.

For living preferences, on average, the internet use frequency by all family members is 33 times per month; the purchase delivery frequency by all family members during a month is 2.2 times, and the frequency of working from home during past month is 0.44. These variables to some extent can represent the adoption of technology by households.

In terms of local land use variables at the residential neighborhood level, residents in the sample on average have 17.5 miles roadway in their 0.75 miles circular buffer. The average connected node ratio for local road is 0.72 and the average link node ratio is about 1.4. The average balanced entropy score is about 0.51 and the score of blended entropy is about 0.34. Descriptive statistics also show a substantial variation in residential environments of survey respondents, as indicated by relatively large ranges and standard deviations of the variables.

6. MODEL RESULTS

After examining correlations among variables and conducting several test runs, the final model specification was selected. The variable LNR was dropped from the specification since it is highly correlated with CNR. Two models are shown in Table 3: the Heckman sample selection regression (Model 1), and a conventional OLS regression (Model 2). The Chi-Square test and F-test show that the Heckman and OLS regression models are both statistically significant, overall.

In Model 1, 1842 samples are censored with zero driving trips, producing zero CO_2 emissions from their vehicles. The upper half of the table shows log-transformed OLS regression coefficients; the bottom half of the table shows the probit regression coefficients. The correlation of the residuals in the selection and outcome equations, ρ, is statistically significant (5% level). This reaffirms the value of the sample selection model. That is, by estimating models with two processes, different factors associated with the two processes can be captured separately providing a clearer assessment of CO_2 emissions. The model with sample selection is methodologically sophisticated and provides a more nuanced interpretation. Note that the magnitudes and signs of variables from the OLS Model 2 without sample selection are not close to that from Model 1; therefore, ignoring sample selection in the data may generate misleading conclusions about associations.

Several variables in the binary probit model are notable for answering which factors are associated with driving and hence carbon emissions from personal vehicle use. In line with expectations, the probability of driving on the assigned travel day is higher when a household has higher income, more workers, and more children younger than 16 years. One more worker at home is associated with 9% higher probability of driving; households with children younger than 16 years have a 3% higher probability of driving compared with households that do not have children.

The key question is whether smart growth is correlated with lower CO_2 emissions. Not all the variables used to describe land use are statistically significantly associated with CO_2 emissions. Smart growth neighborhoods are characterized by greater land use mix, higher density, denser roadway networks, and higher connectivity or accessibility by various modes. The results confirm that residences located in smart growth neighborhoods in Virginia are associated with lower CO_2 emissions on average. More specifically, a neighborhood with a full balanced land use mix is associated with 12% lower CO_2 emissions compared with households located in single land use neighborhoods; controlling for other variables, a neighborhood with a grid network (CNR=1) is associated with 14% lower CO_2 emissions compared to a neighborhood with only cul-de-sacs. Figure 3 shows the changes in CO_2 emissions when CNR and balanced land use mix change together. Households located in neighborhoods with most balanced land use mix and highest connectivity (grid network) are associated with 24% lower CO_2 emissions compared with households located in single land use cul-de-sacs only neighborhoods. More realistically, 9% lower CO_2 emissions are expected for residents in most balanced and connected neighborhoods compared with the current average levels of mix and connectivity, where CNR is 0.7 and balanced land use mix is 0.5. No statistically significant association was found between CO_2 emissions and blended land use mix.

The model results also show that CO_2 emissions are negatively associated with roadway length (in the buffer around the households) but are positively associated with roadway length per node; A 0.1 mile increase in average roadway length per node (means larger separation between intersections) in a neighborhood is associated with nearly 2% higher CO_2 emissions by residents, all else being equal. The marginal effect of number of cul-de-sacs is small in terms of magnitude, although it shows a statistically significant association with CO_2 emissions.

In the CO_2 emissions model, positive associations are observed with higher income, more drivers at home, more vehicles owned, and more automobile trips. One additional vehicle owned is associated with 5% higher CO_2 emissions. This percent is 10% for one additional automobile trip and 6% for one additional driver. Since income is the variable used in both the drive and emissions models, the coefficient of income in the emissions model must be adjusted before calculating the percent change in emissions. An inverse mills ratio of income is calculated before taking the exponent of this variable. After adjusting, an additional ten thousand dollars of household income is associated with 4% higher CO_2 emissions. The association between average age of household head and CO_2 emissions is modeled as non-linear, showing an inverted-U shape.

For the relationship between living preferences and CO_2 emissions, internet usage frequency and working from home frequency are both statistically significantly associated with higher CO_2 emissions; while the product delivery frequency does not correlate with CO_2 emissions significantly. Note that working from home and greater internet use frequency are expected to be associated negatively with driving, this result is somewhat counter to expectations. However, possible explanation can be the time saving due to telecommunication can be used to make more out of home activities which can result in more driving trips. Similarly, studies have also suggested the use of ICT (Information and Communications Technologies) may generate additional time use for out-of-home recreational activities therefore increasing trip-making propensity (Wang and Law 2007, Kim and Goulias 2004).

7. LIMITATIONS

The emissions analyzed in this paper is limited for CO_2 emissions only, therefore other emissions (NO_x, VOCs, and CO) are not analyzed. The results are based on a self-reported trip diary from a behavioral survey, where people may not accurately report all their trips. Another limitation of the method used in this study is that traffic conditions, e.g., congestion, vehicle speeds and variations are not taken into account. Due to data availability restrictions, the regional accessibility, local walkability, and public transit variables were not included in the emission model but they remain in the conceptual framework. If such data becomes available for Virginia, then it can be used to potentially improve model specification.

The study is also limited due to cross-sectional nature of the data. As such it represents a snapshot of households during a time of change, with a relatively high unemployment rate and high energy prices. Exploring the dynamics of travel behavior and resulting carbon emissions will need panel type surveys. Finally, the focus is on smart growth developments, which is a long-term strategy, but we recognize that other solutions (e.g., new vehicle technologies and new forms of eco-information) may also effectively deal with the emission problem.

8. CONCLUSIONS

In the context of increasing carbon emissions from transportation and the goal of implementing environmentally sustainable policies that can reduce such emissions, it is important to understand how different strategies perform, i.e., smart growth, behavioral change through eco-information/ eco-driving technology, alternative vehicle technologies, and fuel efficiency standards. This research focused on the influence of land use and associated travel behavior on the production of CO_2 emissions, and explicitly captures the conditionality inherent in calculation of CO_2 emissions based on household members' decision to use personal automobile at the household level. Analysis of behavioral data coupled with land use data allows in-depth investigation of how smart growth developments may be associated with tailpipe CO_2 emissions. Detailed spatial land use characteristics representing the mix of land use, local roadway density and connectivity on a relatively small geographic scale were generated using geographic information systems. To some extent availability of data constrains the model specification and the implications—although key variables were available for use in the model specification.

The answer to the key question of whether smart growth is correlated with lower CO_2 emissions is a "yes." Specifically, the results confirm that several indicators of smart growth developments, especially balanced mix of land use and roadway connectivity, in Virginia are associated with lower CO_2 emissions. A neighborhood with a "well-balanced land use mix" (i.e., good jobs-housing balance but excluding industrial land uses) is associated with 12% lower CO_2 emissions compared with a single land use; 10% additional intersections that are not cul-de-sacs (a 0.1-unit increase in connected node ratio) are associated with a 1.4% decrease in CO_2 emissions, while controlling for other variables. Compared with current average conditions in the Virginia dataset (when CNR is 0.7 and balanced entropy is 0.5), a fully balanced land use mix and a grid network are expected to reduce CO_2 emissions by 9% because mixed use and connected developments are associated with lower household VMT.

These findings (balanced mixed land use associated with 9% lower CO_2 emissions) is similar to the findings in related smart growth literature, e.g., compact growth in future years (2030-2050) can reduce CO_2 by 5% to 11% compared with no-build scenario (Stone Jr, Mednick et al. 2009, Rodríguez, Morton et al. 2011, Ewing, Nelson et al. 2009). Such long-term strategies may be coupled with short-term strategies of applying dynamic eco-driving technology and alternative vehicle technologies, where 10% to 20% reductions in CO_2 can be expected with different penetration rates of technologies and different congestion levels on roadways (Xia, Boriboonsomsin et al. 2013, Barth and Boriboonsomsin 2009); additionally, hybrid electric vehicles are associated with substantial reductions (18%) in CO_2 emissions (Stone Jr, Mednick et al. 2009). However, a key question to be answered by policy makers is whether multiple strategies for reducing carbon emissions can be additive and how can multiple policies be implemented resulting in greater carbon reductions.

9. ACKNOWLEDGEMENT

The research was supported through the Tran*LIVE* University Transportation Center grant sponsored by US the Research and Innovative Transportation Administration, of US DOT. Some of the data was provided by the Transportation Mobility and Planning Division of Virginia Department of Transportation. Special thanks are extended to the following entities for their support: Transportation Research Institute and the Center for Innovative Transportation Solutions at Old Dominion University and the Transportation Engineering & Science Program at The University of Tennessee. The views expressed in this paper are those of the authors, who are responsible for the facts and the accuracy of the data presented herein.

10. REFERENCES

1. Ewing, R., K. Bartholomew, S. Winkelman, J. Walters, and D. Chen, *Growing cooler, the Evidence on Urban Development and Climiate Change*, 2008, Urban Land Institute.
2. Tiwari, R., R. Cervero, and L. Schipper, *Driving CO_2 reduction by Integrating Transport and Urban Design strategies.* Cities, 2011. **28**(5): p. 394-405.
3. Stone, B., A.C. Mednick, T. Holloway, and S.N. Spak, *Mobile Source CO_2 Mitigation through Smart Growth Development and Vehicle Fleet Hybridization.* Environ. Sci. Technol., 2009. **43**(6): p. 1704-1710.
4. Xia, H., K. Boriboonsomsin, and M. Barth, *Dynamic Eco-Driving for Signalized Arterial Corridors and its Indirect Network-Wide Energy/Emissions Benifits.* Journal of Intelligent Transportation Systems Journal: Technology, Planning, and Operations, 2013. **17**(1): p. 31-41.
5. Barth, M. and K. Boriboonsomsin, *Energy and emissions impacts of a freeway-based dynamic eco-driving system.* Transportation Research Part D: Transport and Environment, 2009. **14**(6): p. 400-410.
6. *Project SHRP C09: Incorporating Greenhouse Gas Emissions into the Collaborative Decision-making Framework, Transportation Research Board of The National Academies*. 2011 Available from:

http://www.transportationforcommunities.com/cases/pdf/SHRP%202%20C09%20Draft%20Final%20Project%20Report.pdf. Accessed Nov. 2012.
7. Rodríguez, A., B. Morton, Y. Song, C. Frey, N. Rouphail, A. Khattak, H. Zhai, E. Shay, and T. Hadden-Loh, *Can Compact Growth Decrease Automobile Emissions? Methodology and its Application to Mecklenburg County, NC.* submitted to Journal of American Planning Association, 2011.
8. *Driving and the Built Environment: The Effects of Compact Development on Motorized Travel, Energy Use, and CO_2 Emissions*. TRB Special Report 298, 2009, Washington, DC: The National Academies Press.
9. Niemeier, D., S. Bai, and S. Handy, *The impact of residential growth patterns on vehicle travel and pollutant emissions.* The Journal of Transport and Land Use, 2011. **4**(3): p. 65-80.
10. Hennessy, J. and J. Tynan. *Assessing Upstate Greenhouse Gas Emissions from Transportation Sources Under Changing Land Use Patterns* 2010 Available from: http://www.upstateforever.org/progCAWdocs/VMTGrowthAndGHG.pdf. Accessed Nov. 2012.
11. Hankey, S. and J.D. Marshall, *Impacts of urban form on future US passenger-vehicle greenhouse gas emissions.* Energy Policy, 2010. **38**(9): p. 4880-4887.
12. Frank, L.D., B. Stone Jr, and W. Bachman, *Linking land use with household vehicle emissions in the central puget sound: methodological framework and findings.* Transportation Research Part D: Transport and Environment, 2000. **5**(3): p. 173-196.
13. Frank, L., J. Chapman, M. Bradley, and T.K. Lawton, *Travel Behavior, Emissions, & Land Use Correlation Analysis in the Central Puget Sound*, 2005, Lawrence Frank & Company, Inc.: Point Roberts, WA.
14. Bandeira, J., T.G. Almeida, A.J. Khattak, N.M. Rouphail, and M.C. Coelho, *Generating emissions information for route selection: Experimental monitoring and routes characterization.* Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 2013. **13**(1): p. 3-17.
15. Guo, L., S. Huang, and A.W. Sadek, *An Evaluation of Environmental Benefits of Time-dependent Green Routing in the Greater Buffalo-Niagara Region.* Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 2013. **17**(1): p. 18-30.
16. Gkritza, K. and M.G. Karlaftis, *Editorial: Intelligent Transportation Systems Applications for the Environment and Energy Conservation (Part 1).* Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 2013: p. 1-2.
17. Sharma, S. and S. Mishra, *ITS Enabled Optimal Emission Pricing Models for Reducing Carbon Footprints in a Bi-Modal Network.* Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 2013. **17**(1): p. 54-64.
18. Boriboonsomsin, K., A. Vu, and M. Barth, *Eco-Driving: Pilot Evaluation of Driving Behavior Changes Among U.S. Drivers*2010: University of California Transportation Center, UC Berkeley.
19. Cervero, R. and M. Duncan, *Walking, Bicycling, and Urban Landscapes: Evidence From the San Francisco Bay Area.* American Journal of Public Health, 2003. **93**(9): p. 1478-1483.
20. Kockelman, K., *Travel Behavior as Function of Accessibility, Land Use Mixing, and Land Use Balance: Evidence from San Francisco Bay Area.* Transportation Research Record: Journal of the Transportation Research Board, 1997. **1607**: p. 116-125.

21. Brownson, R.C., C.M. Hoehner, K. Day, A. Forsyth, and J.F. Sallis, *Measuring the Built Environment for Physical Activity: State of the Science.* American Journal of Preventive Medicine, 2009. **36**(4, Supplement 1): p. 22-123.
22. Wang, X., Khattak, A., J., Chen, *Accuracy of Geo-imputation: An Approach to Capture Micro Environment.* Accepted for publication, Transportation Research Record: Journal of the Transportation Research Board, 2013.
23. Wang, D. and F. Law, *Impacts of Information and Communication Technologies (ICT) on time use and travel behavior: a structural equations analysis.* Transportation, 2007. **34**(4): p. 513-527.
24. Kim, T. and K.G. Goulias. *Cross Sectional and Longitudinal Relationships among Information and Telecommunication Technologies, Daily Time Allocation to Activity and Travel, and Modal Split using Structural Equation Modeling* in *TRB 83rd Annual Meeting CD-ROM*. 2004. Washington, DC.

LIST OF TABLES AND FIGURES

Figure 1: Conceptual framework showing conditionality between decision to drive and emissions 22
Figure 2: A land use buffer surrounding a hypothetical residence 23
Figure 3: CO2 emission reductions associated with balanced land use mix and network connectivity... 24

Table 1: Summary of key studies.. 25
Table 2: Descriptive statistics for variables in the database 26
Table 3: Heckman sample selection models of (log-transformed) CO2 emissions...................... 27

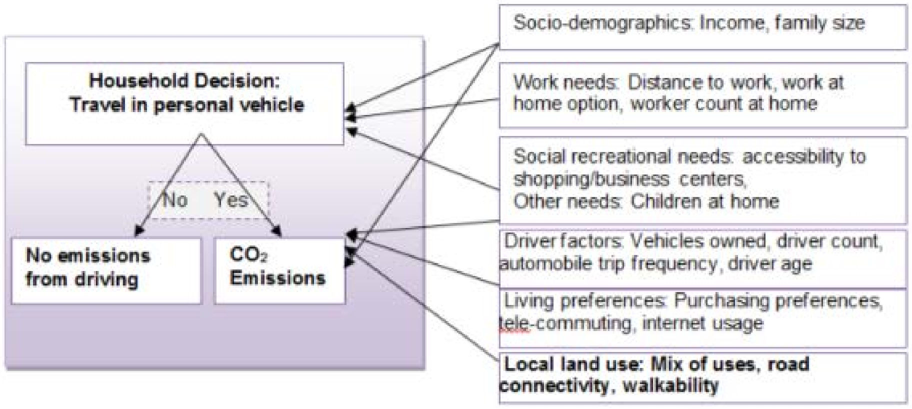


Figure 1: Conceptual framework showing conditionality between decision to drive and emissions.

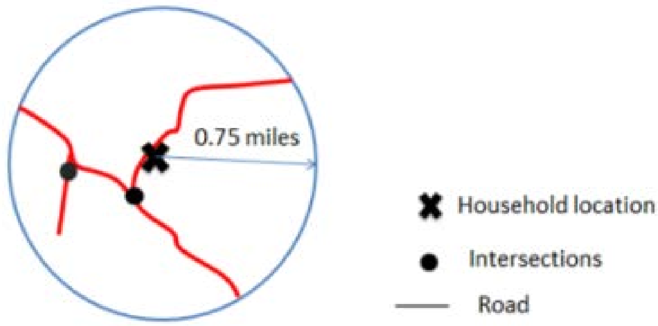


Figure 2: A land use buffer surrounding a hypothetical residence.

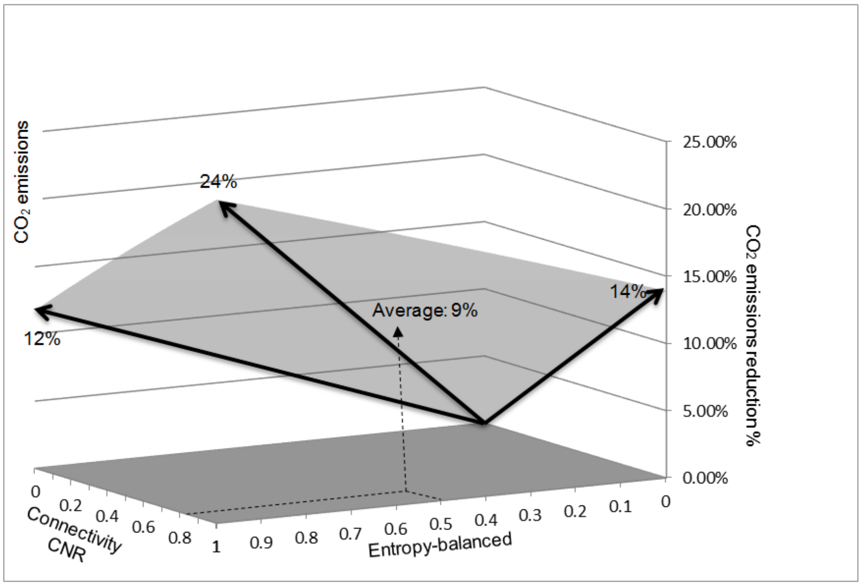


Figure 3: CO2 emission reductions associated with balanced land use mix and network connectivity.

Table 1: Summary of key studies

Paper	*Data used*	*How emissions calculated*	*Scenario design*	*Key Findings*
Stone et al. [3]	US Census 2000, 11 areas in 6 Midwestern states: IL, IN, MI, MN, OH, WI	Calculated based on VMT (NPTS) Tract level CO_2 calculated by EPA 2011 mission rate: 8887*VMT/MPG	3 scenarios: Business-as-usual (BAU) Smart-growth 1(SG1):same growth, new population reallocated to suburban & urban Smart-growth 2 (SG2): lager population growth, new population reallocated to suburban & urban	Compared to BAU, SG1 & SG2 reduce 7.2%, and 16.7% VMT, Compared to BAU, SG1 & SG2, & Hybrid Electric Vehicles reduce 4.7%, 7.8% and 18.0% CO_2 emissions in 2050;
Rodriguez et al.[7]	695,454 individuals in 235,530 households in 2000, Mecklenburg NC	TRANUS Emissions based on link level traffic assignments; Calculated emissions rates.	2 scenarios: Business-as-usual (BAU) Compact growth (CG); Alternative fuels and propulsion system achieved a 27% market penetration by 2050	CG decreases CO_2 emissions by 7.1% (no market penetration) and 10.2% (27% market penetration of alternative technologies) relative to BAU in 2050 respectively.
Niemeier et al. [9]	660,000 population growth between 2000 and 2030, Eight counties of the San Joaquin Valley region	UPlan for land use; Four-step travel demand forecast; UC Drive & MOBLIE6 model for emissions calculation	4 scenarios: Baseline growth (BG): follow current trend Controlled growth (CG): compact growth Uncontrolled growth (UG): (very) low density, roadway expansion with little transit As planned(AP): new road, high speed rail, and medium density	Emission rates: UG>AP>BG>CG. CG: total emissions 6-10% lower than BG. CG: higher residential densities contribute to development patterns that decrease regional vehicle travel and emissions.
Hennessy & Tynan [10]	US Census 1990, US Census 2000, South Carolina Population report	VMT has constant relation with GHG emissions	5 scenarios: Current trend: 5:1 "growth ratio" (land growth /population growth) 4 Compact dev. ratios: 4:1, 3:1, 2:1, 1:1 "growth ratio"	If 1:1 "growth ratio" is implemented, VMT and CO_2 emission will be reduced by approximately 11%.
Frank et al. [13]	Puget Sound Household Travel Survey 1999, Central Puget Sound Region	Speed sensitive emission rates for every link of every trip based on time of day and facility type	Linear regression	Positively associated with emissions: VMT, vehicle/HH, people/HH, & income; Negatively associated with emissions: residential densities, intersection densities, and mixed use
TRB Special Report 298 [8]	US census 2000, NHTS 2001, NPTS 1990, USA Nationwide	Calculated based on VMT EPA 2005 emission rate: 19.4 lb CO_2 per gallon gasoline	3 scenarios: Base scenario: 1.4% VMT increase Scenario 1: 25% compact growth with 12% VMT reduction; Scenario 2: 75% compact growth with 25% VMT reduction	VMT, fuel use, and CO_2 emission reduction: Scenario 1: 1-1.2% in 2030, 1.3-1.7% in 2050 Scenario 2: nearly 8% in 2030, 8-11% in 2050
Hankey & Marshall [11]	US Census 2000 142 US cities	Monte Carlo statistical distribution for future scenario, emission calculated based on VKT and depends on different technology scenario	6 scenarios: Three bounding scenarios: Infill only, Constant density, Suburban nation; Three historical decadal growth: with small (S1), average (S2), large rates (S3) 3 technology scenarios	Comprehensive compact development can reduce US 2000–2020 cumulative emissions by up to 15%–20% of projected cumulative Emissions
Xia, Barth [4]	N/A	Velocity planning algorithm Paramics	3 scenarios: Single vehicle simulation Single network simulation Multiple-lane network simulation	Efficient single vehicle reduce fuel and CO_2 emissions by 12.5% & 13.2% less fuel consumption & CO_2 emissions with short cycle length, lower volume, & higher penetration rates
Barth & Boriboonsomsin [5]	PeMS, TMC, California	Paramics simulation Real world experimentation	Two simulation scenarios: Non-eco-driving Eco-driving: 20% penetration rate of eco-driving	10%-20% CO_2 reduction without increasing travel time dramatically in congested conditions

Table 2: Descriptive statistics for variables in the database

Variable Description	*Name*	*N*	*Mean*	*Std. Dev.*	*Min*	*Max*
CO_2 emissions (grams)	CO_2 (uncensored)	13371	29,170.97	31112.95	61.29	780913.6
	CO_2 (pooled)	15213	26, 021.62	31001.14	0	780913.6
Log-transformed CO_2 emissions	$LNCO_2$	13371	9.82	1.09	4.12	13.568
Did not drive (binary)	NODRIVE	15213	0.12	0.33	0	1
Vehicle miles traveled	VMT	15213	73.84	82.18	0	2468.3
Socio-Demographic Variables						
Household vehicle count	HHVEH	15213	2.25	1.217	0	27
Own hybrid vehicle (binary)	HYBRID	15213	0.04	0.193	0	1
Family income ($x1000)	INCOME	15213	57.48	29.548	5.00	100
Household workers count	WRKCNT	15213	0.98	0.897	0	5
Distance between home and work	DISWH	15213	10.81	66.561	0	2473.19
Having children younger than 16	HAVEKID	15213	0.22	0.413	0	1
Driver-Related Variables						
Auto trip frequency	TAUTO	15213	6.80	5.685	0	52
Household driver count	DRVRCNT	15213	1.82	0.759	0	8
Age of household head	AVGAGE	15213	58.14	15.06	18	92
Living Preference Variables						
Internet use during past month	INTERNET	15213	33.34	27.49	0	186
Product delivery during past month	DELIVERY	15213	2.20	4.43	0	115
Freq. of work from home past month	FWKFH	15213	0.44	2.22	0	40
Local Land Use Variables						
Roadway length (miles)	LENGTH	15213	17.52	11.521	0.07	76.35
Connected node ratio	CNR	15213	0.72	0.135	0.00	1.00
Link node ratio	LNR	15213	1.37	0.262	0.00	5.00
Number of cul-de-sacs	NDANGLE	15213	36.68	31.361	0.00	309
Length per node (miles)	LENGPN	15213	0.21	0.247	0.029	3.756
Mix of land use entropy-Balanced	ENTROPY_BA	15213	0.512	0.180	0.00	0.954
Mix of land use entropy-Blended	ENTROPY_BE	15213	0.338	0.131	0.00	0.837

Note: The household income coded in the dataset is categorical; the categorical income in this paper is recoded with the average income of each category. Those who did not answer the income question were replaced by the mean of household income which is $57,500.

Table 3: Heckman sample selection models of (log-transformed) CO2 emissions

	Model 1 (Heckman)			Model 2 (OLS regression)		
	β	$e^{(\beta)}$	P>z	β	$e^{(\beta)}$	P>z
CO_2 Emissions model						
Auto trip frequency	0.094	1.099	0.000	0.284	1.328	0.000
Household income	0.003	1.004 *	0.000	0.010	1.010	0.000
Household vehicle count	0.049	1.050	0.000	0.177	1.194	0.000
Driver count	0.058	1.060	0.000	0.268	1.307	0.000
Roadway length	-0.008	0.992	0.000	-0.020	0.980	0.000
Connected node ratio	-0.149	0.862	0.042	1.510	4.527	0.000
Length per node	0.182	1.200	0.000	-0.387	0.679	0.000
Mix of land use entropy-Balanced	-0.125	0.882	0.070	-0.392	0.676	0.054
Mix of land use entropy-Blended	0.024	1.024	0.800	0.437	1.548	0.115
Number of cul-de-sacs	-0.001	0.999	0.093	0.004	1.004	0.001
Avg. house head age (years)	0.009	1.009	0.005	0.075	1.078	0.000
Square of avg. house head age	-0.0002	1.000	0.000	-0.001	0.999	0.000
Internet usage freq.	0.002	1.002	0.000	0.001	1.001	0.634
Delivery freq.	0.0001	1.000	0.970	-0.006	0.994	0.245
Work from home freq.	0.007	1.007	0.033	-0.002	0.998	0.823
Constant	8.979	7934.693	0.000	3.071	21.563	0.000
Probit model for decision to drive						
Household income	0.008	0.001 **	0.000			
Distance to work	0.001	0.0002 **	0.145			
Worker count	0.490	0.086 **	0.000			
Have children <16 years	0.164	0.029 **	0.000			
Constant	0.391		0.000			
Summary Statistics						
Observations=15213, Censored Observations=1842 Log likelihood = -21303.26 Prob. >Chi-squared=0.000 ρ= -0.208 Prob. >Chi-squared=0.001				Observations=15213 Prob.>F=0.000 R-squared =0.382		

Notes: *Inverse mills ratios of income are calculated for exp(income), since income is included in both the emissions model and the selection drive-or-not-drive probit model.
**Marginal effect of probit drive or not drive model.

PART 2: ACTIVITY SEQUENCING BEHAVIOR FROM A REGIONAL PERSPECTIVE

A COMPARATIVE STUDY OF TRANSIT-ORIENTED DEVELOPMENT RESIDENTS AND AUTO-ORIENTED DEVELOPMENT RESIDENTS

Sanghoon Son[1] and Asad Khattak[2]

[1]Old Dominion Univ. & Jeju Development Institute, [2]The University of Tennessee, Knoxville

Abstract: Aiming to better understand travel and activity behavior as activity-based modeling effort, this paper explores location choice and sequence of out-of-home activities undertaken by residents in transit-oriented development (TOD) neighborhoods. TOD is a sustainable growth strategy that integrates transportation and land use. It provides greater connectivity and accessibility to those who live in the vicinity, which is unique space in urban and suburban areas. The behavioral data are extracted from the 2008 National Capital Region Household Travel Survey. For the comparative analysis, residents in TOD neighborhoods (N=1,911 households) were identified and closely compared two groups of residents in auto-oriented development (AOD) neighborhoods in Washington D.C. Statistical models that can address individuals' multiple activities were estimated. Results show that TOD residents are much more likely to participate in out-of-home activities within TODs (15%-28%). Also, TOD residents tend to sequence their activities within TODs than AOD residents (20%-33%). Furthermore, the tendency exists among AOD residents depending on their proximity to subway stations. The implications on activity-based modeling are further discussed.

Keywords: Transit-oriented development, travel activity behavior, time use, activity location choice, activity location sequence

1. INTRODUCTION

Transportation planning agencies, e.g., metropolitan planning organizations, generally use travel demand models for transportation planning, especially for making informed infrastructure investment and improvement, as well as transportation policy decisions. Over the past decades, forecasting and analyzing travel demand has been based on individual trips made by households and people as a unit of analysis. However, this has been conceptually criticized by the view that a trip is actually derived from activities. In addition, other limitations are that trip-based models only limitedly account for interaction among travel decision makers (e.g., household members) or travel-related decisions (e.g., destination and mode choice). Moreover, the conventional trip-based approach does not fully account for time/space aspects (or constraints). Although the recent interest in reflecting spatial aspects of land use, trip-based travel demand models weakly account for them. Taken together, represent real world travel decision making are imperfectly represented (1, 2).

An activity-based approach emerged as a new paradigm of travel demand analysis. The activity-based approach more fully takes into account travel as a derived demand, focusing on activity participation decisions with trips viewed as a special case of activity participation. Recently, activity sequencing, household interactions and time-space dimensions has been importantly explored. This behaviorally appealing and broader approach is finding greater application in the field, with development of activity-based land use-transportation model systems such as TRANSIMS (3), UrbanSim (4), and MATSim (5).

Responding to this research trend, this study aims to better understand travel and activity behavior in the context of transit-oriented development (TOD), which has not been investigated intensively. TOD provides more livable and sustainable communities around transit stations with mixed-use and compact development, aiming to mitigate some of the transportation problems and environment concerns. Accordingly, residents of TODs are more exposed to non-motorized modes and to diverse activities. Also, transit systems connect TODs so that accessibility and connectivity are likely different. Thus, TOD is very unique and interesting urban and suburban space.

Among many dimensions of activity behavior, this study particularly explores location choice and sequence behavior for out-of-home activities undertaken by TOD residents, from a regional perspective. A cross-sectional household travel survey of the National Capital Region (N=11,436 households) was used in this study (6). This study identified residents in TODs (N=1,911 households) and analyzed their activity location choice and sequence, in comparison of auto-oriented development (AOD) residents in the Washington D.C. region.

Demand for TOD as a sustainable urban design is increasing worldwide. The findings of this study can help travel demand modelers continues to improve activity-based modeling and provide sounder basis for integration of land use and transportation. A detailed analysis of TOD residents is also beneficial to transit and urban planning agencies in measuring performances and determining policy actions as they face greater interest and demand for TODs across the country and internationally.

2. LITERATURE REVIEW

A large set of studies has empirically analyzed daily activity patterns in different contexts (Table 1). While most empirical studies have focused on general population of urban residents or commuters (7-18), certain population segments have also been the focus, e.g., home-workers (9), non-workers (9, 10), homemakers (14), university students (17), and individual 65+ years (18). Furthermore, weekend activity patterns compared with weekday activity patterns were also analyzed (12, 15). These studies examined out-of-home activities, classified into work, school, shopping, recreation, personal business, etc. Some have grouped them into sub categories such as subsistence, maintenance and recreation (9), obligatory and discretionary (13), or maintenance and discretionary (18). Methodologies used in the literature vary from descriptive analysis (12) to structural equation modeling (9, 14).

Activity behaviors are quite complex to understand, partly because there are many types of daily activities and they take place over time and at different locations. To capture the complexity of observed activity patterns, various measures have been used in earlier studies, e.g., Hanson and Hanson (7) generated and tested 51 measures to explain activity behaviors temporally and spatially, together with travel activity, including number of stops by each activity category, by weekday and weekend, and by locations (e.g., CBD), as well as minutes spent in each activity category (see Table 1). Recent studies have explored activity frequency, duration, sequence of activities, first or last stop (activity) of the day, and number of stops per tour (10). Also, transition matrix of activity types was used to clearly show activity sequence (10, 17). Interestingly, some studies measure daily activity behaviors in terms of space use at the household and individual level (13, 16).

Activity patterns measured in different dimensions are found to be associated with demographic and socioeconomic attributes of individual or household; but the relationships are likely context dependent. For instance, females are positively correlated to frequency, duration, or propensity of shopping activities (7-10) while negatively related to working (7, 9) and recreational (9, 10) activities. Moreover, earlier stage of life cycle is statistically associated with more frequent social activity (7) as well as more time spent or higher propensity for recreation activity (7, 10). As expected, automobile ownership and availability significantly explain frequency or duration of out-of-home activities (7) and some in-home activities (8, 9). In addition to socio-demographics, travel behaviors are both direct and indirect associates of activities (9). Furthermore, from a time budget perspective, in-home and out-of-home activity durations must be traded-off.

Remarkably, there are a few activity behavior studies that are linked to land use patterns. However, the land use variables did not show any statistical significance in the propensity of making specific activities in an earlier study (10). Although a recent study comparing homemakers in New York and suburban areas indicated that travel and activity behavior are related with both built environment and socioeconomic variables. The study also found that homemakers living in New York City spend more time on discretionary activities; but less time on maintenance activities, compared to those in suburbs (14). A gap in the literature is the lack of information about activity patterns of TOD residents. This study examines the travel and activity behavior of TOD in order to develop a comprehensive understanding of their activity location and sequence behavior in regional space.

Earlier studies have indicated that out-of-home activity behavior is associated with socio-demographics and transportation context (7-10), as well as the built environment (14). In line with the earlier studies, this study analyzed out-of-home activity behavior in terms of residents in TOD areas.

Conceptually, TODs provide higher-density and diverse land uses and more walkable/bikable environments. Thus, the residents can have diverse activity opportunities (e.g., work and social activities) with a greater accessibility. In addition, transit systems link local TODs with regional TODs, offering the residents greater commuting/travel options, with a greater connectivity (Figure 1). Consequently, areas near subway stations (TOD areas) can act as “the core of daily activities” for the residents. On the other hand, for the AOD residents, the TODs can act “a routine anchor.” These aspects make TODs unique places with high level of accessibility and connectivity (highlighted by orange box in Figure 1), which is different from conventional auto-oriented and lower density development.

Residents in TODs are likely to have different out-of-home activity patterns compared with residents in AODs. The patterns that are investigated in this study are location choice and sequence at the activity and trip level. Hypotheses in this study include that TOD residents may be more likely to participate in activities and travel in close proximity of TODs, using alternative transportation modes (more frequently). Moreover, it is expected that TOD residents choose TODs for their planned activities, and better sequence their activities centered on and bounded by TODs.

3. STUDY AREA AND BEHAVIORAL DATA

Study Area and Survey Description

The study area is the National Capital Region, with the focus of an inner area of the Washington D.C. metropolitan region surrounded by Capital Beltway (see Figure 2). Notably, intensive transit systems (e.g., subways and buses) run in this region. There are three transit agencies (Washington Metropolitan Area Transit Authority, Maryland Transit Administration, and Virginia Railway Express) for subway and commuter rail services, operating 11 lines over 131 stations. Among the all stations, this study focused on 86 Washington Metropolitan Area Transit Authority transit stations.

The metropolitan region includes 22 jurisdictions, which are home to 5,756,612 people and 2,139,192 households residing in 4,146,132 acres according to 2010 US census (18). According to the TOD database (20), 518,558 people and 248,693 households reside near transit stations, i.e., within 0.5 mile buffer, which account for 9% and 12% of the regional population and households, respectively, over an area of 34,783 acres (about 1% of the total area).

The behavioral data used in this study are from the 2008 National Capital Region household travel survey (N=11,436 households). Notably, this survey adopted a residential mailing address-based list sampling method to deal with non-coverage of substantial number of mobile phone only households (6). The mobile phone only households are more likely to reside near transit stations compared to the general population with relatively smaller size and younger members (21). Also, high density and mixed use areas were oversampled compared with low density areas in the survey. Therefore, samples are relatively well representative of the target population while sufficient numbers of samples residing in TOD areas were available. A relatively low survey response rate of about 10% was reported. However, this is valid in the case of address-based surveys.

Data Preparation and Sample Characteristics

The study area was divided into three groups by a proximity to subway stations: TOD, AOD close to TOD (AOD-C), and AOD far from TOD (AOD-F). Firstly, the TOD was defined as an area bounded by 0.5-mile buffer of each subway station (small inner circles in Figure 2), which is consistent to conventional notion (20). Secondly, additional circular buffers were spatially created from the metro stations, with the distance of 0.5 mile to 1.5 mile (larger outer circles in Figure 2). This is referred to the AOD-C. Lastly, the remaining area is referred to AOD-F, which is 1.5 mile away or father from subway stations.

Notably, the proximity to TODs can play an important role in this context. In other words, the activity behavior can be differently associated with the distance from a subway station. In this regard, a distance of 1.5 mile from a subway station was determined empirically. The empirical distribution of access distance to a subway station shows that 85% of subway users who access to a station, regardless of access modes, fall in the 1.5 mile buffer.

Corresponding to each group, a set of household sample was identified and extracted from survey dataset. As geographical location of household samples is available at the census block level in the survey dataset, any census block intersected with TOD buffer is designated as a TOD census block. Any household samples belonging to the census block are treated as TOD residents; likewise AOD-C and AOD-F, As a result, among the 7,415 identified households in the sample, 1,911 (26%), 2,524 (34%), and 2,980 (40%) households that fall in each group are prepared to compare time use and location choice behavior in the context of residence location, respectively. Detailed 24 hour travel activity profiles and other personal information (e.g., work status, place) were obtained. The data set was error checked and found to be reasonable.

Table 2 summarizes socio-demographic characteristics of the three comparison groups: TOD, AOD-C, and AOD-F. Comparison of household characteristics shows that there are substantially higher proportions of single person-households (49%) and zero vehicle-households (23%) in the TOD, compared to the other two groups. In addition, the TOD households consist of more low-income residents (less than $29,999) and multi-family house dwellers. As a residence location moves from the TOD area to the AOD area, a household has more members, workers, vehicles, and income. Table 2 shows that, at the person level, the individuals residing in the TOD have relatively more individuals aged 19-34 (25%) and employees (71% out of individuals aged 16 and above) than the other groups. Interestingly, in the outer TOD, a relatively higher ratio of African American is found. Overall, the socio-demographics are largely different by residence group.

4. DESCRIPTIVE STATISTICS

Table 3 displays a classification of out-of-home activities. A total of 20 original activity categories are drawn from the survey data. The activity categories are clustered into two broad groups: obligatory and discretionary. Note that any activities undertaken at home are treated as in-home activity even if the activities have something to do with eating a meal or working, e.g., eat/prepare a meal at home and work at home or telecommute.

Descriptive analysis includes comparison of location choice and sequence behavior at the weekday activity level across the three residence groups defined earlier. Using corresponding activity data clustered into seven types, Table 4 compares activity type and location across the residence group. Notably, residents in the TOD participate in moderately more work activities (35%) than the other groups. By contrast, residents in other groups engage in more school activities (8%), while the rest of the activities are quite similarly involved across the residence group. As hypothesized, the TOD residents choose more activities within TOD areas (66%) while the AOD-C and AOD-F residents do for within their local areas (33% and 61%, respectively). However, the AOD-C residents choose more activities in TOD areas not their local areas (33% vs. 43%). Also, the AOD-F residents similarly choose their activity locations between the TOD and AOD-C areas, although the AOD-C areas are closer to them (21% vs. 18%).

Table 5 shows sequence of activity and location. There is moderately more work-shop and work-meal activity sequences by the TOD residents. Notably, one with participation in a work activity in TOD areas is more likely to engage in a discretionary activity. Nevertheless, overall distributions of other activity sequences are fairly similar to each other. Notably, the location sequence behavior is quite different from one group to another. For example, while the TOD residents sequence TOD locations more than 50%, the AOD-F to AOD-F sequence accounts for 47% of all sequences made by the AOD-F residents. However, the sequence of activity locations are quite mixed for the residents of the AOD-C areas Together with findings above, this indicates that time use over urban space is complex, and, in particular, TOD areas are the most frequently used spaces in urban areas.

5. LOCATION CHOICE AND SEQUENCE BEHAVIOR

Statistical Modeling

To understand the location choice and sequence behavior of TOD and AOD residents of the study area, statistical models were estimated. The models can be also used to test the fourth hypothesis on the different location choice and sequence behavior between TOD and AOD residents. The first dependent variable of interest is whether or not the next activity location is TOD. If the next activity location is TOD, the dependent variable was coded with one; otherwise, it was coded with zero. The second dependent variable is whether two consecutive activity locations are TOD areas. Similar to the first dependent variable, in this case, the dependent variable was coded with 1 only if the next two locations are TODs. This study estimated random effect binomial probit models to explore the linkage of location choice and location sequence behavior with associated factors. The random effect can capture the correlation between errors, as an individual can make several location choices and sequences. Econometric models are specified as follows:

$$Y^* = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 TOD + \beta_6 AODC + \varepsilon \quad (1)$$

$$Y^* = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \gamma_5 TOD + \gamma_6 AODC + v + u \quad (2)$$

$Y = 1\ if\ Y^* > 0, and\ 0\ otherwise$

where

Y	=	a binary dependent variable (location choice and sequence)
X_1	=	a set of socio-demographic variables,
X_2	=	a set of work-related variables,
X_3	=	a set of spatial context variables,
X_4	=	a set of temporal context variables,
TOD	=	an indicator variable (1=residing in TOD areas, 0=otherwise),
$AODC$	=	an indicator variable (1=residing in AOD-C areas, 0=otherwise),
β, γ	=	a set of parameters, and
ε, v, u	=	error terms,

To allow for ε to be freely correlated within an individual, but uncorrelated across individuals, the error term in model 2 specifies

$$\varepsilon = v + u$$

Where v is the unobserved individual specific heterogeneity. If v is unrelated to independent variables, this is called the random effect model. In this case, a conditional distribution f(v|X) is not dependent on independent variables, X. If v and X are correlated, this is called the fixed effect model. As noted, this study focused on the random effect model.

The likelihood function is

$$L = \int_{-\infty}^{\infty} \left[\prod P(Y_i = y_i | X\beta + v) \right] f(v) dv.$$

Modeling Results

Table 6 and Table 7 summarize the results of two random effect binary probit models for activity location choice and sequence, with 30,191 and 12,672 observations, respectively. The observations are only for out-of-home activity locations. Both best-fitting models are statistically significant at the 1% level. The model fits are adequate, indicating that location choice and sequence behavior are well captured. In the models, most factors associated with out-of-home activity location choice and sequence behavior are statistically significant and largely consistent with expectations. Marginal effects are presented in order to help interpret the coefficient values in a more intuitive way. As explained in the previous section, the marginal effects were calculated at means of the independent variables, holding all other variables at their means.

The most interesting finding is that TOD residents are strongly associated with participating in activities in TODs and sequencing their locations within TODs. The activity location choice model suggests that TOD residents are 25% more likely to locate themselves to participate in an activity, compared to the AOD-F groups, and 13% more likely than the AOD-C residents, all else being equal (Table 6). In the activity sequence model, the TOD residents are more likely to sequence their activities within TOD; the likelihood is higher by 16% and 10% than the other groups, respectively, holding other variables constant (Table 7). This implies that, for TOD residents, TOD areas play an important role in their daily activities. That is, TOD residents center on TOD areas for their daily activities and travels. The magnitudes estimated can be the impact of high levels of connectivity and accessibility and mix of land uses on location choice and sequence.

In the activity location choice model (Table 6), most socio-demographic variables are statically significant with a moderate magnitude. Individuals with fewer members and vehicles in the household are less likely to participate in activities within TODs. Also, employees, students, and retirees are two largest population groups who choose TOD areas as a location for their activity engagement. Further, males tend to choose their activities around TODs more than females, all else being equal. Several spatial and temporal context variables need to be mentioned. A substantial amount of participants of activities in TOD departs from Fairfax County to Washington, D.C. and Arlington County, which can represent regional traffic flows. Moreover, choosing TOD areas for the activities is slightly higher (about 3%) when one participated in activities from July to September, compared to October to December. Also, main activities in TOD areas are strongly associated when they are work, meal, and shopping related activities.

In the activity location sequence model (Table 7), there are several socio-demographic characteristics significantly associated with engaging activities within TOD areas. They include fewer/no vehicles, male, and more income. Employees, retirees, and interestingly homemakers are more likely to sequence their activities in TOD areas, compared to individuals with other jobs. Many spatial and temporal context variables show statistical significance. Monday and Thursday activities as well as mid-day activities are more likely to be linked around TOD areas. Particularly, strong spatial dependence is statistically found from Washington, D.C. and Arlington County in the context of sequence activity locations. Discretionary and obligatory (e.g., work and school-related) activities are more likely to sequenced within TODs. These findings reflect the characteristics of TOD design, which are higher densities and mixed land use, and regional connectivity through transit system. Also, location and sequence of activities are directly linked with travel distance and mode choice.

6. LIMITATIONS

The results of this study should be interpreted with a caution. The study focuses on a single metropolitan region of Washington, D.C. This region has expansive public transportation systems including subway, commuter rail, and bus services. Also, a relatively large portion of land is higher density and mixed use (by US standards). Thus, the study findings should not be generalized beyond similar environments and contexts. Moreover, this study is limited by the use of cross-sectional travel survey data that only collected a one-day travel and activity profile. Individuals' daily travel and activity patterns may vary from day-to-day. Also, the survey instrument was not able to capture time use over a 24-hour period. Therefore, all daily activities are not measured, e.g., in-home activities. To its credit, the survey included mobile phone only households because it used an address-based sampling method. However, there are still an under-covered population segments, e.g., households with no telephones (about 2% of the population). The relatively high levels of incentives given to some of the respondents may have motivated certain groups (e.g., lower income) to respond to the survey more than others.

7. CONCLUSIONS AND IMPLICATIONS

This study contributes by providing a great understanding of out-of-home activity behavior in the context of TOD, which has not been understood well in the literature. Particularly, activity location and sequence chosen by residents in TOD areas in the Washington D.C. were focused at the trip and activity levels, in comparison of residents in the AOD areas. While an activity-based approach is appealing as a new paradigm in travel demand analysis, more information on travel and activity behavior is needed. Such case includes TOD, which is a sustainable growth and development strategy. At a neighborhood level, TOD providing greater connectivity and accessibility to those who live in the vicinity, by integrating transportation and land use. TODs are regionally connected each other by public transit system; therefore, TOD is unique urban and suburban space.

The activity location and sequence behavior for out-of-home activities was captured with random effect models, which can handle the correlation between errors as an individual can make several location choices and sequences. This study found that TOD residents have a higher propensity to choose the TOD areas for their activity locations, compared with the resident in AOD areas. This study also found that chances of sequencing the out-of-home activities near subway stations are higher when one lives in the TOD areas. Not only there is behavioral difference between TOD and AOD residents, but also the difference is found within AOD residents with dependence on the distance to subway stations. In the context of relatively mature land use in TOD in the study area, they allow relatively easy access to amenities in the vicinity of a TOD without driving and with reasonable distances (i.e., proximity), while the transit system can easily transport people to other TOD areas (i.e., accessibility and connectivity) with other alternative modes.

Two implications are derived from findings in this study. First, from a destination choice standpoint, results show that locating in TODs (which have high levels of connectivity and accessibility and mix of land uses) allows sequencing activities and travel within the constrained space of TODs. This can be translated into strong geographical dependency among local and regional TODs in the time-space use behavior.

Second, a spatial unit of travel demand analysis can be adjusted. In general, a traffic analysis zone is widely used for the unit. However, as mentioned in this study, TOD is geographically confined by about 0.5 mile from stations (therefore, the size of TOD in the study areas accounts for 1% of the total area), but various activities take place in unique built and transportation environments. Moreover, as subway stations are typically located on the borderline between two traffic analysis zones or their intersections, the spatial aspect of TOD as the core of activities are not represented in travel demand modeling. In this regard, travel demand models may consider TOD as separate zone or a center of zone. This can capture travel and activity behavior more realistically and will improve modeling effort in a way that reflect individuals' behavioral of location choice and sequence.

TODs are unique in that they provide livable environments that are transit/alternative mode friendly, higher density, and mixed use. By showing that the TOD resident segment of the population may have different travel and activity patterns, the findings from this study support the work of travel demand modelers who continue to improve activity-based modeling and provide a sounder basis for integration of land use and transportation. A detailed analysis of TOD residents is also beneficial to transit and urban planning agencies in measuring performances and determining policy actions as they face greater interest and demand for TODs worldwide.

ACKNOWLEGEMENTS

The study was supported by the Virginia Department of Transportation and the TranLIVE University Transportation Center grant sponsored by US the Research and Innovative Transportation Administration of US DOT. Special thanks are extended to the following entities for their support: Transportation Research Institute and the Center for Innovative Transportation Solutions at Old Dominion University.

REFERENCES

1. Bhat, C. R., and F. S. Koppelman. Activity-Based Modeling of Travel Demand. In *The Handbook of Transportation Science* (R. W. Hall, ed.), Kluwer Academic Publishers, Norwell, Mass., 1999, pp. 35–61.
2. McNally, M.G. and C.R. Rindt, The Activity-Based Approach, In *Handbook of transport modeling*, K.J. Button and D.A. Hensher, Editors. Elsevier Science B.V.: New York, 2007, pp. 53-69.
3. Transportation Analysis and Simulation System (TRANSIMS). http://code.google.com/p/transims/. Accessed Feb, 1, 2012.
4. UrbanSim. http://www.urbansim.org/Main/WebHome. Accessed Feb, 1, 2012.
5. Multi-Agent Transport Simulation Toolkit (MATSim). http://www.matsim.org/. Accessed Feb, 1, 2012.
6. National Capital Region Transportation Planning Board Metropolitan Washington Council of Governments. *2007/2008 TPB Household Travel Survey: Technical Documentation*. Publication Feb 2010.

7. Hanson, S. and P. Hanson. The Travel-Activity Patterns of Urban Residents: Dimensions and Relationships to Sociodemographic Characteristics. *Economic Geography*, Vol. 57, No. 4, 1981, pp. 332-347.
8. Levinson, D. and A. Kumar. Activity, Travel, and the Allocation of Time. *Journal of the American Planning Association*, Vol. 61. No. 4, 1995, pp. 458-470.
9. Lu, X. and E. I. Pas. Socio-demographics, activity participation and travel behavior. *Transportation Research Part A: Policy and Practice*, Vol. 33, 1999, pp. 1-18.
10. Misra, R. and C. Bhat. Activity-Travel Patterns of Nonworkers in the San Francisco Bay Area: Exploratory Analysis. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1718, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 43-51.
11. Frusti, T., C. Bhat, and K. Axhausen. Exploratory Analysis of Fixed Commitments in Individual Activity—Travel Patterns. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1807, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 101-108.
12. Lockwood, A., S. Srinivasan, and C. Bhat. Exploratory Analysis of Weekend Activity Patterns in the San Francisco Bay Area, California. Transportation Research Record: Journal of the Transportation Research Board, No. 1926, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 70-78.
13. Buliung, R. N. and P. S. Kanaroglou. Urban Form and Household Activity-Travel Behavior. *Growth and Change*, Vol. 37, 2006, pp. 172-199.
14. Cynthia, C. and C. E. McKnight. Does the Built Environment Make a Difference? Additional Evidence from the Daily Activity and Travel Behavior of Homemakers Living in New York City and Suburbs. *Journal of Transport Geography*, Vol. 15, 2007, pp. 380-395.
15. Zhong, M., J. Hunt, X. Lu. Studying Differences of Household Weekday and Weekend Activities: A Duration Perspective. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2054, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 28-36.
16. Fan, Y. and A. Khattak. Urban Form, Individual Spatial Footprints, and Travel: Examination of Space-Use Behavior. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2082, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 98-106.
17. Eom, J. K., J. R. Stone, and S. K. Ghosh. Daily Activity Patterns of University Students. *Journal of Urban Planning and Development*, Vol. 135, 2009, pp. 141-149.
18. Ziems, S., K. Konduri, B. Sana, and R. Pendyala. Exploration of Time Use Utility Derived by Older Individuals from Daily Activity-Travel Patterns. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2156, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 111-119.
19. U. S. Census Bureau. American FactFinder fact sheet, 2010. http://www.census.gov/. Accessed Feb. 1, 2012.
20. TOD Database, Center for Transit-Oriented Development. http://toddata.cnt.org/. Accessed Apr. 1, 2012.

21. Son, S. and A. Khattak. Non-coverage Errors in Surveys Due to Mobile Phone Only Household Samples. Submitted to 92nd Annual Meeting of Transportation Research Board, Washington D.C.

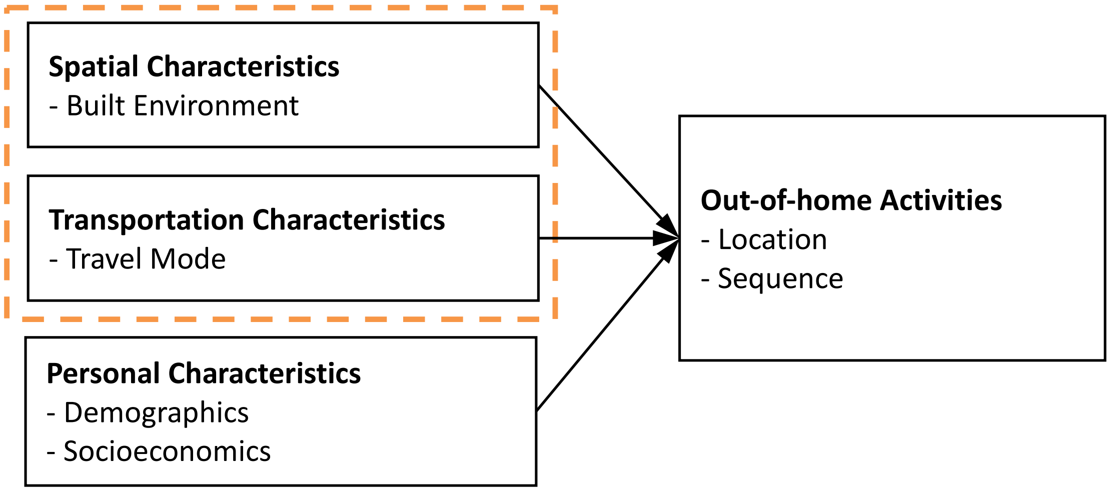


Figure 2: Conceptual Structure

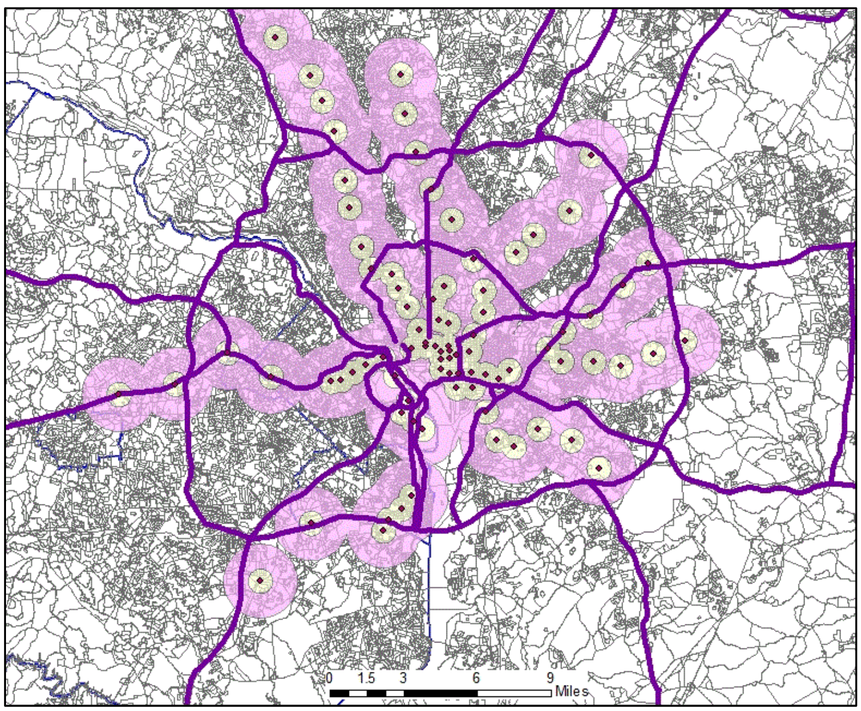


Figure 3: The study area of Washington D.C.

Table 4: Summary of activity behavior studies

Author	Purpose and target	Activity category	Key measure	Data	Method	Key findings
Hanson and Hanson (7)	To relate urban residents' travel and dairy out-of-home activity patterns with socio-demographic status and individual's role situation	• Social • Personal business • Shopping • Work • Recreation	• Number of stops for each activity type • Proportion of stops made by each mode • Time spent in each activity type	1971 Uppsala longitudinal household travel survey, Sweden (N=149 individuals)	Principal component analysis/ OLS regression model	Complex behaviors of travel and activity can be viewed multi-dimensionally. Socio-demographic and individual role attributes are statistically associated with different travel-activity patterns.
Levinson and Kumar (8)	To understand trends in and factors affecting activity patterns among different activities of individuals by different work status and gender	• Work • Home • Shopping • Other • Travel	• Activity duration • Activity frequency • Activity frequency distribution	1968 and 1987/88 metropolitan Washington D.C. household travel surveys (N=36,958, N=10,305 indivi-duals, respectively)	Descriptive analysis/ OLS regression model	Both increases in work and non-work trips lead to less time spent at home. Activity duration for home, shop, other are associated with socio-demographic variables.
Misra and Bhat (10)	To explore out-of-home activity behavior of non-workers, relating with individual and household socio-demographics	• Serve passenger • Personal business/ medical/dental • Shopping • Social/recreation • Home stay • Other	• Number of stops and stops per tour by each activity type • First and last stops (activity) of the day • Transition matrix of activity types	1990 San Francisco Bay area activity travel diary survey, CA (N=3,517 individuals)	Descriptive analysis/ binary logit model	Household and individual characteristics are related with activity participating and chaining, while activity sequencing is mainly determined by current activity types, not by variations of individual or household attributes.
Frusti et al. (11)	To understand fixed commitments in individual activity–travel patterns, relating with socio-demographics, social roles, and work-related characteristics	• Recreation • Personal • Community • Training • Other	• The presence of each fixed commitment	1999 Halle/ Karlsruhe 6-week activity travel survey, Germany (N= 361 individuals)	Binary logit model	The determinants of fixed commitments with statistical significance are found among personal, household, and spouse variables.
Lockwood et al. (12)	To compare weekday with weekend travel-activity patterns in terms of activity participating and activity sequencing/ chaining	• Work/school • Social/recreation • Meals • Shopping • Personal business • Serve passenger • Community/religious	• Frequency/duration of activity episode • Activity episode transitions/chains • First and last activity episodes of the day	2000 2-day San Francisco Bay Area Travel Survey, CA (N= 50,892 individuals)	Descriptive analysis	Weekend activity–travel patterns are different from weekday patterns (e.g., activity purpose and travel distance). Using activity sequencing and trip-chaining behavior, activity–travel on weekends are explained.

Buliung et al. (13)	To examine the spatial characteristics of week-day household activity-travel behavior, associated with location, mobility status, and socio-demographics	• Obligatory • Discretionary	• Household activity space	1994/95 2-day Portland Household Activity-Travel Behavior Survey (N=1,609 households)	Statistical test/spatial regression model	Between urban and suburban, urban households have less daily travel and smaller activity spaces. Statistically significant associates with household activity spaces are also found.
Chen and McKnight (14)	To investigate whether activity and travel behavior of homemakers differ with different types of neighborhoods and they are attributed to the built environment	• Maintenance • Discretionary	• Activity frequency for each type • Time use for each activity type	1997/1998 New York metropolitan area household interview survey	Descriptive analysis/ Structural equation model	Homemakers living in New York City spend more time on discretionary activities; but less time on maintenance activities, compared to those in suburbs. Travel and activity behavior are related with built environment and socioeconomics.
Zhong et al. (15)	To study the differences in weekday and weekend activities in terms of participation frequencies, starting times, and durations	• Travel-related • Work • School • Sociality • Shopping • Eating • Exercise • Entertaining/leisure • Religious, civil, etc. • Out-of-town travel	• Participation frequencies of each activity type • Starting times of each activity type • Durations of each activity type	2001/02 Calgary household activity survey, Canada (N= about 13,000 activities)	Statistical test/model fitting	Weekend activities behaviors are different from their weekday counterparts. For common activity types, they tend to follow different survival functions as well as result in different parameters.
Fan and Khattak (16)	To examine how space uses of individuals is related to urban form factors	N/A	• Individual daily activity space	2006 Greater Triangle region travel survey, NC (N=7,422 individuals)	Spatial regression models	Residents of densely developed neighborhoods with more retail stores and better-connected streets generally have a smaller area of daily activity space.
Eom et al. (17)	To analyze university students' daily activity participation and compare it across the student groups	• School/class • Study/research • Work/volunteer • Social/recreation • Family/personal • Meals • Others	• Average activity frequency/duration • Activity sequencing • Proportion of daily activity profile for each activity /hour	2001 North Carolina State University travel survey, NC (N=843 individuals)	Statistical test	Proportion of daily activity profile (or participation) is not significantly different across the student groups of gender, educational or residential status.
Ziems et al. (18)	To compare the activity time allocation patterns of old individual (age 65+) with other age groups and to quantify satisfaction of derived from the pattern	• Mandatory • Maintenance in-home/out-of-home • Discretionary in-home/out-of-home • Travel • Sleep	• Average time use to in-home and out-of-home activities • Average time use utility	2008 American time use survey (N=12,055 individuals)	Utility regression models	Older individuals show the highest values of time use utility of all age groups. Out-of-home activity engagement is important from the utility perspective.

Table 5: Sample characteristics by distance to subway station (%)

Household level		TOD (N=1,911)	AOD-C (N=2,524)	AOD-F (N=2,980)
Household size	1	49	39	32
	2	34	35	37
	3	10	13	14
	4+	7	13	17
Number of workers	0	23	23	22
	1	47	44	41
	2	28	30	33
	3+	2	3	4
Number of vehicles	0	23	8	4
	1	51	43	36
	2	21	37	42
	3+	5	12	18
Household income	Less than $29,999	13	11	7
	$30,000 - $49,999	16	15	13
	$50,000 - $99,999	33	33	34
	$100,000 - $149,999	24	24	29
	$150,000 or more	14	17	17
Housing type	Single family detached	21	52	58
	Single family attached	21	19	19
	Multi-family	58	29	23
Person level		TOD (N=3,174)	AOD-C (N=4,934)	AOD-F (N=6,291)
Gender	Male	46	45	46
	Female	54	55	54
Age	5-18	9	16	17
	19-34	25	15	15
	35-44	17	15	15
	45-54	16	18	18
	55-64	16	19	18
	65+	17	17	17
Race/ethnicity	African American	21	25	17
	Asian	4	4	7
	Hispanic	5	4	5
	White	68	65	68
	Others	2	2	3
Work status (among age 16+)	Employed	71	66	66
	Retired	18	19	20
	Disabled	3	3	2
	Homemaker	3	5	6
	Unemployed	2	2	2
	Student	3	5	4

Note: Individuals of age 0-4 were not included.
.

Table 6: Out-of-home activity classification

Original activity category	Activity type
Work (regular place)	Work
Work at other location	
Work-related	
Education/ school-related activity	Study
Study / do homework	
Childcare / preschool	
Change mode of transportation	Transportation
Pick up / drop off someone or something	
Eat a meal outside home or work	Meal
Shop in store	Shopping
Quick stop / drive thru	
Visit/ socialize	Social/recreation
Entertainment	
Recreation / exercise	
Personal business at establishment	Personal business
Civic or religious activity	
Mail package or letter or other postal	
Care for children	Other
Sleep/ rest	
Other	

Table 7: Activity Type and Location by Residence Location (%)

Activity Location		TOD (N=7,104)	AOD-C (N=10,401)	AOD-F (N=12,686)
Activity type	Work	35	31	30
	School	4	8	8
	Shopping	22	22	23
	Social/Recreational	12	13	13
	Personal Business	16	17	16
	Meal	9	8	7
	Other	2	2	2
Activity location	TOD	66	43	21
	AOD-C	20	33	18
	AOD-F	14	24	61

Note: Only activities that individuals (age 5+) participated in were included.

Table 8: Activity Sequence and Location by Residence Location (%)

	Category	TOD Resident (N=3,149)	AOD-C Resident (N=4,403)	AOD-F Resident (N=5,120)
Activity type sequence	Work-Shop	16	15	13
	Work-Meal	15	10	11
	Shop-Shop	10	9	12
	PerBus-Shop	8	10	12
	Work-Work	8	10	8
	Work-PerBus	9	8	7
	PerBus-PerBus	4	5	5
	Work-SocRec	5	4	4
	SocRec-Meal	3	2	3
	Shop-SocRec	4	6	6
	PerBus-Meal	3	3	3
	Meal-Shop	3	3	4
	Other	10	14	14
Activity location sequence	TOD – TOD	56	32	14
	TOD – AOD-C	15	19	8
	TOD – AOD-F	6	7	9
	AOD-C – AOD-C	8	15	7
	AOD-C – AOD-F	6	12	14
	AOD-F – AOD-F	9	15	47

Note: Only activities that individuals (age 5+) participated in were included.

Table 9: Activity Location Choice Model Results

Independent variable		Coef.	MFX	z-statistic
Dependent variable		Activity location (1=TOD)		
Constant		-3.599	-	-19.610
Socio-demographic variable	Household size	-0.023	-0.008	-1.570
	Single member (1=yes)	0.038	0.013	0.920
	Num. of vehicles	-0.071 ***	-0.024	-3.880
	No vehicle (1=yes)	0.070 **	0.024	2.120
	Gender (1=male)	0.083 ***	0.028	3.220
	Age (years old)	0.010 **	0.003	2.000
	Age squared (years old)	0.000 ***	0.000	-2.820
	Household Income ($10,000)	0.016 ***	0.006	5.960
	Employed (1=yes)	0.237 ***	0.081	2.670
	Retired (1=yes)	0.164 **	0.056	1.680
	Disabled (1=yes)	-0.167	-0.057	-1.180
	Homemaker (1=yes)	0.030	0.010	0.270
	Unemployed (1=yes)	0.018	0.006	0.140
	Student (16+) (1=yes)	0.249 ***	0.085	2.800
Spatial context variable	Ori: Washington, D.C. (1=yes)	0.137	0.047	1.500
	Ori: Montgomery County (1=yes)	0.042	0.014	0.450
	Ori: Prince George's County(1=yes)	0.030	0.010	0.320
	Ori: Arlington County (1=yes)	0.085	0.029	0.890
	Ori: Fairfax County (1=yes)	0.335 ***	0.114	3.590
	Ori: Alexandria City (1=yes)	0.198 *	0.068	1.890
	Des: Washington, D.C. (1=yes)	3.153 ***	1.077	34.790
	Des: Montgomery County (1=yes)	1.976 ***	0.675	21.330
	Des: Prince George's County(1=yes)	1.312 ***	0.448	13.690
	Des: Arlington County (1=yes)	2.512 ***	0.858	26.580
	Des: Fairfax County (1=yes)	0.064	0.022	0.660
	Des: Alexandria City (1=yes)	1.538 ***	0.525	15.030
Temporal context variable	Jan-Mar (1=yes)	0.016	0.006	0.460
	Apr-Jun (1=yes)	0.024	0.008	0.630
	Jul-Sep (1=yes)	0.075 ***	0.025	2.040
	Monday (1=yes)	-0.023	-0.008	-0.560
	Tuesday (1=yes)	-0.010	-0.004	-0.260
	Wednesday (1=yes)	-0.035	-0.012	-0.870
	Thursday (1=yes)	-0.041	-0.014	-0.950
	Arrive from 5 AM to 9 AM (1=yes)	0.014	0.005	0.210
	Arrive from 9 AM to 2 PM (1=yes)	0.096	0.033	1.410
	Arrive from 2 PM to 8 PM (1=yes)	0.006	0.002	0.090
Activity-related variable	Meal (1=yes)	0.491 ***	0.168	5.350
	Personal business (1=yes)	0.266 ***	0.091	3.030
	School (1=yes)	-0.007	-0.002	-0.060
	Shopping (1=yes)	0.435 ***	0.148	5.010
	Social/Recreational (1=yes)	0.134	0.046	1.500
	Work (1=yes)	0.740 ***	0.253	8.440
Indicator variable	Resident of TOD (1=yes)	0.832 ***	0.284	23.190
	Resident of AOD-C (1=yes)	0.400 ***	0.137	13.360
Summary Statistics	Num. of observations	30191		
	Likelihood ratio $\chi 2$	14,857 ***		
	Log-likelihood (Constant)	-20,210		
	Log-likelihood (Full)	-12,367		
	Pseudo-R^2	0.388		
	Variance term	0.495 (0.703)		

Note: * p<0.10; ** p<0.05; *** p<0.01; A reference for job status is a student (16<); for destinations is the other areas; for months is Oct-Dec; for day of week is Friday; for the arrival time is 8PM-5AM; for the activity is others; for the indicator is a resident of AOD-F.

Table 10: Activity Location Sequence Model Results

Independent variable \ Dependent variable		Location sequence (1=TOD-TOD)		
		Coef.	MFX	z-statistic
Constant		-5.720 ***		-18.288
Socio-demographic variable	Household size	-0.078 **	-0.009	-2.309
	Single member (1=yes)	-0.035	-0.004	-0.405
	Num. of vehicles	-0.159 ***	-0.019	-3.905
	No vehicle (1=yes)	0.149 **	0.018	2.184
	Gender (1=male)	0.091	0.011	1.627
	Age (years old)	0.013	0.002	1.261
	Age squared (years old)	0.000 **	0.000	-2.110
	Household Income ($10,000)	0.029 ***	0.004	4.902
	Employed (1=yes)	0.886 ***	0.106	4.385
	Retired (1=yes)	0.885 ***	0.106	4.018
	Disabled (1=yes)	0.463	0.055	1.452
	Homemaker (1=yes)	0.791 ***	0.094	3.134
	Unemployed (1=yes)	0.476	0.057	1.600
	Student (16+) (1=yes)	0.380 *	0.045	1.673
Spatial context variable	Ori: Washington, D.C. (1=yes)	2.284 ***	0.272	16.031
	Ori: Montgomery County (1=yes)	1.319 ***	0.157	9.044
	Ori: Prince George's County(1=yes)	0.807 ***	0.096	5.196
	Ori: Arlington County (1=yes)	1.975 ***	0.236	13.368
	Ori: Fairfax County (1=yes)	-0.244	-0.029	-1.509
	Ori: Alexandria City (1=yes)	0.716 ***	0.085	4.303
	Des: Washington, D.C. (1=yes)	2.558 ***	0.305	16.995
	Des: Montgomery County (1=yes)	1.633 ***	0.195	10.683
	Des: Prince George's County(1=yes)	0.924 ***	0.110	5.811
	Des: Arlington County (1=yes)	1.967 ***	0.235	12.680
	Des: Fairfax County (1=yes)	-0.279	-0.033	-1.620
	Des: Alexandria City (1=yes)	0.836 ***	0.100	4.954
Temporal context variable	Jan-Mar (1=yes)	0.088	0.011	1.148
	Apr-Jun (1=yes)	0.096	0.011	1.187
	Jul-Sep (1=yes)	0.179 **	0.021	2.270
	Monday (1=yes)	0.226 **	0.027	2.565
	Tuesday (1=yes)	0.075	0.009	0.854
	Wednesday (1=yes)	0.042	0.005	0.466
	Thursday (1=yes)	0.146	0.017	1.583
	Arrive from 5 AM to 9 AM (1=yes)	0.131	0.016	1.363
	Arrive from 9 AM to 2 PM (1=yes)	0.540 ***	0.064	6.199
	Arrive from 2 PM to 8 PM (1=yes)	0.094	0.011	1.106
Activity-related variable	Discretionary-Discretionary (1=yes)	-0.133 **	-0.016	-2.087
	Discretionary-Obligatory (1=yes)	0.219 ***	0.026	3.642
	Obligatory-Discretionary (1=yes)	0.098 *	0.012	1.712
Indicator variables	Resident of TOD (1=yes)	0.917 ***	0.333	12.265
	Resident of AOD-C (1=yes)	0.320 ***	0.125	4.802
Summary statistics	Num. of observations	12,672		
	Likelihood ratio $\chi 2$	6,610 ***		
	Log-likelihood (Constant)	-7,815		
	Log-likelihood (Full)	-4,510		
	Pseudo-R^2	0.422		
	Variance term	1.562 (0.454)		

Note: * p<0.10; ** p<0.05; *** p<0.01; A reference for job status is a student (16<); for destinations is the other areas; for months is Oct-Dec; for day of week is Friday; for the arrival time is 8PM-5AM; for the activity is Obligatory to Obligatory; for the indicator is a resident of AOD-F.

PART 3: MACROSCOPIC DEMAND MODEL

1. DATA USED: HAMPTON ROADS, VIRGINIA

The data used in this study includes:

- Hampton Roads 2000, 2009 and 2034 Socioeconomic Data by TAZ, downed from HRTPO.org
- Hampton Roads 2000 and 2026 Socioeconomic Data (for traditional and smart growth scenarios) by TAZ, provided by HRTPO
- Hampton Roads Regional Travel Demand Model (Cube) provided by VDOT
- Hampton Roads Regional Travel Demand Model TransCAD converted Model (Cube) provided by Caliper
- 1990 and 2000 TAZ for Hampton Roads, provided by HRTPO

Socioeconomic data including residential and employment are based on TAZ (Transportation analysis zones), which is the subdivisions of geographical areas that are delineated for travel analysis and modelling purposes. However, the 1990 TAZ boundaries have significant positional differences when compared to the 2000 boundaries of the same TAZs. Therefore the 2026 socioeconomic data which is based on 1990 TAZ for smart growth scenarios cannot be applied directly to the HR regional travel demand model which is based on 2000 TAZ. In order to using the 2026 socioeconomic data as an input to HR travel demand model, it is important to check the difference between 1990 TAZs and 2000 TAZ. Changes are made mainly in suburban areas where TAZ units were relatively large therefore they are more likely to undergo splitting, merging and redefining with the land use development. After carefully comparison, an equivalent 2000 TAZ ID is assigned to 1990 TAZ ID. Note the boundary of these two TAZ systems does not match completely, appropriate adjustment is made to ensure reasonable accuracy. Figure 1 shows the study areas-Hampton Roads Areas and TAZ boundary.

2. SMART GROWTH SCENARIO DEVELOPMENT IN HAMPTON ROADS, VIRGINIA

2.1 Smart Growth Assumptions

In smart growth development, population and employment in future years are reallocated based on following assumptions: 20% of the projected growth is moved from the suburban communities to the central city areas.

This scenario assumes that all suburban growth should be reassigned according to a 60/20/20 formula with 60 percent of the growth projected in each suburban community to occur in its Greenfield areas. Another 20 percent of each suburban community's projected growth was assigned to its in-fill areas; and finally, 20 percent of each suburban community's projected growth was reallocated to the region's core or central cities. So the methodology arbitrarily reassigned 20 percent of the growth currently projected to occur in our suburban communities to the central cities. It was necessary to identify the central cities to which this reallocation would occur. Based on HRTPO's study, Hampton, Norfolk and Portsmouth were identified as the central cities which would receive the growth reassigned from the suburbs. The new population and employment totals within each locality are allocated based on the principles of Transit Oriented Design (Database). Consideration is given to local and regional accessibility, and proposed highway and transit plans. Figure 2 shows transit additions planned for the region.

2.2 Population and Employment Reallocation

A comparison between the smart and traditional growth scenarios is shown in Table 1. The table shows the reallocation of population growth for the two scenarios in future year. The difference and percent of difference between the two scenarios are also given. The difference between the two scenarios in a suburban city is exemplified by the city of Virginia Beach which is projected to experience an increase in population under the traditional growth scenario of 34,792 people over the 2009 to 2034 period but is projected to add just 16,025 people over the same period in the smart growth scenario. That represented a 4% reduction in the amount of population growth in Virginia Beach. By contrast, the core city of Norfolk is projected to add just 2,776 people to its population in the traditional growth scenario from 2009 to 2034 but will grow by 39,058 in the smart growth scenario. The difference (smart growth population change minus the traditional growth population change) between the two scenarios is shown in Figure 3. Similar to population relocation, Table 2 and Table 3 show the reallocation of retail and non-retail employment for BAU and SG scenarios.

3. MODEL RESULTS FOR HAMPTON ROADS, VIRGINIA

Two development scenarios are tested use TransCAD. The 2034 BAU (Business-as-Usual) scenario represents a relatively dispersed growth pattern. The SG (smart growth) scenario represents an alternative development by encouraging higher density and mixed land use development along transit lines. The light rail lines from Norfolk to Virginia Beach were included in the compact growth scenario, while highway capacity was added to both scenarios as dictated by projected congestion. The travel behavior results for the two future scenarios were examined. Table 4 shows the trip assignment and trip length by trip type for each scenario based on peak hours and daily level. The trip lengths for HBW, HBS and NHB are shorter by 3-6% in SG scenario compared with BAU scenario. Only HBO trips have longer average trip length in SG scenario.

When population, employment and road supply are fixed, the regional vehicle miles traveled decreased marginally by 1.93% in the SG scenario while VHT increased by 1.13% during peak hours with average speed dropping by 3.01% during peak hours. The regional VMT decreased by 0.72 million miles per day with the SG scenario (50.64 million vs 51.36 million). This is 1.41% decrease in regional VMT daily. However, not all counties in the region experience a drop in VMT. Figure 6 shows the impact on VMT for each county. The impact of SG on VMT varies across the region, with the net impact being a decrease in regional VMT. Table 5 shows the mode split on morning and afternoon peak hours, transit's mode share in the compact growth scenario is about twice the mode share in the BAU scenario. For 2034, transit miles travelled (all transit modes) is about 164.61% higher in the CG scenario, while car-based miles travelled, including carpooling and solo driving, is 7.92% lower in the CG scenario relative to the BAU scenario. Overall vehicle miles travelled decreased by 0.44% in the compact growth scenario and walking miles increased by 117%. From the supply side, transit supply is 118.57% higher and vehicle-based supply is 22.25% lower in the CG scenario than in BAU. Although the results are reasonable, the models may not be sufficiently sensitive to quantifying differences in CG and BAU scenarios.

4. REFERENCES

1. Database, T. (Accessed Apr. 1, 2012.). "Center for Transit-Oriented Development." http://toddata.cnt.org/.

LIST OF TABLES AND FIGURES

Figure 1: Study area: Hampton Roads region 52
Figure 2: Transit plans for the Hampton Roads region. 53
Figure 3: 2034 smart growth population allocations 54
Figure 4: 2034 smart growth employment allocations 55
Figure 5: Comparison of traditional and smart growth densities (City of Norfolk)..............56
Figure 6: Change in 2026 daily volume 57

Table 1: Population reallocation in different scenarios 58
Table 2: Retail employment reallocation in different scenarios...59
Table 3: Non-retail employment reallocation in different scenarios 60
Table 4: Non-Retail employment reallocation in different scenarios...................................61
Table 5: Peak-hour mode share and VMT for BAU and SG scenarios................................62

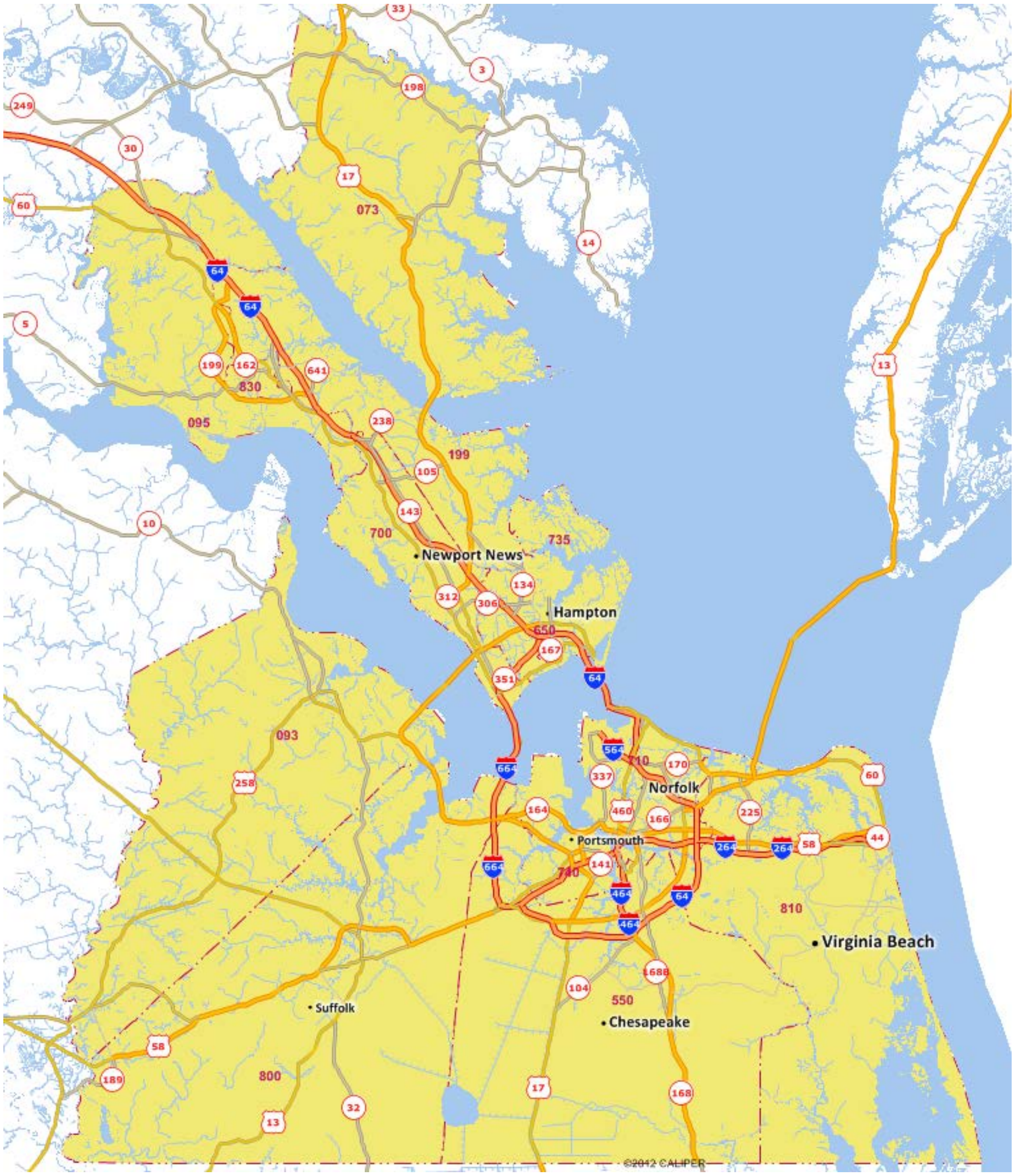


Figure 1: Study area: Hampton Roads region.

Figure 2: Transit plans for the Hampton Roads region.

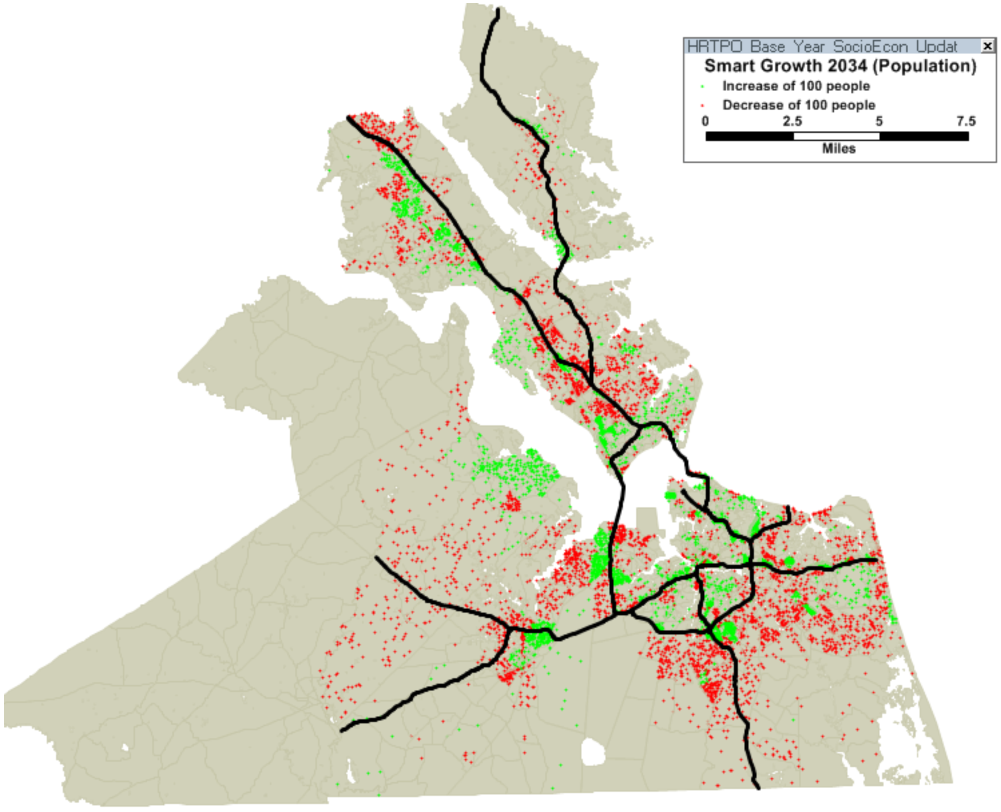
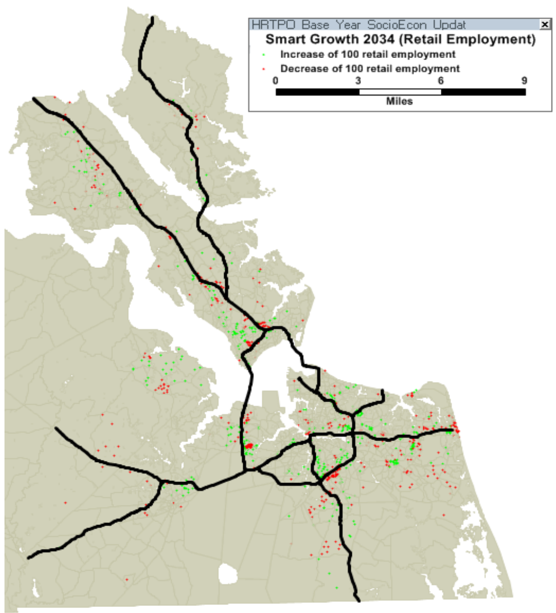


Figure 3: 2034 smart growth population allocations

Note: Green means population increased in those TAZs, Red means population decreased in those TAZs.



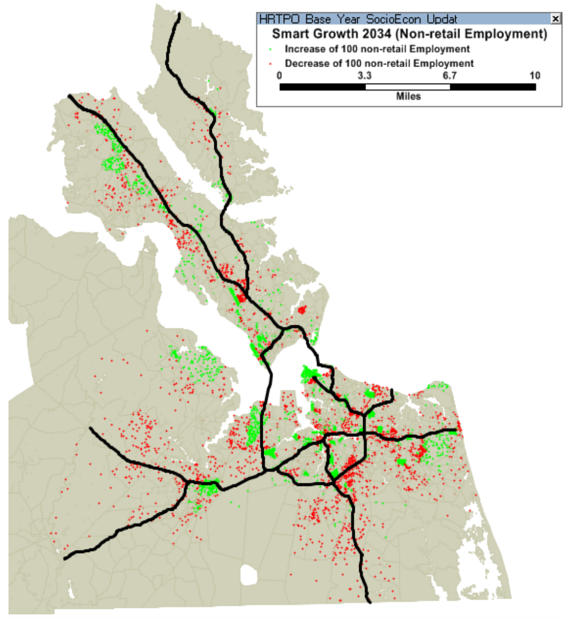


Figure 4: 2034 smart growth employment allocations

Note: Green means employment increased, Red means employment decreased.

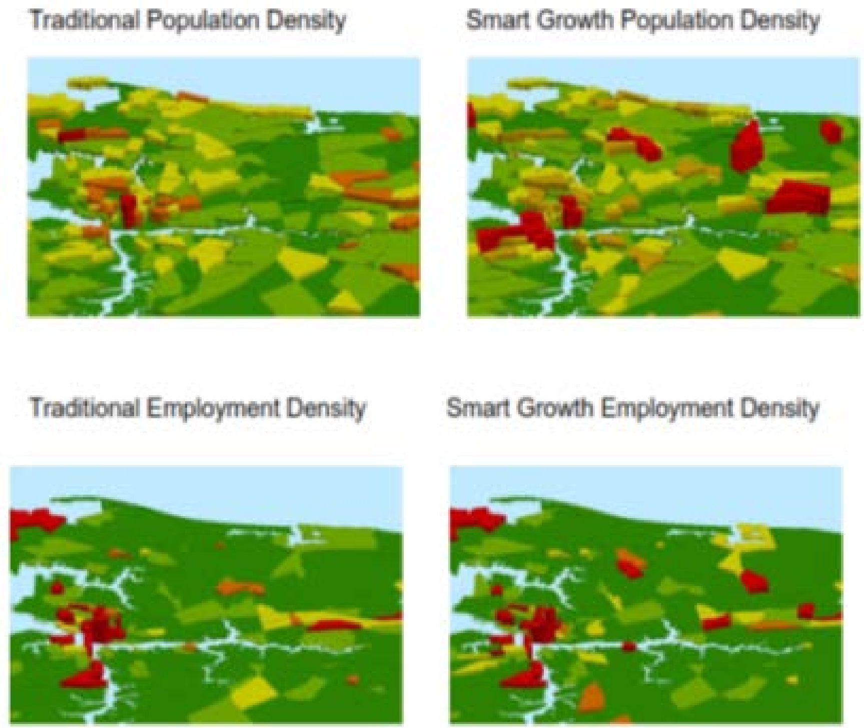


Figure 5: Comparison of traditional and smart growth densities (City of Norfolk).

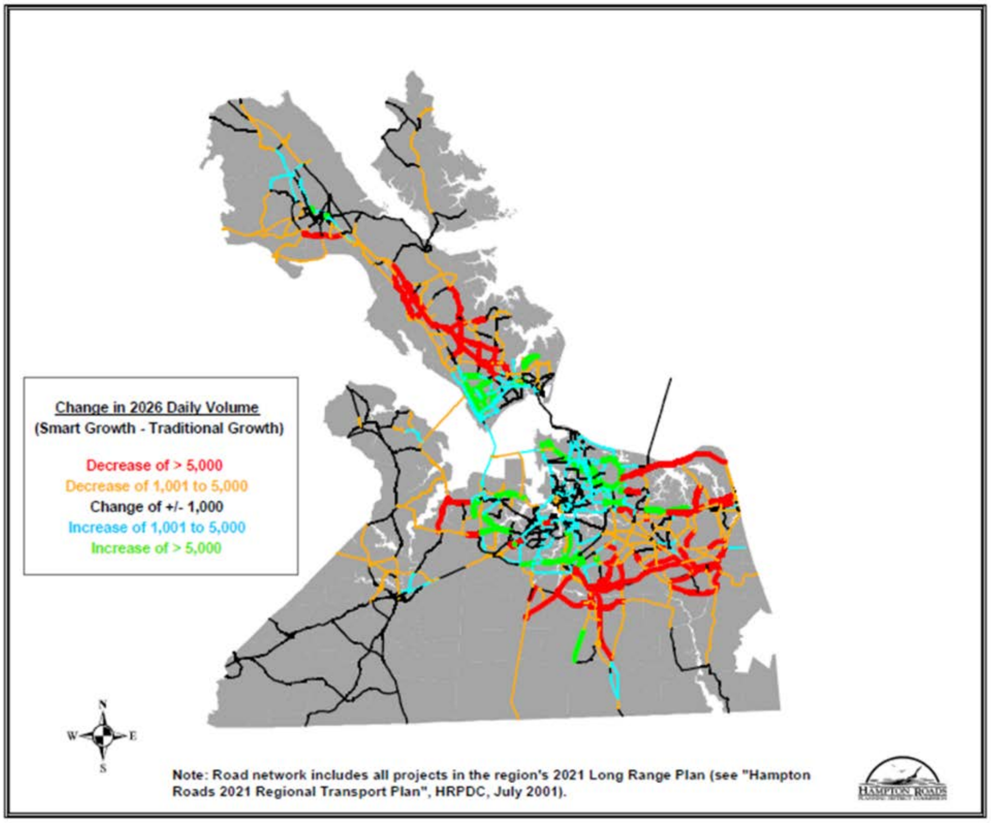


Figure 6: Change in 2026 daily volume (Source: HRPDO).

Table 1: Population reallocation in different scenarios

Population	*2009 Base*	*BAU2034*	*SG2034*	*Change*	*% Change*
Norfolk	237,622	240,400	276,680	36,280	15.09%
Virginia Beach	434,408	469,200	450,433	-18,767	-4.00%
Chesapeake	219,960	313,600	297,633	-15,967	-5.09%
Portsmouth	98,153	104,500	120,278	15,778	15.10%
Suffolk	83,008	180,600	164,974	-15,626	-8.65%
Isle of Wight	34,998	68,600	64,557	-4,043	-5.89%
Hampton	144,750	154,700	165,824	11,124	7.19%
Newport News	182,590	214,100	218,969	4,869	2.27%
Poquoson	11,881	15,000	14,129	-871	-5.81%
Williamsburg	13,568	19,100	17,866	-1,234	-6.46%
James City	63,691	110,100	102,259	-7,841	-7.12%
York	66,005	88,500	83,554	-4,946	-5.59%
Gloucester	24,523	36,700	37,942	1,242	3.38%
TOTAL	1,615,157	2,015,100	2,015,100	0	0

Table 2: Retail employment reallocation in different scenarios

Retail Emp.	***Base 2009***	***BAU2034***	***SG2034***	***Change***	***% Change***
Norfolk	25,172	24,000	27,075	3,075	12.81%
Virginia Beach	52,638	48,500	47,314	-1,186	-2.45%
Chesapeake	29,672	29,500	28,423	-1077	-3.65%
Portsmouth	6,330	4,700	5,304	604	12.85%
Suffolk	6,181	11,500	10,528	-972	-8.45%
Isle of Wight	2,633	3,800	3,485	-315	-8.29%
Hampton	14,014	15,100	15,931	831	5.50%
Newport News	19,442	19,200	19,553	353	1.84%
Poquoson	664	900	846	-54	-6.00%
Williamsburg	7,655	6,400	6,320	-80	-1.25%
James City	10,158	8,700	8,211	-489	-5.62%
York	8,977	6,800	6,457	-343	-5.04%
Gloucester	3,360	3,650	3,303	-347	-9.51%
TOTAL	186,896	182,750	182,750	0	0

Table 3: Non-retail employment reallocation in different scenarios

Non-retail Emp.	*Base 2009*	*BAU2034*	*SG2034*	*Change*	*% Change*
Norfolk	201,493	205,100	211,521	6,421	3.13%
Virginia Beach	207,313	227,590	225,983	-1,607	-0.71%
Chesapeake	96,827	130,110	127,697	-2413	-1.85%
Portsmouth	52,780	45,600	46,942	1342	2.94%
Suffolk	28,730	70,200	68,629	-1571	-2.24%
Isle of Wight	14,658	32,300	29,539	-2761	-8.55%
Hampton	68,073	71,800	73,377	1577	2.20%
Newport News	100,127	121,000	122,088	1088	0.90%
Poquoson	1,853	2,300	2,294	-6	-0.26%
Williamsburg	17,088	22,900	22,690	-210	-0.92%
James City	25,596	44,500	43,807	-693	-1.56%
York	27,489	31,500	31,079	-421	-1.34%
Gloucester	10,370	12,750	12,005	-745	-5.84%
TOTAL	852,397	1,017,650	1,017,650	0	0

Table 4: Non-retail employment reallocation in different scenarios

		Peak			Daily		
		BAU2034	***SG2034***	***Difference***	***BAU2034***	***SG2034***	***Difference***
Avg. Trip Length	HBW	11.30	10.88	-3.72%	10.44	10.01	-4.12%
	HBS	5.25	4.94	-5.90%	5.16	4.85	-6.01%
	HBO	6.05	6.27	3.64%	6.28	6.45	2.71%
	NHB	7.02	6.70	-4.56%	6.74	6.48	-3.86%
Trip Assig.	VMT	10,701,169	10,494,505	-1.93%	51,362,471	50,638,024	-1.41%
	VHT	325,847	329,541	1.13%	1,442,106	1,457,055	1.04%
	Avg. Speed	32.84	31.85	-3.01%	35.6	34.8	-2.25%

Table 5: Peak-hour mode share and VMT for BAU and SG scenarios

	*Mode share**			*VMT*		
	Business-as-usual	***Compact growth***	***% diff***	***Business-as-usual***	***Compact growth***	***% diff***
SOV	48.9%	38.7%	-20.9%	2,652,700	2,461,209	-7.2%
Carpool	34.7%	26.3%	-24.2%	1,975,681	1,800,564	-8.9%
Local bus	5.8%	12.3%	112.1%	154,228	367,290	138.1%
Express bus	0.1%	0.2%	100.0%	6,917	19,994	189.1%
Walk	9.5%	19.5%	105.3%	17,723	38,589	117.7%
Light rail	0.8%	1.5%	87.5%	26,125	57,729	121.0%
Bus rapid transit	--	0.3%		--	7,812	
Commuter rail	0.3%	1.3%	333.3%	22,566	110,240	388.5%

Note: * Because external trips are exogenous to the model, only Internal-internal home-based work trips and home-based other trips are used to calculate mode share.

REGIONAL MICROSCOPIC TRANSPORTATION SYSTEMS ANALYSIS

PART 1: WHAT IS THE LEVEL OF VOLATILITY IN INSTANTANEOUS DRIVING DECISIONS?

Xin Wang[1], Asad Khattak[2], Jun Liu[3], Golnush Masghati-Amoli[4], Sanghoon Son[5]

[1] Virginia Department of Transportaton, [2]The University of Tennessee, Knoxville, [3]The University of Texas, Austin, [2]University of Maryland, [3]Jeju Development Institute

Abstract: Driving styles can be broadly characterized as calm or volatile, with significant implications for traffic safety, energy consumption and emissions. How to quantify the extent of calm or volatile driving and explore its correlates is a key research question investigated in the study. This study contributes by leveraging a large-scale behavioral database to analyze short-term driving decisions and develop a new driver volatility index to measure the extent of variations in driving. The index captures variation in driving behavior constrained by the performance of the vehicle from a decision-making perspective. Specifically, instantaneous driving decisions include maintaining speed, accelerating, decelerating, maintaining acceleration/deceleration, or jerks to vehicle, i.e., the decision to change marginal rate of acceleration or deceleration. A fundamental understanding of instantaneous driving behavior is developed by categorizing vehicular jerk reversals (acceleration followed by deceleration), jerk enhancements (increasing accelerations or decelerations), and jerk mitigations (decreasing accelerations or decelerations). Volatility in driving decisions, captured by jerky movements, is quantified using data collected in Atlanta, GA during 2011. The database contains 51,370 trips and their associated second-by-second speed data, totaling 36 million seconds. Rigorous statistical models explore correlates of volatility that include socioeconomic variables, travel context variables, and vehicle types. The study contributes by proposing a framework that is based on defining instantaneous driving decisions in a quantifiable way using big data generated by in-vehicle GPS devices and behavioral surveys.

Key words: instantaneous driving decision; big data; volatility; acceleration; speed

1. INTRODUCTION

As the most dominant transportation mode in USA, automobile driving has significant impacts on traffic safety, energy, and emissions. With widespread deployment of emerging information and communication technologies, massive amounts of driving data in high resolution are becoming available, allowing researchers to scrutinize driving behavior in far more detail than was possible before. Insights can be obtained by studying instantaneous decisions made during driving in nearly real-time. Also, such “Big data” provides opportunities that support visualization, analysis, and modeling in new ways that could not be imagined before. The combination of data and tools can help create new visions that can potentially transform the way we monitor and evaluate transportation system performance and potential improvement actions. This study takes advantage of the big data collected by in-vehicle Global Positioning System (GPS) devices and survey data to define instantaneous driving decisions as drivers' choices of a set of options during driving. Such choices include maintaining speed, accelerating, decelerating, maintaining acceleration/deceleration, and vehicular jerk, i.e., the decision to change marginal rate of acceleration and deceleration. The sequential chaining of these short-term driving decisions can be volatile because they are intended to respond to the instantaneous changes in surrounding circumstances, such as approach of adjacent vehicles, pavement conditions, geometric transitions in the roadway, and weather conditions. Fluctuations in traffic flow can create challenges for safety, as well as challenges for energy consumption, tailpipe emissions and public health (Ji et al. 2011, Wang et al. 2013). Existing studies have shown that emissions and fuel usage vary significantly with different speed ranges (US EPA 2010). Additionally larger deviations from mean speed can significantly increase crash risk (TRB 1998). Accordingly it is important to understand and quantify variability in drivers' instantaneous decisions and explore the associations with socioeconomic, vehicular, and contextual variables.

Volatility in instantaneous driving decisions can be quantified by variability in vehicular movement, and the variability can be represented by speed and its derivative (acceleration/deceleration) as well as its second derivative (vehicular jerk). Micro level GPS data along with behavioral survey data are used to answer the following fundamental questions:

1) How to develop measures of driving volatility?
2) What is the level of volatility in instantaneous driving decisions?
3) What are the key correlates of driving volatility?

2. LITERATURE REVIEW

Aggressive driving and its impacts on traffic safety has been a concern of the public and many other sectors, including public transportation agencies, policy agencies, insurance companies, various organizations such as American Automobile Association. No consensus exists regarding "aggressive driving" in the literature. Social psychology researchers define it from the perspective of intent (Miles and Johnson 2003); for instance, "road rage" refers to more criminal-oriented offenses (Shinar 1998), while NHTSA classify "aggressive driving" as "driving actions that markedly exceed the norms of safe driving behavior and that directly affect other road users by placing them in unnecessary danger"(NHTSA 2009). Other researchers had a list of "aggressive driving" (James and Nahl 2000) including "weaving in and out of traffic", "driving at speeds far in excess of the norm which results in frequent tailgating, frequent and abrupt lane changes", "passing one or more vehicles by driving on the shoulder and then cutting in", or through certain syndrome of frustration-driven behaviors or negative cognitions such as annoyance, hostility, sustained horn-honking, glaring at others, yelling, gesturing, etc. (Shinar 1998, Underwood et al. 1999, Tasca 2000, Nesbit and Conger 2012). These studies in driving psychology largely depend on self-reported surveys of the driving public (Lajunen and Summala 2003, Miles and Johnson 2003), or video recording which requires manual identifications (Shinar and Compton 2004), with limitations on collecting data systematically and accurately. Critical research issues include: what are the so-called the norms of safe driving behavior; how to define a driver's extent of "aggressive driving" in a precise and quantifiable way.

While the research of "aggressive driving" in social psychology focuses more on peoples' intentions, the above driving behaviors and their cognitive processes in such driving situations are difficult to measure directly and continuously. Nevertheless, the speed profile as a common observable behavior is relatively easy to collect and has the potential of being utilized to characterize driving behavior. Measures used in the literature to identify aggressive or calm driving styles include the ratio of the standard deviation and the average acceleration within a specified time window (Langari and Jong 2005), ratio of standard deviation and vehicular jerk of the normal driving style (Murphey 2008).

Several critical cutoff points for aggressive behavior based on acceleration have been reported; 1.47 m/s^2 (4.82 ft/s^2) and 2.28 m/s^2 (7.47 ft/s^2) were reported as critical estimates of aggressive and extremely aggressive acceleration thresholds in urban driving environments (Kim et al. 2006, Kim and Choi 2013a). However, there is no consensus threshold, for instance, other researchers reported 0.45-0.65 m/s^2 (1.48-2.13ft/s^2) as calm driving, 0.85-1.10m/s^2 (2.79-3.61ft/^{2}) as aggressive driving for urban journeys (De Vlieger et al. 2000).

The percentage of time acceleration exceeds 1.5 m/s^2 (4.92 ft/s^2) was reported as one of the most important parameters (out of 16 parameters) contributing to increases in emissions and fuel consumption (Ericsson 2001). However, researchers argued that using acceleration alone may not represent the driving style accurately; therefore the coefficient of variance were also used as a complementary measurement in order to identify aggressive driving. Accordingly, accelerating at a relative regular rate, along with driving with medium acceleration but high standard deviation of acceleration are both flagged as aggressive driving (Langari and Jong 2005).

Connections between aggressive driving and safety were found in existing studies (Renski et al. 1999, Paleti et al. 2010). Paleti et al. (2010) have explored aggressive pre-crash behaviors and defined aggressive driving to include "speeding, tailgating, changing lanes frequently, flashing lights, obstructing the path of others, making obscene gestures, ignoring traffic control devices, accelerating rapidly from stop, and stopping suddenly." Their results show a positive association between injury severity and aggressive driving (given a crash).

Regarding emissions and fuel consumption, studies have shown that emissions can vary according to the decisions including both strategic decisions (vehicle selection and maintenance tactical decisions (selection of routes, dealing with congestion, and operational decisions (idling, speed selection, and use of cruise control) (Sivak and Schoettle 2012). A large number of studies have linked microscopic "aggressive" driving with emissions. Research has shown that peak emissions are associated with aggressive driving behavior including high speeds and extreme speed-ups or brake-downs (Holmén and Niemeier 1998, De Vlieger et al. 2000, Rouphail et al. 2001, Nam 2003, Nam et al. 2003). Factors describing speed, acceleration, power demand, and gear changing behavior are significantly associated with emissions (HC, NOx, and CO2) as well as fuel consumption (Ericsson 2001). An understanding of speed variation/ speed fluctuation/ driving dynamics, acceleration variation can further benefit research in energy and emissions.

While the literature provides insights, there is still a need to quantify the extent of volatile (aggressive) driving on routine urban journeys using continuous and reliable sources of data. This study is intended to close the gap between psychological studies and crash studies by applying appropriate empirical methods to quantify "volatile driving," and analyze the socio-demographic and travel correlates, which distinguishes this work from other driving behavior studies in social psychology, human factors, and safety fields. This study is also quite different from previous engineering-based aggressive driving studies because unique real-world GPS driving data along with reported behavior data from a survey are used to quantify the extent of variability in driving decisions. Considering that the word "aggressive" contains intent of the person, the use of term "volatility" in driving decisions is preferred in the paper, as it better suits the purpose of measuring the variability in instantaneous driving decisions.

3. DATA DESCRIPTION

Data used in this study come from the Atlanta Regional Commission — A Regional Travel Survey with GPS Sub-Sample conducted in 2011 (survey period covered Feb. 2011 through Oct. 2011). It was a well-executed regional survey using CATI (Computer-assisted telephone interviewing), with 6% final response rate and 34% participate rate. The sample is large-scale, covering about 20 counties in the region of Atlanta, representing various land use types and populations. Overall, the data quality was reasonable and efforts were made to make the sample representative of the region. More details about the survey are available in the report (PTV Nustats in Association with Geostats 2011). Similar to a standard travel behavior survey, the instrument relies on the willingness of households to 1) provide demographic information about the household, its members and its vehicles; 2) have all household members recording all travel-related details for a specific 24-hour period on multiple travel days, including their trip purposes, travel modes and other standard trip diary questions; 3) in the GPS subsample, data were collected by in-vehicle GPS devices for each trip. The device captured travel date, time, latitude and longitude (however this information was removed from the public released database), and the speed data. The GPS data points were collected at a sampling rate of at least 0.25 Hz and the raw GPS data was fed through a processing routine that removed outlying speed values, interpolated missing data and smoothed the speed profile (Atlanta Regional Commission 2011).

The final database contains different levels of data-personal data; household data, trip data, and microscopic second-by-second data for each trip. In all, 51,370 trips made by 1,653 drivers from 850 households were included in the database, which contained a total of more than 36 million seconds of records, covering driving practices on different road types by different type of vehicles.

The data was collected professionally, using state-of-the-art methods and upon examination show that it is reasonable. Specifically, for driving data, the speed data has reasonable ranges, with highest speed of 80 mph, average speed of 37 mph; acceleration changes ranged between -5.2ft/s2 and 7.64ft/s2, which are consistent with the numbers reported in the literature, e.g., 7.47ft/s2 as extremely aggressive driving (Kim and Choi 2013). Vehicular jerk changes ranged between -5.53 ft/s3 and 8.28 ft/s3. For demographics, again the data are reasonable. Specifically, 47.24% of drivers were male; the average age of respondents was 47 years. This fairly represents the driving population in Atlanta. Comparing the sampled data with other data sources such as the census showed that 47.24% of male drivers in the sample is consistent with 47.4% in the Atlanta are population; average age of 47.18 years, this is consistent with Census (49% of population is between 25 to 54); and average vehicle age of 7.9 years is consistent with 33.8% of vehicles in Atlanta area that are between 6-10 years old.

4. METHODOLOGY

4.1 Overall Methodology

Figure 1 shows the overall framework. The purpose of this study is to generate knowledge of short-term driving decisions by taking advantage of real-time in-vehicle speed data, increasingly available through monitoring devices. To do this, the research first defines different instantaneous driving decisions. Speed, acceleration, and vehicular jerk are extracted from the (large-scale) raw data, with decision patterns identified by chaining decisions with different sequences. Next, visualizing the data provides a complete picture of how drivers spend their time on these different driving decisions at different vehicular speeds. Then trip-based measures of short-term driving volatility are created based on acceleration and vehicular jerk profiles. Finally, statistical models are estimated in order to explore the socio-demographic and travel correlates of driving volatility, generating new knowledge about volatility.

Using vehicular jerk to represent instantaneous driving decisions

Distinct from strategic during decisions, instantaneous driving decisions refer to those micro-decisions to accommodate real-time situational changes during their journeys. These instantaneous driving decisions can include: accelerating, decelerating, maintaining constant speed (zero acceleration), jerking the vehicle (change in marginal rate of acceleration or deceleration), or maintaining constant acceleration and deceleration (zero vehicular jerk). As shown in equation 1, vehicular jerk is the derivative of acceleration or the second derivative of speed, representing abrupt movement of vehicles. Therefore, while an acceleration profile shows how fast a driver speeds up and slows down, a vehicular jerk profile shows how fast a driver accelerates and decelerates, which is more suited to capture drivers' abrupt adjustments in speeds. Figure 2 represents the speed, acceleration and vehicular jerk profile for a single sampled driving trip.

$$\begin{aligned} J &= d(a)/d(t) \\ &= d^2(v)/d(t)^2 \\ &= d^3(d)/d(t)^3 \end{aligned} \qquad \text{(Equation 1)}$$

Where,

J is vehicular jerk; a is acceleration; v is velocity; d is distance.

While these three profiles represent the same trip, they show significant differences, especially when speed fluctuates. The spikes in the vehicular jerk profile occur only when there are large changes in the accelerations, negatively or positively. The vehicular jerk profile acts as an amplification of speed changes since it is more sensitive to speed changes.

Different types of vehicular jerk

Different types of vehicular jerk profiles can be observed based on how acceleration and deceleration are chained sequentially. Figure 3 shows six different vehicular jerk styles during driving for illustrative purposes. The upper three graphs show vehicular jerks starting from acceleration and followed respectively by lower acceleration (a), higher acceleration (b), and deceleration (c). The lower graphs show vehicular jerks starting from a vehicle braking and followed respectively by a lower deceleration (d), higher deceleration (e), and acceleration (f). In these graphs, there is a decision point at second 10 when the driver has to decide whether he/she wants to change the current driving situation.

Since vehicular jerk is the second derivative of speed, it can be positive (b, d, f) or negative (a, c, e). Where vehicular jerk is zero, the driver operates the vehicle at a fixed acceleration/deceleration rate or simply maintains the speed. However, generally there can be a greater chance of collisions when negative vehicular jerk happens compared with positive vehicular jerk. In situations where vehicles are followed by other vehicles, negative vehicular jerks can result in abrupt shortening of distance between the vehicles and following vehicles, possibly creating a shockwave under condition c, e and a (a shockwave from strong to weak). Understanding the profiles of different vehicular jerk styles is important for safety and for energy and emissions.

5. EXTENT OF VOLATILITY IN DRIVING

In order to measure volatility in driving behaviors, a set of novel indicators are introduced to create profiles that describe the variance of instantaneous driving decisions from both regional and individual perspectives.

5.1 Acceleration behavior

Driving time use

To understand driving time spent on different instantaneous decisions in a metropolitan environment, the frequency of acceleration, deceleration and zero acceleration by speed bin in 0.5 mph (mile per hour) increments were calculated based on 36 million driving seconds of total 51,370 trips (shown in Figure 4 (a)). On selection of speed bin, we have conducted sensitivity analysis and found that volatility can be somewhat sensitive to the selection of different speed bin widths. There is no ideal bin size, but we know that if the bin size is too large (e.g., 5 mph), then the data are overly aggregated and there is substantial loss of variability (note that there are only 16 bins for speeds ranging from 0 mph to 80 mph). If the bin size is too small (e.g., 0.1 mph), then data noise (random fluctuations) can become an issue, obscuring interpretation (for 0.1 mph speed bins there will be 800 bins for 0 to 80 mph range). The 0.5 mph (equivalent to 0.73 ft/s) speed bin is a reasonable compromise that gives a fairly accurate picture of the acceleration and jerk distributions with respect to driving speeds.

Given that each sample represents one second of driving, the magnitude of frequency bars demonstrate the time used during trips on acceleration, deceleration and maintaining constant speed of the vehicle. Notably, very small accelerations or decelerations (0.03 mph, based on the 5th percentile of speed changes) were considered noise and coded as constant speed. Figure 4 (b) shows the percent of time spent on acceleration, deceleration and constant speed after standardization.

Overall 7% of driving time was spent driving at idling or low speeds (below 5 mph), 47% of driving time was spent on acceleration, 41% of driving time was spent on deceleration and 5% of driving time was spent maintaining constant speed, based on the massive amount of field data from GPS devices. The results can be compared with the Federal Test Procedure (FTP) drive cycle test (known as FTP-75 for the city driving cycle), which involves a decelerating drive mode for 34.5% of the time, and idling mode for 17.9% of the time (Ahn et al. 2002, Berry 2010). Table 1 shows major drive cycles designed to represent typical driving practices in order to certify vehicle fuel economy. The massive field driving data provides first-hand knowledge of real world driving practices, which can inform drive cycle design and provide insights.

Travel time spent at different speeds varies, depending on speed range, with 30-50 mph as the most common speed range. Less driving time was spent on driving at speeds higher than 50 mph. This result depends largely on regional road network structure. Overall greater amounts of driving time were spent on acceleration than deceleration, especially when speed was between 10-50 mph. However, more time was spent on deceleration compared with acceleration in lower speed bins (less than 10 mph). When speed is higher than 50 mph the travel time spent on acceleration and deceleration was nearly equal.

Notably, time spent on maintaining constant speed is much less than time spent on speed alterations. Relatively higher proportion of time is spent on maintaining constant speed when speeds are higher; specifically, more than 10% in speed bins higher than 55 mph and more than 20% at speeds higher than 70 mph. This is reasonable since less stop-and-go traffic is expected on freeways with free flowing traffic, coupled with the use of cruise control on interstates. Notably, neither the data on the use of cruise control nor the road types and second-by-second geo-codes are available in the public use database. This makes it difficult to link the speed profile/bins with specific roadway types, especially when speed is less than 50 mph. For example, the roadway can be a congested interstate or signalized arterial with free flowing traffic. Nevertheless, the graphs reveal useful information that helps understand driving time use. Specifically, the driving time spent on idling (traveling below 5 mph) is below 10% in Atlanta; the time spent on accelerating and braking are roughly equal and substantially higher than time spent on maintaining speed during urban journeys.

Variation in acceleration at different speeds

Most existing studies have applied a single acceleration value as a threshold for identifying aggressive driving. Ahn et al. (2002) have fitted a linear regression line showing that higher accelerations are associated with lower speeds. However, the nonlinear relationships between acceleration and speed in real-life driving situations are largely unexplored. Vehicle engines have to do more work in order to maintain the same acceleration at higher speeds to overcome the increasing air resistance. Therefore the ability to accelerate or decelerate a vehicle decreases naturally at higher speeds.

The speed vs. acceleration/deceleration profile (shown in Figure 5) is consistent with the above expectations. Upper and lower bands represent the means plus/minus one standard deviation bands for accelerations and they denote "typical driving practices." The (red) points that are out of the bands are the "volatile" driving seconds. In general, 15% of the 36 million seconds of driving are volatile (15.73% for acceleration and 14.50% for deceleration). This is reasonable since approximately 68% of the mass will be within one standard deviation for a bell-shaped normal speed distribution. Note that in order to separate the typical behaviors of drivers from moderately and highly risky behaviors, the use of 1 standard deviation threshold is reasonable. Using a 2 or 3 standard deviation threshold instead (i.e., capturing 95% and 99.7% of the observations for normally distributed data), will only leave extreme outliers, that are 5% or even lower (at 0.3%) portion of the data, i.e., high risk behaviors.

Bandwidth is the difference between the upper band value and the lower band value. A falling bandwidth reflects decreasing variation and rising bandwidth reflects increasing variation in speed changes. The largest bandwidth is between 10 mph and 30 mph and it decreases substantially when speed is higher than 40 mph. This confirms that at higher speeds (typically on freeways with a good level of service) drivers usually do not or simply cannot accelerate and decelerate abruptly. When speed is above 55 mph, accelerations scarcely exceed 1.5 feet/sec2, as reflected in the upper band.

A similar trend is observed in the deceleration profile with minor differences. Compared with acceleration, the magnitude of the maximum mean of deceleration is higher. It is -3.0 feet/sec2 for deceleration while the maximum mean value is less than 3.0 feet/sec2 for acceleration. This finding is interesting when combined with information contained in Figure 4. It revealed that in the Atlanta area, on average, drivers spend more time braking and they brake harder compared with accelerations.

5.2 Vehicular Jerk Behavior

Time use for vehicular jerk behavior

To understand how much time drivers spent on different vehicular jerk decisions, the time spent for the speed bins was aggregated by different vehicular jerk types. Then the results were standardized by calculating the percent of time spent on each vehicular jerk style, shown in Figure 6. Similar to the time spent on acceleration, the percent of time spent on zero vehicular jerks remains a small portion, this is especially true when speed is more than 70 mph. Possible reasons are drivers seem to avoid jerks to vehicles at higher speeds, or the use of cruise control is more common at higher speeds. However, the cruise control usage information was not available in the database, otherwise it would have added valuable information to understand instantaneous driving decisions comprehensively.

Different vehicular jerk styles (shown in Figure 3) are observed within different speed bins. Specifically, for the speeding up behaviors (a, b), Style (a) has a very small share when speed is less than 5 mph then reaches its peak (30%) when speed is around 30 mph, after that, it starts to shrink slightly but remains at least 20%. While style (b) has its largest share when speed is around 10 mph then remains at a 20% share constantly. As for slowing down behavior (d, e), style (d) has its largest share (30%) when speed is 5 mph, then remains relatively constant at 20% when speed increases; style (e) has its largest share when speed is close to zero, representing the hard braking behavior when coming to a stop. When speed increases, the percent of style (e) has peaks at 25% with moderate speeds (between 20 mph and 30 mph) and then remains constantly at 20% when speed is higher than 30 mph. As for the other two styles when acceleration and deceleration behavior are chained, both of style

(c) and style (f) account for about 5% and this percentage remains relatively constant at various speeds.

Aggregating the different jerk types by their direction, it was found that in general, more time were spent by drivers on negative vehicular jerk compared with positive vehicular jerk (53.79% vs. 40.83%). The difference between positive and negative vehicular jerk reaches its peak value when speed is between 20 mph and 40 mph. This speed range should perhaps receive more attention since negative vehicular jerk is expected to represent greater risks of rear-end collisions, especially in the absence of free-flow traffic. Notably, the time spent on vehicular jerk style (c), which possibly creates the highest risk to the following traffic, increases gradually with speed, especially when speed is larger than 65 mph. This result points to further investigating the relationship of such behavior with safety outcomes and exploring strategies for mitigating such behavior at higher speeds, e.g., by keeping safe distance when speeds are in this range.

Vehicular jerk variation by speed

Figure 7 (a) shows the distribution of the average vehicular jerk by different types and Figure 7 (b) the mean and standard deviation of vehicular jerk at different speeds. The difference in absolute magnitude of vehicular jerk reveals their intensity. Types (c) and (f) show the highest absolute magnitudes which is reasonable since both of them represent drivers reversing vehicle acceleration, i.e., going from acceleration to deceleration or vice versa. Note that type (f) has a higher absolute magnitude than its negative counterpart, i.e., type (c). This means that on average drivers jerk their vehicles more forcefully to accelerate after braking compared with the opposite. This is especially true when speed is less than 40 mph. The other two positive and the two negative jerk types show similar trends and values.

The upper band and lower band (mean plus/minus one standard deviation) are created respectively for the aggregated positive and negative vehicular jerk. For speed bins higher than 40 mph, the lower band of positive vehicular jerk is below zero and the upper band of negative vehicular jerk is above zero; hence zero were used in calculating the bandwidth in those cases. The upper band of the positive vehicular jerk and lower band of negative jerk collectively create a profile of regular practice for vehicular jerk. In other words, it represents the most typical driving practice on roadways regardless of road type. The bands can also serve as a critical threshold for identifying volatile driving behaviors, which are the red points falling outside the bands in Figure 7(b).

Based on 36 million seconds of driving data, about 13.36% seconds are identified as volatile seconds when using the vehicular jerk profiles. This score represents the average volatility level for typical driving practices for the GPS subsample from the Atlanta Metropolitan Area. More volatile driving practices are found within at lower speeds, as expected. Specifically, 16.4% of the total time drivers are volatile (above 1 standard deviation) when speed is lower than 20 mph, while 13.6% of the time they are volatile when speed is between 20-40 mph. This percentage drops to 12.00% for speed range between 40-60 mph and it is 11.9% for speeds larger than 60 mph.

The critical values of vehicular jerk associated with volatile driving behavior vary by speed. There is a peaking of this measure at speeds of 7.5 mph then it decreases gradually as vehicular speed goes up, until it reaches a steady line with minor fluctuations at speeds between 45-52 mph. In general, the bandwidth is larger at relatively low speeds (less than 20 mph) and it is relatively narrower at higher speeds. This is to say that lower speeds have a boarder range of volatile driving, but this is not the case for higher speeds.

5.3 Volatility Score for Each Trip

A new measure, termed driving volatility score was created after identify the volatile seconds. The idea is to measure individual volatility for each trip using the acceleration or vehicular jerk band. A driver's volatility score is defined as a percentage of time tagged as volatile seconds over the entire trip. In other words, volatility is measured as the percentage of time when the driver's acceleration or vehicular jerk goes beyond the typical driving thresholds (acceleration or vehicular jerk bands).

Figure 8 shows a comparison between the volatility scores generated using acceleration bands versus using vehicular jerk bands for a sampled trip. Less volatile seconds were identified using jerk bands compared with using acceleration bands; volatility score was 8.5% with jerk bands vs. 6.0% with acceleration bands for the trips analyzed. The jerk-based volatile seconds are not always in concordance with volatile acceleration-based volatile seconds. That is to say, sometimes the driver accelerated at a higher than the upper band level but he/she did not jerk the vehicle during this period. Conceptually, it is important to understand and identify key decision points when the driver abruptly changes driving actions, e.g., goes from acceleration to deceleration. Based on the observations shown in Figure 8, jerk seems to capture critical decision points better than acceleration while acceleration has more tolerance for volatility. Vehicular jerk can serve as an effective measurement to identify abrupt instantaneous decision changes. Since the volatility score is calculated for each trip, when data on multiple trips for a single driver are collected, average volatility score can be generated for each driver. This makes it is possible to compare both the intra-trip volatility and volatility between different drivers.

6. CORRELATES OF DRIVING VOLATILITY

After calculating the volatility scores (based on vehicular jerk bands) for each trip in the database, statistical models were estimated to investigate relationships between the volatility and driver demographics, vehicle characteristics and trip specifics. The database contained 51,370 trips made by 1,653 survey respondents in. After removing observations with missing information, the final database sample contained 40,240 trips by 1,486 respondents-—these are unique driver-vehicle pairs, labeled as driver-vehicle ID. Table 2 presents the descriptive statistics for the dependent and independent variables. The average volatility score is 13.84, which means that driving was volatile during 13.84% of the travel time (above or below mean vehicular jerk plus or minus one standard deviation). Some trips show calm driving (minimum score is 0.1%) while some were highly volatile when 55.46% of the time was spent on jerking vehicles at a higher level (outside of the bands).

In the final sample for modeling, 47.24% drivers were male; the mean age of respondent is 47.18, and a broad age range from 15 to 91. The mean vehicle age is 7.91 years and 43.88% of sampled vehicles were auto-sedans, 27.52% SUVs, and 13.59% pick-up trucks. As expected, 96.16% vehicles were gasoline-powered. 46.26% of trips were made during rush hours (6:00 am-10:00 am or 3:00 pm-7:00 pm); 24.37% were made on weekends; 19.49% were commute trips; the average trip duration was 14.17 minutes with an almost equal standard deviation–14.73. Overall, the data seems to be reasonable and in accordance with expectations.

The differences of volatility scores between trips can be result of the driving styles of different drivers (males vs. females, or young vs. older drivers), vehicle performance (new vehicles vs. older vehicles, body type, fuel type), or trip specifics (longer vs. shorter trips, commute vs. non-commute trips, and workday vs. weekend trips). Therefore simple Ordinary Least Squares (OLS) models were first estimated to test their associations. However, the traditional OLS models assume independence of observations and in this case multiple trips were made by the same drivers. Therefore, OLS will violate the independence assumption. One way to deal with correlated observations is to estimate a mixed-effect model, also called the mixed model. This model can capture correlated errors that arise from repeated observations in a group. In this study, the group variable is driver-vehicle pair; repeated variables are personal and vehicular characteristics; non-repeated variables are the measures for each specific trip. A "Driver-Vehicle ID" was created to represent different driver-vehicle pairs in the sample and was used as the random term in the mixed-effects model. The random term quantifies the error due to repeated variables. The mixed-effects regression model can contain both fixed and random terms, as shown in following equations.

$$Y = \beta X + \gamma Z + \varepsilon$$
$$\gamma \sim N(0, G)$$
$$\varepsilon \sim N(0, \sigma^2 I_n)$$

Y is the response vector of volatility score for each trip in the data; *X* is a vector of fixed independent variables (age, gender, vehicle body type, fuel type, vehicle age, trip duration, commute or not, peak hour or off-peak, weekend or not); β is a vector of estimated fixed effects for matrix *X*; and *Z* is a vector of random independent variables (*Driver-Vehicle ID*); γ is a vector of estimated random effects for matrix *Z*; ε is a vector of unknown random errors;

G is an diagonal matrix with identical entries for each fixed effect; I_n is an identity matrix; γ and ε are assumed to be independent.

Table 3 provides the modeling results for mixed models. Given that the distribution of vehicle jerk-based volatility scores is slightly right-skewed, square root transformed volatility score was tested as the dependent variable. However, the transformation improved the statistical properties of the model only marginally, e.g., significance of variables. Therefore, the original volatility score is used as the dependent variable, providing more intuitive parameter interpretation. Overall, the modeling results are reasonable, providing insights about a range of volatility correlates.

A key advantage of the mixed model over OLS model is that the random terms added into the mixed model structure can better model the effects of repeated observations within the group (driver-vehicle pair) by allowing various degrees of freedom for different variables according to their variations within groups. More specifically, all observations are treated equally in the OLS model regardless of their variations within or between groups. In this case, the overall sample size is 40,240 (the total number of trips). However, in the mixed model, only the sample size for generic variables (Ben-Akiva and Lerman 1985), (i.e., trip characteristics) with variations within groups remains the same (40,240), while the sample size for alternative-specific socioeconomic variables (i.e., driver and vehicle characteristics) become 1,486, which is the count of unique driver-vehicle pairs. As a result, larger standard errors are reported for alternative-specific socioeconomic variables in the mixed model. The estimated coefficients in the OLS and mixed models are nearly identical, but with different standard errors for driver and vehicle related terms, as expected. The following modeling interpretation is based on the mixed-effects model using the untransformed volatility score.

Full and final models are presented, with the final model containing only the statistically significant variables (10% level). The results of the final are discussed. The models have a reasonably good fit, explaining 40.3% of the variation in volatility score. As expected, younger drivers exhibit higher volatility in driving (5% level). A ten year increase in driver age is associated with a decrease of 0.57 in volatility scores. However, there is no statistical evidence for association between volatility score and drivers' gender. Driving volatility varies significantly with vehicle characteristics, including vehicle body type, vehicle age and fuel type. The results show that two-seat sports cars are associated with higher volatility, possibly due to their higher horse power. Trips made by two-seat sports cars drivers have 3.28 higher volatility scores, compared with trips made by drivers in the "base" category that includes sedans, RVs, station wagons, and SUVs. While van drivers show 1.82 lower volatility compared with drivers in the base category, perhaps due to their larger size and more sluggish performance. The use of hybrid vehicles shows lower volatility (-1.98) compared with gasoline and diesel vehicles. The volatility scores are lower for older vehicles, perhaps due to their engine performance. A year added to vehicle age is associated with a 0.10 units decline in the volatility score.

Volatility score also shows significant correlation with trip specific factors, including trip duration, time of day, day of the week, and trip purpose. Compared with non-rush hour trips, there is a 0.24 units increase in volatility score during rush hours; compared with workday trips, the decrease in volatility score for weekend trips is 0.30 units; for commute trips, the increase in volatility score is 0.36 units compared with non-commute trips; and a one-minute increase in trip duration is associated with a 0.04 units lower volatility score.

High levels of correlations among explanatory variables were checked and we did not find them to be high. One example is that of commute trips which are typically made during peak hours. In the data, 46.28% of the trips were made during rush hours and 19.51% of the trips were for commute purposes. While these two variables capture different aspects of travel, i.e., time of day and trip purpose, the correlation between them was relatively low (0.156), justifying their joint inclusion in the model.

Examination of the random effects, reported as variance component estimates, shows a sizable variation (34.84%) in the volatility score across driver-vehicle pairs. This further justifies the use of the mixed model. Note that the models presented in this paper show an effort to test whether the measurement of volatility can be used to quantify the relationships between instantaneous driving decisions and other variables that include personal, vehicular, situational context factors. The random effects model confirmed that volatility score varies significantly between different driver-vehicle pairs. However, it does not fully disentangle volatility variations between different driving trips made by the same driver. A more sophisticated hierarchical modeling framework will be needed for answering such questions (Liu et al. 2014a).

7. LIMITATIONS

This study depends heavily on GPS data collected by in-vehicle devices. To some extent the accuracy and availability of location data constrain the analysis. Compared with high industrial sampling rates (e.g. 96 kHz), these data are limited by relatively low sampling frequency which gives only second-by-second speeds. A reasonable question is whether second-by-second speed data are good enough for identifying instantaneous driving decisions. To address this issue, additional analyses were conducted by collecting driving data at 20 Hz using a driving simulator (Liu et al. 2015c). This database includes 35,924 seconds speed data made by 24 drivers, generating 718,481 speed data points, which allows the investigation of micro-driving decision changes within one second. The results show that drivers made no change to their speed for 89.9% of the sampled seconds, i.e., drivers either kept accelerating, decelerating or just maintained speed during a second. Only 10.1% of the sampled seconds involve driver’s decision change. Overall, the analysis found that at least 98.5% instantaneous driving decision changes can be detected using second-by-second data compared with smaller intervals and that the second-by-second data are reasonably accurate for the purposes of this study.

Some other critical information remains unknown to the researchers due to privacy concerns. This includes the type of roads and the geo-codes for each second of driving. Missing geographically referenced information for trips prevents the researchers from extracting useful contextual factors. These include roadway segments used during trips and associated traffic counts, road geometry, traffic operations facilities, and surrounding land uses. Therefore, how the instantaneous decisions are associated with surrounding traffic, facility and land use can be analyzed adding interesting findings. This paper presents an attempt to enhance understanding of volatility in instantaneous driving decisions. More research is needed to investigate the impacts of network attributes, environmental attributes on instantaneous decisions, as shown in the conceptual framework. Expansion of the study can form the basis of future analysis of driver volatility and how it relates to energy, environment and safety.

8. CONCLUSIONS

In the context of using big data for traffic safety improvement, tailpipe emissions and energy use reduction in a driving dominant environment, it is essential to understand drivers' instantaneous driving decisions and their associated impacts. The research takes advantage of large-scale driving databases coupled by second-by-second GPS data to develop a framework for the research agenda in driving behavior studies addressing how to define the instantaneous driving decisions in a quantifiable way and how to quantify explicitly volatile driving in a defensible manner. The answer is to create a volatility indicator to measure the gap between an individual's driving practice and the typical driving practice in that region. Assuming the typical driving practice applied by most people represents the norm of driving culture in that region, the driving practices standing out of that norm could be defined as volatile driving. The paper demonstrates a methodology to measure the volatility, which is based on variance in vehicular jerk between individual drivers and regional sample profiles. The creation of a robust volatility score that is able to quantify the extent of volatility, instead of simply labeling a driver as aggressive or non-aggressive is a key contribution.

To create a typical driving profile for the study metropolitan area, acceleration or vehicular jerk distributions were analyzed using speed bins and enveloped by an upper and lower band (mean plus/minus one standard deviation). While typical driving practices are identified when the acceleration or vehicular jerk fall between the bands, volatile driving is defined as accelerations or vehicular jerks that fall out of the bands range. A volatility score for each trip or each driver can be calculated by the percent of travel time spent on volatile driving. In this sense, developing a regional driving profile is critical since this driving profile serves as a "standard" to define individual's driving volatility. Atlanta's driving profile was developed through an innovative visualization of data, the time spent on each driving behavior was calculated. Specifically, overall 14% of the travel time spent on high vehicular jerk; 7% of driving time was spent on idling or traveling at speeds below 5 mph, 47% of driving time was spent on acceleration, 41% of driving time was spent on deceleration and 5% of driving time was spent on maintaining constant speed. This information can be useful for designing driving cycle in a local context for better emissions estimations. The methodology has great potential to be expanded to measure driving volatility on road infrastructures as an indicator of roadway safety. Roads with higher risk (those experiencing more hard braking and negative jerks) can be identified and proactive strategies can be designed.

Individual level driving volatility also has a practical value. It can be potentially incorporated in advanced traveler information systems applications, e.g., driving behavior monitoring and feedback devices can use volatility information to provide alerts and warnings, network-based microscopic simulations and emission models can use volatility information for more accurate predictions (Bandeira et al. 2013, Bandeira et al. 2014, Khattak et al. 2015). Drivers can check their volatility scores at the end of each day or even instantaneously, knowing where and when they exhibited high volatility. Moreover, statistical models confirmed that volatility scores significantly vary between drivers, which can support the early identification of risk-prone drivers. Volatile practices of risk-prone drivers can be potentially targeted through early warning systems. Applications can be embedded in navigation systems to send drivers warnings when they repeatedly show highly volatile driving during a trip.

9. ACKNOWLEDGEMENT

The research was supported through the TranLIVE University Transportation Center grant sponsored by the Research and Innovative Transportation Administration, of US DOT. The data was provided by the National Renewable Energy Laboratory. Special thanks are extended to the following entities for their support: Transportation Research Institute and the Center for Innovative Transportation Solutions at Old Dominion University and the University of Tennessee Transportation Engineering & Science Program. The views expressed in this paper are those of the authors, who are responsible for the facts and the accuracy of the data presented herein.

10. REFERENCES

1. Ahn, K., Rakha, H., Trani, A. & Van Aerde, M. 2002. Estimating Vehicle Fuel Consumption and Emissions based on Instantaneous Speed and Acceleration Levels. *Journal Of Transportation Engineering*, 128**,** 182-190.
2. Atlanta Regional Commission. 2011. *Atlanta Regional Commission (ARC) Regional Travel Survey* [Online]. Available: http://www.atlantaregional.com/transportation/travel-demand-model/household-travel-survey [Accessed 08-01 2014].
3. Bandeira, J., Almeida, T. G., Khattak, A. J., Rouphail, N. M. & Coelho, M. C. 2013. Generating emissions information for route selection: Experimental monitoring and routes characterization. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 13**,** 3-17.

4. Bandeira, J. M., Carvalho, D. O., Khattak, A. J., Rouphail, N. M., Fontes, T., Fernandes, P., Pereira, S. R. & Coelho, M. C. 2014. Empirical Assessment of Route Choice Impact on Emissions Over Different Road Types, Traffic Demands, and Driving Scenarios. *International Journal of Sustainable Transportation*, null-null.
5. Ben-Akiva, M. & Lerman, S. 1985. *Discrete choice analysis: theory and application to travel demand*, Cambridge, Mass, MIT press.
6. Berry, I. 2010. *The Effects of Driving Style and Vehicle Performance on the Real-World Fuel Consumption of U.S. Light-Duty Vehicles.* Master of Science, Massachusetts Institute of Technology.
7. De Vlieger, I., De Keukeleere, D. & Kretzschmar, J. 2000. Environmental effects of driving behaviour and congestion related to passenger cars. *Atmospheric Environment,* 34**,** 4649-4655.
8. Ericsson, E. 2001. Independent driving pattern factors and their influence on fuel-use and exhaust emission factors. *Transportation Research Part D: Transport and Environment,* 6**,** 325-345.
9. Holmén, B. & Niemeier, D. 1998. Characterizing the effects of driver variability on real-world vehicle emissions. *Transportation Research Part D: Transport and Environment,* 3**,** 117-128.
10. James, L. & Nahl, D. 2000. *Road rage and aggressive driving: Steering clear of highway warfare*, Amherst, New York: Prometheus Books.
11. Ji, S., Cherry, C., Bechle, M., Wu, Y. & Marshall, J. 2011. Electric Vehicles in China: Emissions and Health Impacts. *Environmental Science & Technology,* 46**,** 2018-2024.
12. Khattak, A., Liu, J. & Wang, X. 2015. Supporting instantaneous driving decisions through vehicle trajectory data. Submitted to *Journal of Intelligent Transportation Systems*.
13. Kim, E. & Choi, E. 2013. Estimates of Critical Values of Aggressive Acceleration from a Viewpoint of Fuel Consumption and Emission. *TRB 2013 Annual Meeting.* Washington D.C.
14. Kim, H., Oh, J. & Jayakrishnan, R. 2006. Application of Activity Chaining Model Incorporating a Time Use Problem to Network Demand Analysis. *Transportation Research Record: Journal of the Transportation Research Board,* 1977**,** 214-224.
15. Lajunen, T. & Summala, H. 2003. Can we trust self-reports of driving? Effects of impression management on driver behaviour questionnaire responses. *Transportation Research Part F: Traffic Psychology and Behaviour,* 6**,** 97-107.
16. Langari, R. & Jong, W. 2005. Intelligent energy management agent for a parallel hybrid vehicle-part I: system architecture and design of the driving situation identification process. *Vehicular Technology, IEEE Transactions on,* 54**,** 925-934.
17. Liu, J., Khattak, A. & Han, L. 2015. What is the Magnitude of Information Loss When Sampling Driving Behavior Data? *94th Annual Meeting of the Transportation Research Board.* Washington D.C.

18. Liu, J., Khattak, A. & Wang, X. 2014. The Role of Alternative Fuel Vehicles: Using Behavioral and Sensor Data to Model Hierarchies in Travel. Submitted to *Transportation Research Part C: Emerging Technologies*.
19. Miles, D. & Johnson, G. 2003. Aggressive driving behaviors: are there psychological and attitudinal predictors? *Transportation Research Part F: Traffic Psychology and Behaviour,* 6**,** 147-161.
20. Murphey, Y. 2008. Intelligent Vehicle Power Management: An Overview. *In:* PROKHOROV, D. (ed.) *Computational Intelligence in Automotive Applications.* Springer Berlin Heidelberg.
21. Nam, E. 2003. Proof of Concept Investigation for the Physical Emissions Estimator (PERE) for MOVES, EPA420-R-03-005. Office of Transportation and Air Quality, EPA.
22. Nam, E., Gierczak, C. & Butler, J. 2003. A Comparison Of Real-World and Modeled Emissions Under Conditions of Variable Driver Aggressiveness. *In:* CD-ROM, T. A. M. (ed.) *Annual TRB Meeting.* Washington DC.
23. National Highway Traffic Safety Administration 2009. Countermeasures that work: A highway safety countermeasure guide for state highway safety offices, Report No. DOT HS 811081. Washington, DC: National Highway Traffic Safety Administration.
24. Nesbit, S. & Conger, J. 2012. Predicting aggressive driving behavior from anger and negative cognitions. *Transportation Research Part F: Traffic Psychology and Behaviour,* 15**,** 710-718.
25. Paleti, R., Eluru, N. & Bhat, C. 2010. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. *Accident Analysis & Prevention,* 42**,** 1839-1854.
26. PTV Nustats in Association with Geostats. 2011. *Atlanta Regional Commission Regional Travel Survey Final Report* [Online]. Available: http://www.atlantaregional.com/File%20Library/Transportation/Travel%20Demand%20Model/tp_2011regionaltravelsurvey_030712.pdf [Accessed 08-01 2014].
27. Renski, H., Khattak, A. & Council, F. 1999. Effect of Speed Limit Increases on Crash Injury Severity: Analysis of Single-Vehicle Crashes on North Carolina Interstate Highways. *Transportation Research Record: Journal of the Transportation Research Board,* 1665**,** 100-108.
28. Rouphail, N., Frey, C., Colyar, J. & Unal, A. 2001. Vehicle Emissions and Traffic Measures: Exploratory Analysis of Field Observations at Signalized Intersections. *80th Annual Meeting of the Transportation Research Board.* Washington D.C.
29. Shinar, D. 1998. Aggressive driving: the contribution of the drivers and the situation. *Transportation Research Part F: Traffic Psychology and Behaviour,* 1**,** 137-160.
30. Shinar, D. & Compton, R. 2004. Aggressive driving: an observational study of driver, vehicle, and situational variables. *Accident Analysis & Prevention,* 36**,** 429-437.
31. Sivak, M. & Schoettle, B. 2012. Eco-driving: Strategic, tactical, and operational decisions of the driver that influence vehicle fuel economy. *Transport Policy,* 22**,** 96-99.

32. Tasca, L. 2000. *A Review of the Literature on Aggressive Driving Research* [Online]. Available: http://www.stopandgo.org/research/aggressive/tasca.pdf [Accessed 11-03 2014].
33. Transportation Research Board 1998. TRB Special Report 254: Managing Speed: Review of Current Practice for Setting and Enforcing Speed Limits. Washington D.C.: Transportation Research Board.
34. U.S. Environmental Protection Agency. 2010. *Techinical Guidance on the Use of MOVES2010 for Emission Inventory Preparation in State Implementation Plans and Transportation Conformity* [Online]. Transportation and Regional Programs Division, U.S. Environmental Protection Agency. Available: http://www.epa.gov/otaq/models/moves/420b10023.pdf [Accessed 08-01 2014].
35. Underwood, G., Chapman, P., Wright, S. & Crundall, D. 1999. Anger while driving. *Transportation Research Part F: Traffic Psychology and Behaviour,* 2**,** 55-68.
36. Wang, X., Khattak, A. & Zhang, Y. 2013. Is Smart Growth Associated with Reductions in Carbon Dioxide Emissions? *Transportation Research Record: Journal of the Transportation Research Board,* 2375**,** 62-70.

LIST OF TABLES AND FIGURES

Figure 1: Overview of research methodology 83
Figure 2: Comparison between speed, acceleration and vehicular jerk profiles on a trip 84
Figure 3: Different types of vehicular jerk during driving. 85
Figure 4: Distribution of acceleration, deceleration, and constant speed at different vehicle speeds 86
Figure 5: Average acceleration/deceleration at different speeds 87
Figure 6: Time spent on different vehicular jerk styles 88
Figure 7: Vehicular jerk distribution by speed bins 89
Figure 8: Volatile driving identified by different methods 90

Table 1: United States certification drive cycles compared with Atlanta drive cycle 91
Table 2: Descriptive statistics for dependent and independent variables 92
Table 3: Results of the mixed model using volatility score as the dependent variable 93

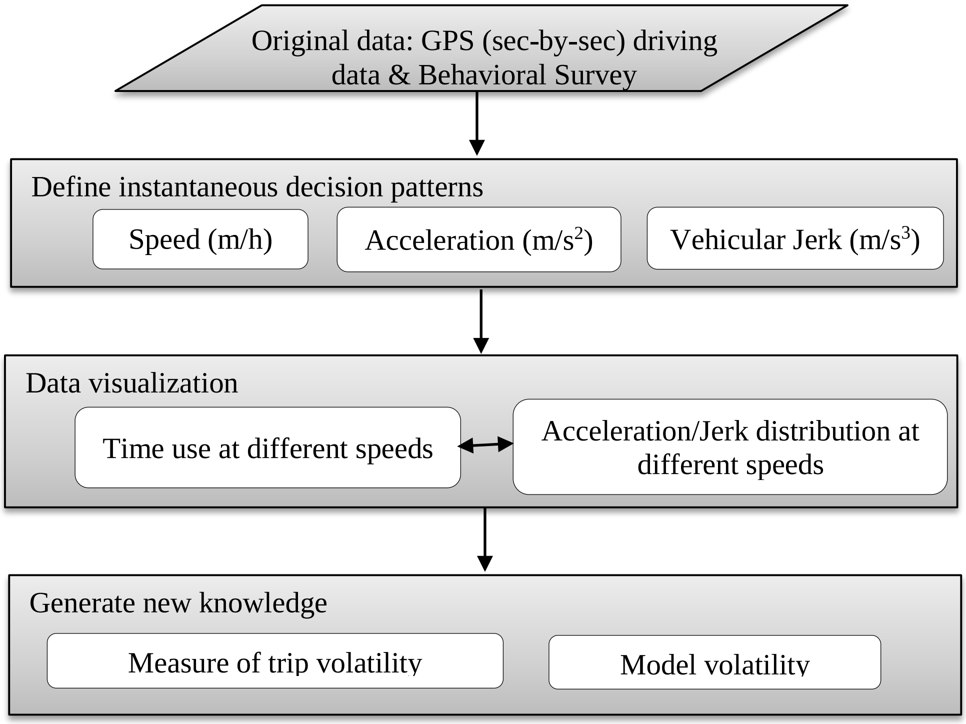


Figure 4: Overview of research methodology.

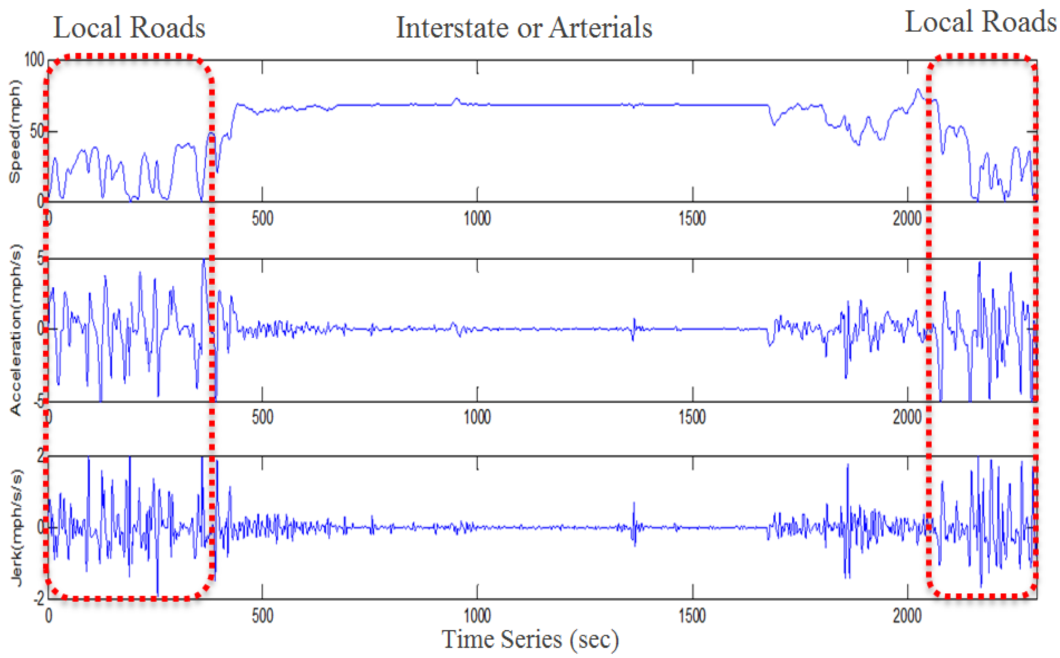
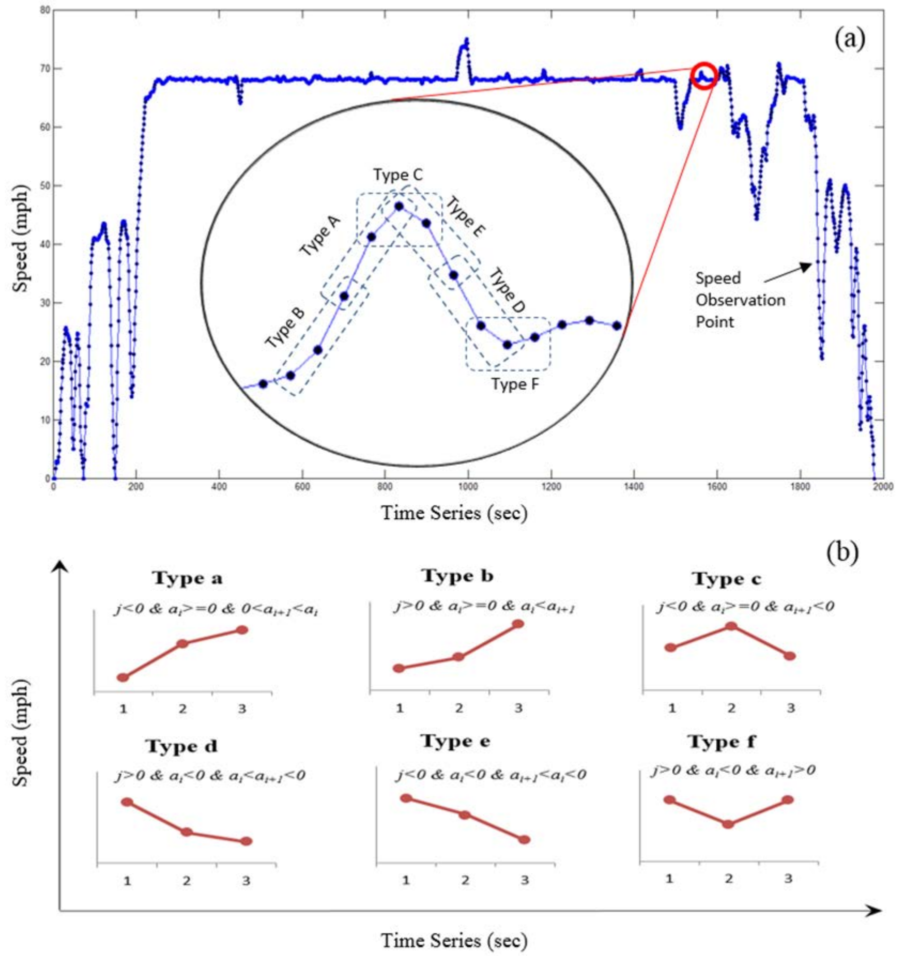
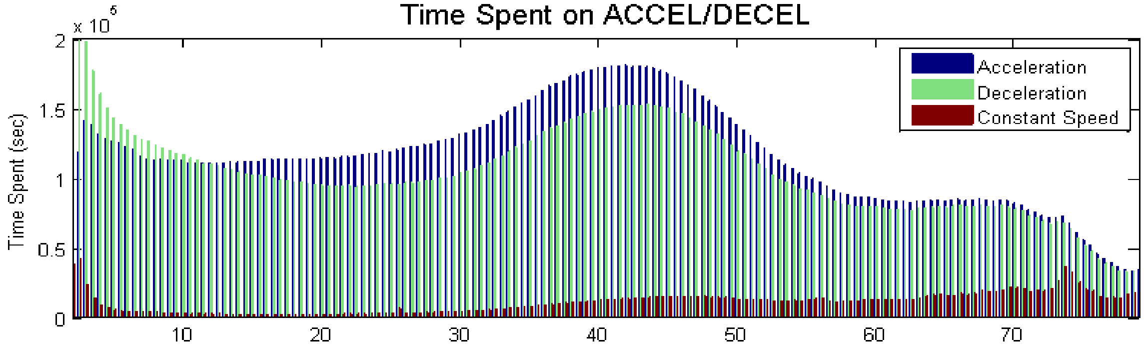


Figure 5: Comparison between speed, acceleration and vehicular jerk profiles on a trip.

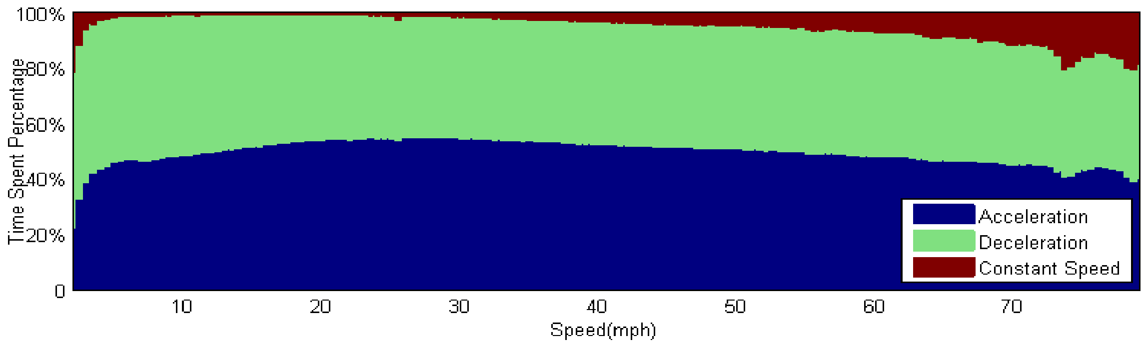


Notes: j=vehicular jerk; a_i=acceleration at time i; a_{i+1}=acceleration at time $i+1$.

Figure 6: Different types of vehicular jerk during driving.



(a) Time spent on acceleration, deceleration and constant speed



(b) Time spent percent (after standardization)

Figure 7: Distribution of acceleration, deceleration, and constant speed at different vehicle speeds. (N= 36,645,599).

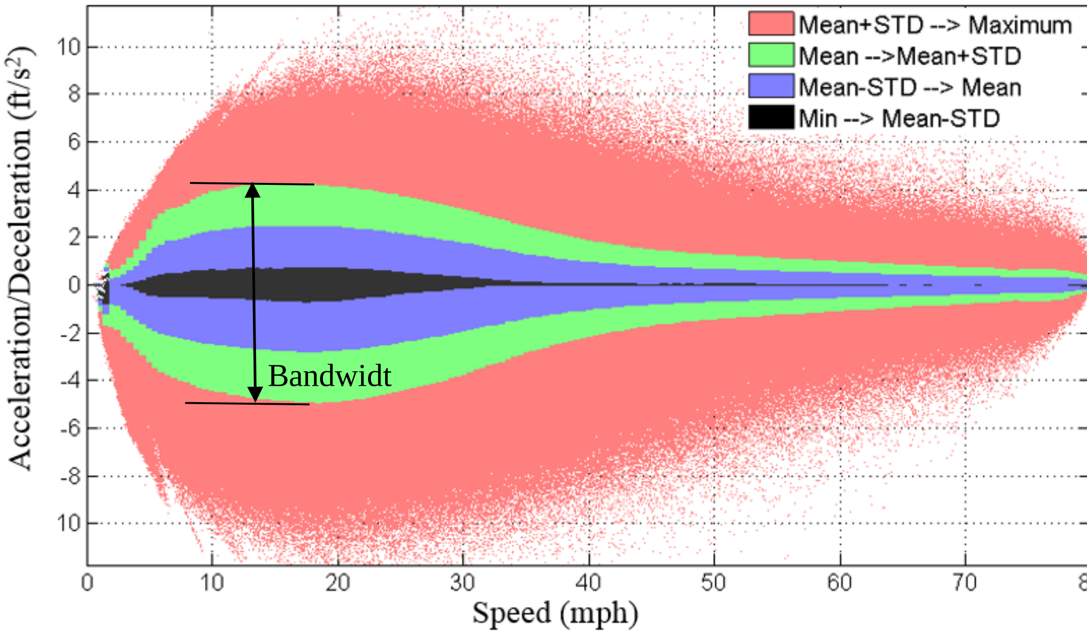


Figure 8: Average acceleration/deceleration at different speeds (N=36,715,308).

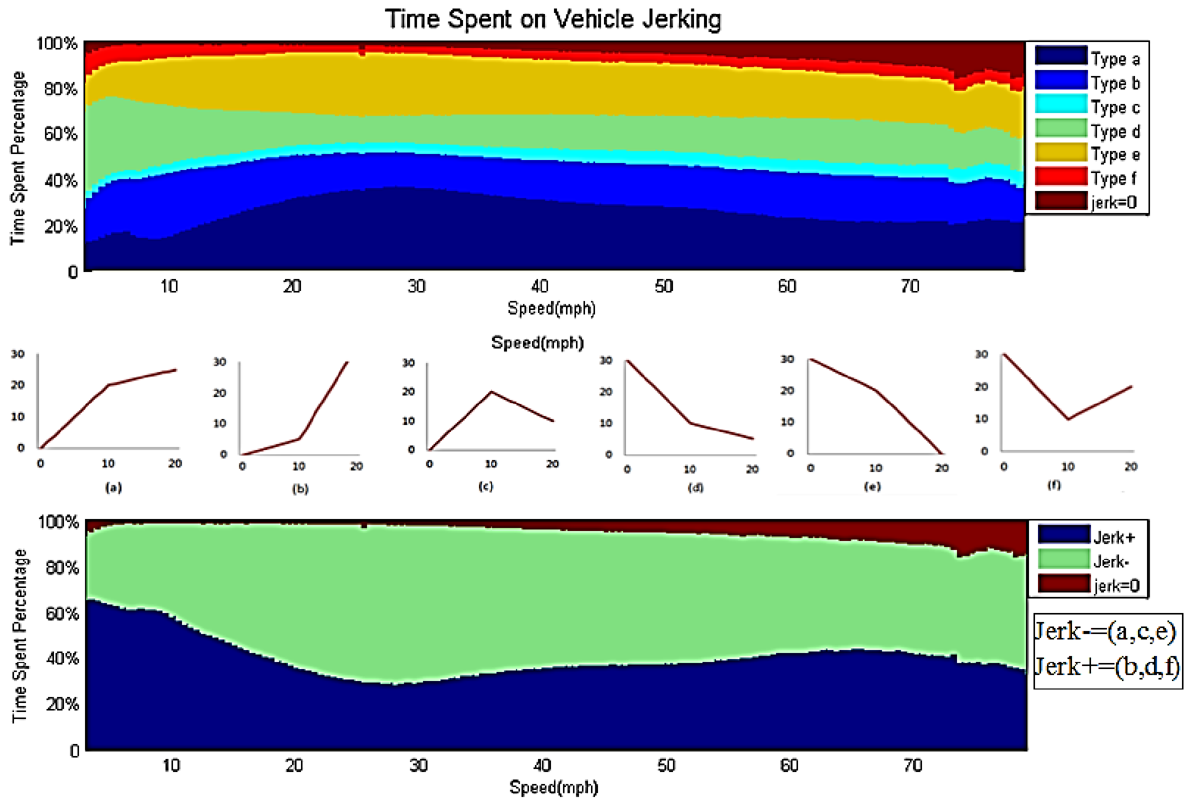
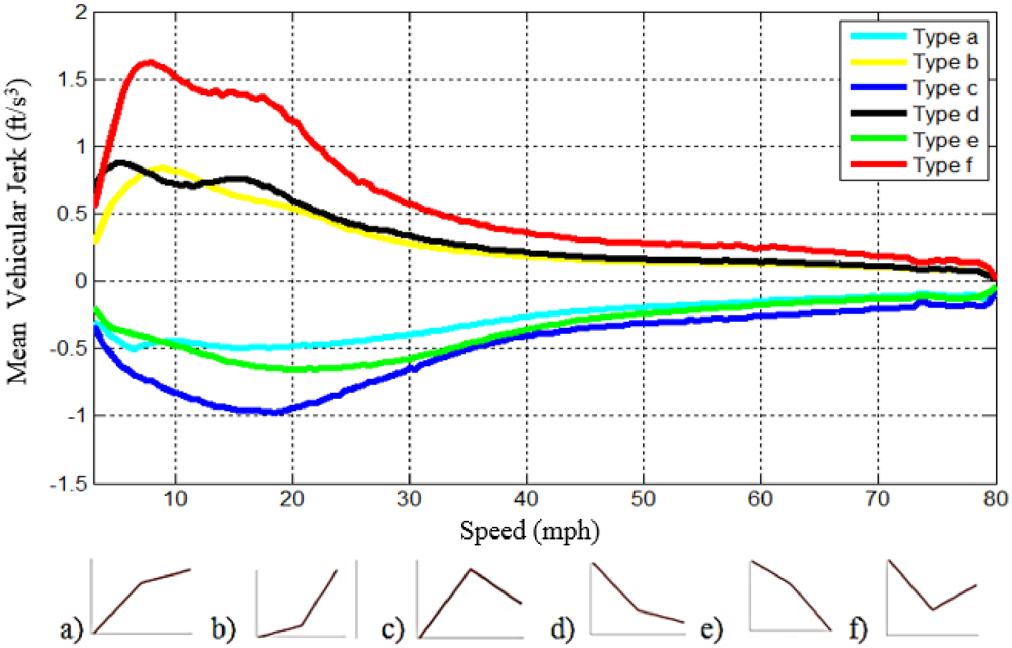
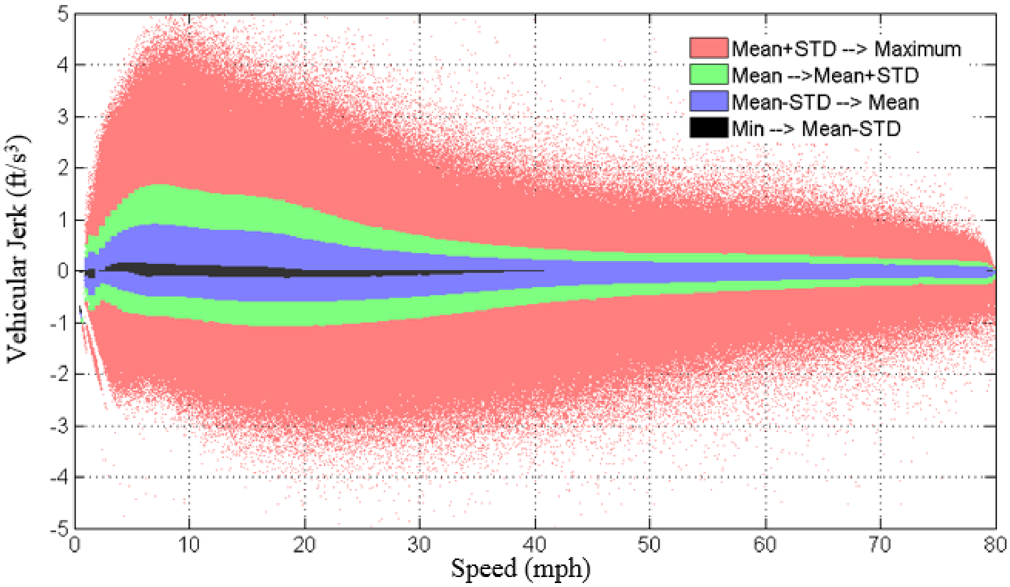


Figure 9: Time spent on different vehicular jerk styles (N=36,715,308).



(a) Typical vehicular jerk distribution by different type



(b) Vehicular jerk distribution and standard deviation (STD)

Figure 10: Vehicular jerk distribution by speed bins (N=36,715,308).

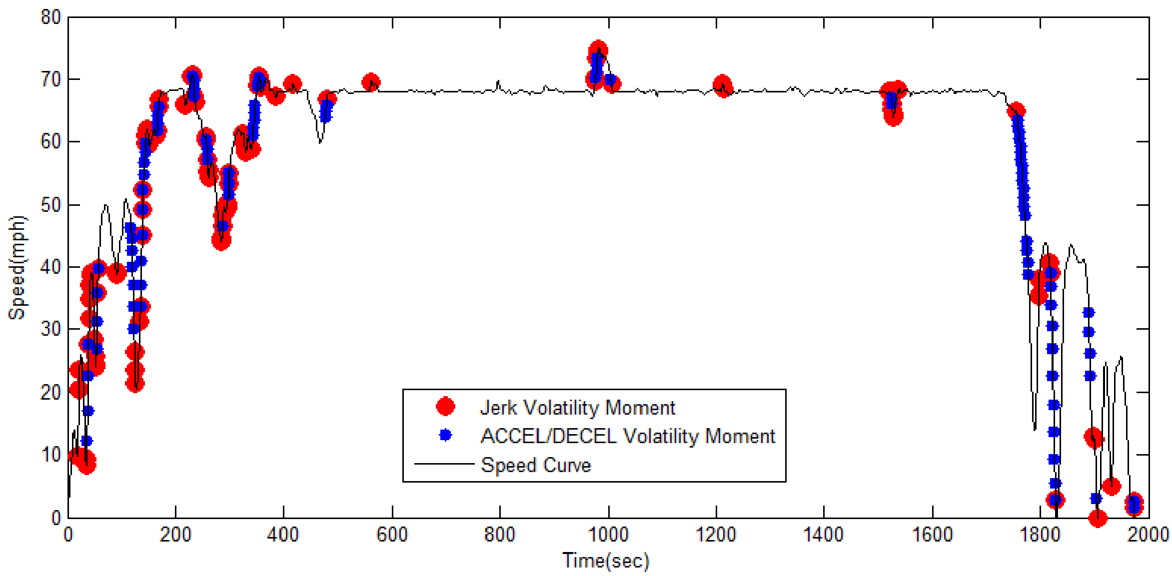


Figure 11: Volatile driving identified by different methods.

Table 1: United States certification drive cycles compared with Atlanta drive cycle. Source (Berry 2010)

Drive Cycle	***Description***	***Data Collection Method***	***Year of Data***	***Top Speed***	***Avg. Speed***	***Max. Acc.***	***Distance***	***Time (min)***	***Idling time***
FTP	Urban/City	Instrumented Vehicles/Specific route	1969	56 mph	20 mph	1.48 m/s^2	17 miles	31 min	18%
C-FTP	city, cold ambient temp	Instrumented Vehicles/ Specific route	1969	56 mph	32 mph	1.48 m/s^2	18 miles	31min	18%
HWFET	Free-flow traffic on highway	Specific route Chase-car/ naturalistic driving	Early 1970s	60 mph	48 mph	1.43 m/s^2	16 miles	12.5 min	None
US06	Aggressive driving on highway	Instrumented Vehicles/ naturalistic driving	1992	80 mph	48 mph	3.78 m/s^2	13 miles	10min	7%
SC03	AC on, hot ambient temp	Instrumented Vehicles/ naturalistic driving	1992	54 mph	35 mph	2.28 m/s^2	5.8 miles	9.9 min	19%
Atlanta	Urban/City	In-veh. GPS devices, Travel survey	2011	80mph	37mph	5.10 m/s^2	7.1 mile^	12.7min^	7%*

Note:
1. FTP: Federal Test Procedure.
2. HWFET: The Highway Fuel Economy Test.
3. US06: The US06 Supplemental Federal Test Procedure (SFTP) for High Speed and High Acceleration Driving behavior.
4. SC03: A Supplemental Federal Test Procedure (SFTP) with Air Conditioning.
5. C- FTP: Federal Test Procedure under cold ambient temperature.
6. ^ mean values are used for Atlanta.
7. * idling & low speeds (below 5 mph).

Table 2: Descriptive statistics for dependent and independent variables

Variables			***N***	***Frequency***	***Mean/Percent***	***Std. Dev***	***Min***	***Max***
Dependent	Volatility Score		40240	-	13.840	6.701	0.1	55.46
Independent	Driver Variable	Gender [Male]	1486	702	47.24%	0.499	0	1
		Driver age (years)	1486	-	47.183	13.319	15	91
	Vehicle Age	Vehicle age (years)	1486	-	7.908	5.417	0	50
	Vehicle Type	Auto-sedan	1486	652	43.88%	0.496	0	1
		Two-seated	1486	58	3.90%	0.194	0	1
		Van	1486	131	8.82%	0.284	0	1
		RV	1486	3	0.20%	0.045	0	1
		SUV	1486	409	27.52%	0.447	0	1
		Station wagon	1486	31	2.09%	0.143	0	1
		Pickup	1486	202	13.59%	0.343	0	1
	Vehicle Fuel Type	Gasoline	1486	1429	96.16%	0.192	0	1
		Diesel	1486	29	1.95%	0.138	0	1
		Hybrid	1486	19	1.28%	0.112	0	1
		Flex fuel	1486	9	0.61%	0.078	0	1
	Trip Variable	Rush hour [Yes]	40240	18616	46.26%	0.499	0	1
		Weekend [Yes]	40240	9805	24.37%	0.429	0	1
		Trip duration (min)	40240	-	14.165	14.738	2.01	374.45
		Commute trip [Yes]	40240	7843	19.49%	0.396	0	1

Note: * Rush hours are AM (6:00 am-10:00 am) or PM (3:00 pm-7:00 pm).

Table 3: Results of the mixed model using volatility score as the dependent variable

Dependent = Volatility Score		***Full model***		***Final model***	
Independent Variables		β	P-value	β	P-value
Constant		16.6983**	<.0001	17.6644**	<.0001
Driver Variables	Gender [Male]	-0.0018	0.9871	-	-
	Driver age (years)	-0.0573**	<.0001	-0.0574**	<.0001
Vehicle Age Variable	Vehicle age (years)	-0.1079**	<.0001	-0.1036**	<.0001
Vehicle Body Type Variable	Auto-sedan	Base		Base	
	Two-seated	3.8554**	<.0001	3.2830**	<.0001
	Van	-1.2621**	0.0084	-1.8231**	<.0001
	Recreational Vehicle-RV	-2.7353	0.1886	Base	-
	Sports Utility Veh.-SUV	0.3291	0.4249	Base	-
	Station wagon	-0.2914	0.6843	Base	-
	Pickup	-0.8836 *	0.0522	-1.5596**	<.0001
Vehicle Fuel Type Variable	Gasoline	Base		Base	
	Diesel	-0.9484	0.1760	Base	-
	Hybrid	-1.7512**	0.0295	-1.9825**	0.0101
	Flex fuel	1.8594*	0.0742	1.5765*	0.0947
Trip Variables	Rush hours [Yes]	0.2375**	<.0001	0.2376**	<.0001
	Weekend [Yes]	-0.3038**	<.0001	-0.3036**	<.0001
	Trip duration (min)	-0.0356**	<.0001	-0.0356**	<.0001
	Commute trip [Yes]	0.3627**	<.0001	0.3630**	<.0001
R^2		0.4028		0.4028	
R^2 Adjusted		0.4026		0.4027	
Root Mean Square Error-RMSE		5.2672		5.2672	
Mean of Response		13.8397		13.8397	
Observations (or Sum Weights)		40240		40240	
Bayesian Information Criterion-BIC		251937		251900	

Variance Component Estimates

	Var. Comp.	Percent of Total	Var. Comp.	Percent of Total
Variance Between Driver-Vehicle Pairs	14.7136	34.66%	14.8319	34.84%
Remaining Variance	27.7429	65.34%	27.7430	65.16%
Total Variance	42.4564	100.00%	42.5749	100.00%

Note:
1. Rush hours: AM (6:00 am-10:00 am), PM (3:00 pm-7:00 pm);
2. ** = significant at a 95% confidence level;
3. * = significant at a 90% confidence level;
4. For mixed model, the random term is Driver-Vehicle ID (N=1486);
5. REML=Restricted Maximum Likelihood;
6. Statistically significant variables (90% level) are kept in the final model.

PART 2: THE ROLE OF ALTERNATIVE FUEL VEHICLES: USING BEHAVIORAL AND SENSOR DATA TO MODEL HIERARCHIES IN TRAVEL

Jun Liu[1], Asad Khattak[2], Xin Wang[3]

[1]The University of Texas, Austin, [2]The University of Tennessee, Knoxville, [3]Virginia Department of Transportation

Abstract: Greater adoption and use of alternative fuel vehicles (AFVs) can be environmentally beneficial and reduce dependence on gasoline. The use of AFVs vis-à-vis conventional gasoline vehicles is not well understood, especially when it comes to travel choices and short-term driving decisions. Using data that contains a sufficiently large number of early AFV adopters (who have overcome obstacles to adoption), this study explores differences in use of AFVs and conventional gasoline vehicles (and hybrid vehicles). The study analyzes large-scale behavioral data integrated with sensor data from global positioning system devices, representing advances in large-scale data analytics. Specifically, it makes sense of data containing 54,043,889 seconds of speed observations, and 65,652 trips made by 2,908 drivers in 5 regions of California. The study answers important research questions about AFV use patterns (e.g., trip frequency and daily vehicle miles traveled) and driving practices. Driving volatility, as one measure of driving practice, is used as a key metric in this study to capture acceleration, and vehicular jerk decisions that exceed certain thresholds during a trip. The results show that AFVs cannot be viewed as monolithic; there are important differences within AFV use, i.e., between plug-in hybrids, battery electric, or compressed natural gas vehicles. Multi-level models are particularly appropriate for analysis, given that the data are nested, i.e., multiple trips are made by different drivers who reside in various regions. Using such models, the study also found that driving volatility varies significantly between trips, driver groups, and regions in California. Some alternative fuel vehicles are associated with calmer driving compared with conventional vehicles. The implications of the results for safety, informed consumer choices and large-scale data analytics are discussed.

Keywords: hierarchical modeling; alternative fuel vehicle; use pattern; driving volatility; travel survey

1. INTRODUCTION

Automobiles are the dominant mode of personal travel in the United States. While they are associated with economic development, automobiles also have adverse impacts on the environment, generate greenhouse gases, and result in dependence on petroleum. One solution to lowering petroleum dependence and reducing emissions is the wider adoption and use of Alternative Fuel Vehicles (AFVs). They are generally more fuel-efficient and environmentally-friendly compared with conventional fuel vehicles (gasoline and diesel) and fulfill expanding individual travel demands of the future (Ji et al. 2012, Lavrenz and Gkritza

2013). Driving behavior in alternative fuel vehicles is of particular interest, if they are to be purchased and used widely. AFVs include plug-in hybrid electric vehicles (PHEVs), battery electric vehicles (BEVs), and compressed natural gas (CNG). While most hybrid electric vehicles are not necessarily AFVs (i.e., are gasoline-based), they are more fuel efficient making use of a smaller engine coupled with electric battery. The key research questions are:

- Whether alternative fuel vehicles and hybrid vehicles have similar use characteristics (trip frequency, vehicle miles traveled, etc.) as conventional vehicles?
- Whether drivers of alternative fuel vehicles are more or less prone to abrupt maneuvers, e.g., aggressive accelerations or vehicular jerk?

The main motivation for the study comes from the potential to learn important lessons from examining the behaviors of early AFV adopters who typically have to overcome adoption barriers such as higher vehicle acquisition costs, shorter driving ranges, scarcity of refueling stations, and potential safety and reliability issues. The study provides a stronger behavioral basis for future tools that can be developed to potentially increase the adoption, diffusion, and use of AFVs and ultimately a large-scale energy transition to alternative fuels. There is an added sense of urgency to examine the use of AFVs as they are gaining greater acceptance and popularity.

Behavioral data used in this study are hierarchical, i.e., they are nested with multiple trips made by different drivers who reside in various regions. Multi-level models have been used for analysis of such data, but not widely in the travel behavior field. This study uses multi-level modeling in a novel way to study whether driving volatility (a key measure of driving performance) varies significantly between trips, driver groups, and regions in California. Relatively new and unique large-scale behavioral data integrated with sensor data from global positioning system devices are used to estimate models and learn from expanded data that has only recently become available (Lohr 2012, Byon and Liang 2013, Siripirote et al. 2014).

2. LITERATURE REVIEW

Vehicle miles/hours traveled, trip frequency, and travel times/distances are often used as measures of performance in transportation. Increasingly, speed and acceleration data are the becoming available and these measures are increasingly used to characterize the driving behavior. Wang et al. used the average speed, average acceleration and the percentage of time in acceleration mode to capture the driving behavior in Chinese cities (Wang et al. 2008). Hung et al. viewed the driving characteristics in a similar way and pointed out the associated factors, including land use, flow density, road width and road network (Hung et al. 2005). Sciarretta et al. investigated the driving behavior of hybrid electric vehicle by collecting their speeds and accelerations. They pointed out that the driving conditions, driver characteristics and vehicle performance are important for understanding the driving experience of hybrid electric vehicle users (Sciarretta and Guzzella 2007). Johannesson et al.

also used the speed and acceleration to quantify driving behavior of hybrid vehicles (Johannesson et al. 2007). Furthermore, the rates of fuel consumption and emissions were used to characterize the driving behavior of internal combustion engine vehicles (Murphey et al. 2009). Generally, hybrid vehicles have higher fuel economy than conventional vehicles (Musardo et al. 2005, Fontaras et al. 2008) and also there are zero-emission electric vehicles in use (Lam and Louey 2006). In order to be somewhat consistent with previous studies, this study uses measures related to the vehicle movement (speed) to characterize the driving behavior.

To understand driving behavior, researchers have defined driving styles, e.g., aggressive driving or calm driving. Typically, cut-off thresholds are used to demarcate driving behavior. Kim et al. gave 1.47 m/s2 (4.82 ft/s2) and 2.28 m/s2 (7.47 ft/s2) as thresholds for aggressive and extremely aggressive accelerations (Kim and Choi 2013b). While De Vlieger et al. pointed out 0.45-0.65 m/s2 for calm driving, 0.65-0.80 m/s2 (2.13-2.62 ft/s2) for normal driving and 0.85-1.10 m/s2 (2.79-3.61 ft/s2) for aggressive driving (De Vlieger et al. 2000). Thresholds suggested in literature are summarized in Table 1. The somewhat arbitrary cut-off points ignore the heterogeneity of driving behavior under different speeds, which has been found in some of the previous studies by the authors (Liu et al. 2014b, Wang et al. [in press]). The results showed that at lower speeds on local/collector roads large acceleration/deceleration values are frequent but at higher speeds (typically on freeways with a good level of service) drivers often do not (or cannot) accelerate and decelerate abruptly. Notably, alternative fuel vehicles may have different performance outcomes because of their different power systems compared with conventional gasoline vehicles (Hori et al. 1998, Moreno et al. 2006).

This study uses the term driving “volatility” instead of “aggressiveness” to measure abrupt accelerations and decelerations, as mentioned in some of our previous studies (Liu et al. 2014b, Wang et al. [in press]). The difference between “aggressiveness” and “volatility” is similar to the terms “accident” and “crash” (Stewart and Lord 2002). Using the term “volatility” is neutral and describes the driving behavior in a more objective and impersonal way. The method for measuring driving volatility is discussed in the next section.

A variety of statistical models have been used to explore links between driving behavior and associated factors, based on the data structure and research purposes. Analysis of variance (ANOVA), Chi-square test and T-tests are the most commonly used methods comparing various groups (Subhashini and Arumugam 1981, Simons-Morton et al. 2005). Ordinary least square (OLS) models including linear and logistic regressions are frequently applied to find the relationships between outcomes and associated factors (McElroy 1967, Khattak et al. 1995, Khattak and Rocha 2003, Dissanayake and Perera 2011). Some studies have noted the hierarchical nature of behavioral data and applied multi-level models to explain relationships (Treno et al. 2003, Kim et al. 2007, Huang et al. 2008). These studies reported the possible variation of predictor effects across groups but did not clearly report whether there are sizable variances at each level. Although data may be structured hierarchically, predictors may not necessarily vary substantially across groups. Therefore, it is very important to report the extent of variations across groups. In this vein, we examine the variances at each level before modeling and report the explained/unexplained variances at each level when predictors are added.

Some studies have applied hierarchical modeling techniques (also called mixed-effects modeling) to handle unobserved heterogeneity (Chin and Quddus 2003, Anastasopoulos et al. 2012) by adding random effects in addition to fixed effects. Notably, mixed-effects models can be characterized as two-level hierarchical models with all predictors (except one random factor) at one level. The data used in this study are more complex and structured at three-levels with three sets of predictors. This study applies a three-level hierarchical model in a novel way to untangle the complex relationships between driving behavior and predictors at various levels.

3. METHODOLOGY

The early AFV adopters are likely to be different from the mainstream consumers in that they are willing to accept the difficulties of adopting alternative fuel vehicles, and likely value the social benefits of AFVs. The issue of self-selection is recognized as important, given that early AFV adopters may represent individuals with higher incomes who are working and traveling longer distances. This study focuses on exploring the differences in use (given adoption) by AFV and conventional vehicle drivers, and not on exploring if a larger market for AFVs exists based on early adopters. Therefore, the issue of self-selection is recognized, but it is not directly addressed in the study.

AFVs are innovations that have some advantages (but also disadvantages) and are diffusing through the system. This study takes advantage of the wealth of information about AFV and conventional vehicle driving contained in behavioral responses coupled with GPS data. It accounts for the hierarchical nature of the data, untangling complex relationships at various levels. The hierarchical model better accounts for lack of independence in explanatory variables and the fact that some independent variables can be different, depending on the level of hierarchy. The data, use measures, and hierarchical modeling structure are discussed in more detail below.

3.1 Data Acquisition

The data used in this study is driving behavioral data collected in a comprehensive travel survey - California Household Travel Survey (CHTS) conducted by California Department of Transportation California during January 2012 through January 2013(Caltrans 2013). The data are large-scale, covering 58 counties across the State of California representing various land use types and populations. This study partitioned the original data into five subsets, including three metropolitan areas (Los Angeles, San Francisco, and Sacramento), California central valley south (Fresno-Stockton), and other areas (mainly suburban and rural areas) in California.

In the CHTS survey, the driving behavior was recorded second-by-second by in-vehicle GPS and OBD (Global Positioning System and On-Board Diagnostics) devices during each trip. The devices captured travel date, time, latitude and longitude (however this geo-code information was removed from the public release database), speed and other standard GPS/OBD variables. Combined with other survey information, the final released data contains driver social demographic data, trip information, and second-by-second driving records for each trip. Table 2 shows the details of the subsets used in this study. The data are structured in a hierarchy—trips are nested within drivers and drivers are nested in regions.

3.2 Driving Volatility Score

In addition to various conventional travel measures, this study uses a relatively new measure of driving volatility to understand how AFVs and conventional vehicles are being used. The driving volatility score is defined as the percentage of abnormal driving seconds (i.e., large vehicular jerk values) over the duration of one entire trip. The value of vehicular jerk is the derivative of acceleration or the second derivative of speed, and is able to capture the instantaneous change of driving decisions (e.g., from accelerating to decelerating). Large values imply abnormal variability of instantaneous driving decisions. To generate the thresholds for recognizing abnormal driving seconds, 54 million driving records collected in CHTS survey are disaggregated in to 0.5 mph speed increment bins. For example, all driving records with speeds from 29.75 ~ 30.25 mph are gathered in the 30 mph bin to generate the mean and standard deviation of vehicle jerk values at 30 mph. If one driving second around 30 mph has a vehicular jerk value greater than the mean +/- 1 standard deviation of this speed range, this second is labeled “volatile driving second”. Thus volatile driving seconds reflect more abrupt driving behavior compared with the majority of driving behavior in the same speed range. Thus, the driving volatility score is a measure of driving behavior during one trip and can be calculated by following equation:

$$Volatility\ Score\ \% = \frac{Seconds\ of\ \text{"Vehicular Jerk"} > Threshold}{Seconds\ of\ Entire\ Trip} \times 100 \quad \text{(Equation 1)}$$

Where,

Threshold = μ (mean) +/- σ(standard deviation) of vehicular jerk values within a speed range *k*.

$$Vehicular\ Jerk, j = \frac{d\boldsymbol{a}}{dt} = \frac{d^2\boldsymbol{v}}{dt^2} = \frac{d^3\boldsymbol{r}}{dt^3} \qquad \text{(Equation 2)}$$

Vehicular jerk is the first derivative of acceleration (a) with respect to time, the second derivative of speed (v) and the third derivative of distance (r). The calculated score is the dependent variable for all models in this study. More details about the driving volatility score calculation are available in previous papers (Liu et al. 2014b, Wang et al. [in press]).

3.3 Hierarchical Linear Modeling

Automobile driving behavior has been linked with a number of factors. Those factors influence driving behavior from different perspectives and form a hierarchical structure of associated factors. For instance, drivers in the same area face similar road network, terrain, and are potentially influenced by the similar driving cultures (Moeckli and Lee 2007). However, drivers also have their own characteristics, such as gender, age, education, income, employment, etc. Further, while drivers are making different trips, their driving behavior is associated with trip features, such as time of day, trip length, trip purpose, etc. Thus, putting these levels (region, driver, and trip) together regardless of the hierarchical features to understand their associations with driving behavior will miss important relationships.

The data used in this study are hierarchical, as shown in Figure.1. Level 1 is the trip level with 65,652 observations; Level 2 is driver level with 2,908 records, and Level 3 has 5 regions. Three levels are involved with three means and variances explained by associated factors in three levels. Level 1 has variables related to trips, such as trip lengths, trip duration, trip average speed, trip purpose, time of day and day of week. Level 2 has variables associated with driver and the vehicle used, such as driver age and gender, vehicle body type, age and fuel type. Note that, since one vehicle corresponds to one driver, the driver level includes both driver- and vehicle- related variables. Level 3 has indicator variables to indicate the region in California.

Since the trips made by the same driver are not independent from each other, assumption of independent observations required for traditional OLS (Ordinary Least Squares) regression models is violated (Hofmann 1997). Therefore, the inter-driver difference and inter-trip difference cannot be estimated accurately without considering the multilevel nature of data and group differences. One method to statistically account for hierarchical structure of data is to use multi-level or hierarchical linear modeling. Hierarchical linear modeling can accommodate non-independence of observations and the heterogeneity of variance across repeated measures (i.e., the same driver made multiple trips). Using hierarchical linear modeling, both the within and between group associations are simultaneously taken into account. The modeling structures are further discussed along with the modeling outputs in next section.

4. RESULTS

4.1 Descriptive Statistics

Socio-Demographics and Travel Characteristics

Tables 3 and 4 present the statistics structured at each level of the hierarchy. Notably, observations with missing information (e.g., no driver age or gender) were removed. The final dataset contains 50,399 trips made by 2,356 drivers from five California regions. Specifically, there are 22,801 trips made by 1,030 drivers from LA, 10,736 trips made by 500 drivers from SF, 5,106 trips made by 255 drivers from SAC, 5,661 trips made by 245 drivers from CCV and 6,095 trips made by the rest of 326 drivers from other areas. For level-1, the trip level, there are no specific clusters. For level-2, driver level, there are 2,365 groups (or drivers); on average each driver made 21 trips (min = 1, max = 79). For level-3, the regional level, the distribution of observation is show in Table 4.

Descriptive Statistics of Hierarchical Linear Model

Descriptive statistics of key variables are shown in Table 5. The numbers seem reasonable and were error checked. Volatility score is measured for 50,399 trips in the database. The average volatility score is 14.31%. The average trip distance was 9.02 mile (min = 0.07 mile, max = 342.78 mile), corresponding to average trip duration of 14.57 minutes and average speed was 29.13 mph; 46.2% of trips were made during rush hours, 22.9% were made on weekends and 16.4% were commute trips (between home and school/work). Among 2,365 respondents, the mean driver age was 48.9 years, ranging from 16 to 87 years; 48.7% were males. The mean vehicle age in the final dataset was 7 years. Trips were made with vehicles with various body types, fuel uses, transmissions and power systems. 42.6% of vehicles were auto sedans, 77.1 % of vehicles were of gasoline fuel type, 85.7% were automatic and 53% were front-wheel drive. Hybrid electric vehicles (HEV) were 13% of the sample, while AFVs were collectively about 5.5%, i.e., PHEV were 0.8%, CNG 1.0%, and BEV 3.7%.

The covariates at level-3 model are dummy variables for CA regions. Notably, for hierarchical modeling, in addition to the fixed-effects parameters in Table 5, there are random-effect parameters (or group variables) which are based on driver ID for level-1 observation groupings and region ID for level-2 observation groupings.

Comparisons of Alternative Fuel Vehicles with Conventional Vehicles

Several comparisons of AFV and conventional vehicle use are shown in Table 6, along with t-tests with conventional vehicles. While the results show some differences, they suggest that AFVs cannot be viewed as monolithic. There are important differences within AFV use and performance that need to be explored, i.e., there are subtle but important differences within AFVs (e.g., PHEV vs. BEV and PHEV vs. CNG).

The key results are summarized below:

- No statistically significant differences ($p<0.05$) were observed between AFVs and conventional vehicles in terms of total daily trips, except drivers of BEVs made significantly fewer trips ($p<0.01$).
- The daily distances traveled are shorter for some AFVs (BEV and PHEV) and longer for other AFVs (HEV and CNG) compared with gasoline vehicles.
- Drivers spent significantly longer time traveling daily in their HEV or CNG vehicles compared with conventional vehicles.
- While slightly more time was spent on deceleration by some AFVs (BEV and PHEV) compared with gasoline vehicles, clear trends did not emerge in terms of time spent on accelerations or deceleration.
- The differences between AFVs and conventional vehicles were not in the same direction when it comes to vehicular jerk.
- HEVs and BEVs had relatively smaller volatility score compared with gasoline vehicles, but PHEV and CNG showed higher volatility scores.

These comparisons have revealed important behavioral differences. A key measure of driving practices-the driving volatility is selected for further modeling. The next step is to use the hierarchical structure of the data to explore associations of AFVs and volatility, while controlling for other factors.

4.2 Multi-Level Modeling

Figure 2 shows the distributions of driving volatility score at each level. At the region level, Los Angeles has a relatively larger mean volatility score than other regions. The driver-level distribution is the distribution of mean volatility scores of 2,356 drivers, since on average one driver made 21 trips in the CHTS survey (as shown in Table 2). At the trip-level, the distribution is right-skewed, as shown in the figure at the top right. While other transformation were tested, a square-root transformation shifted the shape closer to normal, as shown in the bottom right figure. The dependent variable for the hierarchical model was the square-root of volatility score.

Variance-Component Model

Before considering all correlates, which can have both fixed and random effects, this study examined the variances of responses (i.e., square-root driving volatility score) at each level by applying a simple Variance-Component Model, i.e., a constant only model. The model structure used for this is as follows:

$$Y_{ijk} = \beta_{0jk} + e_{ijk}$$

$$\beta_{0jk} = \pi_{00k} + r_{0jk}$$
$$\pi_{00k} = \gamma_{000} + \varphi_{00k}$$

Or,

$$Y_{ijk} = \gamma_{000} + \varphi_{00k} + r_{0jk} + e_{ijk} \quad \text{(Equation 3)}$$

Where,

Y_{ijk} = driving volatility score for *trip (i)* made by *driver (j)* in *region (k)*;
γ_{000} = grand mean (of transformed) driving volatility score of 50,399 trips;
φ_{00k} = standard deviation at level-3 (regional level);
r_{0jk} = standard deviation at level-2 (driver level);
e_{ijk} = standard deviation at level-1 (trip level).

Note that the output of hierarchical modeling generally has three components: 1) fixed-effects parameters, 2) variance estimation of random-effects parameters, and 3) summary statistics. Since the second component is the major focus before the final hierarchical modeling step, the coefficients of fixed-effects parameters are not presented until the last modeling step owing to space limitations.

The outputs of Variance-Component Models are shown in Table 7. Results show that 3.526 is the estimate of grand mean (of the square-root) of driving volatility for 50,399 trips. Averaging across drivers and regions, the expected volatile driving time accounts for 3.526*3.526 =12.43% of the trip duration. The estimates of variance components reveal that there are 0.047 (4.6%), 0.452 (44.5%) and 0.517 (50.9%) variances at regional, driver and trip levels respectively. The standard deviations at each level are 0.216, 0.672 and 0.719 respectively. Clearly, driver level has a sizable variance (44.5%) component, so the use of hierarchal modeling is valuable.

Random Intercept Model

Covariates can be added to explain these variances. Level-related predictors can explain the corresponding variances estimated by variance-component model, as shown in Table 7. At this step, the predictors at higher level explain the variance of the intercept in the lower level model. In other words, only the intercepts are random and coefficients of predictors are fixed. The Random Intercept Model's formulation is as follows:

$$Y_{ijk} = \beta_{0jk} + \beta_{ijk}(Level\ 1\ predictors) + e_{ijk}$$
$$\beta_{0jk} = \pi_{00k} + \pi_{01k}(Level\ 2\ predictors) + r_{0jk}$$
$$\pi_{00k} = \gamma_{000} + \gamma_{001}(Level\ 3\ predictors) + \varphi_{00k}$$

Or,

$$Y_{ijk} = \gamma_{000} + \gamma_{001}(Level\ 3\ predictors) + \varphi_{00k} + \pi_{01k}(Level\ 2\ predictors) + r_{0jk} + \beta_{ijk}(Level\ 1\ predictors) + e_{ijk} \quad \text{(Equation 4)}$$

Where,

Y_{ijk} = driving volatility score for *trip (i)* made by *driver (j)* in *region (k)*;
γ_{000} = grand mean (of transformed) driving volatility score of 50,399 trips;
γ_{001} = coefficients for level-3 predictors (i.e., dummy variable);
π_{01k}= coefficients for level-2 predictors (i.e., driver and vehicle characteristic);
β_{ijk}= coefficients for level-1 predictors (i.e., trip-related factors);
φ_{00k} = root of unexplained variance at level-3 (regional level);
r_{0jk} = root of unexplained variance at level-2 (driver level);
e_{ijk} = root of unexplained variance at level-1 (trip level).

Table 8 presents the unexplained variances at the three levels from the Random Intercept Model. Since the focus is on unexplained variances at three levels, only the random-effects part is presented.

The results in Table 8 show that the variances at three levels became smaller from those reported in Table 7, with predictors explaining some of the variation. Notably, the variance at level-3 (regional level) was explained nearly 100% by the level-3 predictors (nearly zero variance remains). Thus, level-1 and level-2 predictors have constant effects across regions and there is no need to add predictors to explain variances at level-3.

Random Intercept and Slope Model

There is a sizable unexplained variance (0.388, 43.3%) at level-2. Two ways to reduce unexplained variance are: 1) by adding level-2 predictors (driver- and vehicle-related factors); 2) by adding random effects for level-1 (trip level) predictors. For this study, additional level-2 predictors are not available. Random effects of level 1 predictors can be revealed through hierarchical modeling. In addition to the intercepts at level-1 the slopes at level-1 also become the dependent variable at level-2. In this case, the effects of level-1 predictors have two components: fixed effects that explain level-1 variance and random effects that explain level-2 variance. The Random Intercept and Slope Model's formulation is as follows:

$$Y_{ijk} = \beta_{0jk} + \beta_{ijk}(Level\ 1\ predictors) + e_{ijk}$$
$$\beta_{0jk} = \pi_{00k} + \pi_{01k}(Level\ 2\ predictors) + r_{0jk}$$
$$\beta_{ijk} = \pi_{j0k} + \ \pi_{j1k}\ (Level\ 2\ predictors) + r_{ijk}$$
$$\pi_{00k} = \gamma_{000} + \gamma_{001}(Level\ 3\ predictors) + \varphi_{00k}$$
$$\pi_{j0k} = \gamma_{j00} + \gamma_{j01}(Level\ 3\ predictors) + \varphi_{i0k}$$

Or,

$$Y_{ijk} = \gamma_{000} + \gamma_{001}(Level\ 3\ predictors) + \varphi_{00k} + \pi_{01k}(Level\ 2\ predictors) + r_{0jk} + (\gamma_{j00} + \gamma_{j01}(Level\ 3\ predictors) + \varphi_{i0k} + \pi_{j1k}\ (Level\ 2\ predictors) + r_{ijk})\ (Level\ 1\ predictors) + e_{ijk} \quad \text{(Equation 5)}$$

From Table 8, we know that there is nearly zero unexplained variance left at level-3, the level-2 predictors, i.e., driver and vehicle characteristics, have only fixed effects across regions. Only level-1 predictors, i.e., trip attributes, have both fixed effects and random effects that need to be tested further. Thus, Equation (5) can be simplified to:

$$Y_{ijk} = \gamma_{000} + \gamma_{001}(Level\ 3\ predictors) + \varphi_{00k} + \pi_{01k}(Level\ 2\ predictors) + r_{0jk} + (\beta_{ijk} + \tau_{ijk})\ (Level\ 1\ predictors) + e_{ijk} \quad \text{(Equation 6)}$$

Where,

Y_{ijk} = driving volatility score for *trip (i)* made by *driver (j)* in *region (k)*;
γ_{000} = grand mean (of transformed) driving volatility score of 50,399 trips;
γ_{001} = coefficients for level-3 predictors (i.e., dummy variable);
π_{01k}= coefficients for level-2 predictors (i.e., driver and vehicle characteristic);
β_{ijk}= coefficients for level-1 predictors (i.e., trip-related factors);
τ_{ijk}= root of variance of level-1 predictor coefficients across drivers;
φ_{00k} = root of unexplained variance at level-3 (regional level);
r_{0jk} = root of unexplained variance at level-2 (driver level);
e_{ijk} = root of unexplained variance at level-1 (trip level).

Table 9 presents the results of Random Intercept and Slope Model, including estimates of fixed- and random- effects parameters with a reasonable goodness of fit. The results show noticeable variances of level-1 slopes i.e., the variances for weekend vs. weekday travel and commute vs. non-commute trip are relatively large. Finally, the percentage of explained variance is 13.3 % (1-0.448/0.517) at Level-1, 14.8% (1-0.385/0.452) at Level-2, and close to 100% at Level-3.

4.3 Discussion of Key Predictors

In the final hierarchical linear model reported in Table 9, Level-1 predictors about trip characteristics have significant associations (95% confidence level) with the driving volatility but the associations vary across drivers, i.e., same trip level factors may have different estimated coefficients in different groups of drivers. Level-2 predictors, including driver demographics and vehicle features show significant associations with driving volatility except driver gender and vehicle transmission. Slopes of level-2 predictors do not vary substantially across the CA regions. Among level-2 predictors, the fuel types vehicles consume are of particular interest of this study, especially the driving volatility of alternative fuel vehicles. The examination of driving volatility between regions at Level-3 shows significant differences between regions. Note that, interactions among explanatory variables were tested, such as fuel type and gender, fuel type and age, etc., but none were found to be statistically significant (95% confidence level).

Vehicle Fuel Type

The volatility of several alternative fuel vehicles (PHEV, BEV, and CNG), and hybrid vehicles was explored in comparison with gasoline vehicles, which served as the "base." Results show that hybrid electric vehicles are associated with lower volatility scores, by 0.166 units (square-root of driving volatility score). The marginal effects show 8% lower the volatility score magnitude. Note that, one unit lower/higher in volatility score refers to one percent decrease/increase in time spent on volatile/abnormal driving. This result is consistent with an EPA report pointing out that hybrid vehicle drivers tend to be more calm and are able to get better fuel economy (U 2006). The lower volatility score in this study corresponds to less variability of instantaneous driving behavior meaning calmer or smoother driving. In addition to driver attributes and preferences, special vehicle power systems may be part of the reason for the observed lower volatility, i.e., in eco-driving mode, the same acceleration pedal depression for hybrid vehicles generates smoother torque and traction (Toyota 2014). Further, special driving instructions for hybrid vehicles are often provided to drivers. For example, Toyota suggests that when encountering a delay (intersection signal or congested traffic) drivers should release the brake pedal to allow the vehicle to move forward slightly while avoiding overuse of the acceleration pedal (Toyota 2014).

Among AFVs, battery electric vehicles are statistically significantly (95% confidence level) associated with lower volatility scores by 0.315 units (15% lower in terms of volatility score magnitude). While AFV drivers may be less aggressive compared with the same group of conventional vehicle drivers, it is also possible that the engine power of such vehicles may be lower in some instances. Specifically, the engine power of electric vehicles (including plug-in vehicles) depends on the battery level. Depending on the charge in batteries, they cannot always provide full power to the engine required by drivers to do hard accelerations (Chan 2007). Overall, there are clear differences between driver performance (volatility) of conventional and alternative fuel vehicles as revealed by analysis of large-scale behavioral data. While controlling for other factors the results from real-life data show that hybrid vehicles and BEV are associated with calmer driving patterns.

Vehicle Body Type

Vehicle type shows relatively large associations with the driving behavior in this study. Compared with sedans, two-seated vehicles are associated with a 0.191 unit higher (square-root) volatility score average. Convertibles are also linked to an increased score, by 0.287 unit. All other types of vehicles are associated with lower levels of volatility score. Surprisingly, SUVs and pickups are associated with lower scores. The mass of vehicle may have impact on driving behavior. Compared with sedans, two-seated vehicles and convertibles, pickups and SUVs have greater weights and may not be maneuvered as easily as sedans.

Other Vehicle-Related Factors

Older vehicles are also associated with a decreased (square-root) volatility score of 0.019 unit. Vehicles using different transmissions do not show significant differences in terms of volatility. Four-wheel drive vehicles are related to lower driving volatility by 0.119 unit, compared with front-wheel drive vehicles.

Driver Demographics

Older drivers are seen to be less volatile than younger drivers. One year increase in driver age is associated with a 0.008 unit decrease in (square-root) driving volatility score. There is no significant difference between male and female drivers, in terms of driving volatility.

Trip Factors

The negative coefficient of trip distance implies a reverse relationship with driving volatility. Drivers are expected to be less volatile during longer trips and every 1 mile increase in trip distance is associated average 0.008 unit lower (square-root) volatility score with a variance less than 0.001 (standard deviation also less than 0.001). Compared with trips made during non-rush hours, trips made during rush hours are with an increased (square-root) driving volatility score by 0.042 with a variance of 0.039 (or standard deviation 0.198). Commute trips are expected to be with more volatile driving time, average by 0.093 with a variance of 0.107 (or standard deviation 0.327), compared with non-commute trips. Owing to lack of data availability, this study was unable to directly model the association of traffic congestion on driving volatility. However, commute trips and rush-hour trips are often made under congested driving conditions compared with non-commute or non-rush hour trips. This study captures congested driving through proxies of commute and rush hour trips, which are positively associated with higher driving volatility, as expected. Weekend trips are associated with a lower score by 0.077 with a variance of 0.11(or standard deviation 0.332). In short, only trip distance has a clearly negative association with the driving behavior across drivers and the associations of other predictors vary substantially between drivers (i.e., coefficients can be positive or negative across drivers).

Regional Comparisons

At level-3, the results showed that trips made in LA have a 0.529 unit higher (square-root) volatility score than the base (other regions of the CA). The SF and SAC regions have close expected volatility scores. Trips made in the Central Valley areas seem to be more volatile than the base areas, but are less volatile than the three metropolitan areas (i.e., LA, SF and SAC). Overall the magnitude of differences shows particularly volatile driving in LA.

5. LIMITATIONS

The data quality needs to be considered carefully. The response variable, driving volatility score, depends heavily on the second-by-second speed records. The records were generated from in-vehicle GPS and OBD devices and then processed by a professional survey research firm. Thus, the extent of measurement errors in the data is unknown.

Owing to the privacy issues related to driver information, the data sharing system does not release critical information that might help explain some of the variances between trips or drivers. The information includes geo-codes, roadway types used, traffic conditions when traveling, and surrounding land uses, etc. Self-selection in surveys is also a limitation of this study. This is a sample-based study, so reporting and coverage errors may be present.

6. CONCLUSIONS

This study contributes by exploring the use of alternative fuel vehicles by early adopters and comparing their use patterns with conventional vehicles. Firstly, the study uses a large-scale integrated behavioral and sensor database to explore use patterns, especially the short-term decisions made by drivers. Such databases have only recently become available, and also require substantial computing capability. Secondly, the challenge of simultaneously extracting valuable information from complex hierarchically structured data is achieved by the application of hierarchical modeling. Specifically, such modeling better controls for various associated factors, while exploring differences in driver behavior at three levels, i.e., trip level, driver/vehicle level and regional level.

The study answers important research questions about AFV use patterns and driving practices as they gain greater acceptance and popularity. In terms of use, AFV drivers make the same amount of trips as conventional vehicle drivers do, except that drivers of BEV make statistically significantly fewer trips (5% level). The daily distances traveled were shorter for some AFVs (BEV and PHEV) but longer for other AFVs (HEV and CNG) compared with conventional vehicles. Drivers spent significantly longer time traveling daily in their HEV or CNG vehicles compared with conventional vehicles.

The study also found important differences within AFV driving practices. HEV and BEV were found to be associated with calmer driving compared with conventional vehicles, i.e., they are less prone to aggressive accelerations and vehicular jerks. This result is consistent with an EPA study showing that hybrid vehicle drivers tend to be less aggressive. While there is statistical evidence that some AFVs are driven with lower volatility, conclusive evidence that all alternative fuel vehicles are associated with lower driving volatility compared with conventional vehicles was not found.

The implications of this research include:

- Potential benefits from improved safety. By studying driving volatility of individuals who use different vehicle types, implications for safer driving can be anticipated.

Aggressive driving has been linked (statistically significantly) to higher injury severity, given a crash (Paleti et al. 2010). With AFVs driven less aggressively (especially HEV and BEV), safety benefits are expected to accrue. This information can be helpful to public agencies and also to the insurance industry that may offer different rates for AFVs.

- More informed vehicle use decisions. Findings from this study can help potential AFV users make more informed vehicle ownership and use choices. Based on the differences in behaviors highlighted in this study, the AFV industry can make customized marketing plans for promoting the use of AFVs in specific regions. Furthermore, potential buyers can see for themselves how the vehicles purchased are being used by early adopters. For example, this study found that users of BEV made fewer trips. Such information can be provided to potential buyers of BEVs.
- Improvements in accuracy of travel demand models. The study analyzes vehicle miles traveled, and daily trip frequency, etc. for various vehicle types. This has implications for travel demand models and their accuracy. If more AFVs are expected to diffuse through the system in the future then the forecasts can be adjusted accordingly. Specifically, trip generation can be adjusted based on AFV versus non-AFV vehicle ownership. Also, automobile ownership models are used to anticipate demand by regional planning agencies, international organizations (such as the World Bank), and the private sector (automobile manufacturers and oil companies). They are useful in forecasting tax revenues, energy use, and emissions. The results from this study can suggest that automobile ownership models should consider various AFV options available to consumers. The results also inform alternative fuel vehicle policies, given their usage, especially in communities that have (or are considering) favorable local and regional policies toward AFVs.
- Advancing large-scale data analytics. With an explosion in real-world large-scale behavioral and sensor/global positioning system data, this study comprehensively compares the performance of AFVs with conventional vehicles and suggests a timely methodology for analysis of such data.

Finally, more research is needed to further explore differences in AFV purchase and use patterns and how information about such decisions might be used to inform consumers' future adoption decisions.

7. ACKNOWLEDGEMENTS

The National Renewable Energy Laboratory provided the data used in this study. The software R was used for data processing and STATA was for hierarchical modeling. The research was supported through the TranLIVE University Transportation Center grant sponsored by the Research and Innovative Transportation Administration (now in the Office of the Secretary of Transportation), of US DOT. Special thanks are extended to the following entities for their support at University of Tennessee: Transportation Engineering & Science

Program and Initiative for Sustainable Mobility. The views expressed in this paper are those of the authors, who are responsible for the facts and accuracy of information presented herein.

8. REFERENCES

1. Anastasopoulos, P., Mannering, F., Shankar, V. & Haddock, J. 2012. A study of factors affecting highway accident rates using the random-parameters tobit model. *Accident Analysis & Prevention,* 45**,** 628-633.
2. Byon, Y.-J. & Liang, S. 2013. Real-Time Transportation Mode Detection using Smartphones and Artificial Neural Networks: Performance Comparisons between Smartphones and Conventional Global Positioning System sensors. *Journal of Intelligent Transportation Systems,* 18**,** 264-272.
3. California Department of Transportation. 2013. *California Household Travel Survey* [Online]. Available: http://www.dot.ca.gov/hq/tsip/otfa/tab/chts_travelsurvey.html [Accessed May 2nd 2014].
4. Chan, C. 2007. The state of the art of electric, hybrid, and fuel cell vehicles. *Proceedings of the IEEE,* 95**,** 704-718.
5. Chin, H. & Quddus, M. 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis & Prevention,* 35**,** 253-259.
6. De Vlieger, I., De Keukeleere, D. & Kretzschmar, J. 2000. Environmental effects of driving behaviour and congestion related to passenger cars. *Atmospheric Environment,* 34**,** 4649-4655.
7. Dissanayake, S. & Perera, L. 2011. A survey based study of factors related to older driver highway safety. *Journal of Transportation Safety & Security,* 3, 77-94.
8. Ericsson, E. 2001. Independent driving pattern factors and their influence on fuel-use and exhaust emission factors. *Transportation Research Part D: Transport and Environment,* 6**,** 325-345.
9. Fontaras, G., Pistikopoulos, P. & Samaras, Z. 2008. Experimental evaluation of hybrid vehicle fuel economy and pollutant emissions over real-world simulation driving cycles. *Atmospheric environment,* 42**,** 4023-4035.
10. Hofmann, D. 1997. An overview of the logic and rationale of hierarchical linear models. *Journal of management,* 23**,** 723-744.
11. Hori, Y., Toyoda, Y. & Tsuruoka, Y. 1998. Traction control of electric vehicle: Basic experimental results using the test EV "UOT Electric March". *Industry Applications, IEEE Transactions on,* 34**,** 1131-1138.
12. Huang, H., Chin, H. & Haque, M. 2008. Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis. *Accident Analysis & Prevention,* 40**,** 45-54.
13. Hung, W.-T., Tam, K.-M., Lee, C.-P., Chan, L.-Y. & Cheung, C.-S. 2005. Comparison of driving characteristics in cities of Pearl River Delta, China. *Atmospheric Environment,* 39**,** 615-625.

14. Ji, S., Cherry, C. R., J. Bechle, M., Wu, Y. & Marshall, J. D. 2012. Electric vehicles in China: emissions and health impacts. *Environmental science & technology,* 46**,** 2018-2024.
15. Johannesson, L., Asbogard, M. & Egardt, B. 2007. Assessing the potential of predictive control for hybrid vehicle powertrains using stochastic dynamic programming. *Intelligent Transportation Systems, IEEE Transactions on*, 8**,** 71-83.
16. Khattak, A. & Rocha, M. 2003. Are SUVs" Supremely Unsafe Vehicles"?: Analysis of Rollovers and Injuries with Sport Utility Vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, 1840**,** 167-177.
17. Khattak, A., Schofer, J. & Wang, M.-h. 1995. A simple time sequential procedure for predicting freeway incident duration. *Journal of Intelligent Transportation Systems,* 2**,** 113-138.
18. Kim, D.-G., Lee, Y., Washington, S. & Choi, K. 2007. Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. *Accident Analysis & Prevention,* 39**,** 125-134.
19. Kim, E. & Choi, E. Estimates of Critical Values of Aggressive Acceleration from a Viewpoint of Fuel Consumption and Emissions. 2013 Transportation Research Board Annual Meeting, 2013 Washington DC.
20. Lam, L. & Louey, R. 2006. Development of ultra-battery for hybrid-electric vehicle applications. *Journal of power sources,* 158**,** 1140-1148.
21. Langari, R. & Won, J.-S. 2005. Intelligent energy management agent for a parallel hybrid vehicle-part I: system architecture and design of the driving situation identification process. *Vehicular Technology, IEEE Transactions on,* 54**,** 925-934.
22. Lavrenz, S. & Gkritza, K. 2013. Environmental and Energy Impacts of Automated Electric Highway Systems. *Journal of Intelligent Transportation Systems,* 17**,** 221-232.
23. Liu, J., Wang, X. & Khattak, A. 2014. Generating Real-Time Driving Volatility Information. *2014 World Congress on Intelligent Transport Systems.* Detroit, MI.
24. Lohr, S. 2012. The age of big data. *New York Times*, Feburary 11, 2012.
25. McElroy, F. 1967. A necessary and sufficient condition that ordinary least-squares estimators be best linear unbiased. *Journal of the American Statistical Association*, 62**,** 1302-1304.
26. Moeckli, J. & Lee, J. 2007. The making of driving cultures. *Improving Traffic Safety Culture in the United States,* 38**,** 185-192.
27. Moreno, J., Ortúzar, M. & Dixon, L. 2006. Energy-management system for a hybrid electric vehicle, using ultracapacitors and neural networks. *Industrial Electronics, IEEE Transactions on,* 53**,** 614-623.
28. Murphey, Y., Milton, R. & Kiliaris, L. Driver's style classification using jerk analysis. Computational Intelligence in Vehicles and Vehicular Systems, 2009. CIVVS'09. IEEE Workshop on, 2009. IEEE, 23-28.

29. Musardo, C., Rizzoni, G., Guezennec, Y. & Staccia, B. 2005. A-ECMS: An adaptive algorithm for hybrid electric vehicle energy management. *European Journal of Control,* 11**,** 509-524.
30. Paleti, R., Eluru, N. & Bhat, C. 2010. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. *Accident Analysis & Prevention,* 42**,** 1839-1854.
31. Sciarretta, A. & Guzzella, L. 2007. Control of hybrid electric vehicles. *Control systems, IEEE,* 27**,** 60-70.
32. Simons-Morton, B., Lerner, N. & Singer, J. 2005. The observed effects of teenage passengers on the risky driving behavior of teenage drivers. *Accident Analysis & Prevention,* 37**,** 973-982.
33. Siripirote, T., Sumalee, A., Watling, D. & Shao, H. 2014. Updating of travel behavior model parameters and estimation of vehicle trip chain based on plate scanning. *Journal of Intelligent Transportation Systems,* 18**,** 393-409.
34. Stewart, A. & Lord, J. 2002. Motor vehicle crash versus accident: a change in terminology is necessary. *Journal of traumatic stress,* 15**,** 333-335.
35. Subhashini, M. & Arumugam, M. 1981. Analysis of variance (Anova). *CMFRI Special Publication***,** 169-170.
36. Toyota. 2014. Hybrid system features [Online], Toyota Motor Sales, U.S.A., Inc. Available: http://www.toyota.com/t3Portal/document/om/OM47704U/pdf/sec_01-01.pdf [Accessed May 26, 2014].
37. Treno, A., Grube, J. & Martin, S. 2003. Alcohol availability as a predictor of youth drinking and driving: a hierarchical analysis of survey and archival data. *Alcoholism: Clinical and Experimental Research,* 27**,** 835-840.
38. U.S. Environment Protection Agency. 2006. *Fuel Economy Labeling Of Motor Vehicles Revisions to Improve Calculation of Fuel Economy Estimates* [Online]. Available: http://www.epa.gov/carlabel/documents/420r06017.pdf.
39. Wang, Q., Huo, H., He, K., Yao, Z. & Zhang, Q. 2008. Characterization of vehicle driving patterns and development of driving cycles in Chinese cities. *Transportation research part D: transport and environment*, 13**,** 289-297.
40. Wang, X., Khattak, A., Liu, J., Masghati-Amoli, G. & Son, S. [in press]. What is the Level of Volatility in Instantaneous Driving Decisions? *Transportation Research Part C: Emerging Technologies*.

LIST OF TABLES AND FIGURES

Figure 1: Hierarchical data structure used to understand driving behavior......................... 113
Figure 2: Distributions of volatility scores at trip, driver, and regional levels.................... 114

Table 1: Performance thresholds for defining aggressive or calm driving.......................... 115
Table 2: Sample characteristics 116
Table 3: Distributions of observations at each hierarchy 117
Table 4: Distributions of observations at level-3.. 118
Table 5: Descriptive statistics for behavioral data.. 119
Table 6: Comparisons between conventional gasoline vehicles and AFVs 121
Table 7: Outputs of variance-component model 122
Table 8: Outputs of random intercept model.. 123
Table 9: Outputs of random intercept and slope model.. 124

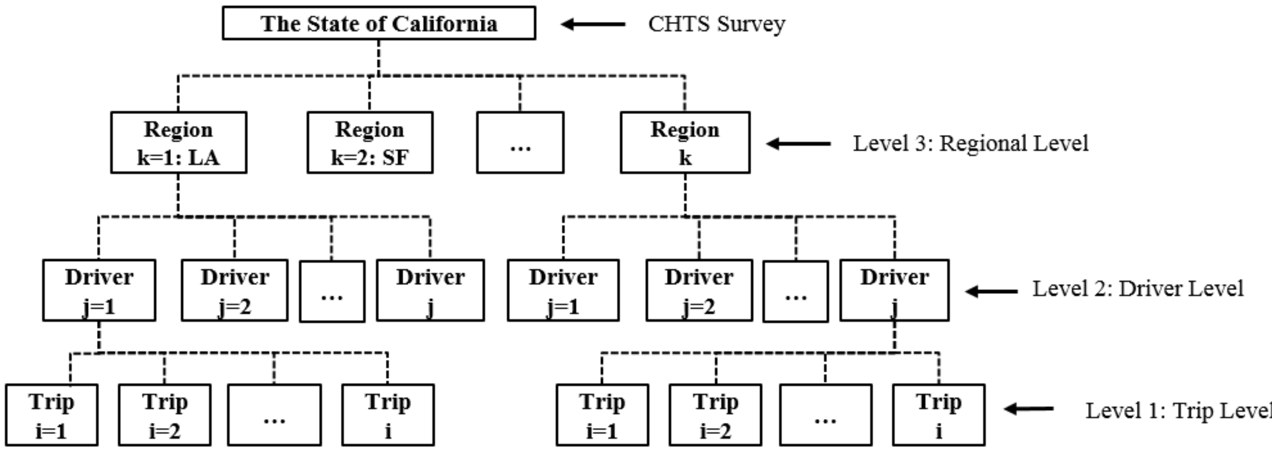
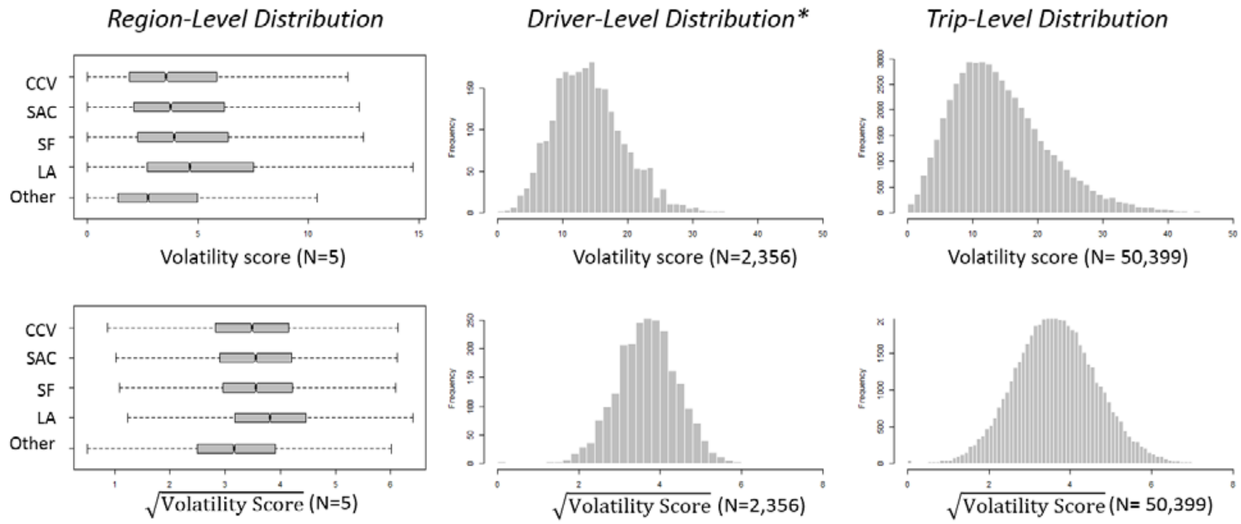


Figure 12: Hierarchical data structure used to understand driving behavior.



* Driver-level distribution is the distribution of mean volatility scores of 2,356 drivers

Figure 13: Distributions of volatility scores at trip, driver, and regional levels.

Table 11: Performance thresholds for defining aggressive or calm driving

Authors	***Study***	***Measures***	***Thresholds***
Kim et al. (Kim and Choi 2013b)	Estimates of critical values of aggressive acceleration from a viewpoint of fuel consumption and emission	Acceleration	1.47 m/s^2 (4.82, ft/s^2) → aggressive driving 2.28 m/s^2 (7.47 ft/s^2) → extreme aggressive driving
Ericsson (Ericsson 2001)	Independent driving pattern factors and their influence on fuel-use and exhaust emission factors	Acceleration	1.5 m/s^2 (4.92 ft/s^2) → aggressive driving
De Vlieger et al. (De Vlieger et al. 2000)	Environmental effects of driving behaviour and congestion related to passenger cars	Acceleration	0.45-0.65 m/s^2 (1.48 -2.13 ft/s^2) → calm driving in city 0.65-0.80 m/s^2 (2.13-2.62 ft/s^2) →normal driving in city 0.85-1.10 m/s^2 (2.79-3.61 ft/s^2) →aggressive driving in city
Langari et al. (Langari and Won 2005)	Intelligent energy management agent for a parallel hybrid vehicle-part I: system architecture and design of the driving situation identification process	Ratio of standard deviation (σ) to average (μ) of acceleration	If $\sigma/\mu>1$ (or 100%)→ aggressive driving
Murphey et al. (Murphey et al. 2009)	Driver's style classification using jerk analysis	Ratio of standard deviation (σ) to average (μ) of jerk	If $\sigma/\mu>0.5$ (or 50%)→normal driving If $\sigma/\mu>1$ (or 100%)→aggressive driving
		Fuel consumption	22.35 miles per gallon → calm driving 20.48 miles per gallon→normal driving 14.93 miles per gallon→aggressive driving

Table 12: Sample characteristics

Region	***Abbreviation***	***Drivers/Vehicles***	***Trips***	***Driving Records (Seconds)***
Los Angeles Metropolitan Area	LA	1,258	29,373	24,185,380
San Francisco Metropolitan Area	SF	636	14,417	12,579,345
Sacramento Metropolitan Area	SAC	315	6,468	5,229,874
California Central Valley (south)	CCV	289	6,878	5,204,840
Other California regions	Other	410	8,516	6,844,450
Total		2,908	65,652	54,043,889

Table 13: Distributions of observations at each hierarchy

Level	No. of Groups	Trips per Group		
		Minimum	Average	Maximum
Level 1	50,399 trips	1	-	1
Level 2	2,356 respondents	1	21.4	79
Level 3	5 regions	5,106	10,080	22,801

Table 14: Distributions of observations at level-3

Region	Trips in Region	Percentage	Drivers in Region	Percentage
LA	22,801	45.24%	1,030	43.72%
SF	10,736	21.30%	500	21.22%
SAC	5,106	10.13%	255	10.82%
CCV	5,661	11.23%	245	10.40%
Other CA	6,095	12.09%	326	13.84%
Total	50,399	100.00%	2,356	100.00%

Table 15: Descriptive statistics for behavioral data

Covariates			N	Mean	Std. Dev.	Min	Max
Dependent	Volatility Score		50,399	14.305	7.534	0.000	67.969
	$\sqrt{\text{Volatility Score}}$		50,399	3.648	1.000	0.000	8.244
Level-1 Predictors	Trip Distance (Mile)		50,399	9.015	15.259	0.077	342.477
	Trip Duration (Minute)		50,399	14.570	17.097	2.000	363.100
	Trip Average Speed (MPH)		50,399	29.129	12.718	2.213	71.255
	*Rush Hour [Yes=1, No=0]		50,399	0.462	0.499	0	1
	Weekend [Yes=1, No=0]		50,399	0.229	0.420	0	1
	Commute Trip [Yes=1, No=0]		50,399	0.164	0.370	0	1
Level-2 Predictors	Gender [Male=1, Female=0]		2,356	0.487	0.500	0	1
	Driver Age (years)		2,356	48.907	13.387	16	87
	Vehicle Age (years)		2,356	7.048	4.722	0	52
	Body Type	Auto-Sedan	2,356	0.426	0.495	0	1
		Two Seated	2,356	0.059	0.235	0	1
		Van	2,356	0.057	0.232	0	1
		Hatchback	2,356	0.081	0.272	0	1
		SUV	2,356	0.193	0.395	0	1
		Station Wagon	2,356	0.040	0.196	0	1
		Pickup	2,356	0.131	0.337	0	1
		Convertible	2,356	0.014	0.118	0	1
	Fuel Type	Hybrid Elec. Vehicles	2,356	0.130	0.337	0	1
		Gasoline Vehicles	2,356	0.771	0.420	0	1
		Diesel Vehicles	2,356	0.037	0.190	0	1
		Plug In Hybrid Elec. Veh.	2,356	0.008	0.089	0	1
		CNG (C. Natural Gas)	2,356	0.010	0.098	0	1
		BEV (Electric) Vehicles	2,356	0.037	0.188	0	1
		Unknown Vehicle type	2,356	0.007	0.082	0	1
	Trans-mission	Automatic	2,356	0.857	0.350	0	1
		Manual	2,356	0.103	0.304	0	1
		Both	2,356	0.035	0.184	0	1
		Unknown	2,356	0.005	0.071	0	1
	Power Train	Front-Wheel	2,356	0.530	0.499	0	1
		Rear-Wheel	2,356	0.174	0.379	0	1
		Four-Wheel	2,356	0.190	0.392	0	1
		Unknown	2,356	0.107	0.309	0	1
Level-3 Predictors #	Region Indicator	LA	5	-	-	0	1
		SF	5	-	-	0	1
		SAC	5	-	-	0	1
		CCV	5	-	-	0	1
		Other	5	-	-	0	1

Note:
*: Rush hours are AM (6:30 am-10:00 am) or PM (3:30 pm-7:00 pm);
#: Level-3 predictors are regional indicators that are indicator variables (0 or 1). They provide information about the region of the driver/vehicles at level-2.

Table 16: Comparisons between conventional gasoline vehicles and AFVs (plus Hybrid)

	Conventional Vehicles		Alternative Fuel Vehicles			
Travel Attributes	**Gasoline Vehicle**	**Diesel Vehicle**	**Hybrid EV**	**PHEV**	**Battery EV**	**CNG (Natural Gas) Vehicle**
Number of vehicles in dataset	1817	88	307	19	86	23
% of large passenger vehicles ^	43.60%	50%	16%	0	2.30%	4.30%
Total trips in dataset	38563	2011	6904	391	1708	484
Mean daily trips	4.23	4.46*	4.2	4.03	3.72***	4.4
Total VMT (mile)	334418	23257.3	70669.57	2984.99	14477.14	5004.41
Mean daily VMT (mile)	35.47	49.14***	43.54***	29.25*	30.73**	47.98***
Total duration (hour)	9054.31	579.64	1876.37	88.6	421.47	129.81
Mean daily duration (hour)	0.96	1.23***	1.14***	0.87	0.9	1.25***
% of short trips (< 3 miles)	41.63%	35.20%	35.81%	37.60%	34.37%	35.96%
% of long trips (> 25 miles)	7.43%	11.79%	9.57%	3.58%	6.85%	9.71%
Mean % time on idling per trip	9.50%	8.99%***	8.77%***	9.06%	8.60%***	8.85%**
Mean % time on extended stable driving per trip [#]	5.16%	7.03%***	5.61%***	4.82%	6.92%***	6.19%***
Mean % time on acceleration per trip	44.89%	44.02%***	44.81%	44.88%	43.46%***	45.08%
Mean % time on deceleration per trip	39.59%	39.21%***	39.81%***	40.50%***	40.22%***	39.13%**
Mean speed (mph)	28.59	30.06***	29.51***	28.85	28.94	31.19***
Maximum speed (mph)	80.16	80.06	80.07	80.01	80.05	80.05
Mean acceleration (ft/s^2)	1.47	1.36***	1.42***	1.49	1.44**	1.48
Maximum acceleration (ft/s^2)	14.37	13.57	12.11	11.95	12.42	9.54
Mean deceleration (ft/s^2)	-1.60	-1.46***	-1.54***	-1.60	-1.49***	-1.65*
Maximum deceleration (ft/s^2)	-16.00	-13.26	-13.76	-12.97	-12.18	-12.48
Mean positive vehicular Jerk (ft/s^3)	0.54	0.48***	0.52***	0.55*	0.52***	0.54
Maximum positive vehicular Jerk (ft/s^3)	13.09	13.57	12.11	9.38	12.42	9.54
Mean negative vehicular Jerk (ft/s^3)	-0.41	-0.38***	-0.40***	-0.43**	-0.40***	-0.41
Maximum negative vehicular Jerk (ft/s^3)	-5.84	-6.51	-5.17	-4.94***	-5.42	-4.49
Mean driving volatility score (%)	14.46	12.92***	13.86***	15.49***	13.69***	15.43***
Maximum driving volatility score (%)	67.97	57.7	51.64	46.33	56.34	48.48

Notes:
1. ^: Large passenger vehicles are VAN, SUV and Pickups, compared with auto-sedan, convertible, hatchback, etc.;
2. [#]: Extended stable driving was defined by speed is above 30 mph and acceleration is less than 0.088 (ft/s2). Acceleration threshold was calibrated using test driving data;
3. Variable in *Italics* show results of t-tests, for comparisons between vehicle group vs. conventional vehicles;
4. *** = t-test significant at a 99% confidence level; ** = t-test significant at a 95%confidence level; * = t-test significant at a 90% confidence level. The base for comparative t-tests are conventional Gasoline Vehicles (GV).

Table 17: Outputs of variance-component model

Effect Type	Terms	Coef.	Std. Err.	P-value	95% Conf. Interval	
					Lower	Upper
Fixed-Effects	Constant, γ	3.526	0.098	0.000	3.334	3.718
Random-Effects	Region ID: Identity					
	Variance (Constant) , φ^2	0.047	0.030		0.013	0.166
	Driver ID: Identity					
	Variance (Constant), r^2	0.452	0.014		0.425	0.481
	Variance (Residual), e^2	0.517	0.003		0.511	0.524
SUMMARY STATISTICS	Number of Observations			50399		
	Number of Groups (Driver ID)			2356		
	Number of Groups (Region ID)			5		
	Log Likelihood			-58188.6		
	Wald χ^2			-		
	Prob > χ^2			-		

Table 18: Outputs of random intercept model

Effect Type	Terms	Coef.	Std. Err.	P-value	95% Conf. Interval	
					Lower	Upper
Fixed-Effects	Constant, γ_{000}	3.947	0.073	0.000	3.804	4.090
	Level-1 Predictors, β_{ijk}	-	-	-	-	-
	Level-2 Predictors, π_{01k}	-	-	-	-	-
	Level-3 Predictors, γ_{001}	-	-	-	-	-
Random-Effects	Region ID: Identity					
	Variance (Constant) , φ^2	1.25E-12	2.47E-11		1.62E-29	9.57E+04
	Driver ID: Identity					
	Variance (Constant), r^2	0.388	0.012		0.365	0.414
	Variance (Residual), e^2	0.509	0.003		0.502	0.515
SUMMARY STATISTICS	Number of Observations	50399				
	Number of Groups (Driver ID)	2356				
	Number of Groups (Region ID)	5				
	Log Likelihood	-57627.017				
	Wald χ^2	1409.78				
	Prob > χ^2	0.000				

Table 19: Outputs of random intercept and slope model

$Y = \sqrt{Driving\ Volatility\ Score}$			Coef.	Std. Err.	P-value	95% Conf. Interval Lower	Upper
Fixed-effects Parameters							
Constant, γ_{000}			3.958***	0.075	0.000	3.812	4.104
Level-1 Predictors β_{ijk}	Trip Distance (Miles)		-0.008***	0.001	0.000	-0.009	-0.007
	Rush Hour [Yes=1, No=0]		0.042***	0.008	0.000	0.026	0.057
	Weekend [Yes=1, No=0]		-0.077***	0.012	0.000	-0.100	-0.054
	Commute Trip [Yes=1, No=0]		0.093***	0.015	0.000	0.063	0.123
Level-2 Predictors π_{01k}	Gender [Male=1, Female=0]		0.006	0.029	0.836	-0.051	0.063
	Driver Age (years)		-0.008***	0.001	0.000	-0.010	-0.006
	Vehicle Age (years)		-0.019***	0.003	0.000	-0.026	-0.013
	Body Type	Auto-Sedan	Base				
		Two Seated	0.191***	0.063	0.002	0.068	0.315
		Van	-0.309***	0.062	0.000	-0.431	-0.187
		Hatchback	-0.103*	0.056	0.068	-0.213	0.007
		SUV	-0.114***	0.043	0.008	-0.198	-0.030
		Station Wagon	-0.024	0.075	0.750	-0.171	0.123
		Pickup	-0.407***	0.052	0.000	-0.508	-0.306
		Convertible	0.287**	0.120	0.017	0.052	0.523
	Fuel Type	Hybrid Elec. Vehicles	-0.166***	0.046	0.000	-0.256	-0.076
		Gasoline Vehicles	Base				
		Diesel Vehicles	-0.119	0.077	0.122	-0.269	0.032
		Plug In Hybrid Elec. V.	-0.107	0.159	0.499	-0.418	0.204
		CNG (C. Natural Gas)	-0.136	0.142	0.336	-0.414	0.141
		BEV (Electric) Vehicles	-0.315***	0.084	0.000	-0.479	-0.151
		Unknown Vehicle type	0.046	0.170	0.788	-0.287	0.379
	Trans-mission	Automatic	Base				
		Manual	-0.070	0.048	0.145	-0.165	0.024
		Both	0.010	0.078	0.896	-0.143	0.163
		Unknown	0.125	0.202	0.535	-0.270	0.520
	Power Train	Front-Wheel	Base				
		Rear-Wheel	0.071*	0.043	0.097	-0.013	0.155
		Four-Wheel	-0.119***	0.044	0.007	-0.205	-0.032
		Unknown	0.002	0.047	0.966	-0.089	0.093
Level-3 Predictors γ_{001}	Region Indicator	LA	0.529***	0.044	0.000	0.442	0.617
		SF	0.358***	0.049	0.000	0.261	0.454
		SAC	0.368***	0.057	0.000	0.257	0.480
		CCV	0.227***	0.057	0.000	0.115	0.339
		Other	Base				
Random-effects Parameters							
Region ID: Identity							
Variance (Constant), φ^2			1.90E-19	4.55E-18		6.80E-40	52.823
Driver ID: Identity							
Variance (Distance), τ_1^2			0.000	0.000		0.000	0.000
Variance (Rush Hour), τ_2^2			0.039	0.004		0.032	0.047
Variance (Weekend), τ_3^2			0.110	0.008		0.096	0.126
Variance (Commute Trip), τ_4^2			0.107	0.010		0.089	0.128
Variance (Constant), r^2			0.385	0.013		0.360	0.411
Variance (Residual), e^2			0.448	0.003		0.442	0.454
SUMMARY STATISTICS							
Number of Observations			50399				
Number of Groups (Driver ID)			2356				
Number of Groups (Region ID)			5				
Log Likelihood			-56580.394				
Wald χ^2			919.89				
Prob > χ^2			0.000				

*** = significant at a 99% confidence level; ** = significant at a 95%confidence level; * = significant at a 90% confidence level.

PART 3: ARE HOV/ECO-LANES A SUSTAINABLE OPTION TO REDUCING EMISSIONS IN A MEDIUM-SIZED EUROPEAN CITY?

Tânia Fontes[1], Paulo Fernandes[1], Hugo Rodrigues[1], Jorge Bandeira[1], Sérgio Ramos Pereira[1], Asad J. Khattak[2], Margarida C. Coelho[1]

[1]University of Aveiro, [2]The University of Tennessee, Knoxville

Abstract: Innovative traffic management measures are needed to reduce transportation-related emissions on arterials and freeways. While in Europe, road lanes management have focused mainly on bus lanes introduction, the conversion of exclusive lanes to High Occupancy Vehicles (HOV) and eco-lanes (lanes dedicated to vehicles running on alternative fuels) has not been studied comprehensively. The objectives of this research are: 1) To develop an integrated microscale modeling platform calibrated with real world data to assess the impact of future Traffic Management Strategies (TMS) in an urban area; 2) To evaluate the introduction of eco-lanes in three different types of roads in European Medium-sized cities and its effects in terms of emissions and traffic performance. The methodology consists of three distinct phases: a) Data collection, b) Traffic and emissions modeling using an integrated platform of microsimulation, and c) Scenarios evaluation. For the baseline scenario, the statistical indicators of the integrated platform show valid results, namely it has been found no significant differences between simulated and VSP modal distributions. Moreover, the methodology applied shows that HOV and eco-lanes in the medium European urban city is feasible. The results shows that on freeways the majority of passengers can reduce their travel time about 5% with a positive impact in terms of total emissions (-3% CO_2,-14% CO, -8% NO_X). On urban arterials, emissions reduction can be achieved only if the average occupancy vehicle increases from 1.37 (current) to 1.50. The implications for eco-lanes are discussed.

Keywords: integrated microscale modeling, eco-lanes, HOV, VSP, emissions

APPLICATIONS AND CASE STUDIES

PART 1: A COMPARATIVE STUDY OF DRIVING PERFORMANCE IN METROPOLITAN REGIONS USING LARGE-SCALE VEHICLE TRAJECTORY DATA

Jun Liu[1], Asad Khattak[1], Xin Wang[3]

[1]The University of Texas, Austin, [1]The University of Tennessee, Knoxville, [1]Viginia Department of Transportation

Abstract: Volatile driving, characterized by hard accelerations and braking, can contribute substantially to higher energy consumption, tailpipe emissions, and crash risks. Drivers' decisions to maintain speed, accelerate or brake rapidly, or jerk their vehicle are largely constrained by their unique regional/metropolitan contexts. These contexts may be characterized by their geography, roadway structure, traffic management, and driving population, etc. This study captures how people generally drive in a region using large-scale vehicle trajectory data, implying how energy are consumed and how emissions are produced in regional transportation systems. Specifically, driving performance in four U.S. metropolitan areas (Los Angeles, San Francisco, Sacramento, and Atlanta) is compared, taking advantage of large-scale behavioral data (78.7 million seconds of speed records) collected by in-vehicle Global Positioning Systems as part of regional surveys. Comparative analysis shows significant regional differences in terms of volatile driving and time spent to accelerate, brake, and vehicular jerk of daily trips. Correlates of higher volatility are also explored, e.g., battery electronic vehicles show low volatility, as expected. This study proposes a novel way to compare regional driving performance by successfully turning Global Positioning System driving data into valuable knowledge that can be applied in practice by developing regional driving performance indices. The new indices can also be used to compare regional performance over time and to imply the levels of sustainability of regional transportation systems.

Keywords: Large-scale data, Driving volatility, Sustainability, Metropolitan region, Mixed-effects model

1. INTRODUCTION

Personal travel involves making decisions about when, where, and how to travel on a daily basis. Traveler behaviors have been extensively discussed from the perspective of transportation sustainability (Kim et al. 2013, Sioui et al. 2013, Khan et al. 2014, Wang et al. 2015a). If automobile is chosen for travel, drivers make micro-level instantaneous decisions about speeds and accelerations that are highly associated with energy consumption, emissions and safety (Brian Park et al. 2009). Automobiles traveling at higher speeds and accelerations generally consume more fuels as compared with those traveling at lower speeds and accelerations (Holm et al. 2007, LeBlanc et al. 2010, Lárusdóttir and Ulfarsson 2015). With new GPS technologies, key aspects of instantaneous driving decisions can be captured in a new performance indicator, known as driving volatility (Wang et al. 2015c). Fundamentally, this is a trip-based index of how a person drives. Driving volatility takes into account of both instantaneous speed, acceleration and vehicular jerk, and can also be used to imply how fuels are consumed and how emissions are produced in a trip. Driving decisions can also be aggregated to the regional level in order to study day-to-day variations in a region and also for comparison across regions. Regional driving performance is of particular interest because it captures how people drive in an area from the perspectives of both regional transportation safety and sustainability. This study compares driving performance across diverse spatial contexts (metro areas) in the United States to explore how the index and its correlates vary and delivers insights in regional transportation sustainability.

The study takes advantage of large-scale data with millions of second-by-second vehicle trajectories, comparing the regional driving performance of four metropolitan areas that include Los Angeles, San Francisco, Sacramento, and Atlanta. Global Positioning System data from regional travel behavior surveys conducted in these regions are used to explore volatility in instantaneous driving decisions. How drivers use their time in accelerating, decelerating or maintaining speed during urban trips is quantified and compared. Then the study investigates correlates of driving volatility and if they vary systematically across regions. The information generated by this study can be used to encourage eco-driving policies, such as calm driving (i.e., driving with low volatility) in a region. The regional driving index compares driving performance across regions, implying different levels of transportation sustainability from the perspective of automobile driving decisions.

2. LITERATURE REVIEW

How people drive in a region can be captured in volatility of their driving. Vehicle speeds and accelerations can be used to characterize driving volatility on trips and aggregated to the region/city level. How people drive will likely be associated with regional factors that may include law enforcement (Ramaekers et al. 2000), road network, road type, and traffic conditions (Ericsson 2000, Wang et al. 2008), geographical context including terrain and slopes (Akamatsu et al. 2003), weather (Keay and Simmonds 2005), driver education and licenses (Shinar et al. 2001), and demographics (Hung et al. 2005).

Researchers have collected driving data to generate representative driving cycles for a city or region; such driving cycles can reflect the regional driving performance, and are used to characterize and compare driving performance across cities or regions (Hung et al. 2005, Wang et al. 2008). Note, a driving cycle is often used for vehicle fuel economy and emission estimations; it is a speed trace over time, representing a vehicular trip under various conditions (EPA 2013, 2014). The US Environmental Protection Agency (EPA) has developed a number of driving cycles that are used to represent the typical driving patterns of a region or a city, such as Los Angeles – LA92 driving cycle and the City of New York – NYCC driving cycle (EPA 2013).

Wang et al. collected driving data from11 Chinese cities and characterized the driving performance in these cities by developing driving cycles. They compared the driving cycles of Chinese cities with European and US driving cycles. They found driving cycles in Chinese cities have higher proportions of time spent on acceleration and deceleration than European and US driving cycles (Wang et al. 2008). Hung et al. compared the average speed and acceleration in three cities of Pearl River Delta, China, with US, European and Japanese driving cycles. Three cities in China (Hong Kong, Macao and Zhuhai) have greater mean travel speeds than US New York City Cycle (NYCC), partly capturing lower congestion levels. Compared with NYCC, Hong Kong has more acceleration/deceleration events, indicating more aggressive driving (Hung et al. 2005). Lin et al. compared average speeds of arterials in three US cities (Metropolitan Bay Area, Sacramento, and Stanislaus). San Francisco Bay Area's arterials had an average speed of 45.9 mph, which is 1.4 mph faster than speeds of Sacramento, and driving data from Stanislaus showed an even lower average speed of 37.3 mph (Lin and Niemeier 2003). Ericsson further developed measures to characterize driving performance, including time use in speed or acceleration ranges and discussed the correlation between driving performance and factors, such as road types, and driver demographics (Ericsson 2000). It is found that the driving performance on roads in residential areas had greater accelerations than arterial roads, while speeds on arterials roads were greater than travel speeds on residential roads. Male drivers spent more time on at heavy acceleration (greater than 1 m/s2 or 3.28 ft/s2) than female drivers (Ericsson 2000).

As mentioned above, driving cycles are often used for fuel economy and emission estimations. Fuel consumption and emissions can be regarded as a function of speed and acceleration (EPA 2010, Lárusdóttir and Ulfarsson 2015). Therefore, information about fuel consumption and emissions in a region may also be used to present the regional driving performance (Lents et al. 2004). In addition to the speed, acceleration and their impact on fuel consumption and emissions, the regional crash records have been used to represent the safety level of the regional driving practice (Alsop and Langley 2001, Kmet and Macarthur 2006). The regional crash records are critical to auto insurance pricing (Cummins and Tennyson 1992, Cohen and Dehejia 2003). Besides, time reliability (Bhat and Sardesai 2006, Sumalee et al. 2006) and driving exposure (including trip distance and duration) (Chipman et al. 1992) are also used as critical measures of the regional driving performance.

Table 1 summarizes previous studies on evaluation of driving performance of a region or a city; they collected data and compared the regional driving performance based various performance measurements, e.g., speed, acceleration/deceleration, time use, kinetic energy, and emissions. A limitation of these studies is the generally small sizes of sampled data, as shown in Table 1. Researchers have given thresholds for defining driving styles, as calm or aggressive (De Vlieger et al. 2000, Kim and Choi 2013b). For example, extremely aggressive acceleration threshold was reported to be 7.47 ft/s2 (Kim and Choi 2013b). Driving decisions take into account differences in driving conditions e.g., local roads vs. interstate or flat vs. rolling roads (Wang et al. 2015c). This study contributes by using new performance criteria (introduced in next section) to characterize driving styles and utilizes “Big Data” to quantify regional driving practices.

3. DATA DESCRIPTION

The data used in this study are large-scale GPS driving data collected in travel surveys, including 1) California Household Travel Survey (CHTS) conducted by California Department of Transportation California during January 2012 through January 2013 (Caltrans 2013), and 2) Atlanta Regional Travel Survey (ARTS) conducted by Atlanta Regional Commission during February 2011 through October 2011 (TSDC). The sample from CHTS covers 58 counties across the State of California representing various land use types and populations and the sample from ARTS covers 20 counties in the region of Atlanta Regional Commission. To generate the regional driving performance, this study partitioned the sample from CHTS into three metropolitan areas and non-metropolitan area. Three metropolitan areas, delineated according to 2013 Census Metropolitan Statistical Areas (Census 2013), include Los Angeles Metropolitan Area (Los Angeles-Long Beach-Riverside), San Francisco Metropolitan Area (San Jose-San Francisco-Oakland) and Sacramento Metropolitan Area (Sacramento-Arden-Arcade-Yuba City). In sum, four metropolitan areas (Los Angeles, San Francisco, Sacramento and Atlanta) are targeted to generate their regional driving performance profiles. The final database contains personal data; household data, trip data, and second-by-second driving data for each trip, and covers driving practices in different road types by different type of vehicles. Table 2 shows the basic features of the final datasets used in this study and the basic information across four metropolitan areas (Sorensen 2010, Caltrans 2011, Schrank et al. 2012, Census 2014).

4. METHODOLOGY

4.1 Conceptual Framework

Figure 1 presents a conceptual framework that describes the relationships between driving performance and various associated factors. The driving performance can be characterized at the regional and individual levels. First, how an individual drives in a region may be a function of their demographics, vehicles, trip attributes, routes and other factors that vary across individuals. Then, how people drive in a region can be correlated with regional factor such as roadway network attributes, economy, population and cultural factors. The measures developed for regional driving performance are an indication of how the population uses the road network in a region. The measures for individual driving performance are an indication of how an individual drives within a region, and the individual driving performances is correlated with both individual and regional factors. All these measures are developed based on large-scale trajectory data in specific regions. The correlates of regional and individual driving performance can provide implications for safety, energy and emissions. For example, larger acceleration rates at a regional level indicate potential high fuel consumption rates, and higher deceleration rates indicate that many drivers have to make hard brakes under existing roadway conditions, e.g., if high levels of congestion exists in a region.

Key objectives of this study are to compare the driving performance between regions and explore how correlates of individual driving performance vary across regions, providing implications for regional transportation sustainability from the perspective of automobile driving. Therefore, this study first compares the regional driving performance by a developing a set of measures that are highly related to fuel consumption and emissions, and then models individual driving performance within a region. The study quantifies how people drive in a region, taking advantage of large-scale data. Note that correlates of driving performance can also be explored by estimating multi-level models (Liu et al. 2015e).

4.2 Driving Performance

Measures of driving performance include speed, acceleration/deceleration and jerk (first derivative of acceleration and second derivative of speed). Jerk is the derivative of acceleration or the second derivative of speed. While the acceleration characterizes changes in speed, vehicular jerk characterizes how a driver changes the acceleration/deceleration. Previous papers by the authors have presented the use of vehicular jerk in quantifying micro/instantaneous driving performance (Liu et al. 2014b, Wang et al. 2015c). To profile the microscopic driving performance, the concept of micro driving performance patterns is introduced in this study. Acceleration and deceleration are two micro patterns. Further, the vehicular jerk corresponds to six different micro patterns, as shown in Figure 2.

This study aggregates all quantified micro driving patterns to provide a macro level of regional driving performance. The results of aggregation include distributions of driving patterns and time use of different patterns. This study proposes two measures to create new regional driving performance indices: Acceleration Index and Vehicular Jerk Index. They are “Speed × Time” weighted mean values. These two indices provide straightforward information about the overall intensity and variability of micro-driving performance. This study also provides a measure to index the individual driving performance, i.e., the Driving Volatility Score. This is the percentage of driving seconds with extreme micro driving decisions (e.g., hard braking/accelerating) over the duration of one trip (Wang et al. 2015c). Any extreme driving decisions are identified based on the regional driving performance. Details are shown in later section. Figure 3 presents a flow chart of driving performance measures developed in this study.

4.3 Modeling Framework

To conduct a comparative study, a model was constructed for each region to explore the correlates of individual driving performance within a region. The differences in correlates between regions are an indication of the in-depth differences of regional driving performance. Considering a driver making multiple trips during the travel survey periods, this study applied mixed-effects models to explore the correlates of individual driving performance. The mixed-effects models take account of inter-driver and inter-trip (but intra-driver) variations of the driving performance (StataCorp 2012, Liu et al. 2015a). The model structure is as follows (SAS 2014):

$$Y = \beta X + \gamma Z + \varepsilon \qquad \text{Equation (1)}$$
$$\gamma \sim N(0, G)$$
$$\varepsilon \sim N(0, \sigma^2 I_n)$$

Where,

Y = individual driving performance in a trip;

X = main explanatory factors that are associated with individual driving performance (i.e., driver demographics, vehicle attributes and trip information);

β = estimated coefficients for the matrix X;

Z = an indicator that indicates whether trips were made by the same driver (i.e., the driver ID);

γ = estimated coefficients for the matrix Z;

ε = random errors, which is assumed to be normally distributed;

G = a diagonal matrix with identical entries for each fixed effect;

I_n = identity matrix; and

γ and ε are assumed to be independent.

5. REGIONAL DRIVING PERFORMANCE COMPARISONS

5.1 Time Use Distributions

Figure 4 presents the time spent in the four regions on acceleration, deceleration and constant speed by speed range in 0.5 mph increments, as well as standardized time shares, i.e., time shares for speed bins. Time spent accelerating or braking varies with speeds. The findings regarding regional comparisons are:

- Regions in California show similarity in terms of larger time shares in high speeds ranges (60 ~70mph), which is different from Atlanta. This likely depends on a host of different factors that include road networks and trip lengths. Trips in regions of California are relatively longer (average-14 minutes) in the sampled datasets than those in Atlanta (average- 12 minutes). Longer urban trips are typically associated with a greater possibility of driving on interstates or expressways. Highway and arterial density of a region might be another reason for regional differences. Compared with California regions, Atlanta has the least density of highways and arterials (Sorensen 2010, Caltrans 2011).

- For normal local driving practices (10~50 mph), there is no clear peak time in Los Angeles. Traffic congestion is a likely contributing reason. According to the 2012 Urban Mobility Plan (Schrank et al. 2012), Los Angeles region is ranked highest in terms of congestion index among these four regions.
- Atlanta has a clear peaking of time spent at around 40 mph, and this region has the smallest congestion index (Schrank et al. 2012).
- Driving in San Francisco Bay Area seems shows more time spent in lower speed ranges, partly because of the hilly terrain and strong grades in the region.

Note that, very small acceleration or deceleration rates (0.04 ft/s^2, based on the 5th percentile of speed changes in one second) were considered noise and coded as constant speed. Standardized time shares show that driving time is mostly devoted to accelerating or decelerating rather than maintaining speed. Acceleration and deceleration have about equal time share. Increasing time is spent on maintaining constant speed when speeds are higher (>60 mph). Notably, Sacramento region has a large time share percentage of constant speed at speeds around 76 mph. The large shares of constant speed possibly associate with traffic congestion on freeways, and to some extent the use of cruise control. Unfortunately, the information about the use of cruise control was not recorded or released to the public in the database. Comparisons between regions reveal that Atlanta shows lower times spent on driving at freeway speeds of 70 mph or above. Time use distributions of vehicular jerk patterns are available from the authors. Types (a), (b), (d) and (e) dominate the driving time (over 80%). Differences in vehicular jerk between four regions are not very clear.

5.2 Acceleration and Vehicular Jerk Distributions

The distributions of micro driving patterns reveal the magnitude/intensity of micro driving decisions (accelerating, decelerating, changing acceleration/deceleration). Figure 5 shows the distribution of quantified acceleration/deceleration and vehicular jerk patterns along with the speeds. Owing to the large number of observations (N= 78.7 million records), the four regions have similar distributions. The overlapping distributions are presented but with the means and two-standard-deviations separately indicated for each region in the Figure. The gradual change of color shows the concentration extent of magnitude to zero (i.e. constant speed or zero acceleration/jerk). The mean and two-standard-deviations are for each speed range (0.5 mph increments) is plotted. Major findings are:

- San Francisco Bay Area has smaller magnitudes of accelerations (closer to zero) than other areas. This may be associated with the hilly terrain and strong grades in the region.
- Los Angeles and Sacramento regions have closer means and standard deviations across speed bins. Both regions have grid road patterns with similar terrains and road densities, so similar driving performance in terms of acceleration is expected and observed.
- Atlanta region seems to have larger means and standard deviations.

- In general, large accelerations and decelerations occur at lower speeds (10 ~ 30 mph) and after speeds reach higher ranges (> 40 mph) the magnitude tends to zero, as shown in Figure 5(i).
- Acceleration/deceleration rates are not distributed homogeneously along with the speeds.
- There are no apparent regional differences at speeds less than 15 mph or above 60 mph, in terms of mean and standard deviation shown in Figure 5(ii).

Six types of vehicular jerk patterns show different distributions. Owing to the limited space, only Types © and (f) are presented in this paper, as shown in Figure 4 (iii) and (iv). Types © and (f) may represent the most volatile micro-driving patterns with acceleration chained with deceleration in a very short time frame. Unlike the acceleration distribution, vehicular jerk does not have a symmetric distribution. The positive jerk patterns have larger magnitudes than negative ones. The large magnitudes of jerk are mainly observed at speeds of 5~20 mph. After speeds reach 30 mph, the vehicular jerk magnitudes are relatively constant around zero.

Type (f) has positive values and SF has smaller magnitudes than other three regions at speeds 15~30 mph. The small magnitudes are associated with smooth or careful driving. The hilly terrain in SF may encourage drivers to change their micro-driving decisions (i.e. Type (f), from decelerating to accelerating) more smoothly. LA and SAC have close magnitudes in Type (f) and ATL has the largest magnitudes. There are no significant differences between regions in terms of the Type © driving pattern. SF and SAC have close magnitudes that are smaller than regions LA and ATL. Overall, this study observed and quantified heterogeneity of micro driving patterns at different speeds.

5.3 Combined Distribution of Time Use and Accelerations/Vehicular Jerk

Due to space limitations, only the overall distributions of speeds, accelerations and time use can be presented (N=78.7 million records). Figure 6 shows a wealth of information about vehicle performance; distributions for each region are available from the authors. The height shows the number of driving records with corresponding driving status (i.e., speed and acceleration/deceleration or vehicular jerk).

At speeds 10 ~30 mph there are fewer driving records with zero acceleration or deceleration (see the trough in Figure 5); for higher speeds (> 60 mph), a large portion of time is spent in maintaining speed with small acceleration or deceleration (see the ridge in Figure 5). Differing from acceleration distributions, vehicular jerk distributions are more concentrated at zero. This implies that any quantified jerk patterns that are different from zero can be easily identified as abnormal micro driving patterns, e.g., sudden braking or accelerating.

5.4 Regional Driving Index

To generate a new measure of driving performance at the regional level, similar to Congestion Index in the Urban Mobility Report (Schrank et al. 2012), this study created the following two indices:

$$Regional\ Accleration\ Index = \sum_{i=1}^{n} (\frac{v_i * t_i}{\sum_{i=1}^{n}(v_i * t_i)} * |a_i|) \qquad \text{Equation (2)}$$

$$Regional\ Vehicular\ Jerk\ Index = \sum_{i=1}^{n} (\frac{v_i * t_i}{\sum_{i=1}^{n}(v_i * t_i)} * |j_i|) \qquad \text{Equation (3)}$$

Where,

v_i= speed record of sampled vehicle during i^{th}time slice in selected region, *i*=1, 2, 3, …, *n*, *n* is the total driving records for one region (e.g., for LA: 24,185,380 seconds);

t_i = duration of i^{th}time slice, i.e., one second if using the second-by-second data;

a_i= acceleration during i^{th}time slice;

j_i= vehicular jerk during i^{th}time slice;

$\sum_{i=1}^{n}(v_i * t_i)$ = total distance traveled in the sample in one region.

These two indices represent the intensity and variability of instantaneous driving decisions respectively. They can be used to compare the driving patterns across metropolitan areas. If Time is sliced equally, the formals can be simplified to speed-weighted mean. Using the data, LA has 24,185,380 time slices (n=24,185,380); SF has n=12,579,345; SAC has n=5,229,874 and ATL has n=36,715,308 slices. Each time slice consists of one second and the speed and acceleration for each second is known. The results are shown in table 3.

Hypothetically, the acceleration index can range from 0 up to 15 or even 20 ft/s2, and similarly for vehicular jerk index. Acceleration index captures the intensity of micro-driving decisions and vehicular jerk index captures the variability in micro-driving decisions in a region. To compare regional driving performance, Los Angeles was used as the base for calculating acceleration and vehicular jerk indices. That is, LA was assumed to be at 100% and the percentages for other regions were calculated relative to LA. The indices for other regions can range from 0% to 100% theoretically. The results show that overall LA has the largest values for both indices and ATL is ranked second. SAC has a larger acceleration index than SF but the vehicular jerk index is smaller.

Congestion Index in Table 1 reveals the amount or level of delay in a regional road network and it represents the relationships between network capacity and regional travel demand. Acceleration and jerk indices proposed in this paper provide overall measures of driving performance and giving how people drive in a region. They represent the interactions between road conditions/configuration (e.g., grade, pavement, acceleration/deceleration lanes etc.) and the driver responses, as well as the aggressiveness of drivers. Larger acceleration index values in LA may be associated with relatively older highway infrastructure higher frequency of shorter acceleration/deceleration lanes, the higher population density (relative to the other regions examined in this study) see Table 1. Thus the intense land use might be a reason that not enough space is available for acceleration or deceleration lanes. The larger vehicular jerk index is associated with the transportation system, as vehicular jerk events (changes of acceleration rates) are more likely to occur when the traffic flow is congested. The Vehicular Jerk Volatility Index may be correlated with road congestion levels, as indicated by the literature summarized in Table 1. The interaction of between driving performance indices and such factors, e.g., socio-demographics, economic and geographic contexts need further exploration when more data are available.

Although this study aims to quantify and compare regional driving performance using large-scale data, the indices can also be used in other ways. Specifically, they can be used to monitor how driving performance of a specific region fluctuates over time, e.g., day to day, week to week, or season to season. Therefore, reports of daily, weekly and seasonal regional driving performance can provide useful information to planners and engineers. The fluctuations can also be associated with special event, e.g., higher fluctuations in these indices may be observed for in-town football games, or before and after the implementation of a traffic management strategy such as advanced traveler information systems. Overall, these new indices can be used much in the same way that congestion or safety indices are used.

5.5 Driving Volatility Score

Considering different regions have different driving performance, this study suggests generating regional thresholds based on regional driving performance. Driving performance thresholds in one region cannot be used to judge performance of drivers in another, given the differences in the socio-demographic, economic and geographic contexts. This study uses the μ +/- 2σ as the threshold to create the volatility score, as shown in Figure 3.

The μ +/- 2σ are calculated based on the distribution of the same patterns (i.e., acceleration, declaration and six types of vehicular patterns) in a specific speed ranges. During driving, a person can freely select speeds. As long as the intensity of making such patterns is within the threshold generated based on the driving performance of all drivers in the region at a particular speed, this driver will be deemed as behaving “normally,” while this pattern is recognized as abnormal driving. This study uses the percentage of driving time with abnormal driving patterns over the duration of the entire trip, termed driving volatility score, to index the extent of variability of instantaneous driving performance during one trip. Both methods of acceleration and vehicular jerk can be used to generate the driving volatility score, as an index of individual driving performance. Our previous studies have pointed out vehicular jerk method clearly points out the critical patterns in trips (Liu et al. 2014b, Wang et al. 2015c). Thus, this study still uses vehicular jerk as a key measure of driving volatility.

6. MODELING DRIVING VOLATILITY

Table 4 presents the descriptive statistics. Observations with missing information are removed. The final datasets used for modeling contains 20,608 trips made by 933 drivers in LA, 9,276 trips made by 426 drivers in SF, 4,642 trips made by 235 drivers in SAC, and 40,201 trips made by 1,486 drivers in ATL. The response variable is the driving volatility score. The mean volatility score for LA is 5.15% with a standard deviation – 3.901 %.

Key explanatory variables include demographics, vehicle body type, fuel type and trip characteristics. The demographics of drivers in the sample have a good representation to corresponding regions (based on a comparison of age and gender in Tables 1 and 2). Vehicle body types and fuel uses cover the fleets on roads in recent years. Hybrid and electric vehicles are of particular interest in this study.

Table 5 presents the results of the mixed-effects model. Since one driver could make multiple trips, there are repeated observations for one driver. A grouping factor (Driver ID) was entered in the model to capture the errors caused by repeated observations. Results show that there is a substantial variation (38.44% ~ 42.89%) captured by the grouping factor. Thus, the grouping is necessary in the model.

The modeling results convey two-folded insights in correlates of driving volatility across regions:

1) Similarities across regions. In general, the variances of driving volatility between and within drivers are similar across the four regions. About 40% of variation in driving volatility can be attributed to the driver- and vehicle-related factors and 60% of variation is attributed to trip-level differences, e.g., trip distance, time of day, trip purpose and weekday. Regarding correlations between driving volatility and various factors, all significant estimates have the same signs and most have similar magnitudes, implying that some correlates of driving volatility (e.g., weekend, commute trip and driver age) do not vary substantially across regions. Modeling results show that older drivers are less volatile in driving; a ten year increase in driver age is associated with approximately 0.3 units decline in driving volatility score. Notably, one unit in volatility score refers to one percent of above normal threshold of driving time during one trip. Trips made on weekends seem to be calmer; commutes trips tend to be more volatile; and longer trips have smaller volatility scores. The similarities in correlates of driving volatility across regions imply that the individual driving performance may be homogeneous from certain angles (such as weekend, commute trip and driver age) under different driving contexts.

2) Dissimilarities across regions. Regarding variances of driving volatility between and within drivers, LA and SAC have the larger share of variance between drivers (>42%) than SF and ATL, implying that, drivers in LA and SAC are more likely to be different from each other (in terms of driving volatility), if they are in different driver groups

(characterized by age and gender) and they are likely to drive different types of vehicles.

In addition, some of the volatility correlates have uneven magnitudes across regions, i.e., rush hour, trip distance, vehicle age, vehicle body type and fuel type. Trips made during rush hours seem to be more volatile than non-rush hour trips, while the volatility difference between rush and non-rush hour trips is greater for the Atlanta region than the three California regions. Long trips are expected to be associated with lower driving volatility, since long trips typically have a greater possibility of driving on a freeways or major arterial roads than short trips do. LA and SF appear to have stronger associations between driving volatility and trip length than drivers in ATL do.

The modeling showed that older vehicles are associated with lower driving volatility; SAC seems to have an even greater association than other three regions. Two-seated vehicles are associated with higher driving volatility. ATL has the largest magnitude, and two-seated vehicles have on average 1.861 units higher volatility scores than sedans. One unit in volatility score refers to one percentage of above normal driving time during one trip. LA has the smallest effect, 1.095 units higher than sedans in LA. For two-seated vehicle drivers, road conditions may influence their volatility. The congestion index in 2012 Urban Mobility Report shows that LA has the largest index and ATL has the smallest one among these four regions. Better driving conditions may provide a chance to two-seated vehicle drivers higher chances to brake and accelerate faster. Vans are associated with lower scores and SAC has the largest coefficient, 1.872 units lower than sedans in SAC. Convertibles have 1.504 units higher volatility scores compared with the “base,” but only in SF the coefficient is statistically significant (5% level). SUVs show lower scores across all four regions, with Sacramento residents showing a large negative value of 1.280 units lower volatility than sedans. Pickups also show lower volatility than sedans. The mass of these larger vehicles may have impact on driving performance. Compared with sedans, pickups and SUVs have greater weights and may not have the accelerations or braking capability that are comparable with sedans.

Interestingly, hybrid vehicles show lower volatility scores across all regions, especially for ATL and SAC. For instance, in ATL, hybrid vehicles show substantially lower volatility scores (1.496 units) than vehicles consuming gasoline. A report from EPA mentioned aggressiveness of hybrid vehicles and conventional vehicles (EPA 2006) and they pointed out that hybrid drivers tend to be less aggressive, improving the fuel economy. Using a different set of data and measures, this study confirmed that hybrid vehicle drivers are less volatile than drivers of conventional vehicles. Plug-in hybrid vehicles also seem to be calmer than conventional vehicles, based on the results. Electric vehicles are associated with lower volatility scores, and Sacramento has the largest coefficient, with 3.03 units lower volatility than conventional vehicles. The engine power of electric vehicles (including plug-in vehicles) heavily depends on their charge/battery level. Sometimes, lower level of batteries cannot fully provide the engine with power required by drivers (Chan 2007). As a result, the vehicles tend to be less volatile. This study indicates calmer driving performance by alternative fuel and hybrid vehicles, showing relatively lower volatility scores.

The dissimilarities in correlates of driving volatility across regions may be attributed to several factors that include differences in behavioral processes across regions, inherent heterogeneity, measurement errors, and differences in spatial, network, socioeconomic, regional driving context, and technology factors. While these factors were not investigated directly in this study, a future study can address them.

7. LIMITATIONS

The quantified regional driving performance in this study depends heavily on the second-by-second speed records. These records were generated from in-vehicle GPS+OBD devices and then processed by a professional survey research firm. Thus, the extent of measurement errors (e.g., GPS accuracy and missing data interpolation) in the data is unknown (TSDC). Improvements in data quality measured through multiple devices can better capture regional driving performance in the future.

Driving performance can be influenced by a variety of factors. Factors discussed in this study have shown their significant impacts on driving volatility. However, owing to the privacy issues, some critical information remains unknown, e.g., road types used by the respondents, traffic conditions at the time of travel, surrounding land use, etc. Inclusion of additional factors in future studies can provide a more complete the picture of associations between driving performance and other factors.

8. CONCLUSIONS

This study quantifies volatility in regional driving performance using “Big Data” that contains millions of driving records with detailed travel-related information, such as driver, vehicle and trip features, delivering insights in regional transportation sustainability from the perspective of automobile driving. The study extends the fundamental understanding of instantaneous driving performance in various regions in the United States, relative to our earlier studies (Liu et al. 2014b, Wang et al. 2015c). A comparative analysis of regional driving performance shows significant regional differences in driving time use and of instantaneous driving decisions captured through driving volatility and includes accelerations, decelerations or changing acceleration/deceleration rates for drivers in a regional sample. Policies for promoting sustainable transportation (e.g., calm driving) can be suggested, as calm driving is associated with less fuel consumption and emissions (Holm et al. 2007, Beardsley 2010, LeBlanc et al. 2010, Lárusdóttir and Ulfarsson 2015). Importantly, with the new data available from GPS, it is possible to create new indices of regional driving performance (similar to congestion index).

This study suggests the Regional Acceleration Volatility Index and Regional Vehicular Jerk

Volatility Index as complementary measures to those used in the Texas Transportation Institute studies (Schrank et al. 2012). These new indices are overall measures of driving performance. They give a sense of how people drive in a region, representing interactions between road conditions/configuration (e.g., grade, pavement, acceleration/deceleration lanes) and the driver responses, as well as the volatility of drivers. They also reflect how fuels are consumed and how emissions are produced in a region. These two indices represent the intensity and variability of instantaneous driving decisions in a region respectively. They can be used to compare the driving patterns across metropolitan areas, e.g., when comparing regional driving performance, Los Angeles had the largest volatility index values and ATL is ranked second, implying that Los Angeles may need an eco-driving policy to promote an environmentally friendly automobile driving environment. The indices can also be used to monitor how the driving performance of one region fluctuates over time, e.g., from day to day, week to week, or season to season. Reports of daily, weekly and seasonal regional driving performance can be useful for planners and engineers who are concerned about transportation safety, efficiency and sustainability.

Further, this study explored associations of regional driving performance with driver-, vehicle- and trip-related factors that are nested in a region. The mean + 2 standard deviations are used as the threshold to create a driving volatility score (capturing "unusual" micro-driving patterns). The thresholds vary by region, depending on how people drive in a particular region. The driving volatility score captures changes in driving performance during one trip. To explore correlates of volatility across regions, mixed-effects regression models were estimated separately for four regions. Analysis of correlates showed that older people are less volatile in their driving; trips on weekends seem to be calmer; commutes trips tend to be more volatile; and longer trips have smaller volatility scores. Dissimilarities in correlates across regions were quantified. Modeling results also offer implications for developing policies that promote transportation sustainability. For example, hybrid vehicles were found to be associated with calm driving (i.e., lower volatility) across four metropolitan regions, implying that expanding the penetration rate of hybrid vehicles in fleets may be able to increase the driving safety as well as to reduce the fuel consumption and emissions.

Overall, this study contributes by demonstrating how to quantify regional driving performance in data rich environments. Specifically, driving performance can be quantified by establishing regional thresholds, and generating indices for regional and individual driving performance, as demonstrated in this study. The indices can be used to compare regions or monitor changes in volatility over time in a specific region and to imply the levels of transportation sustainability in regions from the perspective of automobile driving. Future studies may be focused on examining how the driving volatility scores and indices proposed and explored in this study are associated with the fuel consumption and emissions if relevant consumption and emission data (e.g., the dashboard fuel economy displays in recent year models) are available, providing a direct link between the driving performance and regional transportation sustainability.

9. ACKNOWLEDGEMENTS

The National Renewable Energy Laboratory provided the data used in this study. The software MATLAB was used for data processing and visualization. Statistical software STATA was used for modeling. The research was supported through the TranLIVE University Transportation Center (US DOT grant number DTRT12-G-UTC17) and Southeastern Transportation Center (US DOT grant number DTRT13-G-UTC34) both sponsored by the Office of the Secretary of Transportation, US Department of Transportation. Special thanks are extended to the following entities for their support at University of Tennessee: Transportation Engineering & Science Program and Initiative for Sustainable Mobility. The views expressed in this paper are those of the authors, who are responsible for the facts and accuracy of information presented herein.

10. REFERENCES

1. Akamatsu, M., N. Imachou, Y. Sasaki, H. Ushiro-Oka, T. Hamanaka, and M. Onuki. 2003. Simulator study on driver's behavior while driving through a tunnel in a rolling area. In: *Proceedings of the Driving Simulation Conference. Michigan, USA*
2. Alsop, J., and J. Langley. 2001. Under-reporting of motor vehicle traffic crash victims in New Zealand. *Accident Analysis & Prevention* 33, 353-359.
3. Beardsley, M. 2010. MOVES2010: Information for Transportation Modelers. *Transportation Research Board*.
4. Bhat, C. R., and R. Sardesai. 2006. The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B: Methodological* 40, 709-730.
5. Brian Park, B., I. Yun, and K. Ahn. 2009. Stochastic optimization for sustainable traffic signal control. *International Journal of Sustainable Transportation* 3, 263-284.
6. Caltrans. 2011. 2010 California Public Road Data, California Department of Transportation, Sacramento, CA.
7. Caltrans. 2013. California Household Travel Survey. http://www.dot.ca.gov/hq/tsip/otfa/tab/chts_travelsurvey.html Accessed May 2nd 2014.
8. Census. 2013. Metropolitan and Micropolitan Statistical Areas Main. http://www.census.gov/population/metro/ Accessed May 3rd 2014.
9. Census. 2014. State & County QuickFacts. U.S. Census Bureau.
10. Chan, C. 2007. The state of the art of electric, hybrid, and fuel cell vehicles. *Proceedings of the IEEE* 95, 704-718.
11. Chipman, M. L., C. G. MacGregor, A. M. Smiley, and M. Lee-Gosselin. 1992. Time vs. distance as measures of exposure in driving surveys. *Accident Analysis & Prevention* 24, 679-684.
12. Cohen, A., and R. Dehejia. 2003. The Effect of Automobile Insurance and Accident Liability Laws in Traffic Fatalities, National Bureau of Economic Research.
13. Cummins, J. D., and S. Tennyson. 1992. Controlling automobile insurance costs. *The Journal of economic perspectives*, 95-115.

14. De Vlieger, I., D. De Keukeleere, and J. Kretzschmar. 2000. Environmental effects of driving behaviour and congestion related to passenger cars. *Atmospheric Environment* 34, 4649-4655.
15. EPA. 2006. Fuel Economy Labeling Of Motor Vehicles Revisions to Improve Calculation of Fuel Economy Estimates, U.S. Environment Protection Agency.
16. EPA. 2010. Technical Guidance on the Use of MOVES2010 for Emission Inventory Preparation in State Implementation Plans and Transportation Conformity. EPA-420-B-10-023, Transportation and Regional Programs Division , Office of Transportation and Air Quality, U.S. Environmental Protection Agency.
17. EPA. 2013. Dynamometer Drive Schedules. http://www.epa.gov/nvfel/testing/dynamometer.htm Accessed March 3rd 2014.
18. EPA. 2014. Fuel Economy Testing and Labeling. EPA-420-F-14-015, U.S. Environmental Protection Agency, Ann Arbor, MI.
19. Ericsson, E. 2000. Variability in urban driving patterns. *Transportation Research Part D: Transport and Environment* 5, 337-354.
20. Holm, P., D. Tomich, J. Sloboden, and C. Lowrance. 2007. Traffic analysis toolbox volume iv: guidelines for applying corsim microsimulation modeling software.
21. Hung, W.-T., K.-M. Tam, C.-P. Lee, L.-Y. Chan, and C.-S. Cheung. 2005. Comparison of driving characteristics in cities of Pearl River Delta, China. *Atmospheric Environment* 39, 615-625.
22. Keay, K., and I. Simmonds. 2005. The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia. *Accident analysis & prevention* 37, 109-124.
23. Khan, S., H. Maoh, C. Lee, and W. Anderson. 2014. Towards Sustainable Urban Mobility: Investigating Non-Work Travel Behavior in a Sprawled Canadian City. *International Journal of Sustainable Transportation*.
24. Kim, E., and E. Choi. 2013. Estimates of Critical Values of Aggressive Acceleration from a Viewpoint of Fuel Consumption and Emissions. In: *2013 Transportation Research Board Annual Meeting*, Washington DC
25. Kim, J., S. Fujii, and B. Lee. 2013. Strategies to promote sustainable mobility management incorporating heterogeneity. *International journal of sustainable transportation* 7, 107-124.
26. Kmet, L., and C. Macarthur. 2006. Urban–rural differences in motor vehicle crash fatality and hospitalization rates among children and youth. *Accident Analysis & Prevention* 38, 122-127.
27. Lárusdóttir, E. B., and G. F. Ulfarsson. 2015. Effect of Driving Behavior and Vehicle Characteristics on Energy Consumption of Road Vehicles Running on Alternative Energy Sources. *International Journal of Sustainable Transportation* 9, 592-601.
28. LeBlanc, D. J., M. Sivak, and S. Bogard. 2010. Using naturalistic driving data to assess variations in fuel efficiency among individual drivers.

29. Lents, J. M., N. Davis, M. Osses, R. M. Nikkila, and M. J. Barth. 2004. Comparison of on-road vehicle profiles collected in seven cities worldwide. In: *Transport and Air Pollution 13th International Scientific Symposium*
30. Lin, J., and D. Niemeier. 2003. Regional driving characteristics, regional driving cycles. *Transportation Research Part D: Transport and Environment* 8, 361-381.
31. Liu, J., B. Bartnik, S. H. Richards, and A. J. Khattak. 2015a. Driver Behavior at Highway Rail Grade Crossings with Passive Traffic Controls: A Driving Simulator Study. *Journal of Transportation Safety and Security*. doi:http://dx.doi.org/10.1080/19439962.2015.1043478.
32. Liu, J., A. Khattak, and X. Wang. 2015b. The role of alternative fuel vehicles: Using behavioral and sensor data to model hierarchies in travel. *Transportation Research Part C: Emerging Technologies* 55, 379-392. doi:10.1016/j.trc.2015.01.028.
33. Liu, J., X. Wang, and A. Khattak. 2014. Generating Real-Time Driving Volatility Information. In: *2014 World Congress on Intelligent Transport Systems*, Detroit, MI
34. Ramaekers, J. G., H. Robbe, and J. O'Hanlon. 2000. Marijuana, alcohol and actual driving performance. *Human Psychopharmacology Clinical and Experimental* 15, 551-558.
35. SAS. 2014. Mixed and Random Effect Model Reports and Options. http://www.jmp.com/support/help/Mixed_and_Random_Effect_Model_Reports_and_Options.shtml Accessed Feb. 14 2014.
36. Schrank, D., B. Eisele, and T. Lomax. 2012. TTI's 2012 urban mobility report. *Proceedings of the 2012 annual urban mobility report. Texas A&M Transportation Institute, Texas, USA.*
37. Shinar, D., E. Schechtman, and R. Compton. 2001. Self-Reports Of Safe Driving Behaviors in Relationship to Sex, Age, Education and Income in the US Adult Driving Population. *Accident Analysis & Prevention* 33, 111-116.
38. Sioui, L., C. Morency, and M. Trépanier. 2013. How carsharing affects the travel behavior of households: a case study of montréal, Canada. *International Journal of Sustainable Transportation* 7, 52-69.
39. Sorensen, P. 2010. Reducing Traffic Congestion and Improving Travel Options in Los Angeles.
40. StataCorp. 2012. Multilevel mixed-effects linear regression.
41. Sumalee, A., D. P. Watling, and S. Nakayama. 2006. Reliable network design problem: case with uncertain demand and total travel time reliability. *Transportation Research Record: Journal of the Transportation Research Board* 1964, 81-90.
42. TSDC. 2014. Secure Transportation Data Project. Transportation Secure Data Center, National Renewable Energy Laboratory
43. Wang, H., W. H. Lam, X. Zhang, and H. Shao. 2015a. Sustainable transportation network design with stochastic demands and chance constraints. *International Journal of Sustainable Transportation* 9, 126-144.

44. Wang, Q., H. Huo, K. He, Z. Yao, and Q. Zhang. 2008. Characterization of vehicle driving patterns and development of driving cycles in Chinese cities. *Transportation research part D: transport and environment* 13, 289-297.
45. Wang, X., A. J. Khattak, J. Liu, G. Masghati-Amoli, and S. Son. 2015b. What is the level of volatility in instantaneous driving decisions? *Transportation Research Part C: Emerging Technologies* 58, 413-427. doi:10.1016/j.trc.2014.12.014.

LIST OF TABLES AND FIGURES

Figure 1: Conceptual framework: correlates of driving performance 146
Figure 2: Vehicular jerk patterns for speed data.. 147
Figure 3: Measures of driving performance used in this study 148
Figure 4: Time use distribution of acceleration, deceleration and constant speed in study regions 149
Figure 5: Distributions of acceleration and jerk at various speeds..................................... 150
Figure 6: Combined distributions of speed, acceleration (or vehicular jerk), and time use... 151

Table 1: Summary of selected studies on regional driving performance 152
Table 2: Characterization of study regions and sample statistics.. 153
Table 3: Volatility indices for regional driving performance.. 154
Table 4: Descriptive statistics for samples from the four regions 155
Table 5: Mixed-effects regression models for vehicular jerk in the four regions 156

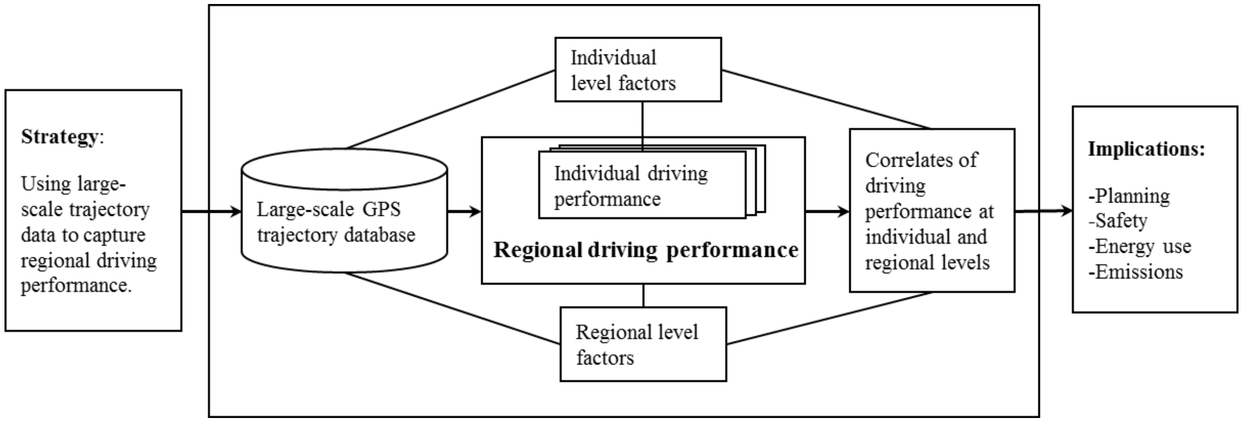
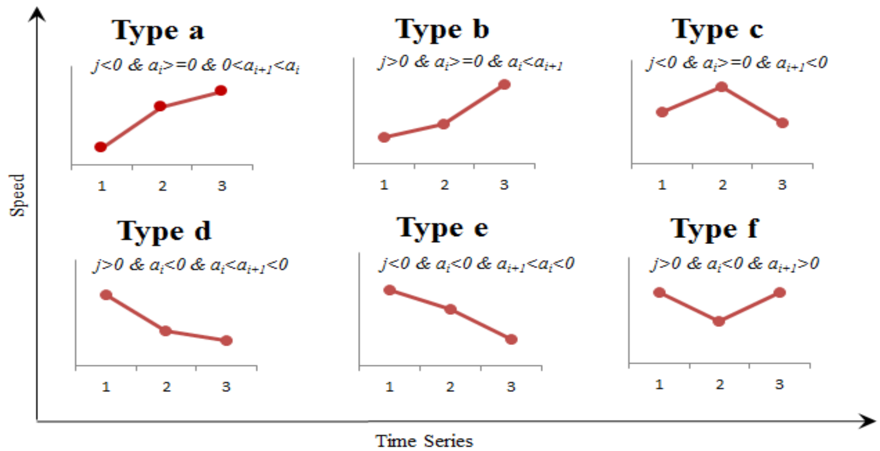


Figure 1: Conceptual framework: correlates of driving performance.



Note: *j=jerk;* a_i*=acceleration at time i;* a_{i+1}*=acceleration at time i+1*

Figure 2: Vehicular jerk patterns for speed data. Source: (Wang et al. 2015c).

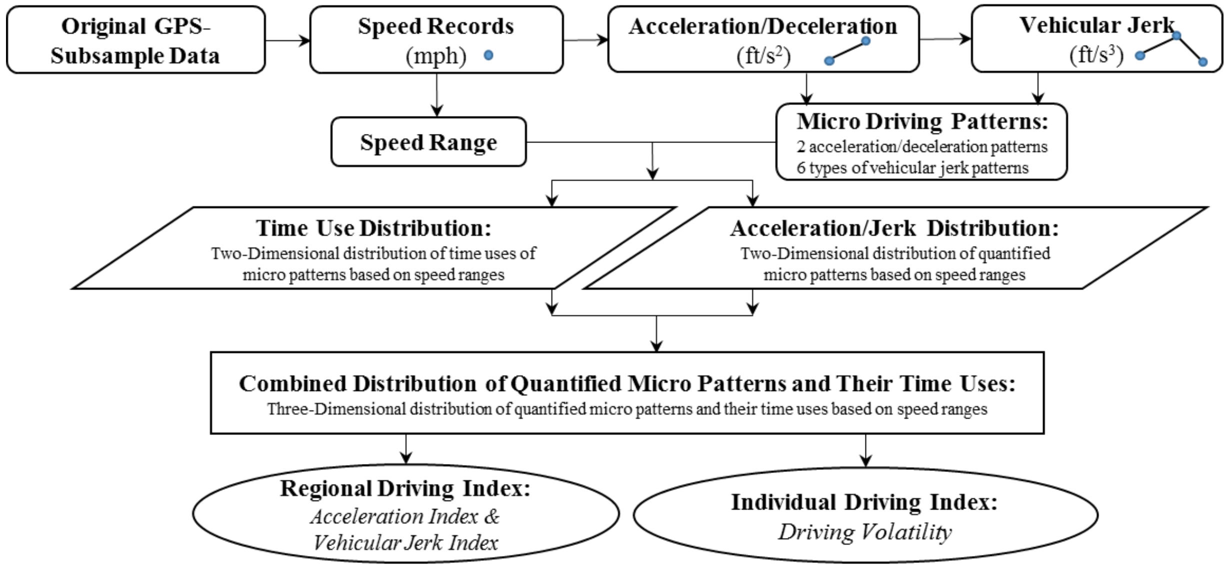


Figure 3: Measures of driving performance used in this study.

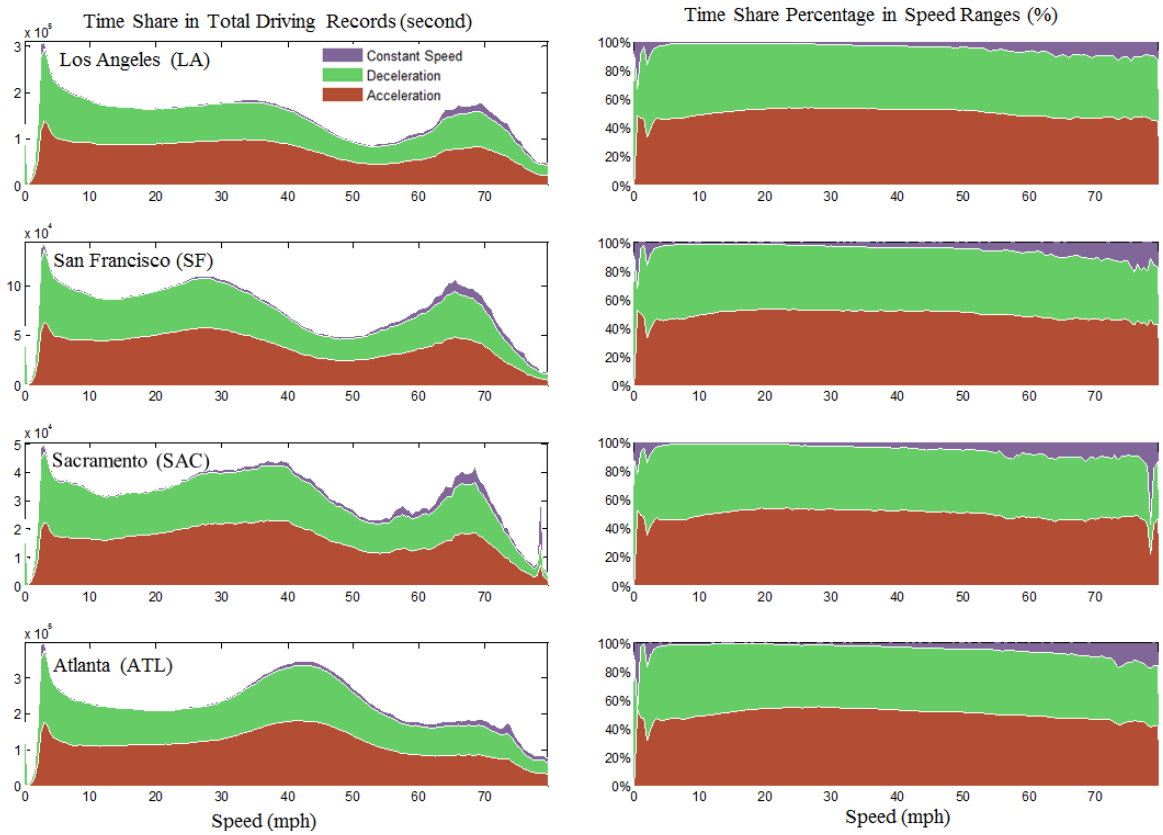


Figure 4: Time use distribution of acceleration, deceleration and constant speed in study regions.

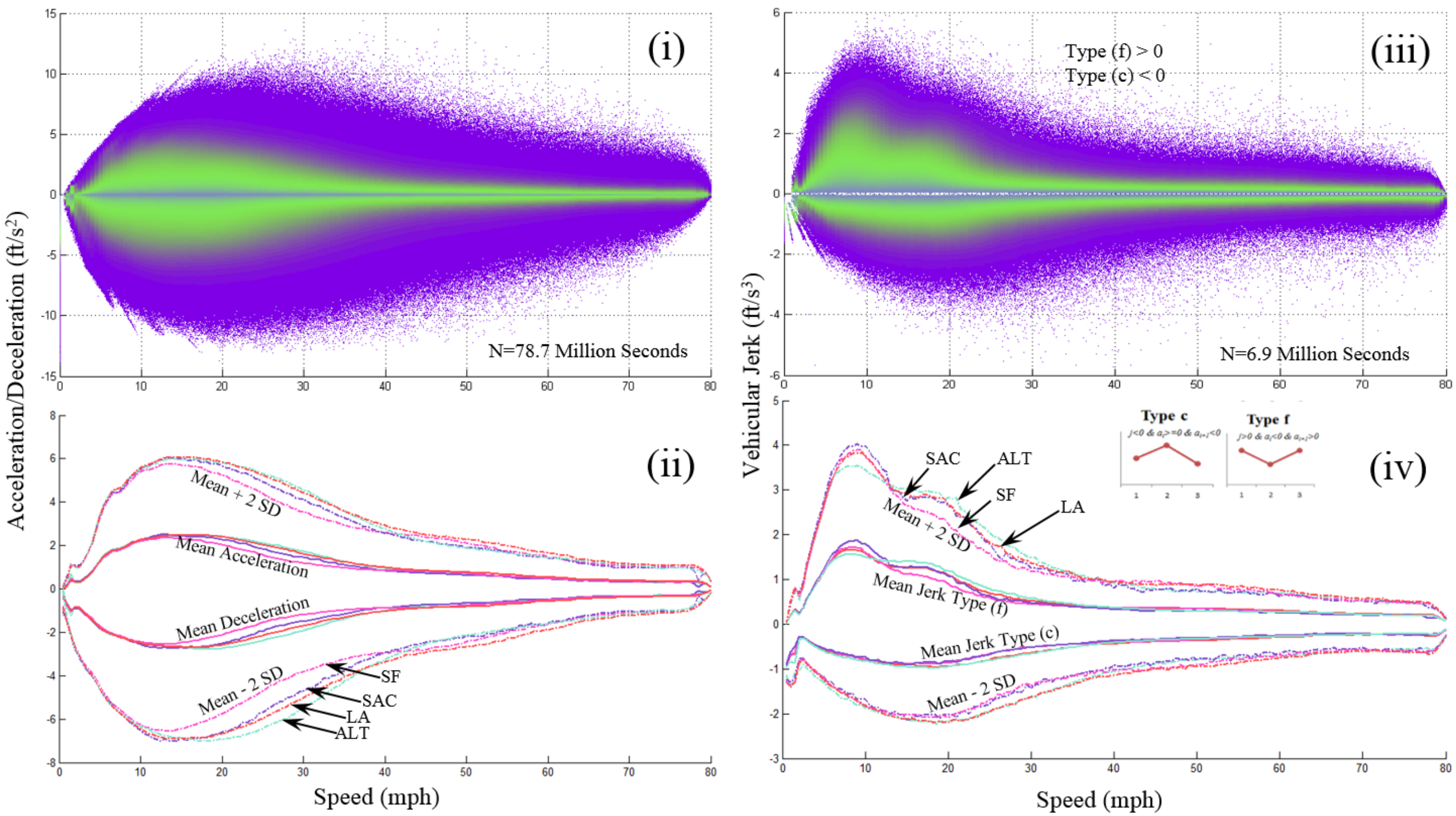


Figure 5: Distributions of acceleration and jerk at various speeds.

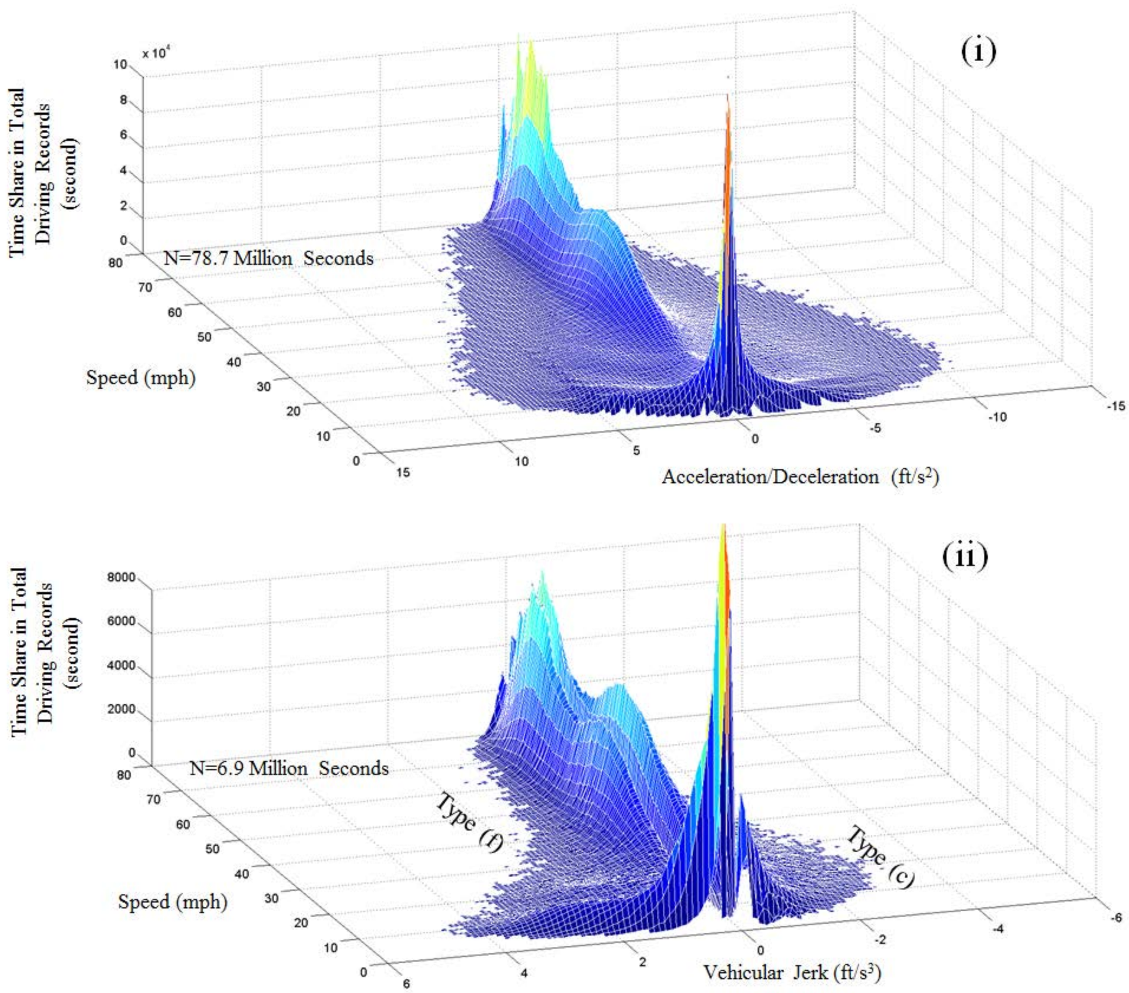


Figure 6: Combined distributions of speed, acceleration (or vehicular jerk), and time use.

Table 1: Summary of selected studies on regional driving performance

Author(s)	*Year*	*Investigation location(s)*	*Sample size*	*Performance measures*	*Key findings*
Wang et al., 2008	2003-2005	11 cities in China (Beijing, Shanghai, etc.)	310 trips	Speed, acceleration/deceleration, kinetic energy, acceleration/deceleration events	Chinese cities → more time spent on acceleration and deceleration than European and US cities.
Hung et al., 2005	2002-2003	Hong Kong, Macao and Zhuhai in China	68 trips	Speed, acceleration/deceleration, kinetic energy, cruising	Hong Kong, Macao and Zhuhai (in China) → greater mean speeds than New York City (NYC). Hong Kong → more acceleration/deceleration events than NYC.
Lin et al. 2003	2000	Metropolitan Bay Area, Sacramento and Stanislaus	529,901 seconds	Speed, acceleration/deceleration frequency	Bay Area's arterials → greater mean speed than Sacramento. Stanislaus data → lower mean speed than Sacramento.
Lents et al. 2004	2001-2004	7 cities from different countries (Los Angeles-US, Mexico City-Mexico, Sandiago-Chile, etc.)	Not reported	Average daily driving speed, emission	Mexico City → lowest mean travel speed Los Ageless → highest mean travel speed Pune → highest VOC emission rate Lima → highest CO2 emission rate
Yao et al. 2007	2003-2005	7 Chinese cities (Beijing, Shanghai, etc.)	49 vehicles, 918,400 seconds	Emission (CO, HC and NO_x), fuel consumption	No comparisons between cities was made. Greater emissions (CO, HC, NO_X) occurred on residential roads than freeways.
Saleh et al. 2010	2008-2009	Edinburgh-UK, New Delhi-India	44 trips of motorcycles	Speed, acceleration/deceleration, kinetic energy	Edinburgh → greater acceleration and deceleration rates than Delhi. Edinburgh → higher kinetic energy than Delhi.
This study	2011-2013	Los Angeles , San Francisco, Sacramento, Atlanta	3,861 vehicles, 101,628 trips, 78,709,907 seconds	Speed, acceleration, vehicular jerk, time use distribution	See sections below.

Table 2: Characterization of study regions and sample statistics

Metropolitan Region	***Los Angeles***	***San Francisco***	***Sacramento***	***Atlanta***
Abbreviation	LA	SF	SAC	ATL
Sample Data Features				
Sample Vehicles	1,258	636	315	1,652
Sample Trips	29,373	14,417	6,468	51,370
Driving Records (second)	24,185,380.00	12,579,345.00	5,229,874.00	36,715,308.00
Characterization of Study Regions				
Road Pattern	Gridiron	Non-Gridiron	Gridiron	Non-Gridiron
Terrain	Flat	Hilly	Flat	Rolling
Road Density (lane miles per square mile)	12.64	6.80	11.39	3.20
Roadway Congestion Index	1.57	1.41	1.30	1.18
Population Density (people per square mile)	8092.30	17179.10	4764.20	3154.30
Female Percentage	0.50	0.49	0.51	0.50
Percentage of Youth (15~29 yrs)	24.20%	22.70%	24.30%	28.00%
Percentage of Older (> 65 yrs)	10.50%	13.60%	10.60%	9.80%

Sources: 2010 Census Data and 2012 Urban Mobility Report (Sorensen 2010, Caltrans 2011, Schrank et al. 2012, Census 2014).

Table 3: Volatility indices for regional driving performance

Region	*Acceleration Index*	*LA-based Acceleration index*	*Vehicular Jerk Index*	*LA-based Vehicular Jerk Index*
LA	0.951	100%	0.268	100%
SF	0.826	87%	0.25	93%
SAC	0.847	89%	0.241	90%
ATL	0.909	96%	0.254	95%

Table 4: Descriptive statistics for samples from the four regions

			Los Angeles (LA)					San Francisco (SF)					Sacramento (SAC)					Atlanta (ATL)				
Variables			*N*	*Mean*	*Std. Dev.*	*Min*	*Max*	*N*	*Mean*	*Std. Dev.*	*Min*	*Max*	*N*	*Mean*	*Std. Dev.*	*Min*	*Max*	*N*	*Mean*	*Std. Dev.*	*Min*	*Max*
Dependent	Volatility Score (%)		20,608	5.15	3.90	0.00	40.84	9,276	5.11	3.91	0.00	48.88	4,642	5.10	4.09	0.00	44.53	40,201	4.45	3.71	0.00	37.86
Driver-related Factors	Gender [Male]		933	0.47	0.50	0	1	436	0.53	0.50	0	1	235	0.46	0.50	0	1	1,486	0.47	0.50	0	1
	Driver Age (years)		933	47.10	13.56	16	87	436	49.73	12.37	16	82	235	50.69	12.85	17	78	1,486	47.18	13.32	15	91
Vehicle-related Factors	Body Type	Auto-Sedan	933	0.50	0.50	0	1	436	0.46	0.50	0	1	235	0.46	0.50	0	1	1,486	0.44	0.50	0	1
		Two Seated	933	0.07	0.25	0	1	436	0.09	0.28	0	1	235	0.03	0.18	0	1	1,486	0.04	0.19	0	1
		Van	933	0.06	0.24	0	1	436	0.07	0.26	0	1	235	0.06	0.23	0	1	1,486	0.09	0.28	0	1
		Convertible	933	0.01	0.12	0	1	436	0.02	0.15	0	1	235	0.01	0.11	0	1	-	-	-	-	-
		SUV	933	0.20	0.40	0	1	436	0.21	0.41	0	1	235	0.26	0.44	0	1	1,486	0.28	0.45	0	1
		Station Wagon	933	0.04	0.20	0	1	436	0.06	0.24	0	1	235	0.03	0.17	0	1	1,486	0.02	0.14	0	1
		Pickup	933	0.11	0.31	0	1	436	0.10	0.30	0	1	235	0.15	0.36	0	1	1,486	0.14	0.34	0	1
		RV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,486	0.00^	0.04	0	1
	Vehicle Age (years)		933	6.96	4.62	0	44	436	7.21	4.75	0	34	235	7.62	4.60	0	32	1,486	7.91	5.42	0	50
	Fuel Type	Gasoline Vehicle	933	0.80	0.40	0	1	436	0.71	0.45	0	1	235	0.82	0.38	0	1	1,486	0.96	0.19	0	1
		Hybrid Vehicle	933	0.11	0.31	0	1	436	0.17	0.38	0	1	235	0.11	0.31	0	1	1,486	0.01	0.11	0	1
		Diesel Vehicle	933	0.04	0.19	0	1	436	0.05	0.21	0	1	235	0.02	0.14	0	1	1,486	0.02	0.14	0	1
		Plug-in Hybrid	933	0.00^	0.07	0	1	436	0.01	0.10	0	1	235	0.01	0.11	0	1	-	-	-	-	-
		CNG (C.Natural Gas)	933	0.02	0.14	0	1	436	0.01	0.10	0	1	-	-	-	-	-	-	-	-	-	-
		Electric Vehicle	933	0.02	0.14	0	1	436	0.05	0.21	0	1	235	0.02	0.13	0	1	-	-	-	-	-
		Other Vehicles	933	0.00^	0.07	0	1	436	0.00^	0.07	0	1	235	0.02	0.13	0	1	1,486	0.01	0.08	0	1
Trip-related Factors	*Rush Hour [Yes]		20,608	0.46	0.50	0	1	9,276	0.46	0.50	0	1	4,642	0.47	0.50	0	1	40,201	0.46	0.50	0	1
	Weekend [Yes]		20,608	0.23	0.42	0	1	9,276	0.23	0.42	0	1	4,642	0.23	0.42	0	1	40,201	0.24	0.43	0	1
	Trip Distance (Mile)		20,608	8.76	13.66	0.08	321.06	9,276	9.44	16.05	0.13	304.79	4,642	8.57	12.64	0.19	251.01	40,201	7.94	13.18	0.01	431.52
	Trip Duration (Minute)		20,608	14.45	15.58	2.02	330.73	9,276	15.40	17.93	2.02	274.58	4,642	13.87	14.01	2.02	260.08	40,201	14.18	14.74	2.01	374.45
	Trip Average Speed (MPH)		20,608	29.21	13.75	2.21	75.59	9,276	28.47	11.93	3.89	73.87	4,642	30.06	11.31	4.89	71.19	40,201	27.22	11.13	4.18	79.37
	Commute Trip [Yes]		20,608	0.16	0.36	0	1	9,276	0.17	0.38	0	1	4,642	0.14	0.35	0	1	40,201	0.19	0.40	0	1

Note: This study uses the μ +/- 2 σ as the threshold to create the volatility score.

*: Rush hours are AM (6:30 am-10:00 am) or PM (3:30 pm-7:00 pm);

^: Values less than 0.01.

Table 5: Mixed-effects regression models for vehicular jerk in the four regions

Regions			*Los Angeles (LA)*		*San Francisco (SF)*		*Sacramento (SAC)*		*Atlanta (ATL)*	
Variables			*β*	P>|z|	*β*	P>|z|	*Β*	P>|z|	*β*	P>|z|
Constant			6.535***	0.000	7.154***	0.000	9.301***	0.000	6.446***	0.000
Driver-related Factors	Gender [Male]		0.153	0.378	0.092	0.705	-0.119	0.732	0.203*	0.097
	Driver Age (years)		-0.028***	0.000	-0.027***	0.005	-0.030**	0.023	-0.030***	0.000
Vehicle-related Factors	Body Type	Auto-Sedan	Base		Base		Base		Base	
		Two Seated	1.095***	0.001	1.203***	0.007	-1.456	0.121	1.861***	0.000
		Van	-1.134***	0.001	-1.191**	0.014	-1.872**	0.014	-1.296***	0.000
		Convertible	1.504**	0.034	1.238	0.116	-0.493	0.738	-	-
		SUV	-0.424*	0.054	-0.386	0.222	-1.280***	0.002	-0.436***	0.002
		Station Wagon	-0.819*	0.056	-0.225	0.683	-1.560	0.125	-0.703*	0.089
		Pickup	-1.359***	0.000	-1.507***	0.001	-1.672***	0.003	-0.950***	0.000
		RV	-	-	-	-	-	-	-2.162	0.110
	Vehicle Age (years)		-0.042**	0.027	-0.052*	0.053	-0.205***	0.000	-0.054***	0.000
	Fuel Type	Gasoline Vehicle	Base		Base		Base		Base	
		Hybrid Vehicle	-0.609**	0.029	-0.907***	0.007	-1.296**	0.020	-1.496***	0.005
		Diesel Vehicle	0.133	0.77	-0.173	0.771	-1.998	0.104	-0.949**	0.027
		Plug-in Hybrid	-0.733	0.558	-0.642	0.621	-1.458	0.340		
		CNG (Comp. Nat.	-0.448	0.446	0.880	0.462	-	-	0.616	0.409
		Electric Vehicle	-0.522	0.395	-0.671	0.274	-3.030**	0.018	-	-
		Other Vehicles	-1.441	0.248	-0.430	0.798	1.373	0.304	-	-
Trip-related Factors	Rush Hour [Yes]		0.136***	0.002	0.152**	0.023	0.059	0.551	0.218***	0.000
	Weekend [Yes]		-0.268***	0.000	-0.273***	0.001	-0.371***	0.002	-0.277***	0.000
	Trip Distance (Mile)		-0.014***	0.000	-0.012***	0.000	-0.005	0.217	-0.006***	0.000
	Commute Trip [Yes]		0.373***	0.000	0.497***	0.000	0.430***	0.009	0.336***	0.000
Random-effects Parameters			Estimate	Pct. of Total	Estimate	Pct. of Total	Estimate	Pct. of Total	Estimate	Pct. of Total
Variance between drivers/vehicles			6.380	42.01%	5.931	39.13%	7.330	42.89%	5.317	38.44%
Variance between trips (within driver)			8.808	57.99%	9.227	60.87%	9.759	57.11%	8.516	61.56%
SUMMARY STATISTICS										
Number of observations			20608		9276		4642		40201	
Number of groups			933		436		235		1486	
Percent of explained variance between drivers			10.53%		12.76%		24.04%		13.45%	
Percent of explained variance between trips			0.72%		0.86%		0.48%		0.62%	
Wald χ^2			241.98		132.98		86.02		450.28	
Prob. > χ^2			0.000		0.000		0.000		0.000	
Log likelihood			-52783.864		-23949.483		-12136.443		-101887.03	

Note: *** significant at a 99% confidence level; ** significant at a 95% confidence level; * marginally significant at a 90% confidence level.

PART 2: SUPPORTING CALMER INSTANTANEOUS DRIVING DECISIONS: USE OF VEHICLE TRAJECTORY DATA TO GENERATE ALERTS AND WARNINGS

Asad Khattak[1], Jun Liu[2], Xin Wang[3]

[1]The University of Tennessee, Knoxville, [2]The University of Texas, Austin, [3]Virginia Department of Transportation

Abstract: Drivers make short-term steering and speed decisions based on incoming information from external sources. To avoid collisions, they may accelerate or brake hard by applying greater than "normal" pressure on the accelerator or brake. Some hard accelerations and braking may be avoidable through driver feedback in real-time. Alerts and warnings can result in calmer driving, and potentially reduce traffic flow disturbances, energy use, emissions, safety problems, and wear and tear on brakes and the engine. This study explores how real-time vehicle trajectory data can be used to generate driver feedback through actionable alerts and warnings. The study provides a novel framework for how acceleration and braking monitoring can generate alerts and warnings, provided through advanced traveler information systems. Extreme driving patterns under seemingly normal conditions are the key to generating actionable personalized feedback. Using data from a behavioral survey coupled with second-by-second speed data from automobile trips, statistical models are estimated to distinguish between groups of individuals who will likely receive a substantial amount of alerts and warnings versus those who will receive none. The implications of the findings and potential applications to fleet vehicles and driving population are discussed.

Keywords: advanced traveler information systems, driving decisions, acceleration, traveler behavior, model selection

1. INTRODUCTION

Automobile driving is of particular interest, because of its large impact on safety, energy consumption, and emissions (Evans 1979). Driving involves making decisions based on information perceived by drivers instantaneously. The information perceived while driving can be roughly divided into two sets: a) Driving context, such as road condition, traffic flow, and weather, and b) Driving situation, such as vehicle speed, engine rotation speed, direction of vehicle, fuel consumption, etc. Currently, modern technologies are able to provide such real-time driving status information, to communicate how drivers behave while driving and react to the changing context. In some instances, driving can be volatile (hard acceleration and braking) as drivers respond to changes in the context (Wang et al. 2015b). Abrupt acceleration and sudden hard braking can be associated with fuel wastage, greater emissions, safety problems, and avoidable wear and tear on brakes and the engine (Won and Langari 2005, Murphey et al. 2009, Kim and Choi 2013b). Notably, variations in speed and acceleration have significant associations with crash rates (TRB 1998). This study explores variations in short-term driving behaviors and how real-time driving behavior information can be used to generate actionable alerts and warnings for drivers, resulting in calmer driving. Note that travel alerts are typically milder than travel warnings.

As emerging technologies facilitate data collection (e.g., in-vehicle Global Position System

(GPS) device allows speed to be captured), and this study takes advantage of such data for the purposes of driver decision support. The questions to be answered in this study are:

1) How to identify extreme short-term driving decisions?
2) How to provide real-time alerts and warnings to the driver?
3) What groups of individuals are more likely to receive alerts and warnings?
4) What are the potential applications of driver feedback for traffic control?

2. LITERATURE REVIEW

Studies have provided instructions to assist en-route driving behaviors through in-vehicle alerts or warnings to drivers (Erlichman 1992, Jovanis et al. 1997, Lee et al. 2002, Wiese and Lee 2004, Scott and Gray 2008, Chan and Ng 2009). Many systems are designed to provide drivers with short-term warnings/alerts about potential collisions (Horowitz and Dingus 1992, Suetomi and Kido 1997, Ben-Yaacov et al. 2002, Lee et al. 2002, Maltz and Shinar 2004, Shladover and Tan 2006, Sengupta et al. 2007), lane changes (Shladover and Tan 2006), congestion (van Driel et al. 2007) and adverse weather (Gupta et al. 2002, Goodwin and Pisano 2003, Ji et al. 2006). Some systems are designed to monitor driver behaviors and warn drivers about speeding (Jimenez et al. 2008), fatigue (Bishop 2000, Brown and Noy 2004, Schleicher et al. 2008) and traffic signal/sign violations (Chen et al. 2010). Studies have provided various ways of characterizing extreme driving behaviors, distinguishing extreme driving behaviors from normal driving behaviors(Lajunen et al. 1997b) and using acceleration thresholds for categorizing aggressive and calm driving (Rakha et al. 2001, Langari and Won 2005, Kim and Choi 2013b). Rate of change in accelerations (or called vehicular jerk) has also been suggested for monitoring driving behaviors because it better captures the change of instantaneous driving decisions, e.g., abruptly accelerating after deceleration (Murphey et al. 2009). Other measures for defining extreme/aggressive driving behaviors have been discussed in the literature, such as horn honking (Shinar 1998), tailgating and running red traffic lights (Parker et al. 1998), traffic rule compliance (Lajunen et al. 1997a, 1998), frequent or unsafe lane changes, failing to signal, tailgating, failing to yield right of way, and disregarding traffic controls (NHTSA 2000).

Extreme driving behaviors also correlate with various factors, such as driver demographics, driver personality, and time pressure or value of time. Previous studies show that young drivers (as compared with drivers > 45 years old) are more likely to make extreme driving behaviors, such as cutting across multiple lanes and passing on shoulders; compared with females, males are more likely to commit aggressive actions on road (Parker et al. 1995, Krahé and Fenske 2002, Shinar and Compton 2004). A study revealed that the possibility of being an aggressive driver is positively correlated with macho personality and the horsepower of a vehicle (Krahé and Fenske 2002). Drivers sensing the time pressures during rush hours are also found to have an increased likelihood of making aggressive driving actions (Parker et al. 1995, Shinar and Compton 2004).

Studies, suggesting cut-off points as thresholds for distinguishing extreme driving behaviors from normal driving behaviors, seem to have an unstated assumption that driving behaviors do not vary under various driving contexts. However, this assumption may not hold in real-world driving situations, which involve various driving contexts, e.g., local roads, arterials and freeways. Previous studies by the authors have shown significantly varying distributions of acceleration and vehicular jerk under different speeds indicating changing driving contexts (Liu et al. 2014b, Liu et al. 2015d, Wang et al. 2015b). This leads us to using a new approach to identify extreme short-term driving decisions from real-world driving practices and generating warnings/alerts taking into account heterogeneous driving under various driving contexts.

3. DATA AND METHODOLOGY

3.1 Data Description

To provide an empirical study, a large-scale trajectory dataset is used to demonstrate how to identifying extreme short-term decisions and when a warning/alert should be generated for drivers. The dataset contains over 36 million seconds of trajectory records collected by in-vehicle GPS devices during 2011 Atlanta Regional Travel Survey (Geostats 2011).The trajectories represent various driving practices in 20 counties in Atlanta Metropolitan area. The data collection was done for each motorized travel trip. In all, the data used in this study include over 50 thousand trips made by 1653 drivers residing in the survey area. The trajectories provide second-by-second records with time, vehicle velocity and location coordinates. Unfortunately, information about coordinates was removed from the publicly released dataset, because of privacy concerns.

In addition to trajectories used in the study, travel behavioral information, including driving demographics, vehicle characteristics and trip features, is also utilized to explore the correlates of extreme driving behaviors, uncovering what groups of users will be potentially receiving more warnings/alerts. The findings may suggest what groups of users are prioritized to implement the warning/alert systems proposed in this study.

3.2 Key Measures

Our previous papers have discussed measures that can be used for quantifying short-term driving behaviors (Liu et al. 2014b, Liu et al. 2015d, Wang et al. 2015b). Most are commonly used physical measures for movements, including speed, acceleration and vehicular jerk. In this study, the concept of vehicular jerk was expanded from 3 consecutive seconds to 4. Generally, expanded vehicular jerk literally can cover a short-term driving pattern within n successive seconds. Thus there are (n+1)*n! types of possible patterns of short-term driving decisions. For example, if we observe the decision patterns in three seconds (four data points), the patterns may include: 1) three acceleration decisions, 2) two acceleration and one deceleration decisions, 3) one acceleration and two deceleration decisions, or 4) three deceleration decisions. For each case, there are 3!=6 possible combinations. In all, there are 4*6=24 possible patterns of driving decisions for three sequential seconds. For expanded vehicular jerk, the study focuses on the patterns shown in Figure 1.

To quantify expanded vehicular jerk patterns, two measures are introduced as below:

1) The sum of absolute vehicular jerk values,

$$JE = \sum_{i=1}^{i=n-1} |J_i| \tag{1}$$

Where,

J_i-the *i*th jerk value in *n* sequential seconds.

2) Variance of vehicular jerk.

Figure 2 shows the calculations for acceleration, vehicular jerk, sum of absolute vehicular jerk and variance of jerk, and tells the difference between sum of absolute vehicular jerk and the variance of jerk.

Figure 3 presents the profiles of speed, acceleration, jerk, sum of absolute vehicular jerk, and jerk variance in one sampled trip. The spikes in the three jerk profiles occur only when there are large changes in the accelerations. Importantly, these measures further amplify substantial changes in speed choice.

4. IDENTIFICATION OF EXTREME SHORT-TERM DECISIONS

4.1 Thresholds for Extreme Short-Term Decisions

To develop thresholds for identifying extreme driving decision patterns, this study suggests thresholds that change with speeds. They can be further customized to an individual driver, based on their own previous performance (and similar population in the area). To illustrate the basic idea, Figure 4 presents distributions of quantified driving decision patterns within 40±0.25 mph range. The distribution indicates how drivers behave while travel speeds are around 40 mph. Whether a driving decision pattern is extreme (classified as hard acceleration or braking) depends on driving performance in this speed range and not at all other speeds. Mean and standard deviation can be calculated and the mean plus two (or three or even higher) standard deviations for a specific speed range can be used as threshold. The example in Figure 6 shows that driving decision patterns with acceleration rates exceeding about 2 ft/s2 (0.61 m/s2) or deceleration rates beyond -2 ft/s2 (-0.61 m/s2) will be considered extreme driving behavior.

Note that, for acceleration and single vehicular jerk profiles, there are two thresholds respectively for acceleration (or positive jerk) and deceleration (or negative jerk). Thus, the threshold for deceleration (or negative jerk) is mean minus two standard deviations.

The study developed thresholds for all speed ranges by 0.5 mph increments to identify extreme driving decision patterns, as shown in Figure 5. The inner edges of areas are the thresholds varying across speeds. The 3D plots show how many extreme driving decision patterns at each speed range. Clearly, a large portion of driving records are in normal or stable areas (yellow and blue). Figure 5(ii) indicates that extreme variances are close to the bottom of the 3D plot and thus they do not occur frequently, as expected. Note that, thresholds shown in Figure 5 are established based on 36 million driving records in Atlanta Metropolitan area. Driving records from different areas may form different thresholds. Somehow, thresholds can reflect the performance of regional driving practices. In this study, extreme driving patterns are identified through comparing individual driving moments with the majority of driving practices from the same region under identical speed ranges (indicating similar driving contexts).

4.2 Alerts and Warnings

This study introduces critical areas (shown in Figure 5), established at a regional level, to generate alerts and warnings to drivers. Figure 6 presents a sample trip with time-speed trajectory and the alerts and warning points using different methods. The acceleration method gives alerts and warnings for driving seconds with the large changes of speed, illustrated by steep slopes. Vehicular jerk related methods work similarly by regarding large changes of driving decisions (i.e., acceleration and deceleration), but jerk variance method supplements the variations of driving decision changes within short periods. Using expanded vehicular jerk method plus the jerk variance to generate alerts and warnings is the approach proposed in this study. The driving alerts and warnings are given to drivers if they are having both extreme jerk values and extreme jerk variances. The alerts and warnings tell extreme variations of short-term driving decisions. Figure 6(v) illustrates the locations where feedback and warnings may be given to inform the person about their extreme driving. There are about 14 warnings/alerts generated for this sample trip. In the Atlanta Metropolitan area, on average 9.8 warnings or alerts (with a standard deviation -12.8) are given to drivers per 10 minutes. The warnings and alerts are given, if basing them on regional driving performance.

5. CORRELATES OF ALERTS AND WARNINGS

5.1 Descriptive Statistics

This study explores correlates of alerts and warnings with associated factors. After observations with incomplete information were removed, the final database for correlation analysis contained 40,201 trips from 1,486 drivers. Given the available travel behavioral information from 2011 Atlanta Regional Travel Survey, factors that are examined include driver socio-demographics (gender and age), vehicle characteristics (vehicle age, body type and fuel type) and trip features (purpose, length, duration and time). Table 2 presents descriptive statistics of variables examined. On average, drivers during each trip would receive about 9 warnings/alerts (with a standard deviation – 16.23) generated using regional thresholds. The dependent variable is number of warnings/alerts during a trip, which is a count data. Given the over-dispersion (i.e., standard deviation > mean), negative binomial (NB) regression is applied for untangling correlates of warnings/alerts with associated factors (Gardner et al. 1995, Hilbe 2011).

5.2 Multi-level Negative Binomial Modeling

Table 2 shows that there are different number of observations for trip-related factors and driver- or vehicle- related factors, because a two-level hierarchical structure is embedded in the data, i.e., trip-level and driver-level. Each driver may make multiple trips during the survey. Traditional Ordinary Least Square (OLS) models regression models assume that observations are independent from each other. However, trips made by the same driver are not completely independent at the driver-level. For example, if a driver who likes jerking his or her car (personal driving style), all trips made by this driver may be associated with more warnings/alerts than those made by other drivers. Thus the assumption for OLS regression models is violated. This study applies a multi-level regression model, which allows us to observe the inter-driver difference and inter-trip difference. Finally, a multi-level negative binomial regression (MNB) model is used to explore correlates of alerts and warnings in a hierarchy.

A two-level structure is explored in this study: driver-level and trip-level nested in drivers. For a two-level MNB model, the probability function is given by (StataCorp 2013):

$$\Pr(y_{ij} = k) = \frac{\Gamma[(1/\alpha) + k]}{\Gamma(1/\alpha)\Gamma(k+1)} \left[\frac{1/\alpha}{(1/\alpha) + \mu_{ij}}\right]^{1/\alpha} \left[\frac{\mu_{ij}}{(1/\alpha) + \mu_{ij}}\right]^{k} \qquad (2)$$

Where,

y_{ij} = observed number of warnings/alerts for j^{th} trip made by i^{th} driver, j = 1, 2, …, m, i = 1, 2, …,n;

k = possible number of warnings/alerts for a trip;

μ_{ij} = predicted number of warnings/alerts for j^{th} trip made by i^{th} driver;

α= over-dispersion parameter;

Note that, Equation 2 is extended from the standard negative binomial probability function (StataCorp 2013) to incorporate normally distributed random effects at different levels. The number of warnings/alerts can be predicted by (Booth et al. 2003)

$$\mu_{ij} = \exp(X_{ij}\beta + Z_{ij}\gamma_i + \varepsilon_{ij}) \qquad (3)$$

Where,

X_{ij} = explanatory factors with fixed-effects;

β = fixed effects/coefficients;

Z_{ij} = explanatory factors with random-effects;

γ_i = random effects/coefficients, $\gamma \sim N(0, G)$, G is a $q \times q$ variance matrix and q is the number of variables for a driver;

$\exp(\varepsilon_{ij})$ = a gamma-distributed error term.

In this study, there is one explanatory variable with random effects γ which capture variations between drivers. This variable is featured by driver-vehicle pair (one driver corresponds to a unique vehicle in this sample).

5.3 Model Selection

Theoretical plausibility and empirical properties were used to guide the development of models. Outputs of the model with all plausible variables (shown in Table 2) may show that some variables are not statistically significant. Appropriate methods for model selection were used to provide the best statistical model, based on the prediction values, using standard information criteria, i.e., Akaike information criterion (AIC) and Schwarz's Bayesian information criterion (BIC). Given models estimated with the same data and that they are statistically significant overall, the ones with smallest values of AIC and BIC are considered statistically superior model. For any statistical model, AIC is given by (Akaike 1974, Bozdogan 1987)

$$AIC = -2\mathrm{Ln}L + 2k \qquad (4)$$

And BIC is measured by (Schwarz 1978)

$$BIC = -2\mathrm{Ln}L + k\mathrm{Ln}N \qquad (5)$$

Where,

$\mathrm{Ln}L$ = maximum log-likelihood of a model;

k = the number of parameters in a model;
N = the number of the sample size.

Note that, compared to AIC, BIC always favors a simpler model (i.e., smaller degrees of freedom) for data with a large number of observations. Given the large number of observations (N=40,201 in the database, at the trip-level), this study uses AIC as the major information criterion when AIC and BIG do not agree on which model is the best, in terms of goodness of fit.

Table 3 presents the results of model selection. Since trip duration and length are highly correlated (correlation = 0.93), only one of them should be included in the model along with other factors. Results show that a model with trip duration (Model 1) is better-fitting than a model with trip distance (Model 2). Then, using Model 1, backward elimination method was used to remove factors that do not have a statistically significant estimate (Sutter and Kalivas 1993), starting from the factors with highest p-value and stopping when all p-values are less than a given criterial p-value (normally 0.1). The remaining factors in the model should have an estimate/effect with a p-value less than 0.1. Notice that, some levels of attributions of categorical variables have a p-value larger than 0.1. During the elimination process, these levels were added to the "base" level.

Finally, Model 6 provides the smallest AIC value where all factors have an estimate less than 0.1. Model 7 was estimated to test if dropping station wagon will be appropriate, as the p-value was in the marginal range. However, AIC still indicated that Model 6 is better-fitting and was the final model for correlates of warnings/alerts.

5.4 Modeling Results

Table 4 presents the modeling results of the full model –Model 1 and the final selected model – Model 6. Incidence Rate Ratios (IRRs) are also reported along with coefficients for model interpretation. Variables that do not have a significant estimate in Model 1 are successfully excluded in Model 6. Signs of estimated coefficients are as expected. The likelihood ratio test for MNB vs. ordinary NB regression shows that there is sizable variance between groups (driver-vehicle pairs) and multi-level regression is significantly better than ordinary NB regression (p<0.001). The Alpha (=0.866) is also significantly larger than zero, indicating negative binomial model is better than Poisson model. IRRs in the model indicate how many times greater the expected number of warnings/alerts would be for one unit increase in explanatory variables, while hold the other variables constant in the model. The following interpretation is based on outputs of the final model.

Modeling results show that older drivers need fewer warnings/alerts, by a factor of 0.992 for each added year in driver age. Compared with the base level (mainly auto sedans), trips by drivers of two-seated vehicles are expected to have about 1.559 times more warnings/alerts. Vehicles in other body types are associated with less warnings/alerts. Older vehicles seem to be associated with less warnings/alerts as well. Drivers of older vehicles will receive fewer warnings/alerts, by a factor of 0.985. In addition, compared with the base (mainly gasoline vehicles), hybrid vehicles as well as vehicles using diesel are associated with fewer warnings/alerts, by 0.629 and 0.654, respectively. Furthermore, warnings/alerts are also correlated with trip features. During morning and afternoon rush hours, more warnings/alerts (by a factor of 1.09 and 1.034 respectively) may be given to drivers. Trips made on weekends are associated with less warnings/alerts, by 0.953. Long trips are expected to have more warnings/alerts, by 1.063. Commute trips are associated with more warnings/alerts, by a factor of 1.160, compared with non-commute trips.

The results from modeling indicate what groups of users will be potentially receiving more warnings/alerts. The findings may suggest what groups of users are prioritized to implement the warning/alert systems proposed in this study: young drivers, new vehicles and two-seated vehicles.

6. INCORPORATING WITH DRIVING ASSIST SYSTEMS

Regional thresholds can be used to account for the driving context and highlight extreme driving. Based on this study, at least two types of warning/alerting information can be provided to drivers:

- Real time driving behavior information: Drivers may be alerted or warned when they exceed certain thresholds of acceleration or vehicular jerk, providing them with dynamic feedback on their behavior through Advanced Traveler Information Systems (ATIS). Displays can be designed to inform drivers their real-time extreme driving behavior, without overly distracting them, e.g., through a light on the dashboard that turns yellow or red from green. This can also be supplemented via email notifications.
- Daily/monthly/yearly driving behavior summary information. Long-term advice on driving patterns can be provided to the driver based on analysis of their daily, monthly or yearly driving performance. Such information can be provided through websites, and may contain a record, analysis of driving patterns and customized advice on improving accelerations, braking, speeds, and turns, etc.

Thresholds of identifying extreme driving patterns can be based on combinations of accelerations, single vehicular jerk, expanded vehicular jerk and variance in these parameters. While this study used the mean plus/minus two standard deviation thresholds for identifying extreme patterns, other threshold criteria can also be used, e.g., mean plus three standard deviations. Note that, the thresholds may be further adjusted based on time of day, weather, terrain, and roadway classification. They can be personalized based only on trips undertaken by the individual or use regional data to calculate thresholds (Liu et al. 2015d). Adding these functions to current mobile devices has the potential for calmer driving.

7. LIMITATIONS

The speed records in this study were generated from in-vehicle GPS devices and then processed by a professional survey research firm. Thus, the extent of measurement errors in the data is unknown, although the results from statistical modeling are reasonable. Another threat is whether the second-by-second data are good enough to monitor the instantaneous driving decisions. Compared with high industrial sampling rates, the sampling rate of the data used in this study was quite low. Owing to the current technology, surveys including the driving data collection have to face this constraint. Knowing that the driver reaction time includes the time for driver perception, identification, judgment and reaction (TRB 1998) usually takes more than one second (AASHTO 2011). Thus, there is a small chance that a driver will make multiple micro driving decisions within one second (Liu et al. 2015b). Note that due to data limitations, the accelerations analyzed are based only on speed and no consideration is given to the direction of vehicular movements.

Factors discussed in this study have shown their significant correlations with extreme driving patterns, while some critical information remains unknown to the researchers, e.g., road types, traffic conditions surrounding the vehicle, and the built environment.

8. CONCLUSIONS

This study extends the fundamental understanding of short-term driving decisions from previous papers (Liu et al. 2015e, Wang et al. 2015b) by introducing the expanded vehicular jerk as a measure of short-term driving behavior and processing the data in a novel way to generate actionable alerts and warnings. Hard accelerations and braking are important for safety monitoring. This study shows that sequence of acceleration and braking events and vehicular jerk are key driving behavior factors to consider. A methodology to establish thresholds was provided to identify extreme behaviors. The thresholds used to generate alerts and warnings can be based on the distribution of quantified patterns in a specific speed range within a region or they can be based on historical performance of an individual driver (or a combination of the two). Driving seconds with quantified patterns beyond the mean plus two standard deviations were tagged as critical moments in this study for demonstration purposes. Other thresholds can also be used to identify extreme behaviors. Warnings or alerts about extreme driving decisions can be provided to drivers in real time (as feedback) or at the end of the day, month, etc., to help them shift to calmer driving.

Results from rigorous statistical modeling reveal that the amount of warnings/alerts varies significantly between driver groups and it is highly associated with young drivers, new vehicles, two-seat vehicles, AM and PM rush hours, and commute trips. The findings suggest what groups of users are prioritized to implement the warning/alert systems proposed in this study.

The findings are useful for potential applications to fleet vehicles and the general driving population. Warnings and alerts based on accelerations and vehicular jerk can be incorporated in driving assist systems, e.g., advanced traveler information systems (ATIS). Current traveler information systems (such as 511 in the US) are largely meant to support more macro driver decisions (e.g., route choice and route diversion) and do not provide much instantaneous information that can help drivers make more short-term driving decisions. The real-time warnings and alerts, reflecting driving performance based on performance of fellow fleet vehicles or neighbors or just their own performance, can support short-term micro decisions. This in turn can benefit the community or fleets in several ways: 1) calmer driving; 2) safer driving in general (especially on icy or slippery road surfaces where alert thresholds can be lowered); 3) lower fuel consumption and emissions; and 4) identification of dangerous road segments (such as poor sight distance) that are associated with extreme driving.

9. ACKNOWLEDGEMENTS

The National Renewable Energy Laboratory provided the data used in this study. The research was supported through the TranLIVE University Transportation Center (US DOT grant number DTRT12-G-UTC17) and Southeastern Transportation Center (US DOT grant number DTRT13-G-UTC34) both sponsored by the Office of the Secretary of Transportation, US Department of Transportation. Special thanks are extended to the following entities for their support at University of Tennessee: Transportation Engineering & Science Program and the Initiative for Sustainable Mobility. The views expressed in this paper are those of the authors, who are responsible for the facts and accuracy of information presented herein.

10. REFERENCES

1. Evans, L., Driver behavior effects on fuel consumption in urban driving. Human Factors: The Journal of the Human Factors and Ergonomics Society, 1979. **21**(4): p. 389-398.
2. Wang, X., A. Khattak, J. Liu, G. Masghati-Amoli, and S. Son, What is the Level of Volatility in Instantaneous Driving Decisions? Transportation Research Part C: Emerging Technologies, 2015.
3. Kim, E. and E. Choi. Estimates of Critical Values of Aggressive Acceleration from a Viewpoint of Fuel Consumption and Emissions. in 2013 Transportation Research Board Annual Meeting. 2013. Washington DC.
4. Won, J.-S. and R. Langari, Intelligent energy management agent for a parallel hybrid vehicle-part II: torque distribution, charge sustenance strategies, and performance results. Vehicular Technology, IEEE Transactions on, 2005. **54**(3): p. 935-953.
5. Murphey, Y., R. Milton, and L. Kiliaris. Driver's style classification using jerk analysis. in Computational Intelligence in Vehicles and Vehicular Systems, 2009. CIVVS'09. IEEE Workshop on. 2009. IEEE.
6. TRB, Managing Speed: Review of current practice for setting and enforcing speed limits, in Transportation Research Board, National Research Council. 1998: Washington D.C.
7. Erlichman, J. A pilot study of the in-vehicle safety advisory and warning system (IVSAWS) driver-alert warning system design. in Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 1992. SAGE Publications.
8. Lee, J.D., D.V. McGehee, T.L. Brown, and M.L. Reyes, Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. Human Factors: The Journal of the Human Factors and Ergonomics Society, 2002. **44**(2): p. 314-334.
9. Jovanis, P., J. Campbell, K. Klaver, and W. Chen. Driver preferences for visual and auditory collision warning alerts. in Mobility For Everyone 4th World Congress on Intelligent Transport Systems, 21-24 October 1997, Berlin.(Paper No. 1076). 1997.
10. Wiese, E.E. and J.D. Lee, Auditory alerts for in-vehicle information systems: The effects of temporal conflict and sound parameters on driver attitudes and performance. Ergonomics, 2004. **47**(9): p. 965-986.
11. Scott, J. and R. Gray, A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. Human Factors: The Journal of the Human Factors and Ergonomics Society, 2008. **50**(2): p. 264-275.
12. Chan, A.H. and A.W. Ng, Perceptions of implied hazard for visual and auditory alerting signals. Safety science, 2009. **47**(3): p. 346-352.
13. Horowitz, A.D. and T.A. Dingus. Warning signal design: A key human factors issue in an in-vehicle front-to-rear-end collision warning system. in Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 1992. SAGE Publications.

14. Suetomi, T. and K. Kido, Driver behavior under a collision warning system: a driving simulator study. ITS technology collection on CD-ROM: SAE's essential resource for ITS vehicle applications, 1998, 1997.
15. Ben-Yaacov, A., M. Maltz, and D. Shinar, Effects of an in-vehicle collision avoidance warning system on short-and long-term driving performance. Human Factors: The Journal of the Human Factors and Ergonomics Society, 2002. **44**(2): p. 335-342.
16. Maltz, M. and D. Shinar, Imperfect in-vehicle collision avoidance warning systems can aid drivers. Human Factors: The Journal of the Human Factors and Ergonomics Society, 2004. **46**(2): p. 357-366.
17. Shladover, S.E. and S.-K. Tan, Analysis of vehicle positioning accuracy requirements for communication-based cooperative collision warning. Journal of Intelligent Transportation Systems, 2006. **10**(3): p. 131-140.
18. Sengupta, R., S. Rezaei, S.E. Shladover, D. Cody, S. Dickey, and H. Krishnan, Cooperative collision warning systems: Concept definition and experimental implementation. Journal of Intelligent Transportation Systems, 2007. **11**(3): p. 143-155.
19. van Driel, C.J.G., M. Hoedemaeker, and B. van Arem, Impacts of a Congestion Assistant on Driving Behaviour and Acceptance Using a Driving Simulator. Transportation Research Part F: Traffic Psychology and Behaviour, 2007. **10**(2): p. 139-152.
20. Gupta, N., A.M. Bisantz, and T. Singh, The effects of adverse condition warning system characteristics on driver performance: an investigation of alarm signal type and threshold level. Behaviour & Information Technology, 2002. **21**(4): p. 235-248.
21. Goodwin, L.C. and P. Pisano, Best practices for road weather management. Road Weather, 2003.
22. Ji, Q., P. Lan, and C. Looney, A probabilistic framework for modeling and real-time monitoring human fatigue. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 2006. **36**(5): p. 862-875.
23. Jimenez, F., F. Aparicio, and J. Paez, Evaluation of in-vehicle dynamic speed assistance in Spain: algorithm and driver behaviour. Intelligent Transport Systems, IET, 2008. **2**(2): p. 132-142.
24. Bishop, R., Intelligent vehicle applications worldwide. Intelligent Systems and their Applications, IEEE, 2000. **15**(1): p. 78-81.
25. Schleicher, R., N. Galley, S. Briest, and L. Galley, Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? Ergonomics, 2008. **51**(7): p. 982-1010.
26. Brown, C.M. and Y.I. Noy, Behavioural adaptation to in-vehicle safety measures: Past ideas and future directions. Traffic and transport psychology: theory and application, 2004: p. 25-46.
27. Chen, R., W.-L. Jin, and A. Regan, Broadcasting safety information in vehicular networks: issues and approaches. Network, IEEE, 2010. **24**(1): p. 20-25.
28. Lajunen, T., J. Karola, and H. Summala, Speed and acceleration as measures of driving style in young male drivers. Perceptual and motor skills, 1997. **85**(1): p. 3-16.
29. Langari, R. and J.-S. Won, Intelligent energy management agent for a parallel hybrid vehicle-part I: system architecture and design of the driving situation identification process. Vehicular Technology, IEEE Transactions on, 2005. **54**(3): p. 925-934.
30. Rakha, H., I. Lucic, S.H. Demarchi, J.R. Setti, and M.V. Aerde, Vehicle dynamics model for predicting maximum truck acceleration levels. Journal of transportation engineering, 2001. **127**(5): p. 418-425.

31. Shinar, D., Aggressive driving: the contribution of the drivers and the situation. Transportation Research Part F: Traffic Psychology and Behaviour, 1998. **1**(2): p. 137-160.
32. Parker, D., T. Lajunen, and S. Stradling, Attitudinal predictors of interpersonally aggressive violations on the road. Transportation Research Part F: Traffic Psychology and Behaviour, 1998. **1**(1): p. 11-24.
33. Lajunen, T., A. Corry, H. Summala, and L. Hartley, Impression management and self-deception in traffic behaviour inventories. Personality and individual differences, 1997. **22**(3): p. 341-353.
34. Lajunen, T., A. Corry, H. Summala, and L. Hartley, Cross-cultural differences in drivers' self-assessments of their perceptual-motor and safety skills: Australians and Finns. Personality and Individual Differences, 1998. **24**(4): p. 539-550.
35. NHTSA. Resource Guide Describes Best Practices For Aggressive Driving Enforcement 2000 [cited 2014 May 15th]; Available from: http://www.nhtsa.gov/About+NHTSA/Traffic+Techs/current/Resource+Guide+Describes+Best+Practices+For+Aggressive+Driving+Enforcement.
36. Shinar, D. and R. Compton, Aggressive driving: an observational study of driver, vehicle, and situational variables. Accident Analysis & Prevention, 2004. **36**(3): p. 429-437.
37. Parker, D., J.T. Reason, A.S.R. Manstead, and S.G. Stradling, Driving errors, driving violations and accident involvement. Ergonomics, 1995. **38**(5): p. 1036-1048.
38. Krahé, B. and I. Fenske, Predicting aggressive driving behavior: The role of macho personality, age, and power of car. Aggressive Behavior, 2002. **28**(1): p. 21-29.
39. Liu, J., X. Wang, and A. Khattak, Generating Real-Time Driving Volatility Information, in 2014 World Congress on Intelligent Transport Systems. 2014: Detroit, MI.
40. Liu, J., A. Khattak, and X. Wang, Creating Indices for How People Drive in a Region: A Comparative Study of Driving Performance, in 94th Annual Meeting of the Transportation Research Board 2015: Washington D.C.
41. Geostats, P.N.i.A.w., Atlanta Regional Commission Regional Travel Survey Final Report. 2011.
42. Gardner, W., E.P. Mulvey, and E.C. Shaw, Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychological bulletin, 1995. **118**(3): p. 392.
43. Hilbe, J., Negative binomial regression. 2011: Cambridge University Press.
44. StataCorp, Multilevel mixed-effects negative binomial regression. 2013.
45. Booth, J.G., G. Casella, H. Friedl, and J.P. Hobert, Negative binomial loglinear mixed models. Statistical Modelling, 2003. **3**(3): p. 179-191.
46. Akaike, H., A new look at the statistical model identification. Automatic Control, IEEE Transactions on, 1974. **19**(6): p. 716-723.
47. Bozdogan, H., Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika, 1987. **52**(3): p. 345-370.
48. Schwarz, G., Estimating the dimension of a model. The annals of statistics, 1978. **6**(2): p. 461-464.
49. Sutter, J.M. and J.H. Kalivas, Comparison of forward selection, backward elimination, and generalized simulated annealing for variable selection. Microchemical journal, 1993. **47**(1): p. 60-66.
50. AASHTO, A Policy on Geometric Design of Highways and Streets, 6th Edition, 2011. American Association of State Highway and Transportation Officials, Washington, DC, 2011. **1**: p. 990.

51. Liu, J., A. Khattak, and L. Han, How much information is lost when sampling driving behavior data?, in 94th Annual Meeting of the Transportation Research Board. 2015: Washington D.C.
52. Liu, J., A. Khattak, and X. Wang, The role of alternative fuel vehicles: Using behavioral and sensor data to model hierarchies in travel. Transportation Research Part C: Emerging Technologies, 2015. **55**: p. 379-392.

LIST OF TABLES AND FIGURES

Figure 1: Profiles of expanded vehicular jerk in second-by-second trajectories...................171
Figure 2: Quantification of expanded vehicular jerk patterns 172
Figure 3: Profiles of speed, acceleration, jerk, sum of vehicular jerk, and jerk variance in one sampled trip...173
Figure 4: Defining thresholds for 40±0.25 mph speed range 174
Figure 5: Thresholds for identifying extreme driving decision patterns................................175
Figure 6: Alerts/warnings given by different methods 176

Table 1: Descriptive statistics of variables in the model 177
Table 2: Model selection...178
Table 3: Modeling results for frequency of warnings and alerts 179

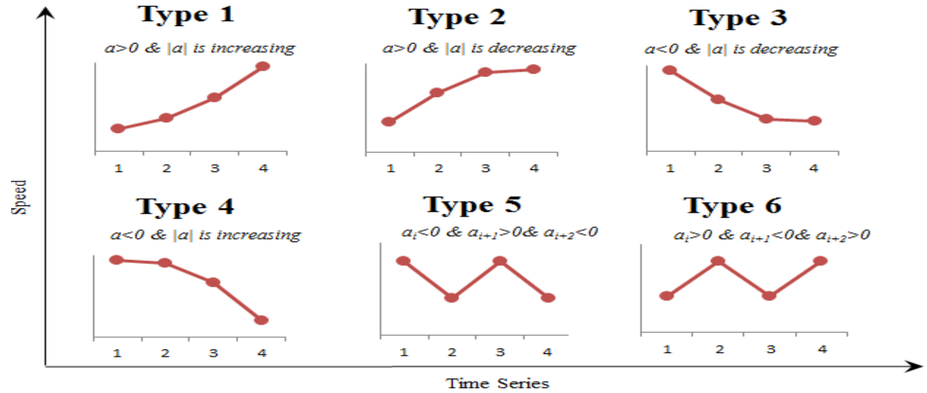


Figure 1: Profiles of expanded vehicular jerk in second-by-second trajectories.

Time Series (sec)	Speed (mph)	Accel/Decel (mph/s)	Jerk (mph/s/s)	Sum Jerk (mph/s/s)	Variance of Jerk
		Case 1			
1	30				
2	25	-5			
3	30	5	10		
4	25	-5	-10	20	200
		Case 2			
1	30				
2	25	-5			
3	30	5	10		
4	45	15	10	20	0

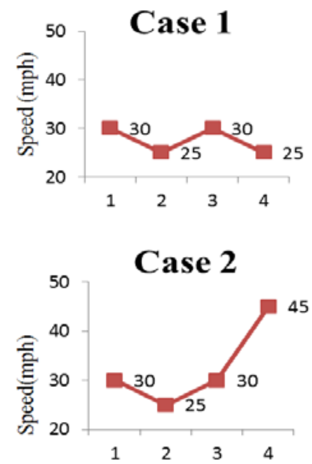


Figure 2: Quantification of expanded vehicular jerk patterns.

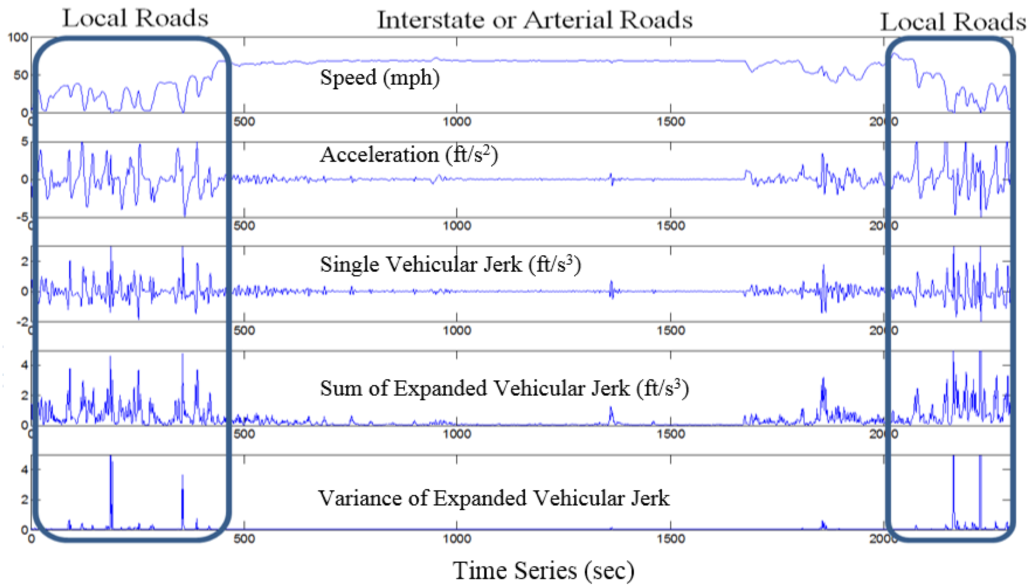


Figure 3: Profiles of speed, acceleration, jerk, sum of vehicular jerk, and jerk variance in one sampled trip.

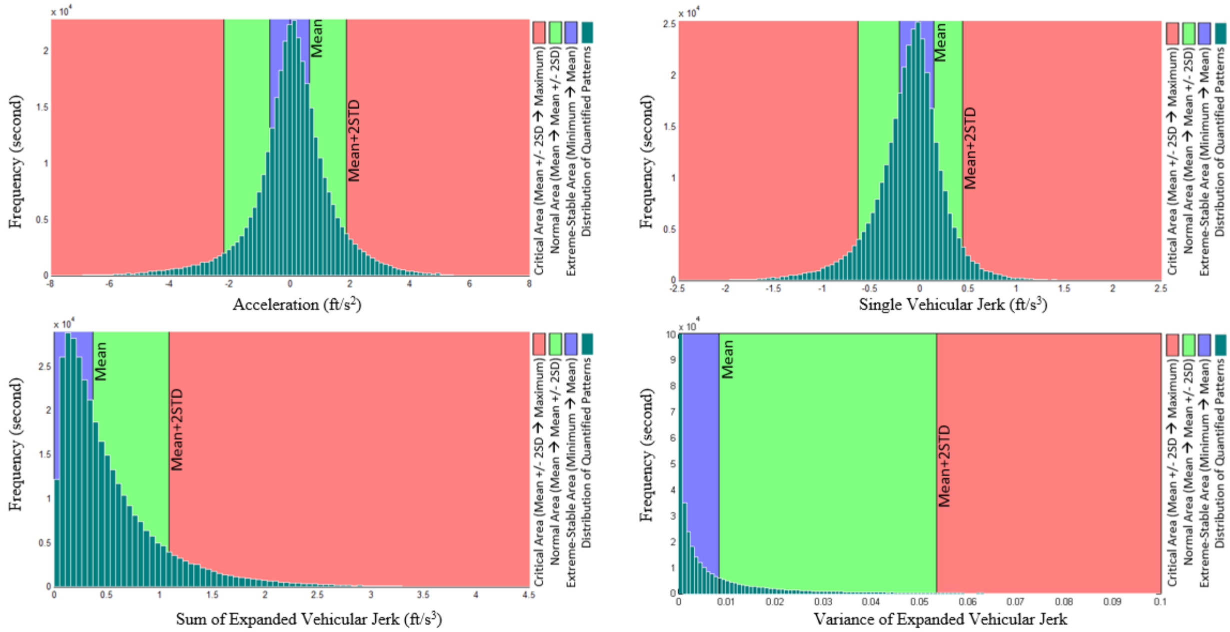


Figure 4: Defining thresholds for 40±0.25 mph speed range.

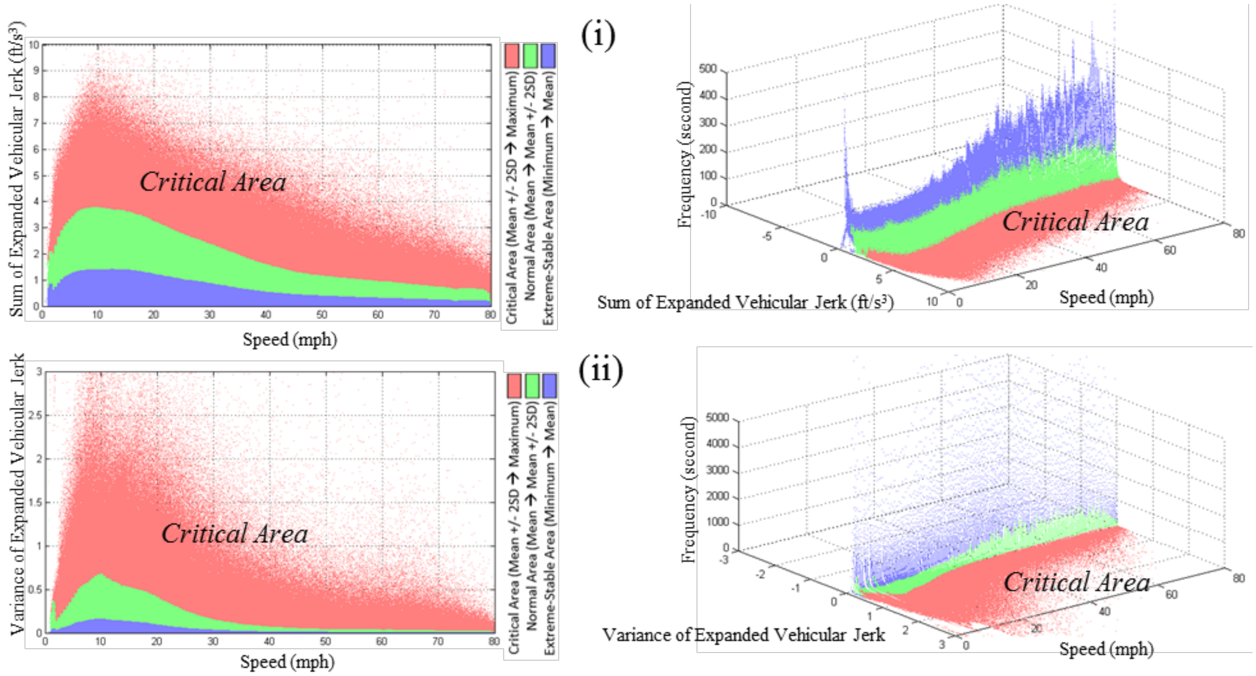


Figure 5: Thresholds for identifying extreme driving decision patterns (n=36 million).

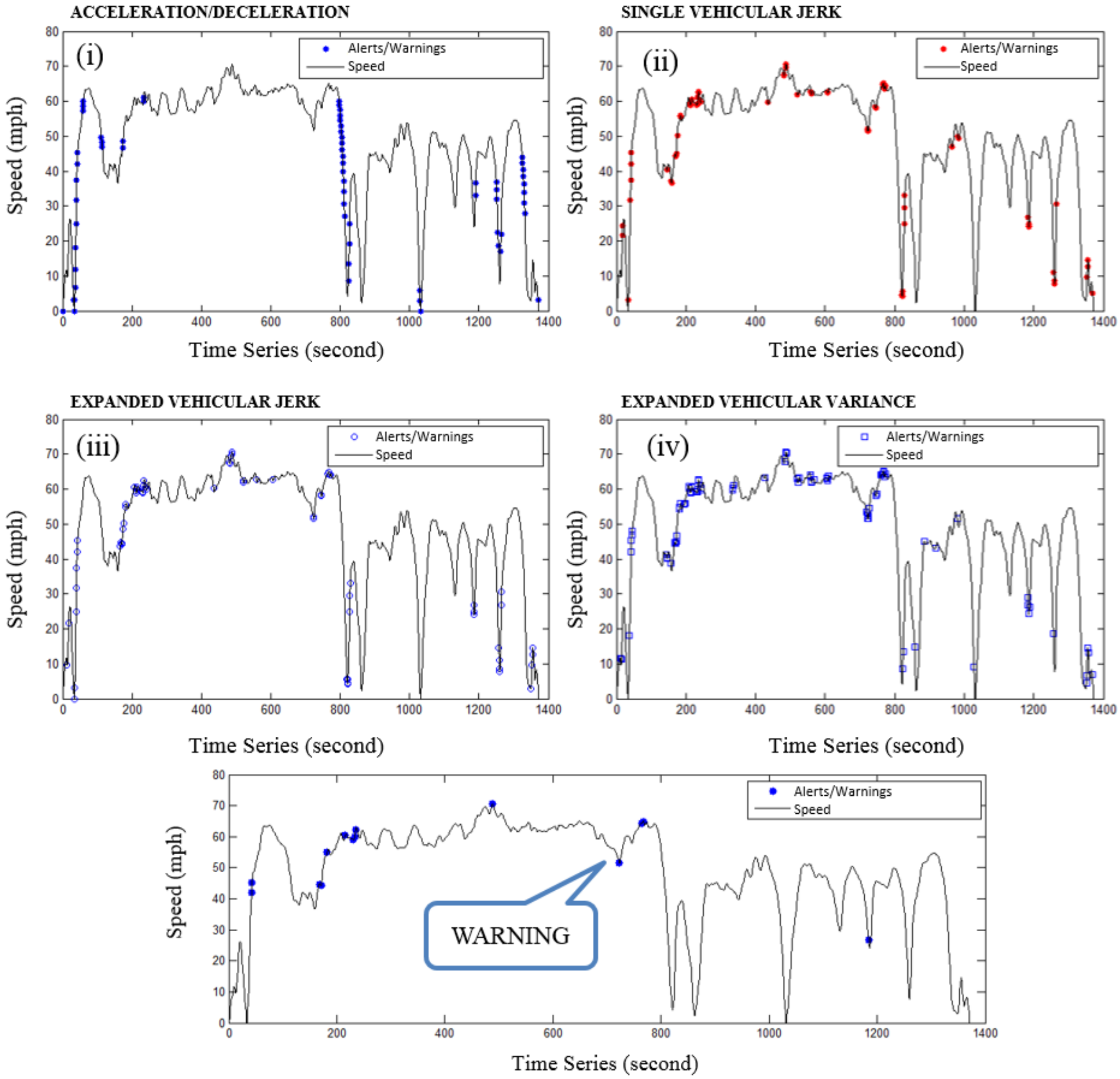


Figure 6: Alerts/warnings given by different methods.

Table 1: Descriptive statistics of variables in the model

Variables			*Obs.*	*Frequency*	*Mean/Percentage*	*Std. Dev.*	*Min*	*Max*
Number of Warnings/Alerts			40201		8.29	17.524	0	1098
Driver-related Factors	Gender [Male]		1486	702	47.24%	0.499	0	1
	Driver Age (years)		1486		47.183	13.319	15	91
Vehicle-related Factors	Body Type	Auto-Sedan	1486	652	43.88%	0.496	0	1
		Two Seated	1486	58	3.90%	0.194	0	1
		Van	1486	131	8.82%	0.284	0	1
		RV	1486	3	0.20%	0.045	0	1
		SUV	1486	409	27.52%	0.447	0	1
		Station Wagon	1486	31	2.09%	0.143	0	1
		Pickup	1486	202	13.59%	0.343	0	1
	Vehicle Age (years)		1486		7.908	5.417	0	50
	Fuel Type	Gasoline	1486	1429	96.16%	0.192	0	1
		Diesel	1486	29	1.95%	0.138	0	1
		Hybrid	1486	19	1.28%	0.112	0	1
		Flex Fuel	1486	9	0.61%	0.078	0	1
Trip-related Factors	Rush Hour*	Non Rush	40201	18654	46.40%	0.499	0	1
		AM Rush	40201	7325	18.22%	0.386	0	1
		Lunch Rush	40201	2946	7.33%	0.261	0	1
		PM Rush	40201	11276	28.05%	0.449	0	1
	Weekend [Yes]		40201	9795	24.37%	0.429	0	1
	Trip Distance (Mile)		40201		7.941	13.184	0.010	431.520
	Trip Duration (Minute)		40201		14.176	14.740	2.010	374.450
	Commute Trip [Yes]		40201	7837	19.49%	0.396	0	1

Note: * Rush hours are AM (6:30 am-10:00 am), Lunch (12:00 pm-1:00 pm) and PM (3:30 pm-7:00 pm).

Table 2: Model selection

Variables			***Model 1***	***Model 2***	***Model 3***	***Model 4***	***Model 5***	***Model 6***	***Model 7***
Driver-related Factors	Gender [Male]		√	√	√	√	√	×	×
	Driver Age (years)		√	√	√	√	√	√	√
Vehicle-related Factors	Body Type	Auto-Sedan	Base	Base	Base	Base	Base	Base	Base
		Two Seated	√	√	√	√	√	√	√
		Van	√	√	√	√	√	√	√
		RV	√	√	√	√	Base	Base	Base
		SUV	√	√	√	√	√	√	√
		Station Wagon	√	√	√	√	√	√	Base
		Pickup	√	√	√	√	√	√	√
	Vehicle Age (years)		√	√	√	√	√	√	√
	Fuel Type	Gasoline	Base	Base	Base	Base	Base	Base	Base
		Diesel	√	√	√	√	√	√	√
		Hybrid	√	√	√	√	√	√	√
		Flex Fuel	√	√	Base	Base	Base	Base	Base
Trip-related Factors	Rush Hour	Non Rush	Base	Base	Base	Base	Base	Base	Base
		AM Rush	√	√	√	√	√	√	√
		Lunch Rush	√	√	√	Base	Base	Base	Base
		PM Rush	√	√	√	√	√	√	√
	Weekend [Yes]		√	√	√	√	√	√	√
	Trip Duration (Minute)		√	×	√	√	√	√	√
	Trip Distance (Mile)		×	√	×	×	×	×	×
	Commute Trip [Yes]		√	√	√	√	√	√	√
Log Likelihood (Model)			-109828.8	-110726.1	-109829.1	-109829.5	-109830.1	-109830.8	-109833
Degrees of Freedom			21	21	20	19	18	17	16
AIC			219699.7	221494.1	219698.2	219697.1	219696.2	219695.7	219697.3
BIC			219880.3	221674.7	219870.3	219860.5	219851.1	219841.9	219834.9

Notes: √ = variables included in the model; × = variables excluded from the model; Base = Base category for unordered categorical variables.

Table 3: Modeling results for frequency of warnings and alerts

Model			***Full Model***			***Final Model***		
Variables			*β*	IRR	P>|z|	*β*	IRR	P>|z|
Constant			1.272	3.567	0.000	1.278	3.590	0.000
Driver-related Factors	Gender [Male]		0.043	1.044	0.244	-		
	Driver Age (years)		-0.008	0.992	0.000	-0.008	0.992	0.000
Vehicle-related Factors	Body Type	Auto-Sedan	Base			Base		
		Two Seated	0.434	1.543	0.000	0.444	1.559	0.000
		Van	-0.411	0.663	0.000	-0.415	0.660	0.000
		RV	-0.440	0.644	0.288	-		
		SUV	-0.119	0.888	0.006	-0.118	0.889	0.007
		Station Wagon	-0.241	0.786	0.056	-0.241	0.786	0.056
		Pickup	-0.220	0.802	0.000	-0.204	0.815	0.000
	Vehicle Age (years)		-0.015	0.985	0.000	-0.015	0.985	0.000
	Fuel Type	Gasoline	Base			Base		
		Diesel	-0.433	0.649	0.001	-0.425	0.654	0.001
		Hybrid	-0.471	0.624	0.004	-0.464	0.629	0.004
		Flex Fuel	0.168	1.183	0.458	-		
Trip-related Factors	Rush Hour	Non Rush	Base			Base		
		AM Rush	0.083	1.087	0.000	0.086	1.090	0.000
		Lunch Rush	-0.021	0.980	0.356	-		
		PM Rush	0.031	1.032	0.019	0.034	1.034	0.010
	Weekend [Yes]		-0.047	0.954	0.001	-0.048	0.953	0.001
	Trip Duration (Minute)		0.061	1.063	0.000	0.061	1.063	0.000
	Commute Trip [Yes]		0.149	1.161	0.000	0.148	1.160	0.000
Alpha			0.866			0.866		
Variance between groups (driver-vehicle pair)			0.408			0.409		
Likelihood-ratio test of MNB vs. NB			8690.46			8795.81		
SUMMARY STATISTICS								
Number of Observations			40201			40201		
Number of Groups			1486			1486		
Log Likelihood (Model)			-109829			-109831		
Degree of Freedom			21			17		
AIC			219699.7			219695.7		
BIC			219880.3			219841.9		
Wald χ^2			14654.7			14648.72		
Prob > χ^2			0.000			0.000		

PART 3: GENERATING EMISSIONS INFORMATION FOR ROUTE SELECTION: EXPERIMENTAL MONITORING AND ROUTES CHARACTERIZATION

Jorge Bandeira[1], Tiago G. Almeida1, Asad J. Khattak[2], Nagui M. Rouphail[3], Margarida C. Coelho[1]

[1] University of Aveiro, [2]University of Tennessee, [3]Institute for Transportation Research and Education

Abstract: Drivers make short-term steering and speed decisions based on incoming information from external sources. To avoid collisions, they may accelerate or brake hard by applying greater than “normal” pressure on the accelerator or brake. Some hard accelerations and braking may be avoidable through driver feedback in real-time. Alerts and warnings can result in calmer driving, and potentially reduce traffic flow disturbances, energy use, emissions, safety problems, and wear and tear on brakes and the engine. This study explores how real-time vehicle trajectory data can be used to generate driver feedback through actionable alerts and warnings. The study provides a novel framework for how acceleration and braking monitoring can generate alerts and warnings, provided through advanced traveler information systems. Extreme driving patterns under seemingly normal conditions are the key to generating actionable personalized feedback. Using data from a behavioral survey coupled with second-by-second speed data from automobile trips, statistical models are estimated to distinguish between groups of individuals who will likely receive a substantial amount of alerts and warnings versus those who will receive none. The implications of the findings and potential applications to fleet vehicles and driving population are discussed.

Keywords: advanced traveler information systems, driving decisions, acceleration, traveler behavior, model selection

[ABSTRACT ONLY]

CONCLUSIONS

This report provides a deeper understanding of the impacts of growth and technology strategies on reducing transportation energy use and emissions. As an effort to enhancing livability and sustainability, a modelling and simulation framework capable of addressing interactions between land use, transportation, and emissions as the foundation for research on sustainable urban development strategies was created.

First, the NHTS behavioral data was used to model whether smart growth land use strategy is associated with reductions in CO_2 emissions. This helped us understand the impacts of smart growth on travel decisions, especially pre-trip decisions. The results showed that smart growth is correlated with lower CO_2 emissions. This study further assessed the behavioral impacts of another smart growth strategy, which relies on comparing Transit Oriented Developments (TODs) with Automobile Oriented Developments (AODs). Travel behavior of people who live in TODs was compared with traditional AOD communities. Results show that TOD residents have a higher propensity to choose the TOD areas for their activity locations, compared with residents in AOD areas. This study also found that chances of sequencing the out-of-home activities near subway stations are higher for residents of TODs. Using macro-level models, smart growth scenarios were developed for the Hampton Roads area (with a population of 1.6 million). The regional travel demand model in TransCAD was used for this purpose. The regional vehicle miles traveled decreased during peak hours in the smart growth scenario compared with the base scenario, which was business as usual.

Second, micro-level behaviors were extensively explored in this study. A large-scale GPS travel survey data (containing 51,371 trips and their associated second-by-second total 36 million seconds) from Atlanta, GA was used to study instantaneous driving decisions, which are essential to accurately predicting energy consumption and tailpipe emissions. This study introduced a new index named “driving volatility” to describe individual driver’s instantaneous decisions during driving, e.g., their choice regarding speed, and acceleration/deceleration. For the sampled trips, on average, 10% of the accelerations and decelerations exceeded the mean plus 2 standard deviations threshold. The correlates of driving volatility were found to vary by gender and age, and also trip attributes, e.g., trip lengths.

The vehicle use behaviors of Alternative Fuel Vehicle owners was explored, using a behavioral database from California. Their use patterns (e.g., trip frequency and daily vehicle miles traveled) and driving practices were analyzed. It was found that AFV drivers make the same amount of trips as conventional vehicle drivers, except that drivers of Battery Electric Vehicles (BEVs) make statistically significantly fewer trips (5% level). The daily distances travelled were shorter for some AFVs (BEV and PHEV) but longer for other AFVs (HEV and CNG) compared with conventional vehicles. Drivers spent significantly longer time traveling daily in their HEV or CNG vehicles compared with conventional vehicles. HEV and BEV were found to be associated with calmer driving compared with conventional vehicles, i.e., they are less prone to aggressive accelerations and vehicular jerks.

Micro-simulations were used to demonstrate eco-friendly infrastructure applications. The associations of individual trip decisions, featured by eco-lanes, on the emissions were explored. An integrated platform for modelling traffic and emissions using the VISSIM model and the Vehicle Specific Power (VSP) methodology was developed. After a rigorous calibration process it was possible to validate the model in terms of speeds, volumes, travel times, and VSP modal distributions. This platform was then used to simulate the inclusion of HOV and eco-lanes in an urban area. Findings for a corridor in a small city show that on a freeway the majority of passengers can reduce their travel time by about 5% accompanied by reductions in total emissions (-3% CO_2, -14% CO,- 8% NO_X). In the urban corridors, reductions of up to 20% in total emissions can be achieved if the average occupancy is in the range of 1.50 passengers/vehicle relative to lower occupancies. Adding an extra eco-lane can have a positive impact on travel time and emissions if the traffic volumes remain similar to the before period. However, adding eco-lanes did not show substantial time savings compared with the no eco-lane scenario. The incorporation of green vehicles in the HOV lanes also did not show a significant impact on the corridor's performance in terms of travel time.

Lastly, potential applications that can reduce energy use and emissions are explored. From a macroscopic perspective, a new regional index was created, based on large-scale data that is becoming available through new technologies. This helps us understand how people drive in a city and across different cities. Specifically, travel behavior in four metropolitan areas including Atlanta, Sacramento, San Francisco, and Los Angeles was analyzed. A comparative analysis of regional driving performance shows significant regional differences in driving time use and of instantaneous driving decisions captured through driving volatility and includes accelerations, decelerations or changing acceleration/deceleration rates for drivers in a regional sample. Importantly, with the new data available from GPS devices, it is possible to create new indices of regional driving performance (similar to congestion index). This study suggests the Regional Acceleration Index and Regional Vehicular Jerk Index as complementary measures to those used in (Schrank et al. 2012). These two indices represent the intensity and variability of instantaneous driving decisions in a region respectively. They can be used to compare the driving patterns as well as precursor behaviors that are likely to be associated with emissions, energy use, and safety across metropolitan areas. Different driving practices were found in these metropolitan areas given their different development density, road network structure, and socio-economics.

Several applications can emerge from the knowledge generated in this study. From a behavioral perspective, drivers can be assisted in their trip making. Monitoring acceleration and braking of drivers can generate alerts and warnings, provided through advanced traveler information systems. Furthermore, a methodology was developed to generate information about emissions for various routes and for drivers to decide if they wish to take the most eco-friendly route, when faced with a choice of routes. GPS equipped-vehicles were used to traverse various paths between origins and destinations to collect second-by-second trajectory data required for micro-scale emission analysis.

Driving volatility information based on accelerations and vehicular jerk can be incorporated in driving assist systems, e.g., advanced traveler information systems. The real-time driving volatility information reflecting driving performance based on patterns of fellow fleet vehicles or neighbors or just on a driver's own performance can support short-term micro level decisions in terms of eco-driving. This in turn can benefit the community or fleets in several ways: 1) calmer driving; 2) safer driving in general, especially on icy or slippery road surfaces where alert thresholds can be lowered; 3) lower fuel consumption and emissions; and 4) identification of dangerous road segments (such as intersections with poor sight distance) that may result in volatile driving.

The study provides an application of eco-friendly traffic assignment, which gives a clear sense of environmental benefits, i.e., in terms of reducing energy use and emissions. Selection of fuel and/or emission-optimal routes and faster intercity routes with less congestion are typically associated with lower energy use and emissions. Changes in drivers' route selection behavior can have a substantial impact on the amount of CO and NOx emissions. Besides the potential to reduce energy by choosing appropriate eco-friendly routes, there is also potential to reduce emissions (e.g., CO more than 50%) based on the drivers' awareness of eco-friendly routes.

The project provides insights that growth and technology strategies offer for significant savings in fuel and emissions: 1) Smart growth strategies of more compact developments that are transit oriented are associated with lower energy use and emissions. Changing the urban form of city through growth policies can result in different activity participation patterns and travel demand distributions; 2) Intelligent Transportation Systems can be helpful in terms of safer driving and lowering energy and emissions. Real-time alerts and warnings have the potential for increasing calmer eco-driving. The knowledge and tools developed in the project will be made available for use by engineers and transportation planners.

APPENDIX-PROJECT PUBLICATIONS AND PRESENTATIONS

Several individuals have participated in writing relevant papers and completing this project. Their names appear on the relevant papers in this report. Publications and Presentations supported by this TranLIVE US DOT UTC grant are listed below:

Project Related Refereed Journal Publications

1. Wang, X., & A. Khattak. Is smart growth associated with reductions in CO_2 emissions? *Transportation Research Record*, 2375, 2013, pp. 62-70.

2. Wang X., A. Khattak, J. Liu, G. Amoli, & S. Son. What is the level of volatility in instantaneous driving decisions? *Transportation Research Part C*, Volume 58, Part B, September 2015, pp. 413-427.

3. Liu J., A. Khattak, X. Wang. The role of alternative fuel vehicles: using behavioral and sensor data to model hierarchies in travel. *Transportation Research Part C: Emerging Technologies*, Issue 55, 2015, 379-392.

4. Bandeira J., T. Almeida, A. Khattak, N. Rouphail, & M. Coelho. Generating emissions information for route selection: Experimental monitoring and routes characterization. *Journal of Intelligent Transportation Systems*, 17:1 Taylor & Francis Publishers, 2013, pp. 3-17.

5. Bandeira, J., D. Carvalho, P. Fernandes, T. Fontes, S. Pereira, N. Rouphail, A. Khattak, M. Coelho. Empirical assessment of route choice impact on emissions over different road types, traffic demands, and driving scenarios. *International Journal of Sustainable Transportation*, Volume 10, Issue 3, 2016, pp. 271-283.

6. Bandeira J., S. Pereira, T. Fontes, P. Fernandes, A. Khattak, M. Coelho. An eco-traffic management tool, computer-based modelling and optimization in transportation. *Advances in Intelligent Systems and Computing*, Volume 262, 2014, pp 41-56.

7. Bandeira, J. M., Fontes, T., Pereira, S. R., Fernandes, P., Khattak, A., & Coelho, M. C. (2014). Assessing the importance of vehicle type for the implementation of eco-routing systems. *Transportation Research Procedia*, 3, 800-809.

Project Related Presentations

1. Wang, X., & A. Khattak, Is smart growth associated with reductions in CO_2 emissions? Presented at 2013 Transportation Research Board Annual meeting, Washington, D.C.

2. Wang X., A. Khattak, J. Liu, G. Amoli, S. Son. What is the level of volatility in instantaneous driving decisions? Presented at the Transportation Research Board 2014 annual meeting (TRB Paper 14-2780); Presented as plenary Session Invited Talk, 13th COTA International Conference of Transportation Professionals, CICTP, Shenzhen, China, August 2013; presented at Hong Kong University of Science and Technology, and at PacTrans Seminar Series at University of Washington.

3. Liu J., X. Wang, A. Khattak. Providing real-time driving volatility information, presented at ITS World Congress 2014 conference (Detroit, Sept 7-11).

4. Khattak A. & J. Liu. Supporting calmer instantaneous driving decisions: use of vehicle trajectory data to generate alerts and warnings, TRB paper # 15-1345, presented at 2015 Transportation Research Board Annual meeting, Washington, D.C.

5. Wang X, J. Liu, and A. Khattak. Generating fuel economy information to support cost effective vehicle choices: comparing standard and customized driving cycles, TRB paper # 15-4548, presented at 2015 Transportation Research Board Annual meeting, Washington, D.C.

6. Liu J., A. Khattak, & L. Han. What is the magnitude of information loss when sampling driving performance data? TRB paper # 15-0968, presented at 2015 Transportation Research Board Annual meeting, Washington, D.C.

7. Liu J., A. Khattak & X. Wang. A comparative study of driving performance in metropolitan regions using large-scale vehicle trajectory, TRB paper # 15-0966, presented at 2015 Transportation Research Board Annual meeting, Washington, D.C.

8. Son S., A. Khattak & K. Choi. Comparing travel behavior between transit-oriented developments and automobile-oriented developments: matched pair analysis, TRB Paper 14-2327, presented at the 2014 Transportation Research Board Annual meeting, Washington DC.

9. Bandeira J., P. Fernandes, T. Fontes, S. Pereira. A. Khattak, & M. Coelho. Exploring multiple eco-routing guidance strategies in a commuting corridor, presented at 2016 Transportation Research Board Annual meeting, Washington, D.C.