



University Transportation Research Center - Region 2

Final Report



Feasibility Assessment for Battery Electric Vehicles based on Multi-Day Activity-Travel Patterns

Performing Organization: State University of New York (SUNY)



April 2017



Sponsor:
University Transportation Research Center - Region 2

University Transportation Research Center - Region 2

The Region 2 University Transportation Research Center (UTRC) is one of ten original University Transportation Centers established in 1987 by the U.S. Congress. These Centers were established with the recognition that transportation plays a key role in the nation's economy and the quality of life of its citizens. University faculty members provide a critical link in resolving our national and regional transportation problems while training the professionals who address our transportation systems and their customers on a daily basis.

The UTRC was established in order to support research, education and the transfer of technology in the field of transportation. The theme of the Center is "Planning and Managing Regional Transportation Systems in a Changing World." Presently, under the direction of Dr. Camille Kamga, the UTRC represents USDOT Region II, including New York, New Jersey, Puerto Rico and the U.S. Virgin Islands. Functioning as a consortium of twelve major Universities throughout the region, UTRC is located at the CUNY Institute for Transportation Systems at The City College of New York, the lead institution of the consortium. The Center, through its consortium, an Agency-Industry Council and its Director and Staff, supports research, education, and technology transfer under its theme. UTRC's three main goals are:

Research

The research program objectives are (1) to develop a theme based transportation research program that is responsive to the needs of regional transportation organizations and stakeholders, and (2) to conduct that program in cooperation with the partners. The program includes both studies that are identified with research partners of projects targeted to the theme, and targeted, short-term projects. The program develops competitive proposals, which are evaluated to insure the most responsive UTRC team conducts the work. The research program is responsive to the UTRC theme: "Planning and Managing Regional Transportation Systems in a Changing World." The complex transportation system of transit and infrastructure, and the rapidly changing environment impacts the nation's largest city and metropolitan area. The New York/New Jersey Metropolitan has over 19 million people, 600,000 businesses and 9 million workers. The Region's intermodal and multimodal systems must serve all customers and stakeholders within the region and globally. Under the current grant, the new research projects and the ongoing research projects concentrate the program efforts on the categories of Transportation Systems Performance and Information Infrastructure to provide needed services to the New Jersey Department of Transportation, New York City Department of Transportation, New York Metropolitan Transportation Council, New York State Department of Transportation, and the New York State Energy and Research Development Authority and others, all while enhancing the center's theme.

Education and Workforce Development

The modern professional must combine the technical skills of engineering and planning with knowledge of economics, environmental science, management, finance, and law as well as negotiation skills, psychology and sociology. And, she/he must be computer literate, wired to the web, and knowledgeable about advances in information technology. UTRC's education and training efforts provide a multidisciplinary program of course work and experiential learning to train students and provide advanced training or retraining of practitioners to plan and manage regional transportation systems. UTRC must meet the need to educate the undergraduate and graduate student with a foundation of transportation fundamentals that allows for solving complex problems in a world much more dynamic than even a decade ago. Simultaneously, the demand for continuing education is growing – either because of professional license requirements or because the workplace demands it – and provides the opportunity to combine State of Practice education with tailored ways of delivering content.

Technology Transfer

UTRC's Technology Transfer Program goes beyond what might be considered "traditional" technology transfer activities. Its main objectives are (1) to increase the awareness and level of information concerning transportation issues facing Region 2; (2) to improve the knowledge base and approach to problem solving of the region's transportation workforce, from those operating the systems to those at the most senior level of managing the system; and by doing so, to improve the overall professional capability of the transportation workforce; (3) to stimulate discussion and debate concerning the integration of new technologies into our culture, our work and our transportation systems; (4) to provide the more traditional but extremely important job of disseminating research and project reports, studies, analysis and use of tools to the education, research and practicing community both nationally and internationally; and (5) to provide unbiased information and testimony to decision-makers concerning regional transportation issues consistent with the UTRC theme.

Project No(s):

UTRC/RF Grant No: 49198-22-27

Project Date: April 2017

Project Title: Market Potential for Battery Electric Vehicles based on Multi-Day Activity-Travel Patterns

Project's Website:

<http://www.utrc2.org/research/projects/market-potential-battery>

Principal Investigator(s):

Jee Eun Kang

Assistant Professor
Department of Industrial and Systems Engineering
University at Buffalo
Buffalo, NY 14228
Tel: (716) 645-2357
Email: jeeunka@buffalo.edu

Anpeng Zhang

Department of Industrial and Systems Engineering
University at Buffalo
Buffalo, NY 14228

Changhyun Kwon

Department of Industrial and Management Systems Engineering
University of South Florida
Tampa, FL 33620
Tel: (813) 974-5588
Email: chkwon@usf.edu

Performing Organization(s):

State University of New York (SUNY)

Sponsor(s):

University Transportation Research Center (UTRC)

To request a hard copy of our final reports, please send us an email at utrc@utrc2.org

Mailing Address:

University Transportation Research Center
The City College of New York
Marshak Hall, Suite 910
160 Convent Avenue
New York, NY 10031
Tel: 212-650-8051
Fax: 212-650-8374
Web: www.utrc2.org

Board of Directors

The UTRC Board of Directors consists of one or two members from each Consortium school (each school receives two votes regardless of the number of representatives on the board). The Center Director is an ex-officio member of the Board and The Center management team serves as staff to the Board.

City University of New York

Dr. Hongmian Gong - Geography/Hunter College
Dr. Neville A. Parker - Civil Engineering/CCNY

Clarkson University

Dr. Kerop D. Janoyan - Civil Engineering

Columbia University

Dr. Raimondo Betti - Civil Engineering
Dr. Elliott Sclar - Urban and Regional Planning

Cornell University

Dr. Huaizhu (Oliver) Gao - Civil Engineering

Hofstra University

Dr. Jean-Paul Rodrigue - Global Studies and Geography

Manhattan College

Dr. Anirban De - Civil & Environmental Engineering
Dr. Matthew Volovski - Civil & Environmental Engineering

New Jersey Institute of Technology

Dr. Steven I-Jy Chien - Civil Engineering
Dr. Joyoung Lee - Civil & Environmental Engineering

New York University

Dr. Mitchell L. Moss - Urban Policy and Planning
Dr. Rae Zimmerman - Planning and Public Administration

Polytechnic Institute of NYU

Dr. Kaan Ozbay - Civil Engineering
Dr. John C. Falcocchio - Civil Engineering
Dr. Elena Prassas - Civil Engineering

Rensselaer Polytechnic Institute

Dr. José Holguín-Veras - Civil Engineering
Dr. William "Al" Wallace - Systems Engineering

Rochester Institute of Technology

Dr. James Winebrake - Science, Technology and Society/Public Policy
Dr. J. Scott Hawker - Software Engineering

Rowan University

Dr. Yusuf Mehta - Civil Engineering
Dr. Beena Sukumaran - Civil Engineering

State University of New York

Michael M. Fancher - Nanoscience
Dr. Catherine T. Lawson - City & Regional Planning
Dr. Adel W. Sadek - Transportation Systems Engineering
Dr. Shmuel Yahalom - Economics

Stevens Institute of Technology

Dr. Sophia Hassiotis - Civil Engineering
Dr. Thomas H. Wakeman III - Civil Engineering

Syracuse University

Dr. Riyad S. Aboutaha - Civil Engineering
Dr. O. Sam Salem - Construction Engineering and Management

The College of New Jersey

Dr. Thomas M. Brennan Jr - Civil Engineering

University of Puerto Rico - Mayagüez

Dr. Ismael Pagán-Trinidad - Civil Engineering
Dr. Didier M. Valdés-Díaz - Civil Engineering

UTRC Consortium Universities

The following universities/colleges are members of the UTRC consortium.

City University of New York (CUNY)
Clarkson University (Clarkson)
Columbia University (Columbia)
Cornell University (Cornell)
Hofstra University (Hofstra)
Manhattan College (MC)
New Jersey Institute of Technology (NJIT)
New York Institute of Technology (NYIT)
New York University (NYU)
Rensselaer Polytechnic Institute (RPI)
Rochester Institute of Technology (RIT)
Rowan University (Rowan)
State University of New York (SUNY)
Stevens Institute of Technology (Stevens)
Syracuse University (SU)
The College of New Jersey (TCNJ)
University of Puerto Rico - Mayagüez (UPRM)

UTRC Key Staff

Dr. Camille Kamga: *Director, Assistant Professor of Civil Engineering*

Dr. Robert E. Paaswell: *Director Emeritus of UTRC and Distinguished Professor of Civil Engineering, The City College of New York*

Herbert Levinson: *UTRC Icon Mentor, Transportation Consultant and Professor Emeritus of Transportation*

Dr. Ellen Thorson: *Senior Research Fellow, University Transportation Research Center*

Penny Eickemeyer: *Associate Director for Research, UTRC*

Dr. Alison Conway: *Associate Director for Education*

Nadia Aslam: *Assistant Director for Technology Transfer*

Nathalie Martinez: *Research Associate/Budget Analyst*

Tierra Fisher: *Office Assistant*

Bahman Moghimi: *Research Assistant; Ph.D. Student, Transportation Program*

Wei Hao: *Research Fellow*

Andriy Blagay: *Graphic Intern*

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. The contents do not necessarily reflect the official views or policies of the UTRC or the Federal Highway Administration. This report does not constitute a standard, specification or regulation. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Feasibility Assessment for Battery Electric Vehicles based on Multi-Day Activity-Travel Patterns		5. Report Date April 11, 2017	6. Performing Organization Code
7. Author(s) Anpeng Zhang, Jee Eun Kang, Changhyun Kwon		8. Performing Organization Report No.	
9. Performing Organization Name and Address University at Buffalo, the State University of New York Buffalo, NY 14260		10. Work Unit No.	11. Contract or Grant No. 49198-22-27
12. Sponsoring Agency Name and Address UTRC The City College of New York 137th Street and Convent Avenue New York, NY 10031		13. Type of Report and Period Covered Final Report	
14. Sponsoring Agency Code		15. Supplementary Notes	
16. Abstract A Battery Electric Vehicle (BEV) feasibility considering State Of Charge (SOC) level is assessed using multiday activity-travel patterns to overcome the limitations of using one-day activity-travel patterns. Since multi-day activity-travel patterns are not readily available, we generate multi-day activity-travel patterns through sampling from readily available single-day household travel survey data with considerations of day-to-day intrapersonal variability. One of the key observation we make is that the distribution of interpersonal variability in single-day travel activity datasets is similar to the distribution of intrapersonal variability in multi-day datasets. Thus, interpersonal variability observed in cross-sectional single-day data of a large population can be used to generate the day-to-day intrapersonal variability. The proposed sampling method is based on activity-travel pattern type clustering, travel distance and variability distribution to extract such information from single-day data. Validation and stability tests of the proposed sampling methods are presented. BEV feasibility assessment results show that our sampling method combined with trivial method provide better estimation of population wide BEV feasibility than using cross-sectional data only.			
17. Key Words Activity-Travel Patterns, Day-to-day variability, Battery Electric Vehicles Feasibility Assessment		18. Distribution Statement	
19. Security Classif (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No of Pages 36	22. Price

Acknowledgement

Authors would like to thank Dr. Kay Axhausen (Swiss Federal Institute of Technology) for providing Mobidrive data set that enabled validation of the proposed method. Authors would also like to thank Dr. Joan Lucas (SUNY Brockport) for her support and Dr. Will Recker (UC Irvine) for insightful comments and discussions on this work.

Feasibility Assessment for Battery Electric Vehicles based on Multi-Day Activity-Travel Patterns

Anpeng Zhang[†], Jee Eun Kang^{*†}, and Changhyun Kwon[‡]

[†]Department of Industrial and Systems Engineering, University at Buffalo, SUNY

[‡]Department of Industrial and Management Systems Engineering, University of South Florida

Contents

1	Executive Summary	3
2	Background and Objectives	4
3	Introduction and Literature Review	5
4	Summary of Work Performed	10
4.1	Proposed Variability Measurement	10
4.1.1	Data Description	10
4.1.2	Activity-Travel Pattern Sequence	10
4.1.3	Measures of variability	12
4.2	Connection between multi-day intrapersonal variability and one-day population-wide interpersonal variability	14
4.3	Sampling Procedure	18
4.3.1	Clustering	19
4.3.2	Transition probability	20
4.3.3	Sampling method process	21
4.3.4	Summary of sampling method	21
4.4	Validation of MIV variability generated from sampling	22
4.4.1	Intrapersonal variability distribution	22
4.4.2	MIV error (MIVE) for personal multi-day activity-travel pattern	23
4.4.3	Stability of multi-day sampling method	25
4.5	Multi-Day Battery Electric Vehicle (BEV) Feasibility Assessment	26

*Corresponding author; jeeeunka@buffalo.edu

List of Figures

1	Multi-day travel distance variability	14
2	Adjusted PIV and MIV	15
3	Standard activity-travel pattern of Person No.23	16
4	Samples of person with low MIV	17
5	Samples of person with medium MIV	18
6	Samples of person with high MIV	18
7	Activity travel pattern visualization for all clusters	19
8	MIV results of the sample	23
9	MIV error validation	25
10	Multiple sample comparison results	26
11	Example SOC curves based on different samples	27
12	SOC error validation	28
13	SOC positive rate with work charging only	28

List of Tables

1	Results of various measure of day-to-day intrapersonal variability in previous studies	7
2	Original data sample	11
3	Uni-Dimensional Sequence Representation of Activity-Travel Patterns	12
4	Transition probability	20

1 Executive Summary

For Battery Electric Vehicles (BEVs), the biggest challenge is whether a traveler’s intended activity-travel patterns can be served by a BEV, since BEVs have a shorter range and longer charging times. Although the coming generation of luxury BEVs as well as newest Tesla models should have ranges much higher than 100 miles, most of the BEVs on the market can drive around 60 miles to 100 miles with full battery. In order to assess feasibility of using a BEV, State Of Charge (SOC) level is derived using activity-travel patterns with assumed charging opportunities and behaviors. However, often times, only one-day activity-travel patterns are available since multi-day activity-travel pattern data are not readily available. That is, day-to-day variations of travelers activity-travel patterns are ignored and the BEV feasibility assessment inherently is over-estimated as travelers will consider purchasing a BEV only when all of their activity-travel patterns can be met by BEVs limited range.

To overcome the limitations of using only one-day activity travel patterns, we generate multi-day activity-travel patterns through sampling from readily available single-day household travel survey data with considerations of day-to-day intrapersonal variability. One of the key observation we make is that the distribution of interpersonal variability in single-day travel activity datasets is similar to the distribution of intrapersonal variability in multi-day datasets. Thus, interpersonal variability observed in cross-sectional single-day data of a large population can be used to generate the day-to-day intrapersonal variability. The proposed sampling method is based on activity-travel pattern type clustering, travel distance and variability distribution to extract such information from single-day data. Validation and stability tests of the proposed sampling methods are presented.

BEV feasibility assessment results show that our sampling method combined with trivial method provide better estimation of population wide BEV feasibility than using cross-sectional data only. The proposed method assesses that around 40/50 person have over 80% SOC positive rate for both original data and our sample data, meaning that BEV will be able to serve most of their trips while the trivial method of using only one-day activity-travel patterns will show more difference and tend to overestimate the availability of BEV for travelers. While around 10/50 person have less SOC positive rate indicating that they are less appropriate for BEVs, our method tends to overestimate the availability for this group of people and trivial method tends to underestimate them. Given the results, we can use both the trivial method of one-day activity travel patterns and the proposed method of generating multi-day activity-travel patterns as the upper bound and lower bound to better estimate the BEV availability rate. Thus, we can have more accurate results even if we only have single-day cross-sectional data by extracting more day-to-day intrapersonal variability information with our sampling method.

2 Background and Objectives

In the last few years we have witnessed the success of Alternative Fuel Vehicles (AFVs) in the automobile market, mainly Hybrid Electric Vehicles (HEVs). This is a phenomenon that has no precedent in the transportation system. The domination of Internal Combustion Engine Vehicles (ICEVs) is beginning to be challenged by more sustainable types of vehicles. With the governments legislative efforts, technical maturity, public acceptance, and success in the competitive market, achieving greener transportation has never seemed more promising. The next generation of electric vehicles, Plug-in Hybrid Electric Vehicles (PHEVs) and Battery Electric Vehicles (BEVs) will deliver even greater reduction of greenhouse gas emissions through the plug-in charging capability, drawing more energy from electricity. In particular, BEVs fully convert gasoline-based mobility to electricity-based mobility. BEVs are Zero Emission Vehicles (ZEVs) and are considered as one of the most environmentally friendly AFVs.

The challenges associated in adopting BEVs are fundamentally different from those of other AFVs. For HEVs, the main challenges were public acceptance and technical feasibility as they can be refueled at existing gasoline stations for ICEV-comparable ranges. For Battery Electric Vehicles (BEVs), the biggest challenge is whether a traveler's intended activity-travel patterns can be served by a BEV, since BEVs have a shorter range and longer charging times. Although the coming generation of luxury BEVs as well as newest Tesla models should have ranges much higher than 100 miles, most of the BEVs on the market can drive around 60 miles to 100 miles with full battery. Thus, a traveler with travel needs over 60 miles will hardly be served by a BEV without within-day charging. But the traveler may be able to achieve their intended travel if charging opportunities are sufficiently available within the time frame of the travel activities. Due to these reasons, an individual traveler's activity-travel patterns serve as a crucial input for assessing the feasibility, market potential, and promotion of BEVs.

Full-day activity-travel patterns allow us to identify potential vehicle usage profiles, which would include vehicle operations and battery status under different scenarios with varying charging opportunities and travel/charging behaviors of the travelers. Full-day activity-travel patterns allow us to create potential vehicle usage profiles, which would include vehicle operations and battery status under different scenarios with varying charging opportunities based on travel needs and charging availability/behaviors. Previous studies based on activity-travel patterns have focused on the potential impact on energy consumption, emissions profiles, and potential changes of operating PHEVs and BEVs (Axsen et al., 2011; Dong and Lin, 2012; Gonder et al., 2007; Kang and Recker, 2009; Zhang et al., 2011; Dong et al., 2014; Kang and Recker, 2014; Khayati and Kang, 2015). In these works, scenario analysis assumed varying charging availability and charging behaviors. These analyses generated temporal vehicle operations and charging profiles that served to understand the electricity demand and emissions. Since PHEVs do not have range limitations, studies showed the change in energy consumption and electricity charging demand without altering the activity-travel patterns. When the same approach is applied for travel range limited BEVs, travel behavioral assumptions are made to deal with trips when battery is depleted. Dong et al. (2014) assumed

that a traveler would miss out on activities and travels if the battery is depleted (Dong et al., 2014). Kang and Recker (2014) assumed that 1) delay would occur for the duration of time that is sufficient to make the next trip possible, or 2) travelers would be willing reschedule their intended activity-travel patterns within their scheduling flexibility (Kang and Recker, 2014). Khayati and Kang (2015) assumed that intra-household interactions of vehicle allocation and activity allocation would occur to compensate for limited range of BEVs as well as to utilize lower operating cost of BEVs (Khayati and Kang, 2015).

While it is possible to identify travelers that can be served by BEVs and assess feasibility of BEVs using full-day activity-travel patterns, such data is available for only one travel day through national or state Household Travel Survey. That is, day-to-day variations of travelers activity-travel patterns are ignored and the BEV feasibility assessment inherently is over-estimated as travelers will consider purchasing a BEV only when all of their activity-travel patterns can be met by BEVs limited range. The limitations of relying on single-day data from longitudinal activity-travel pattern data for assessing the potential adoption of PHEVs or BEVs have been argued. Despite the commonly accepted claim based on single-day survey data that 78% of the commuters will be satisfied with a 40 mile range BEV, only 9% of vehicles did not exceed daily travel distance less than 100 miles for one year period (Pearre et al., 2011). Another study states that a BEV with a 100 mile range will be sufficient for 50% of one-vehicle households considering multi-day (Khan and Kockelman, 2012). Other studies also highlighted the need for using multi-day data (Dong et al., 2014; Smith et al., 2011). The commonly accepted claim based on one-day survey data argues that 78% of the commuters will be satisfied with a 40 mile range BEV. However, based on one-year data, only 9% of vehicles did not exceed daily travel distance less than 100 miles which is well beyond the BEV range (Pearre et al., 2011).

The objective of this project is to assess feasibility of BEVs based on multi-day activity-travel patterns, incorporating day-to-day variations in activity-travel patterns. Data collection is an expensive task and data sets with multiday activity-travel patterns are rarely available. We propose to define a methodology to generate multi-day activity-travel patterns via sampling from readily available sets of one-day activity-travel patterns, which encompasses likely day-to-day variations of each traveler. Based on the multi-day activity-travel patterns generated by this sampling methodology, we will analyze whether all intended activity-travel patterns can be served by a BEV within the acceptable tolerance of delay. These results will provide a more realistic perspective on the adoptability of BEVs compared to analysis based on one travel day.

3 Introduction and Literature Review

Intrapersonal variability, also known as day-to-day variations, of activity-travel patterns are found to show strong repetitions, yet with considerable variations (Hanson and Huff, 1981, 1988; Pas and Sundar, 1995; Pendyala and Pas, 2000). Observations of day-to-day variations of activity-travel patterns have been studied to understand activity-travel behavior of adaptation, habit, and

symmetry. Both stability and variability have been observed at intrapersonal levels as well as at both spatial and temporal levels (Buliung et al., 2008; Koppelman and Pas, 1984; Pendyala et al., 2001; Pas and Koppelman, 1986; Pas and Sundar, 1995; Susilo and Axhausen, 2014). Variations of travel behavior have also been explained by day-of-week factors.

In previous studies, Pas and Koppelman (1986) utilized daily trip generation rates to measure the intrapersonal variability. According to their observation, employment status, household role, social class and daily travel resource could all affect intrapersonal variability, thus different population groups are likely to have huge differences in day-to-day travel activity. Later, Pas (1988) categorized activity-travel patterns into five types with cluster analysis and calculated the probability of selecting each pattern type for day-of-week. They mentioned that day-of-the-weeks differences are highly related to sociodemographic characteristics while day-of-week would not affect weekday travel behavior for workers. Then, by including trip chaining and daily travel time, Pas and Sundar (1995) extended the trip generation rate day-to-day variation analysis with similar formulations of the total sum of squares in travel behavior. Their results indicated that intrapersonal variability could vary according to different sample data, but it significantly affects the total variability in day-to-day travel behavior of individuals. Recently, Elango et al. (2008) introduced delta trips as the measurement of day-to-day trip making variability. Their experiment results show that day-to-day intrapersonal variability based on household trip number is greatly affected by demographic variables, including income, person number and etc without considering seasonal affects. We provide a summary of variability measurement in previous studies in the following Table1. In this table, we summarize the dataset used in the paper, the standard of intrapersonal variability considered as well as the major conclusion of the paper. We also show if the paper concludes that intrapersonal variability occupies a large proportion of the total individual travel variability. Although diverse measurements are used in previous work and distinct numerical results are presented, intrapersonal variability has been proved to be closely related to the variation of people's travel activity patterns.

Table 1: Results of various measure of day-to-day intrapersonal variability in previous studies

Paper	Data Source	Intra-var standard	Large proportion of total variance	Major findings
Pas and Koppelman (1986)	Reading, England (1973)	daily trip frequency	Yes	sociodemographic groups dependent
Pas (1988)	Employed people in Reading, England (1973)	day of week, daily pattern type	Yes	day-of-week independent for daily pattern, sociodemographic characteristic important for weekly pattern
Pas and Sundar (1995)	North King County, Washington (1989)	trip frequency, daily travel time, trip chaining	Yes	similar magnitude under different standards
Schlich and Axhausen (2003)	Mobidrive in Germany (1999)	trip based method, time budgets based method	Not mentioned	trip-based better, complexity needs consideration, additional proofs for previous works
Susilo and Kitamura (2005)	Mobidrive in Germany (1999)	action space(the second moment of the out-of-home activity locations)	No	workers and students more stable
Elango et al (2008)	Commute Atlanta study (2004)	delta trips	Not mentioned	significant demographic variables effects, day-of-week effects, less seasonal effects
Buliung et al (2008)	Toronto Activity Survey	Travel Panel Survey minimum convex polygon metric(smallest convex polygon containing all activity locations within a respondents activity-travel pattern)	Not mentioned	spatial variety in travel behaviors, typically conduct activities at repeated locations, human spatial behavior sensitive to temporal scale of analysis

Stopher et al. (2008)	28-day GPS survey in Sydney	daily travel distance, daily number of trips, average travel time per trip, etc	Yes	1 or 2 days of data overestimate variances, stable at 18 or 19 days
Kang and Scott (2010)	Toronto Activity Panel Survey with GIS toolkit (2003)	Travel-activity time-use patterns	Yes for weekdays, No for weekends	joint activities(interact with other household members) have higher proportion of intra-var
Chikaraishi et al. (2011)	German Mobility Panel survey data (1999-2008)	travel time expenditure	Yes	situational attributes dependent, longer observation time is important
Susilo and Axhausen (2014)	Mobidrive in Germany many (1999)	repetitiveness of activity travel patterns	Not mentioned	individual's behavioral choices dependent, activity type affects reputation pattern
Moiseeva et al. (2014)	GPS data in Eindhoven, Netherlands (2010)	activity travel pattern	No	sociodemographic characteristics dependent, intra-variability reflect environment learning speed

Despite the evidence, these day-to-day intrapersonal variabilities are often ignored in studies analyzing activity-travel behaviors and estimating travel demand due to the unavailability of data. National-, State-, and Regional-Level Household Travel Surveys collect detailed information of activity-travel along with household socio-demographics. Governments, industries, and researchers rely on these data sets for travel forecasting, planning, traffic management, etc. These surveys, generally include only one weekday activity-travel information. Recently, with various types of IT technologies, collecting multi-day data has become more readily available and affordable. These data sets greatly enable us to understand day-to-day intrapersonal variability of certain travel choices. However, these data are often passively collected and therefore miss information such as travel/activity purpose, specific travel modes (carpooling, specific services used), cost of travel, accompanying passengers, etc.

Cross-sectional data that contains detailed information of various aspects of activity-travel decisions and interpersonal variability, can be used to generate intrapersonal variability thanks to a large data set. The assumption is that single-day cross-sectional data contains a diverse set of activity-travel patterns that is sufficient to be used as surrogate for multi-day activity-travel patterns. For instance, suppose several people travels for work only on some days and travels for work and shopping on other days. Then suppose we collect the single-day travel activity pattern data of these people, we are likely to observe some days of work only activity travel patterns and some other days of work and shopping activity travel patterns, and the ratio between work only pattern and work and shopping pattern could be similar to the ratio of work days as well as work and shopping days of the whole population in average. In other words, multi observations over a large collection of presumed homogeneous observations can be used as a surrogate for repeated observations over a single individual.

Cross-sectional data is widely used in the previous studies related to travel behavior due to the limited access of longitudinal data resources. Since cross-sectional data contains little time-related information and the various measurements of intrapersonal variability are applied, the results based on those data ranges from 20% to 80% (Chen et al., 2016). We need a good measurement of intrapersonal variability in order to indicate the social characteristic and travel activity pattern of a person more accurately.

Based on the above consideration, we propose a sampling procedure based on transition probabilities between clustered activity-travel pattern types of day-to-day basis. The distribution of day-to-day intrapersonal variability and travel distance are also considered to provide more accurate samples. We show that using the proposed sampling, individual-level day-to-day person-to-person match creates the the variability distribution close to intrapersonal multi-day variability for a five-day period. In section 2, we introduce the measurements of variability used in this paper. In section 3, we explain the similarity between interpersonal variability and intrapersonal variability. In section 4, we present our sampling method as well as the step-to-step procedure. In section 5, we discuss the validation method of our defined day-to-day intrapersonal variability. In section 6, we test a case study of Battery Electric Vehicle (BEV) feasibility assessment and market potential

based on state of charge (SOC).

4 Summary of Work Performed

First, we analyze day-to-day variability of general activity-travel patterns. In Section 4.1, we define measures that we use to describe day-to-day variability. In Section ref:connection, we show the connection between day-to-day variability measure and one-day interpersonal variability. Based on this understanding, we propose a sampling method in Section 4.3, and validate the performance of the proposed sampling method in Section 4.4.

Then, we apply the proposed sampling method for the BEV feasibility assessment using multiday activity-patterns in Section 4.5.

4.1 Proposed Variability Measurement

In this section, we introduce the definitions of important concepts in our research, including activity-travel pattern sequence as well as measurements of variabilities.

4.1.1 Data Description

We use the data from the project Mobidrive which is funded by the German ministry of Research and Education, which has been used in several previous studies (Schlich and Axhausen, 2003; Susilo and Kitamura, 2005; Susilo and Axhausen, 2014). They developed a continuous travel diary of six weeks, hoping to find the behavior pattern of the respondents. The travel diary survey was conducted in two German cities of Karlsruhe and Halle with about 300 thousand inhabitants in the fall of 1999. The main study includes information from 317 persons over 6 years of age in 139 households. More details on the project Mobidrive and its six-week continuous survey are present in (Axhausen et al., 2002). In order to exclude a great level of sociodemographic influences, we focus on the population group of employed people traveling with vehicles only. Considering the potential difference between weekdays and weekends, we only choose 5 weekday data for each available person. Based on these considerations, we chose a sample of 50 travelers, and each person's data contains 5-day travel information. The sample data contains 927 daily trips, including 353 trips going home, 166 trips going to work, 87 trips for leisure and 96 trips for shopping. The average daily travel distance is 13.79 km, and the average number of trips traveled on each day is 3.708. The detailed travel attributes can be found in later sections as an example table.

4.1.2 Activity-Travel Pattern Sequence

An activity-travel pattern is a complex output of activity-travel decisions that contains the following information: activity decisions (e.g. activity type, durations, etc.), travel decisions (e.g. travel times, mode, accompanying persons, distances, etc.), and interacting activity/travel decisions (e.g. departure time, activity start times, locations, etc.). Several categories of measurements

have been used to represent these complex patterns: vector of descriptive attributes, stop-based measurement, trip-link measurements, Herfindahl-Hirschman Index, and uni/multi-dimensional sequence representation as a time-space path (Allahviranloo et al., 2014; Hanson and Huff, 1981; Joh et al., 2001; Pas, 1988; Recker et al., 1985; Susilo and Axhausen, 2014; Wilson, 1998).

We will use a uni-dimensional activity-travel sequence as the basic representation of the data. Sequence analysis has been widely used in various fields to understand features, functions, structures, or evolution. Sequencing representation was first used for activity-travel patterns by Wilson (1985) to analyze variability of one-dimensional activity-travel patterns (Wilson, 1998). This type of activity-travel sequence is also used in Allahviranloo et al. (2014), Xu et al. (2016) and Ebadi et al. (2017). Later, multi-dimensional representation was used to include information of mode choice, location, and accompanying persons (Joh et al., 2001, 2002). For this research project, we follow the representation seen in Wilson (1998), Allahviranloo et al. (2014), Xu et al. (2016) and Ebadi et al. (2017) and define a representation. We include “Home”, “Work”, “Shopping”, “Leisure”, “School”, “Personal business” and “Other” as different types of activity type, and the time spend on traveling would be “Trip” activity types. These activity types are identified based on the trip purpose from data, and abbreviated as H, W, S, L, C, P, O, T , which serve as elements in the activity-travel pattern sequence array.

Since we have daily travel data as well as trip purposes for each person, we know the activity type and the time it happened. Given the unit time stamp for 24-hour period, each time stamp (e.g. 6 min) labeled as the eight defined activity types, thus we achieve a vector of activity-travel pattern with 240 elements of activity types. We would include only the 180 elements denoting the activities from 6:00 to 23:59 in order to exclude the sleeping time with few activities. Activity type and activity participation duration include information of potential charging opportunities and durations. “Trips” and their durations are used to infer travel distance that is critical for battery status.

Here is a sample for the original data containing travel activity data in single day for one person in Table 2. We can see various attributes of the original data, including household no (hh_nr), person no (pr_nr), trip purpose (t_pur), departure time (t_dep), arrival time (t_arr), day of week (d_o_w) and trip distance (t_dist). An illustration of the converted pattern is shown in Table 3.

ID	hh_nr	p_nr	t_nr	t_pur	t_dep	t_arr	d_o_w	t_dist	Employ
696	16	4	1	School	25800	26400	Tuesday	13	1
697	16	4	2	Home	42900	43500	Tuesday	13	1
698	16	4	3	Leisure	46800	68400	Tuesday	60	1

Table 2: Original data sample

Time	1	2	3	...	12	13	14	15	...	179	180
P1D1	H	H	H	...	H	T	C	C	...	L	L

Table 3: Uni-Dimensional Sequence Representation of Activity-Travel Patterns

4.1.3 Measures of variability

Given the fact that we have the personal single-day activity-travel pattern for each day, we can define the variability to measure the difference between two different activity-travel pattern sequences. Based on uni-dimensional activity-travel representation, Sequence Alignment Method (SAM) can be used to compare two patterns and produce a score of variability (Kruskal, 1983). As in Allahviranloo et al. (2014), these SAM scores are based on the number of operations needed to convert the source pattern to the target activity-travel pattern. We use Levenshtein distance $L(S_1, S_2)$ to measure the variability between two activity-travel pattern sequences S_1, S_2 , and the value is also referred to as variability. The cost of “insert”, “delete” are set to 1 and “substitution” are set to 2, which allows the maximum variability between two activity-travel patterns to be 360. For further details, readers are referred to Allahviranloo et al. (2014).

One-day population-wide inter-Personal variability (PIV) Since we have the measurement of variability between two activity-travel pattern sequences, we are able to define the population-wide interpersonal variability (PIV) for single-day data. Obviously with single-day data, we are always comparing data between different people. The PIV would be the variability between one of the single-day activity-travel patterns and the *standard activity-travel pattern* in the whole dataset. The *standard activity-travel pattern* is the single-day activity-travel pattern with smallest variability to all other single-day activity travel patterns in the whole dataset, and it could help identify the most average activity-travel pattern for whole population.

Suppose we have a single-day activity-travel pattern dataset of N persons, denoted as $P_{\text{single day}} = [S_1, S_2, \dots, S_N]$. Here, S_i is single-day activity-travel pattern sequence. We can calculate the Levenshtein distance $L(S_i, S_j)$ denoting the variability between S_i and S_j ($\forall i, j$ pair). We take p as the index of the standard activity-travel pattern for the whole population if it is closest to all other activity-travel patterns. Thus, we get standard activity-travel pattern as S_p and population-wide single-day variability of person i as:

$$PIV(i) = L(S_i, S_p) \tag{1}$$

$$\text{where } p = \arg \min_{1 \leq j \leq N} \sum_{i=1}^N L(S_i, S_j) \tag{2}$$

Obviously, we always have $PIV(p) = 0$.

Multi-day intrapersonal variability (MIV) Different from data with single-day activity-travel pattern, multi-day dataset would provide more information on traveler’s behaviors given each person’s day-to-day intrapersonal variability. When this data is available, we use multi-day intrapersonal variability to measure how different the activity-travel patterns are for one person. Similar as PIV, we will find a standard activity-travel pattern in the multi-day activity-travel pattern sequence data for each person, and the summation of variability between each single-day pattern and standard pattern as the multi-day intrapersonal variability. With this definition, we would be able to measure how driver’s behavior varies from his normal activity-travel pattern in different days.

Suppose we have a multi-day activity-travel pattern dataset $P_{\text{multi-day}}$ of N persons, and M -day data are recorded for each person. We can reshape data to form a $N \times M$ matrix where each element $S_{i,j}$ in the matrix denotes the single-day activity of person i on day j .

For one person n , we can define the multi-day intrapersonal variability as follows.

Since we have the multi-day activity-travel pattern data $S_{n,j}, j = 1, 2, \dots, M$, we can calculate the Levenshtein distance $L(S_{n,m_1}, S_{n,m_2})$ between day m_1 and day m_2 for this person n . Similar as PIV, we will take $p(n)$ as the index of the standard activity-travel pattern for this person if it is closest to other activity-travel patterns in multiple days for person n . Then, we define multi-day intrapersonal variability for person n in M days as:

$$MIV(n) = \sum_{i=1}^M L(S_{n,i}, S_{n,p(n)}) \quad (3)$$

$$\text{where } p(n) = \arg \min_{1 \leq j \leq m} \sum_{i=1}^m L(S_{n,i}, S_{n,j}) \quad (4)$$

Multi-day intrapersonal distance variability In addition to the intrapersonal variability of activity travel pattern, travel distance has been identified as one of the most important metrics and the traveler’s daily travel distance also contribute to the total variability of travel behavior. We derive the relationship between average travel distance of multi-day and maximum/minimum travel distance based on multi-day travel data, hoping to constraint the travel distance range of multi-day. The ranges will be used as one of sampling condition. The objective of defining this is when single-day travel distance is given, we can set a range of travel distances that may occur for next consecutive days.

We plot the average 5-day travel distance, maximum travel distance as well as minimum travel distance of our testing dataset in the following Figure 1, and the trend lines plot could reveal the relationship between different variables. We would like to mention that several out layer in the original data has been discard. As we can see in the figure, suppose we have average travel distance d_{avg} of multi-days, we can estimate the maximum travel distance d_{max} and the minimum travel distance d_{min} with the trend line equations as follows:

$$d_{\text{max}} = 1.5197d_{\text{avg}} + 13.646 \quad (5)$$

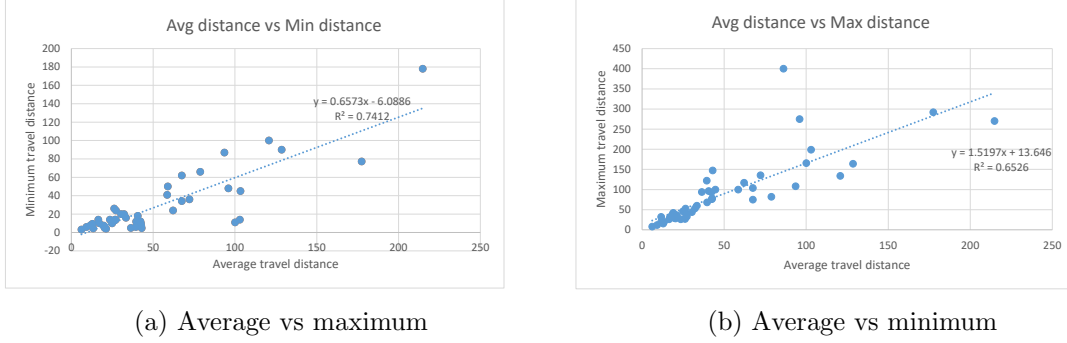


Figure 1: Multi-day travel distance variability

$$d_{\min} = 0.6573d_{\text{avg}} - 6.0886 \quad (6)$$

We can see the R -square value is both higher than 0.6, showing that our estimated equation fits well to the observed data. These can be used as bounds during sampling process.

4.2 Connection between multi-day intrapersonal variability and one-day population-wide interpersonal variability

We have mentioned that the connection between intrapersonal variability and interpersonal variability is helpful while we estimate driver’s behavior and generate multi-day samples from single-day travel data. We argue that **the variability distribution in One-day PIV is similar to the variability distribution in MIV** given the assumption that cross-sectional data for a large population contains information about day-to-day intrapersonal variability. Therefore given a single-day data set, we can create the multi-day variability by mimicking the variability from the single-day data set.

Two adjusted variability measurements (PIV and MIV) on the above section are plotted for the whole population in the following Figure 2. Here, we use adjusted constant coefficient c to unitize PIV and MIV value so that they will fall in the range of $[0, 1]$. We take $1/360$ as the adjusted constant coefficient for PIV and $1/1080$ for MIV. We can validate the similarity between the variability measurements with the Kolmogorov-Smirnov (KS) test on the two curves from the data. We achieved p -value of the KS test with value of 0.711 failing to reject the null hypothesis that the two samples are from the same distribution.

The connection between the distributions of MIV and PIV can be attributed to each traveler’s MIV contributing to PIV distribution. In order to help explain this, we divide the population into three groups based on the value of adjusted MIV. Low MIV group contains 15 people with adjusted MIV less than 0.2; medium MIV group contains 23 people with adjusted MIV larger than 0.2 and less than 0.4; high MIV group contains 12 people of the rest population.

Multi-day intrapersonal variability is comparing the variability between different days for each person. If MIV is low for a person, then his daily activity-travel pattern would be quite similar. However, if one person has high MIV, his activity-travel pattern will vary a lot from day to day.

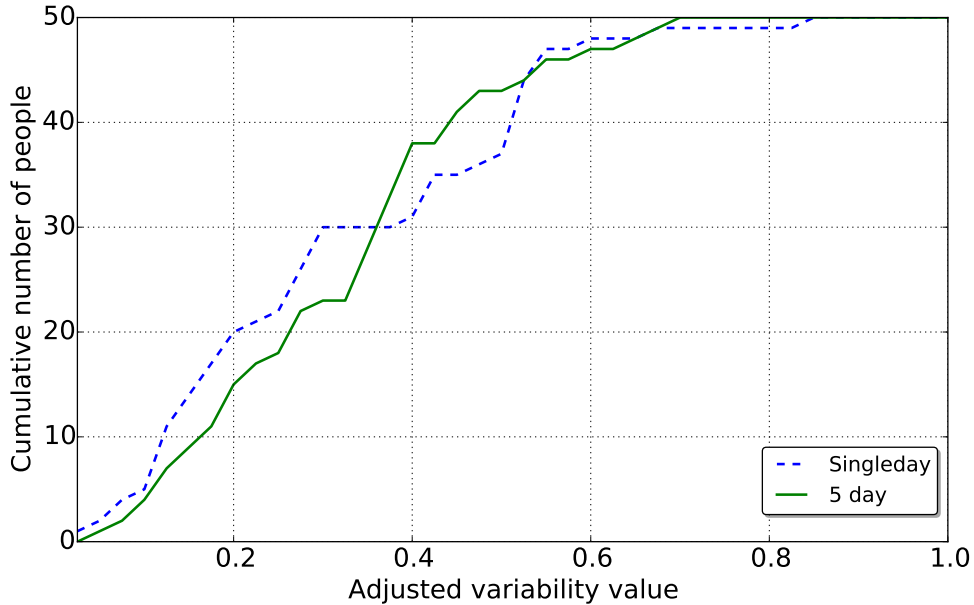


Figure 2: Adjusted PIV and MIV

On the other hand, PIV gives us the difference between various people comparing to the *standard activity-travel pattern*. Given our definition, the *standard activity-travel pattern* we get will be some sort of “average” among all people. The standard is the one that’s closest to others. In other words, most activity-travel patterns should have same activity types in the same or close time intervals as the PIV standard pattern. For example, in our case of both full-time and part-time workers, we can easily know that most of them will go to work in most of the days. And it’s very likely that our “standard” is actually a simple work activity-travel pattern. The standard activity-travel pattern for a given data set is the person No.23. As we can see in Figure 3, this person goes working from home in the morning and goes back home in the evening. A quite standard worker, and this is also the type of activity-travel pattern that appears the most in the data.

For people with low intrapersonal variability in Figure 4, they follow similar activity-travel pattern of a standard worker. They might have some difference on the time of departing and returning to work, which should slightly increase the variability between different people. They might go other places than working, this will increase both PIV and MIV slightly. Thus, we can explain the feature in both cumulative curves on the left side of x-axis. They have **low MIV**, and low PIV comparing to standard pattern because they show standard working pattern. All people are employed in our data, giving us most activity-travel patterns going to work even if we only have single day data, so we can observe standard activity-travel pattern as work only pattern. Thus, the correlation between PIV and MIV among low MIV group is positive since the activity pattern has little difference comparing to the standard pattern that affects PIV and MIV.

Conclusion: Typical full-time workers will have very low MIV since they spend most of their time working. They also have small PIV comparing to standard pattern of working activity type.

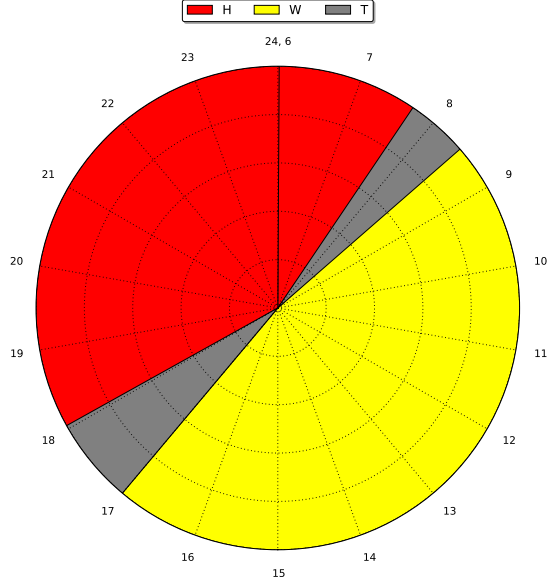


Figure 3: Standard activity-travel pattern of Person No.23

Note: It doesn't mean the "standard" pattern would belong to a low-var person since part-time workers can follow standard pattern on single day and focus on other activities on other days.

Although most people go to work, people still need to go other places like shopping and leisure. Thus, only a few of them have quite low MIV. However, for these people, their MIV would not be high since they only go other places after work or on some working-free days. We show two samples of **medium MIV** in Figure 5 where one person didn't go to work on some days and the other person have longer working hours on some days and other activities on some days. Similarly, while most people go to work, people could spend different time on their trip to work. Thus, most of the activity-travel pattern would be not significantly different from "standard" since they spent most of the time on working activity type. For people with medium MIV, they can either have one day very distinct from standard pattern or several days different from standard pattern. Although the difference falls in appropriate range, the correlation between PIV and MIV is still largely affected by random factors and the choice of standard pattern. Thus, we observe negative correlation among medium MIV group denoting the dependence between PIV and MIV.

Conclusion: Most people will have moderate variability comparing to PIV standard pattern due to the flexibility in different people's activities, and moderate MIV due to shopping or leisure needs.

In Figure 6, these people have **high MIV**, and they have activity-travel patterns that are very different from PIV standard pattern on most days. Since one of the days out of five days are picked for each person in single-day data, there is high chance that we observe a very different pattern from PIV standard pattern. Thus, we should also be able to observe those very different patterns in single day data as high PIV and MIV people in the cumulative distribution figure. If a person has higher intrapersonal variability, he is likely to have more activity-travel patterns going out or he

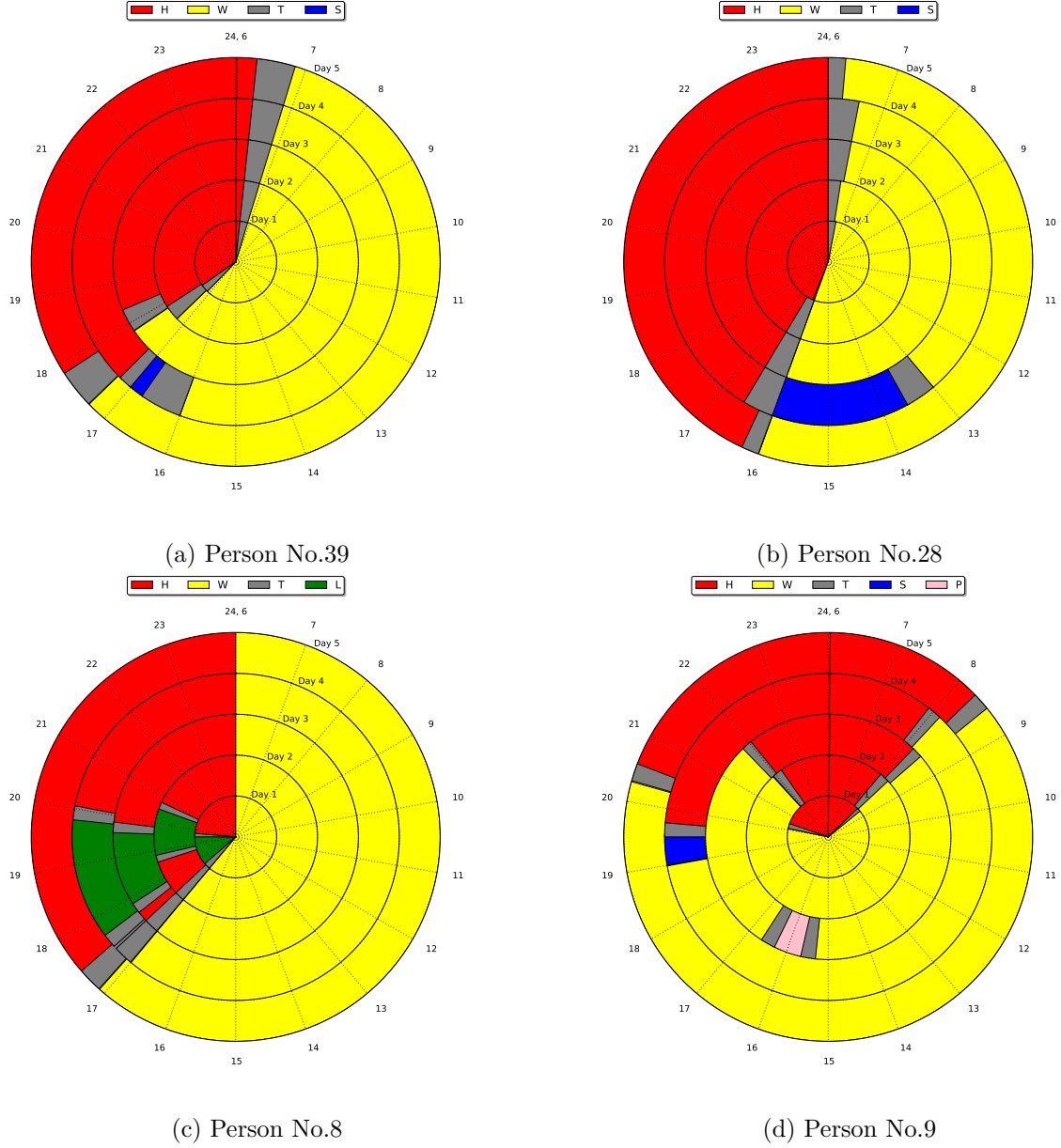
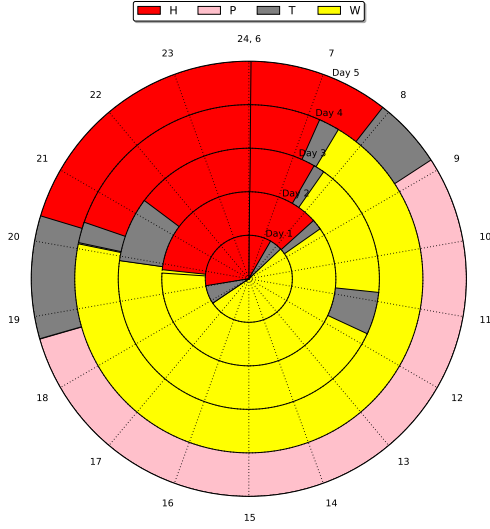


Figure 4: Samples of person with low MIV

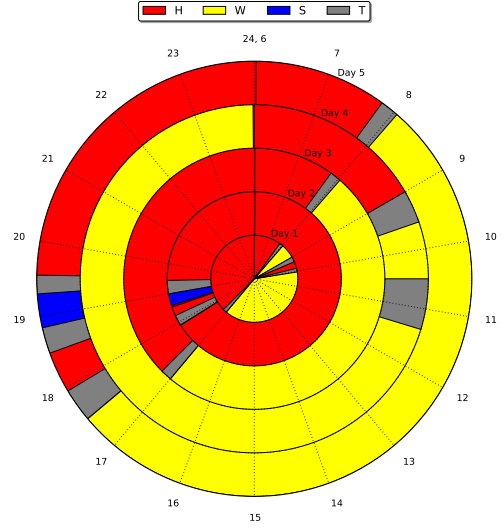
goes different places after work. For the former, he is likely to show a higher variability pattern in single day data. For the later, it is possible to observe a medium variability pattern in single day data. In general, higher MIV leads to higher PIV for the same person and the positive correlation value validate the conclusion.

Conclusion: People with high MIV is likely to have single-day activity-travel pattern that is very different from each other and PIV standard pattern, which will also lead to high PIV.

In summary, MIV is represented in PIV with a large sample, and we should be able to see a good distribution of intrapersonal variability in single day data. Since intrapersonal variability denotes average value, the distribution looks similar. Statistically, we have correlation between adjusted

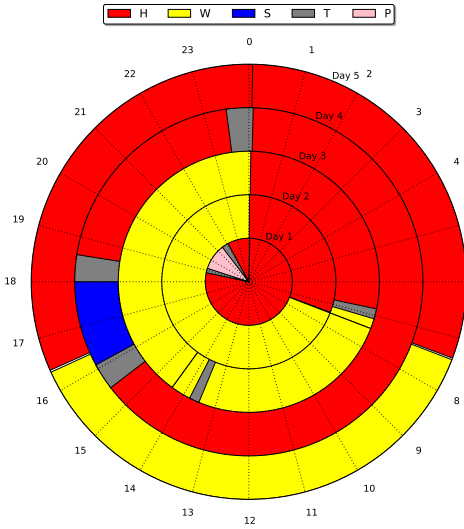


(a) Person No.13

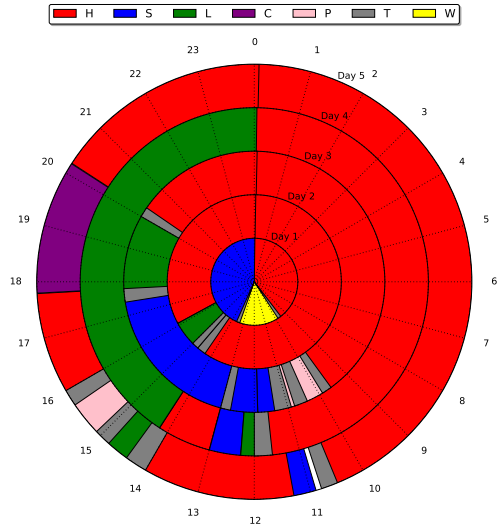


(b) Person No.31

Figure 5: Samples of person with medium MIV



(a) Person No.32



(b) Person No.46

Figure 6: Samples of person with high MIV

PIV and MIV of 0.365 among the whole population, correlation of 0.388 among low MIV group, correlation of -0.129 among medium MIV group and correlation of 0.371 among high MIV group.

4.3 Sampling Procedure

Our sampling method is designed to generate single-day activity-travel pattern data to multi-day activity-travel pattern data, and the idea is to pick different single-day activity-travel patterns

from whole population based on the given personal single-day activity-travel pattern to construct reasonable personal multi-day activity-travel pattern.

4.3.1 Clustering

Clustering is a well-known machine learning technique that can be used to partition the input activity-travel patterns into groups, or clusters, based on their degree of similarity. The best known clustering technique is k -means clustering. We adopt the method proposed by Allahviranloo et al. (2014) which defines an attribute vector of activity-travel pattern as the similarity/variability score against all other patterns for clustering analysis (Allahviranloo et al., 2014). That is, for N number of travelers, there are N attributes which are the SAM scores against all patterns, including itself. We can also apply k -meteoroid algorithm, which is similar to k -means, so that we can have a better understanding of the cluster centers as daily activity-travel patterns.

We visualize the clustering result of the frequency of different types of activities denoted by different colors in Figure 7. Cluster 1 and cluster 2 show similar patterns while working time is longer in cluster 2. Cluster 3 is quite different from the other three clusters due to the large number of leisure activities in addition to work activities. Most people in cluster 4 have activity types as Home, and they have quite different daily activities of other purposes.

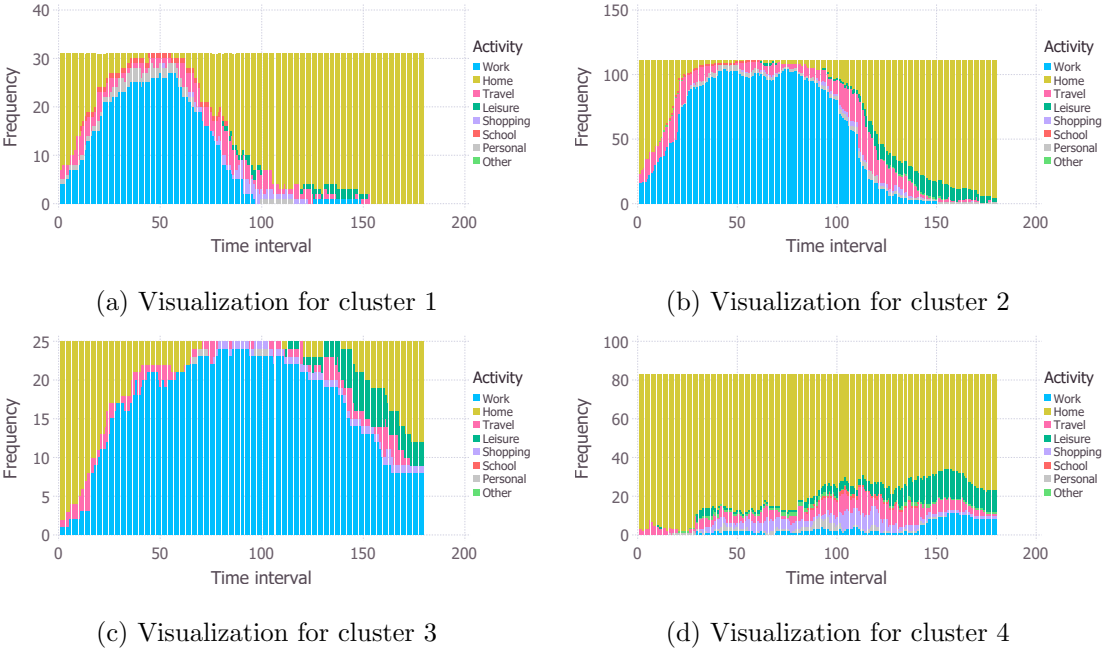


Figure 7: Activity travel pattern visualization for all clusters

Given clustering results, we have 4 groups of people with distinct travel activities and different MIV distribution for each group of people. Thus, we can obtain information about MIV distribution if we know the cluster group of a person. Since cross-sectional data contains intrapersonal variability information and adjusted PIV distribution is similar to adjusted MIV distribution, we can use

PIV distribution of each cluster to estimate the MIV distribution in order to extract intrapersonal variability information from single-day travel activity datasets. Since multi-day dataset is available, cluster MIV distribution is applied during sampling instead of cluster PIV distribution.

4.3.2 Transition probability

With cluster results, day-to-day intrapersonal activity pattern type transition probabilities can be calculated. GPS-based multi-day travel pattern data can be used to draw these probabilities. If activity purpose surveys are available for multi-day travel data, these probabilities can be simply calculated. If they only include travel information, we can infer some information: based on locations, home and work activities can be identified. The rest of the activities will be categorized as “Other”. The activity type “Others” will be used to additionally categorized all non-home, non-work activities. With no other information available, the conditional probabilities of selecting each activity type are assumed to follow the frequencies observed in single-day activity-travel survey data. If any form of multi-data is not available, the number of frequencies in on-day data will be used as probabilities for each day.

Given a k -clustered result vector $C = [C_1, C_2, \dots, C_n]$, the value of $C_i, i \in 1, \dots, n$ denotes which cluster element i of the vector falls in. Suppose we have k clusters, we can define a $k \times k$ matrix M and each element M_{pq} denote the total counted number of $C_i = p, C_{i-1} = q, i \in 1, \dots, n$ in C , meaning that the former day falls in cluster q while the current day falls in cluster p . Then, we can get the k -clustered transition probability matrix $M'_{k \times k}$ with the following equation.

$$M'_{pq} = \frac{M_{pq}}{\sum_{k=1}^k M_{kq}} \quad (7)$$

The transition probability will give us the probability of transferring from one cluster type to another cluster type in general. The transition probability extracted from 5-day travel activity dataset of 50 employed people is shown in Table 4. Here, for element a_{ij} in the matrix, it means that the probability of transferring from the cluster of previous day (C_j) to the cluster of the current day C_i is a_{ij} and we use C_i to denote cluster i .

Table 4: Transition probability

Cluster	Given C_1	Given C_2	Given C_3	Given C_4
C_1	0.24	0.155	0	0.101
C_2	0.56	0.619	0.455	0.144
C_3	0	0.083	0.273	0.058
C_4	0.2	0.143	0.273	0.696

Obviously, transition probability can only be obtained from multi-day travel activity datasets. In this paper, we assume that the transition probability will hold for the different populations regardless of single-day or multi-day dataset to simplify the problem without considering the sociodemographic

characteristics of the population. We might focus on estimating transition probability with personal or household information given only single-day travel data is available in future studies.

4.3.3 Sampling method process

Based on the definition we provided, we introduce our method for sampling activity-travel patterns.

Suppose we have single-day activity-travel pattern sequence data for N persons as $P_{single} = [S_1, S_2, \dots, S_N]$. We calculate the variability $L_{ij} = L(S_i, S_j)$ between all pairs of activity-travel pattern sequences S_i and S_j , thus generate the $N \times N$ variability matrix L . Then, this variability matrix L is input as the cost matrix for k-meteoroid algorithm for the clustering of all activity-travel patterns. We can also choose initial points manually by major activity type to provide more accurate clustering results. Major activity type would be the type of activity that a person spend time on other than staying at home. Thus, we can divide all activity-travel pattern vector S_i into k different clusters, and a cluster type K_i will be assigned to S_i . Here, we have $K_i \in 1, 2, \dots, k$ and $i = 1, 2, \dots, N$.

We can then generate a $k \times k$ transition probability matrix with the given defined method based on the activity-travel pattern clustering result K from multi-day activity-travel pattern data. We will only include transition counts from the same person, and take summation of counted value from whole population.

With the given single-day activity-travel pattern S_n for each person n , we can achieve the corresponding clustering result K_n . With the cumulative distribution of MIV for cluster K_n , we can randomly generate the intrapersonal variability $MIV(n)$ for each person n by inverse of the cdf function. Suppose we will generate $m - 1$ days of activity-travel pattern $S'_{in}, i = 2, 3, \dots, m$, and we can combine them with the original single-day activity-travel pattern $S'_1 = S_n$ to construct the multi-day sample $S'_{in}, i = 1, 2, 3, \dots, m$. Since we have the clustering result K_n for S'_1 , we can generate all $K'(n)_i, i = 1, 2, \dots, m$ based on the transition probability and the clustering result $K'(n)_{i-1}$. With the clustering result $K'(n)_i$ for multi-day sample and intrapersonal variability $MIV(n)$ for person n , we can set limits the sample pool from original data. Only activity-travel patterns that fall in cluster $K'(n)_i$ with MIV smaller than $MIV(n)$ are allowed in the sample pool. We can also set additional distance limit based on the daily travel distance from original single-day data. We use $f(\cdot)$ and $g(\cdot)$ to evaluate the maximum distance and minimum distance allowed in the sample pool. Then, we can randomly choose a single-day activity-travel pattern as our sample S'_{in} from the limited sample pool. After repeated this process for all people, we can convert an N -person-single-day dataset to an N -person- m -day dataset of multi-day activity-travel patterns.

4.3.4 Summary of sampling method

We summarize the sampling method that we use in this paper as follows.

Step 1. Preprocessing of raw data to get single-day activity-travel pattern sequences of all person.

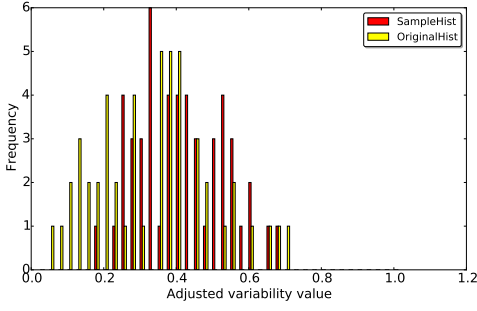
- Step 2.** Use k -meteoroid algorithm to cluster the activity-travel patterns, taking Levenshtein distance matrix as input cost. Initial points could be chosen manually based on major activity type.
- Step 3.** Determine multi-day cluster result for sample based on transition probability and original single-day cluster result for each person.
- Step 4.** Determine MIV for each person based on single-day data clustering result as well as the corresponding MIV cdf or PIV cdf depending on the available data.
- Step 5.** Determine the sample pool for each person on each day based on the given MIV, clustering result and corresponding MIV cdf.
- Step 6.** Additional limits could be applied on activity-travel patterns in sample pool based on the original single-day travel distance and constraint function of $f(\cdot)$ and $g(\cdot)$ for the maximum and minimum travel distance, respectively.
- Step 7.** Randomly choose a single-day activity-travel pattern for each person on each day until our multi-day dataset is fully constructed.

4.4 Validation of MIV variability generated from sampling

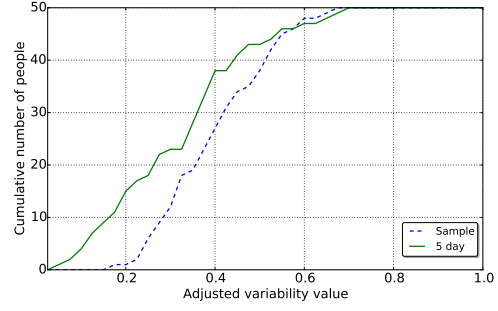
In order to validate the goodness and stability of our sampling method, we compare our generated multi-day sample data with the original multi-day data in various standards including MIV and MIV error. It is natural to compare daily activity-travel pattern of one specific person to the corresponding ones in the original data since we have the single-day data as well as its corresponding multi-day data. However, achieving exact day-to-day match is implausible due to the randomness of sampling and limited information available from single-day data. For example, one person goes to work from Monday to Thursday and then goes shopping on Friday in original data, comparing the sampled data that he goes shopping on Monday and working on the other days. We get errors for this case if we do day-to-day match, however we believe this is inevitable for any sampling without sociodemographic information. As our goal is to create a multi-day dataset that includes variabilities for the population instead of one individual person, we think the distribution of clusters is more important than the cluster order. So we provide two methods to validate the performance of our sampling method to show whether the generated 5-day activity-travel pattern represents the variability observed in original 5-day data.

4.4.1 Intrapersonal variability distribution

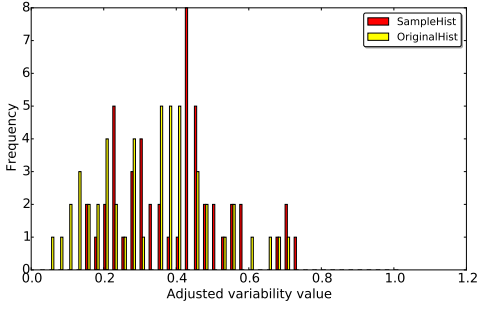
This method is to compare the general MIV variability distribution of whole population between original data and sample data. This will provide us with the insight of the distribution of different types of people. Since we have multi-day data for both original data and sample data, we can



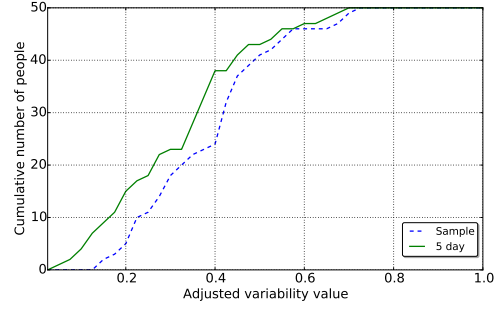
(a) MIV Histogram with distance limit



(b) MIV Cumulative distribution with distance limit



(c) MIV Histogram without distance limit



(d) MIV Cumulative distribution without distance limit

Figure 8: MIV results of the sample

compare the distribution of MIV for whole population to have a general view. The results are shown in the following Figure 8, and our sample data has similar MIV distribution as the original data.

In addition, we applied Kolmogorov-Smirnov test on the data to compare our sampled data and the original data. For the comparison of MIV based on sampling with distance limit, we achieved p-value of the test with value of 0.1122 failing to reject the null hypothesis that the two data samples are from the same distribution. For the comparison of MIV based on sampling without distance limit, we achieved p-value of the test with value of 0.1777 failing to reject the null hypothesis that the two data samples are from the same distribution. These statistic evidences show that our sampling method performs well to consider intrapersonal variability while generating samples from single-day travel dataset.

4.4.2 MIV error (MIVE) for personal multi-day activity-travel pattern

Since we want to compare the difference between original multi-day data and sampled multi-day data based on the intrapersonal variability, it is reasonable to define the difference between one person's original multi-day activity-travel pattern and another person's sample multi-day activity-travel pattern as *multi-day intrapersonal variability error* (MIVE).

Given original M -day activity-travel pattern data $P_{Mday,i} = [S_{i,1}, S_{i,2}, \dots, S_{i,M}]$ for person i and sample M -day activity-travel pattern data $P'_{Mday,j} = [S'_{j,1}, S'_{j,2}, \dots, S'_{j,M}]$ for person j , we can generate the permutation of P' with a total number of $M!$. We use $per(P')$ to denote all sequence ele-

ments of P'_{Mday} , where $per(P')_1 = [S'_{j,1}, S'_{j,2}, \dots, S'_{j,M}]$, $per(P')_2 = [S'_{j,2}, S'_{j,1}, \dots, S'_{j,M}]$, \dots , $per(P')_{m!} = [S'_{M,j}, S'_{M-1,j}, \dots, S'_{M,1}]$. Thus, we can define $VD(\cdot)$ between two multi-day activity-travel pattern P and P' as the summation of Levenshtein distance between each activity-travel pattern pair of P and P' . We use $L(\cdot)$ to denote the Levenshtein distance. Then, we can easily define MIV error between P and P' as:

$$MIVE(P, P') = \min_{1 \leq i \leq m!} VD(P, per(P')_i) \quad (8)$$

$$\text{where } VD(P, P') = \sum_{k=1}^M L(S_{i,k}, S'_{j,k}) \quad (9)$$

Suppose we have an M -day activity-travel pattern dataset $P_{Mday,N}$ of N people as well as the corresponding multi-day sample $P'_{Mday,N}$. Since our new validation method wants to find a day-to-day-person-to-person match with least MIV error difference, we can first formulate a MIV error matrix $E_{N \times N}$, where each element $E_{ij} = MIVE(P_{Mday,i}, P'_{Mday,j})$ denoting the MIV error between person i and person j .

Then, we can easily formulate our new validation method as an assignment problem, to match each person in original data to one person in sample data. The cost of matching original person i to sample person j is E_{ij} . Thus, we can define a variable $x_{ij} = 1$ to denote original person i matching sample person j , otherwise $x_{ij} = 0$.

$$\min \sum_{i=1}^n \sum_{j=1}^n E_{ij} x_{ij} \quad (10)$$

$$\sum_{i=1}^n x_{ij} = 1 \quad \forall j \quad (11)$$

$$\sum_{j=1}^n x_{ij} = 1 \quad \forall i \quad (12)$$

$$x_{ij} \in \{0, 1\} \quad \forall i, j \quad (13)$$

We can solve this problem with CPLEX, and get the minimum MIV error match of sample data and original data. The MIV comparison results are as follows in Figure 9. Here, we have *our method* as our proposed sampling method and error considering population wide distribution and *trivial method* as duplicate single-day data for five days as the multi-day sample. In Figure 9a and 9b, we order the person by absolute MIV error while we show the trend of relative MIV error (absolute MIV error/MIV) in Figure 9a and 9b with person ordered by MIV.

Here, we can see MIV error is smaller using our sampling method for most people with higher MIV during their travels since trivial sampling method without considering day-to-day intrapersonal variability is not able to well estimate MIV. For some people with low MIV, their travel can be well estimated by the trivial method since their activity patterns are similar each day, and our proposed

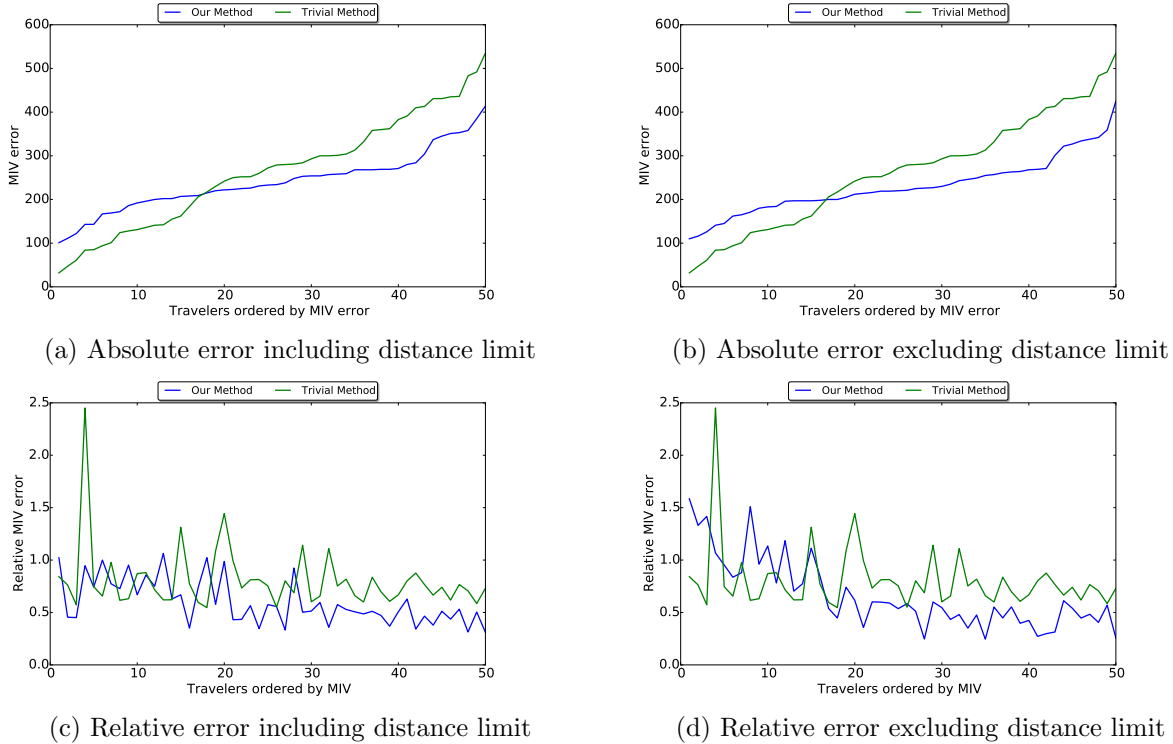
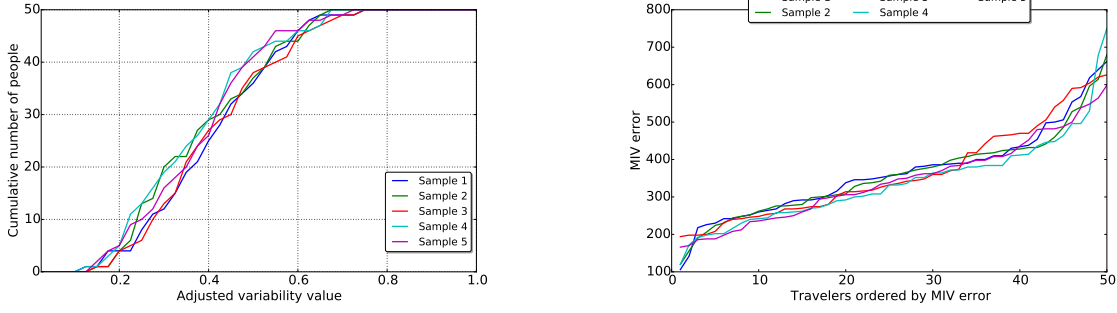


Figure 9: MIV error validation

sampling method may overestimate MIV due to random selection of single-day trips.

4.4.3 Stability of multi-day sampling method

We present our sampling method to generate multi-day travel activity data based on single-day data. Although we applied clustering, transition probability as well as other factors to estimate day-to-day intrapersonal variability, the sampling process itself is still random. Thus, it is essential to make sure that our sampling method can generate stable multi-day sample data instead of random distinct samples. Thus, we generate multiple samples with our sampling method considering distance limit to compare the MIV and MIV error so that we can visualize the stability of our sampling method. In Figure 10, we show the comparison of adjusted MIV and MIV error between 5 generated multi-day samples and stable results are shown in both figures. For adjusted MIV, the difference between various samples is approximately less than 0.04 for the same cumulative number of people and the distribution of MIV are similar for all samples. For MIV error, the gap between different samples is less than 100 and the overall trend of the curve is similar for all samples. Thus, our sampling method is able to provide random samples with stable MIV given same single-day travel activity data.



(a) MIV Cumulative distribution without distance limit (b) Absolute error excluding distance limit

Figure 10: Multiple sample comparison results

4.5 Multi-Day Battery Electric Vehicle (BEV) Feasibility Assessment

State of charge (SOC) is the equivalent of a fuel gauge for the battery pack in a BEV, which we consider as the remaining battery level in a BEV for one person in this project. In the following case study, we assume that all vehicles have the range of 100 miles, and all charging stations are level 2 charging stations that requires 7 hours to fully recharge a BEV from empty battery level.

Suppose we define the SOC of a person i at time t as SOC_t^i , finish a trip of travel distance d or stop and charge the BEV starting at time t_1 and ending at time $t_1 + n\Delta t$ where n is the number of time intervals Δt between t_1 and t_2 , the BEV range is R , and the charging rate is C , we can get the following SOC relationship for traveling and SOC relationship for charging:

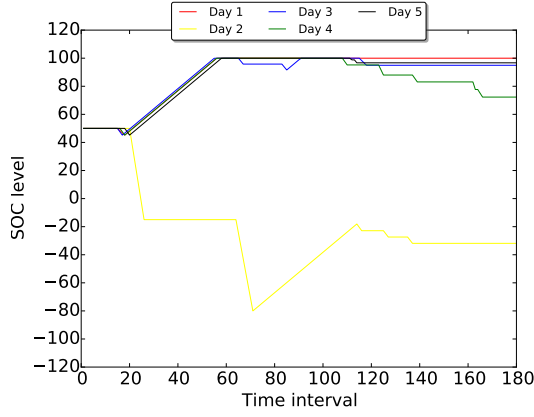
$$\text{Traveling: } SOC_{t_1+k\Delta t}^i = SOC_{t_1}^i + \frac{k-n}{n}d, \quad k = \{1, 2, \dots, n\} \quad (14)$$

$$\text{Charging: } SOC_{t_1+k\Delta t}^i = \min(SOC_{t_1}^i + Ck\Delta t, R), \quad k = \{1, 2, \dots, n\} \quad (15)$$

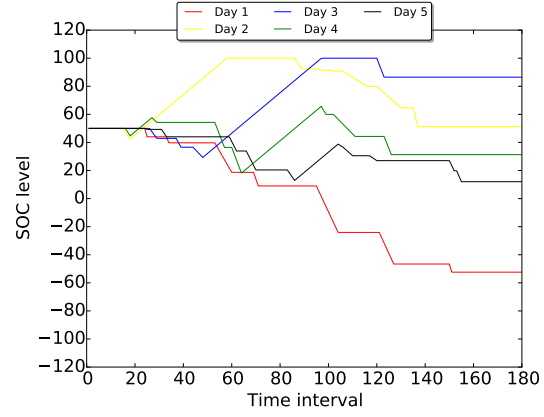
Since we have time interval as 6 mins for activity-travel pattern sequences, we can generate SOC level every 6 mins for each person on each day to create a 240-element SOC array. We assume that the vehicle will be at half-full battery at the beginning of each day, and different charging availability will construct different scenarios. We visualize the SOC curves of two person as an example in 5 days based on original data, our sample and trivial sample as follows in Figure 11 with work charging scenario. Here, our sample data is generated based on the original single-day data with our proposed sampling method, and trivial sample is generated by duplicate the given single-day data for all 5 days.

Obviously, we can observe that we couldn't generate a multi-day sample that is similar as the original multi-day data at individual level due to the randomness of the sample pool. However, we can consider population wide difference between original multi-day data and our multi-day sample data comparing to the difference between original multi-day data and trivial sample data.

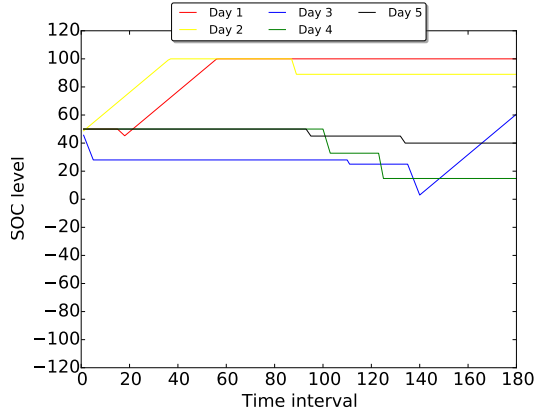
We can apply **similar methods** as that of MIVE to calculate SOC error for personal multi-day activity-travel pattern. We use Levenshtein distance to calculate the difference between two activity-



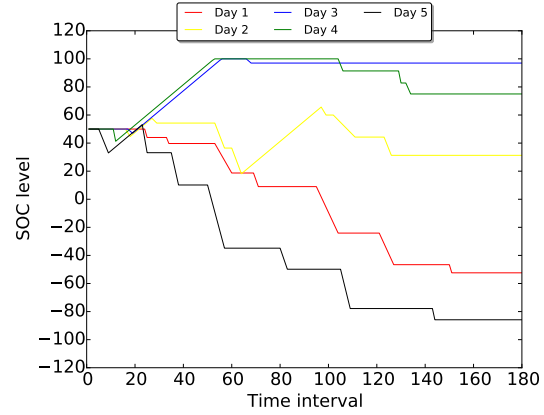
(a) Person No.24 based on original data



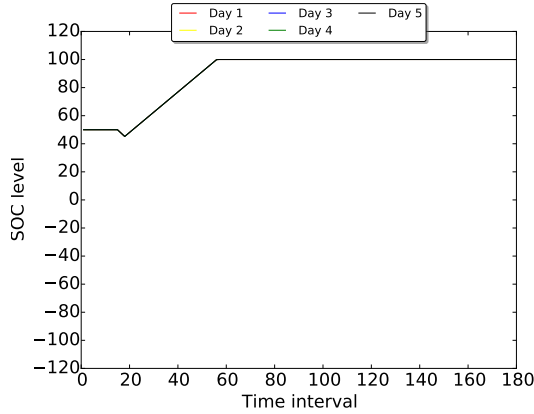
(b) Person No.40 based on original data



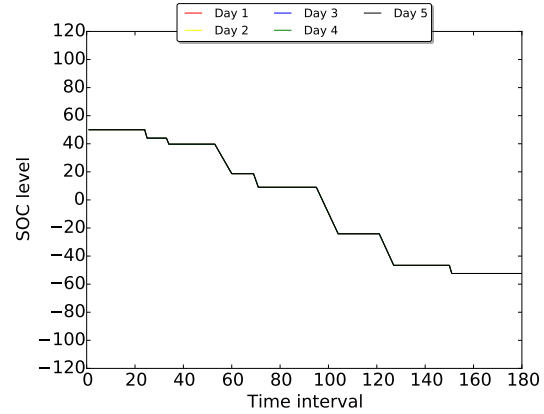
(c) Person No.24 based on our method



(d) Person No.40 based on our method



(e) Person No.24 based on trivial method



(f) Person No.40 based on trivial method

Figure 11: Example SOC curves based on different samples

travel patterns for each person, while we will use the 1-norm distance ($\sum_{i=1}^n |x_i - y_i|$) to calculate the difference between two SOC vectors for each person. With sampling method considering distance limit, we have average SOC error of 63.13 based on our sampling method and average SOC error of 74.70 based on trivial method. While distance limit is not included, we have average SOC error of

60.54 based on our sampling method and average SOC error of 74.70 based on trivial method.

The SOC comparison results are as follows in Figure 12 based on relative SOC error (absolute SOC error/MIV):

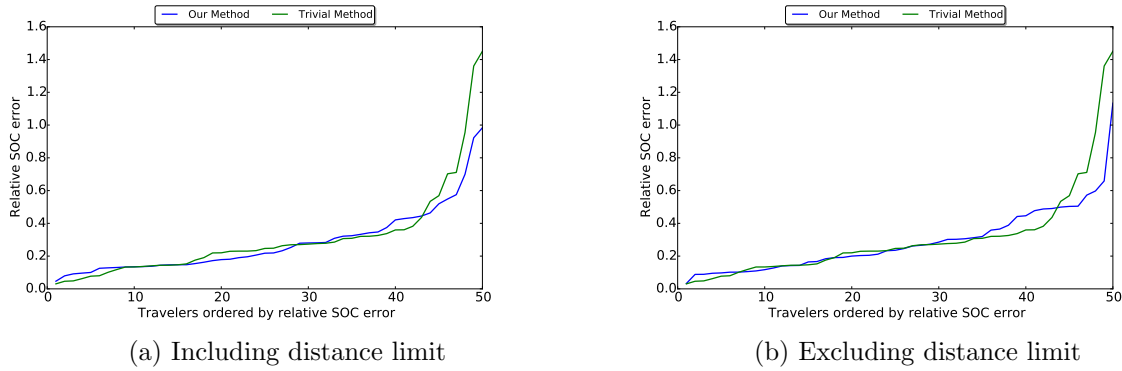


Figure 12: SOC error validation

Our method will be able to reduce the SOC error comparing to trivial method. The method of comparing the SOC level of the same person between sample data and original data is not considered here since we would like to show the difference of the distribution of whole population instead of each individual.

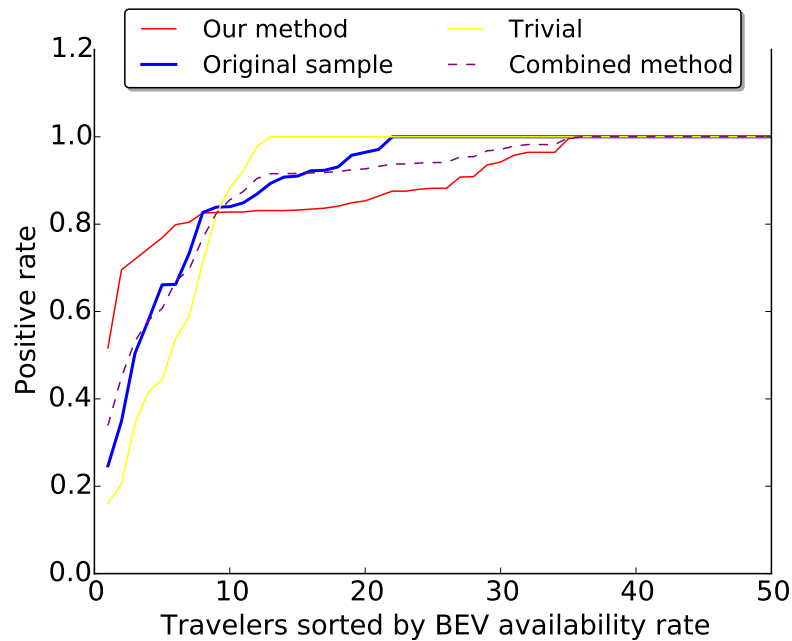


Figure 13: SOC positive rate with work charging only

It could include different standards while considering if a user is available for BEV, and we are considering to finish all trips within BEV range. That means we always want to have positive SOC level since negative SOC level means running out of battery. Thus, we can define *SOC positive*

rate as the percentage of time that the vehicle has positive SOC level during 5 days’ travel as the standard of whether a person is available for BEV. Considering the fact that we will not be able to provide 1 on 1 comparison given the randomness of the sample, we will sort all travelers in original data and sample data by their 5-day SOC positive rate and compare the distribution in the whole population. The results comparing samples generated from our method and trivial method as well as the original multi-day data are shown in the following Figure 13 where y-axis show the percentage value of the SOC positive rate and x-axis show the index of the person.

In Figure 13, around 40 person have over 80% SOC positive rate for both original data and our sample data, meaning that BEV will be able to serve most of their trips while the trivial method will show more difference and tend to overestimate the availability of BEV for travelers. While around 10 person have less SOC positive rate indicating that they are less appropriate for BEVs, our method tends to overestimate the availability for this group of people and trivial method tends to underestimate them. Given the fact that the original sample curve falls in the middle of trivial method curve and our method curve, we can use the sample result of our method as well as the trivial method as the upper bound and lower bound to better estimate the BEV availability rate of the original sample. We show the average estimation result of our method and trivial method as the combined method as the purple dashed line in the figure, which fits better to the original data estimation than using our method or trivial method only. Thus, we can have more accurate results even if we only have single-day cross-sectional data by extracting more day-to-day intrapersonal variability information with our sampling method.

5 Conclusions and Recommendations

This research incorporates multiday activity-travel patterns for assessing Battery Electric Vehicle (BEV) feasibility. BEV is considered to be feasible if the State of Charge (SOC) level stays positive given assumptions of charging availability and behavior. In order to generate multiday activity-travel patterns that are not readily available, we sample from readily available one-day activity-travel patterns.

First, we introduce several measurements of activity travel pattern variability, including single-day interpersonal variability and multi-day intrapersonal variability. We also explain the similarity between these two variability measurements for single-day data sample and multi-day data sample. Given such evidence, we develop our sampling method to generate multi-day sample based on single-day travel data by clustering traveler into different groups. We also include transition probability of cluster type as well as the daily travel distance to provide more accurate sampling results. This method allows us to provide a case study based on Mobidrive data to estimate the BEV availability by considering the SOC level of all travelers given a broader view of the variability distribution of whole population by multi-day sampling. The results shows that considering our multi-day sampled data along with the original single-day data provides more accurate traveler behavior estimations.

These generated multi-day activity-patterns can be used represent day-to-day intrapersonal

variability in activity-travel decisions. Since it is impossible to accurately predict the intrapersonal variability without social characteristic information, we consider the distribution of intrapersonal variability of whole population instead. Our multi-day sample data performs well comparing to the estimation results of single-day data although the multi-day sample data tends to overestimate the intrapersonal variability of some people. Our sampling method helps extracting more information of day-to-day intrapersonal variability contained in single-day cross sectional data. In addition, more accurate estimation of BEV availability can be provided by combining our sampling method with trivial method since both methods help to bound the range of potential original multi-day estimations. This paper provides an ample potential for further study. In order to provide more accurate multi-day sample data with limited information, social characteristic attributes might be considered to provide better clustering results of travelers.

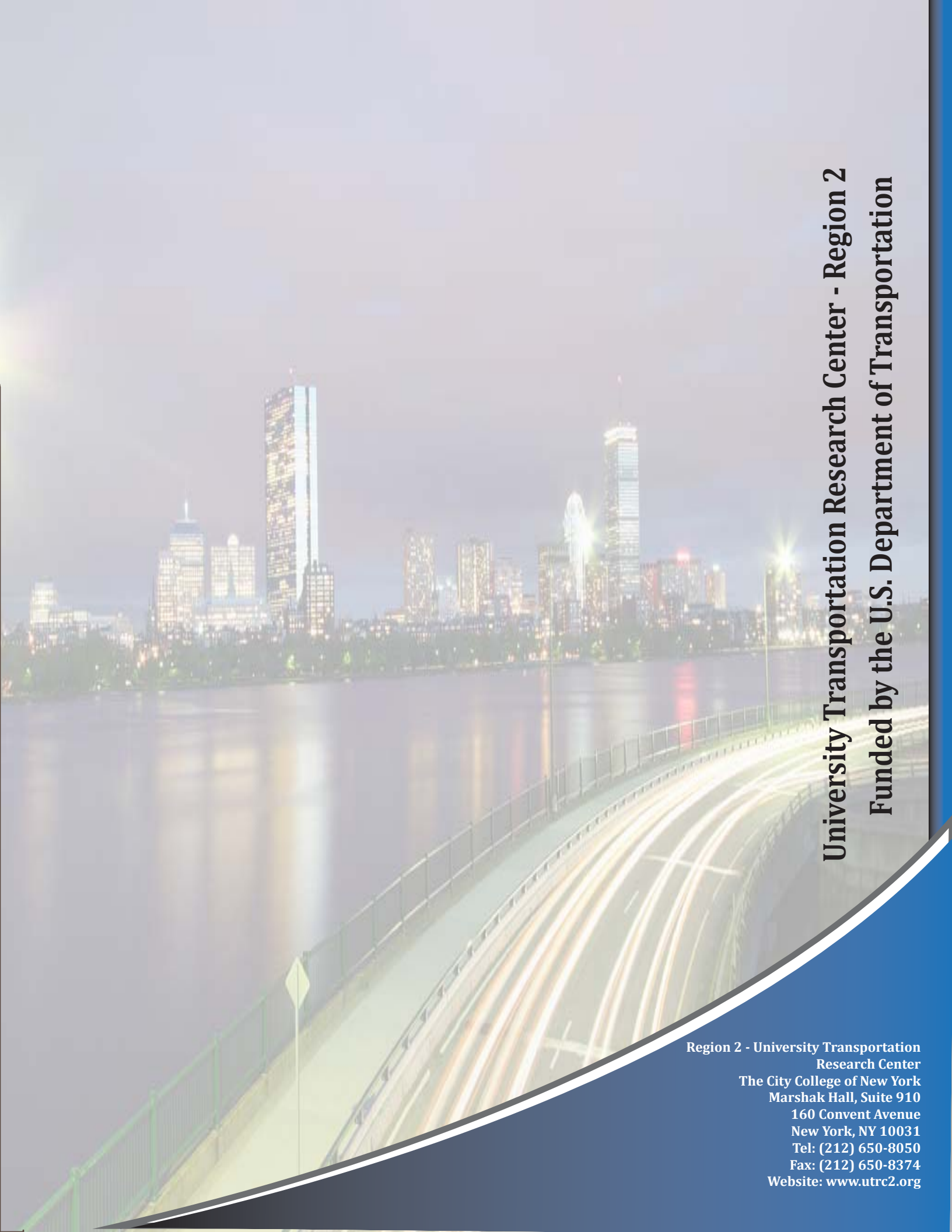
References

- Allahviranloo, M., R. Regue, W. Recker. 2014. Pattern clustering and activity inference. *Transportation Research Board 93rd Annual Meeting*. 14-1274.
- Axhausen, K. W., A. Zimmermann, S. Schönfelder, G. Rindsfuser, T. Haupt. 2002. Observing the rhythms of daily life: A six-week travel diary. *Transportation* **29**(2) 95–124.
- Axsen, J., K. S. Kurani, R. McCarthy, C. Yang. 2011. Plug-in hybrid vehicle ghg impacts in california: Integrating consumer-informed recharge profiles with an electricity-dispatch model. *Energy Policy* **39**(3) 1617–1629.
- Buliung, R. N., M. J. Roorda, T. K. Rummel. 2008. Exploring spatial variety in patterns of activity-travel behaviour: initial results from the toronto travel-activity panel survey (ttaps). *Transportation* **35**(6) 697–722.
- Chen, C., J. Ma, Y. Susilo, Y. Liu, M. Wang. 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies* **68** 285–299.
- Chikaraishi, M., A. Fujiwara, J. Zhang, K. Axhausen, D. Zumkeller. 2011. Changes in variations of travel time expenditure: some methodological considerations and empirical results from german mobility panel. *Transportation Research Record: Journal of the Transportation Research Board* (2230) 121–131.
- Dong, J., Z. Lin. 2012. Within-day recharge of plug-in hybrid electric vehicles: energy impact of public charging infrastructure. *Transportation Research Part D: Transport and Environment* **17**(5) 405–412.
- Dong, J., C. Liu, Z. Lin. 2014. Charging infrastructure planning for promoting battery electric

- vehicles: An activity-based approach using multiday travel data. *Transportation Research Part C: Emerging Technologies* **38** 44–55.
- Ebadi, N., J. E. Kang, S. Hasan. 2017. Inferring activity-mobility behavior of college students based on smartcard transaction data. Tech. rep.
- Elango, V., R. Guensler, J. Ogle. 2008. Day-to-day travel variability in the commute atlanta, georgia, study. *Transportation Research Record: Journal of the Transportation Research Board* .
- Gonder, J., T. Markel, M. Thornton, A. Simpson. 2007. Using global positioning system travel data to assess real-world energy use of plug-in hybrid electric vehicles. *Transportation Research Record: Journal of the Transportation Research Board* (2017) 26–32.
- Hanson, S., J. O. Huff. 1981. Assessing day-to-day variability in complex travel patterns. *Transportation Research Record* **891** 18–24.
- Hanson, S., O. J. Huff. 1988. Systematic variability in repetitious travel. *Transportation* **15**(1-2) 111–135.
- Joh, C.-H., T. Arentze, F. Hofman, H. Timmermans. 2002. Activity pattern similarity: a multidimensional sequence alignment method. *Transportation Research Part B: Methodological* **36**(5) 385–403.
- Joh, C.-H., T. A. Arentze, H. J. Timmermans. 2001. Multidimensional sequence alignment methods for activity-travel pattern analysis: A comparison of dynamic programming and genetic algorithms. *Geographical Analysis* **33**(3) 247–270.
- Kang, H., D. M. Scott. 2010. Exploring day-to-day variability in time use for household members. *Transportation Research Part A: Policy and Practice* **44**(8) 609–619.
- Kang, J. E., W. Recker. 2009. An activity-based assessment of the potential impacts of plug-in hybrid electric vehicles on energy and emissions using 1-day travel data. *Transportation Research Part D: Transport and Environment* **14**(8) 541–556.
- Kang, J. E., W. W. Recker. 2014. Measuring the inconvenience of operating an alternative fuel vehicle. *Transportation Research Part D: Transport and Environment* **27** 30–40.
- Khan, M., K. M. Kockelman. 2012. Predicting the market potential of plug-in electric vehicles using multiday gps data. *Energy Policy* **46** 225–233.
- Khayati, Y., J. E. Kang. 2015. Modeling intra-household interactions for the use of battery electric vehicles. *Transportation Research Board 94th Annual Meeting*. 15-4052.
- Koppelman, F., E. Pas. 1984. Estimation of disaggregate regression models of person trip generation with multiday data. *Papers presented during the Ninth International Symposium on Transportation and Traffic Theory held in Delft the Netherlands, 11-13 July 1984..*

- Kruskal, J. B. 1983. An overview of sequence comparison: Time warps, string edits, and macromolecules. *SIAM review* **25**(2) 201–237.
- Moiseeva, A., H. Timmermans, J. Choi, C.-H. Joh. 2014. Sequence alignment analysis of variability in activity travel patterns through 8 weeks of diary data. *Transportation Research Record: Journal of the Transportation Research Board* (2412) 49–56.
- Pas, E. I. 1988. Weekly travel-activity behavior. *Transportation* **15**(1-2) 89–109.
- Pas, E. I., F. S. Koppelman. 1986. An examination of the determinants of day-to-day variability in individuals' urban travel behavior. *Transportation* **13**(2) 183–200.
- Pas, E. I., S. Sundar. 1995. Intrapersonal variability in daily urban travel behavior: some additional evidence. *Transportation* **22**(2) 135–150.
- Pearre, N. S., W. Kempton, R. L. Guensler, V. V. Elango. 2011. Electric vehicles: How much range is required for a days driving? *Transportation Research Part C: Emerging Technologies* **19**(6) 1171–1184.
- Pendyala, R., E. Pas. 2000. Multiday and multiperiod data for travel demand modeling. *Invited Resource Paper in Transport Surveys: Raising the Standard, Proceedings of an International Conference on Transport Survey Quality and Innovation. Transportation Research Board, National Research Council II-B/1 – II-B/22.*
- Pendyala, R., A. Parashar, G. Muthyalagari. 2001. Measuring day-to day variability in travel characteristics using gps data. *79th annual meeting of the Transportation Research Board.*
- Recker, W. W., M. G. McNally, G. S. Root. 1985. Travel/activity analysis: pattern recognition, classification and interpretation. *Transportation Research Part A: General* **19**(4) 279–296.
- Schlich, R., K. W. Axhausen. 2003. Habitual travel behaviour: evidence from a six-week travel diary. *Transportation* **30**(1) 13–36.
- Smith, R., S. Shahidinejad, D. Blair, E. Bibeau. 2011. Characterization of urban commuter driving profiles to optimize battery size in light-duty plug-in electric vehicles. *Transportation Research Part D: Transport and Environment* **16**(3) 218–224.
- Stopher, P., C. FitzGerald, J. Zhang, T. Bretin. 2008. Analysis of 28-day global positioning system panel survey: Findings on variability of travel. *Transportation Research Record: Journal of the Transportation Research Board* .
- Susilo, Y., R. Kitamura. 2005. Analysis of day-to-day variability in an individual's action space: exploration of 6-week mobidrive travel diary data. *Transportation Research Record: Journal of the Transportation Research Board* (1902) 124–133.

- Susilo, Y. O., K. W. Axhausen. 2014. Repetitions in individual daily activity–travel–location patterns: a study using the herfindahl–hirschman index. *Transportation* **41**(5) 995–1011.
- Wilson, W. C. 1998. Activity pattern analysis by means of sequence-alignment methods. *Environment and Planning A* **30**(6) 1017–1038.
- Xu, Z., J. E. Kang, R. B. Chen. 2016. A random utility based estimation framework for the household activity pattern problem .
- Zhang, L., T. Brown, G. S. Samuelson. 2011. Fuel reduction and electricity consumption impact of different charging scenarios for plug-in hybrid electric vehicles. *Journal of Power Sources* **196**(15) 6559–6566.

A long-exposure photograph of a city skyline at night, reflected in a body of water. In the foreground, a bridge or highway has light trails from moving vehicles. The sky is dark, and the city lights are bright and colorful.

University Transportation Research Center - Region 2
Funded by the U.S. Department of Transportation

**Region 2 - University Transportation
Research Center**
The City College of New York
Marshak Hall, Suite 910
160 Convent Avenue
New York, NY 10031
Tel: (212) 650-8050
Fax: (212) 650-8374
Website: www.utrc2.org