Multi-day Activity-Travel Pattern Sampling Based on Single-Day Data: Application of Battery Electric Vehicles Feasibility Assessment

Anpeng Zhang†, Jee Eun Kang∗†, Kay Axhausen♭, and Changhyun Kwon‡

†Department of Industrial and Systems Engineering, University at Buffalo, SUNY
♭Institute of Transport Planning and Systems, Swiss Federal Institute of Technology, Switzerland
‡Department of Industrial and Management Systems Engineering, University of South Florida

May 23, 2017

Abstract

Although it is important to consider multi-day activities in transportation planning, multi-day activity-travel data are expensive to acquire. In this study, we propose to generate multi-day activity-travel data through sampling from readily available single-day household travel survey data with considerations of day-to-day intrapersonal variability. One of the key observations 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. A case study on the battery electric vehicle (BEV) feasibility considering the state-of-charge level is also conducted based on the proposed methods, and results show that our sampling method combined with single-day based trivial method provides better estimation of the population wide BEV feasibility than using cross-sectional data only.

1 Introduction and Background

Intrapersonal variability, also known as day-to-day variations, of activity-travel patterns is 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

*Corresponding author; jeeunka@buffalo.edu
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 observations, 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-week 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 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. 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. In Table ??, we show whether previous works conclude that intrapersonal variability occupies a large proportion of the total individual travel variability. We only include four major factors (trip frequency, travel time, activity location and activity travel pattern) in variability measurements while detailed measurements related to the same factor could vary depending on different studies. A more detailed summary of variability measurement in previous studies is given in Appendix A. 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 activity-travel patterns.

Despite these evidences, day-to-day intrapersonal variability is often ignored in studies analyzing activity-travel behaviors and estimating travel demand due to the unavailability of multi-day data. Several studies showed the limitations of relying on single-day data from longitudinal travel pattern data. Since single-day data contain little time-related information and various measurements of intrapersonal variability are applied, the results based on those data range from 20% to 80% (Chen et al., 2016). In case of battery electric vehicle (BEV) feasibility, working with one-day data only could result in distinct conclusions.
Table 1: Variability proportion and measurements in previous studies

<table>
<thead>
<tr>
<th>Large proportion of total variance</th>
<th>Paper</th>
<th>Intrapersonal variability measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Trip frequency</td>
</tr>
<tr>
<td>Yes</td>
<td>Pas and Koppelman (1986)</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Pas (1988)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pas and Sundar (1995)</td>
<td>✓                                                           ✓</td>
</tr>
<tr>
<td></td>
<td>Kang and Scott (2010) (Weekdays)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chikaraishi et al. (2011)</td>
<td>✓</td>
</tr>
<tr>
<td>No</td>
<td>Susilo and Kitamura (2005)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chikaraishi et al. (2011)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kang and Scott (2010) (Weekends)</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Moiseeva et al. (2014)</td>
<td>✓</td>
</tr>
<tr>
<td>Not mentioned</td>
<td>Schlich and Axhausen (2003)</td>
<td>✓                                                           ✓</td>
</tr>
<tr>
<td></td>
<td>Elango et al. (2008)</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Buliung et al. (2008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Susilo and Axhausen (2014)</td>
<td>✓</td>
</tr>
</tbody>
</table>
However, data collection is an expensive task and data sets with multi-day activity-travel patterns are rarely available. 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 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.

Thus, we intend to generate multi-day data samples based on single-day datasets with consideration of day-to-day variation. Cross-sectional data (single-day data) in statistics and econometrics is usually collected by observing many subjects (such as individuals or regions) without regard to differences in time. For contrary panel data (multi-day data), researchers conduct several observations of the same subjects over a period of time in a longitudinal study. Large-scale cross-sectional datasets contain detailed information of various aspects of activity-travel decisions and interpersonal variability. We intend to extract intrapersonal variability given rich travel activity information in such datasets. We assume that single-day data contains a diverse set of activity-travel patterns that is sufficient to be used as surrogate for multi-day activity-travel patterns. Suppose several people have similar travel activities, they may travel for work only on some days and travel for work and shopping on other days. When we collect the single-day travel-activity pattern data of these people, we are likely to observe work-only activity travel patterns on some days and work-and-shopping activity travel patterns on other days. The distribution of work only pattern and work-and-shopping pattern among the chosen single-day samples could be similar to the distribution of work days as well as work-and-shopping days of each person on average. In other words, multiple observations over a large collection of presumed homogeneous observations can be used as a surrogate for repeated observations over a single individual.

We also need a well-defined measurement of intrapersonal variability in order to indicate the travel activity pattern of a person more accurately. Given such consideration, we measure intrapersonal variability based on the similarity of activity cluster types. Given our measurement of variability, correlation between the distribution of person-to-person variation on single day and day-to-day variation over whole population can be observed during the study. Given such connection between interpersonal and intrapersonal variability, we propose a sampling procedure to generate multi-day travel data samples with available single-day travel datasets. In addition to distribution of day-to-day intrapersonal variability, our sampling method is based on day-to-day basis transition probabilities between clustered activity-travel-pattern types. Travel distance are also considered to provide more accurate samples. Our major contribution focus on the proposed sampling method since there is few previous work studying multi-day dataset sampling based on single-day datasets with consideration of intrapersonal variability. In empirical experiments, our generated sample has a
variability distribution that is similar to true multi-day intrapersonal variability under individual-level day-to-day person-to-person matching. A case study on multi-day BEV feasibility is also presented to utilize the intrapersonal variability information in single-day datasets. With the combination of our proposed sampling method and single-day trivial sampling method, we can bind BEV feasibility results in a limited range, avoiding overestimations or underestimations of the potential market.

In Section 2, we introduce the measurements of variability used in this paper. In Section 3, we explain 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 present a case study of the battery electric vehicle (BEV) feasibility assessment and market potential based on the state of charge (SOC) with our proposed sampling method.

2 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.

2.1 Data Description

We use the data from the project Mobidrive funded by the German Ministry of Research and Education, and the data has been used in several previous studies (Schlich and Axhausen, 2003; Susilo and Kitamura, 2005; Susilo and Axhausen, 2014). The data contain a continuous travel diary of six weeks, helping 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 choose 5 weekday data for each available person to avoid distinct activities during weekends. Based on these considerations, we choose a sample of 50 travelers, and each person’s data contain 5-day travel information. Our sample data contain 927 daily trips, including 353 trips going home, 166 trips going to work, 87 trips for leisure, 96 trips for shopping and the rest with other purposes. 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.

2.2 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 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 (1998) to analyze variability of one-dimensional activity-travel patterns. 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 activity types, and the time spent on traveling would be ‘Trip’ activity type. These activity types are identified based on the trip purposes from data, and abbreviated as \( H, W, S, L, C, P, O, \) and \( T \), 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) is labeled as one of the eight defined activity types. Thus, we achieve a vector of activity-travel pattern with 240 elements of activity types. We 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 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 1. There are 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 2, and \( P1D1 \) is used to denote data of person 1 on day 1.

<table>
<thead>
<tr>
<th>ID</th>
<th>hh_nr</th>
<th>pr_nr</th>
<th>t_pur</th>
<th>t_dep</th>
<th>t_arr</th>
<th>d_o_w</th>
<th>t_dist</th>
<th>Employ</th>
</tr>
</thead>
<tbody>
<tr>
<td>696</td>
<td>16</td>
<td>4</td>
<td>1</td>
<td>25800</td>
<td>26400</td>
<td>Tuesday</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>697</td>
<td>16</td>
<td>4</td>
<td>2</td>
<td>42900</td>
<td>43500</td>
<td>Tuesday</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>698</td>
<td>16</td>
<td>4</td>
<td>3</td>
<td>46800</td>
<td>68400</td>
<td>Tuesday</td>
<td>60</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2: Original data sample
Table 3: Uni-Dimensional Sequence Representation of Activity-Travel Patterns

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>...</th>
<th>179</th>
<th>180</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1D1</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td></td>
<td>H</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>L</td>
<td></td>
<td>C</td>
<td>C</td>
<td>L</td>
<td>L</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.3 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) could be used to compare two patterns and produce a score of variability (Kruskal, 1983). As in Allahviranloo et al. (2014), these SAM scores were based on the number of operations needed to convert the source pattern to the target activity-travel pattern. Levenshtein (1966) introduced a Levenshtein distance as the smallest number of operations required to change one sequence to another with substitutions, insertions and deletions. Here, the basic element operations of insertion is to insert a character into a sequence, deletion is to delete a character from a sequence and substitution is to replace a character in a sequence. 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 both insertion and deletion 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).

2.3.1 single-day population-wide inter-personal variability (PIV)

Since we have the measurement of variability between two activity-travel pattern sequences, we can define the population-wide interpersonal variability (PIV) for single-day data. Obviously, we always compare data between different people in single-day data. 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 $P_{\text{single-day}} = [S_1, S_2, ..., S_N]$, where $S_n$ represents the single-day activity-travel pattern sequence of $n$-th person. We can calculate the Levenshtein distance $L(S_n, S_{n'})$ for the variability between $S_n$ and $S_{n'}$ for each pair of $n$ and $n'$. We let $i$ denote the index of the standard activity-travel pattern for the whole population, if the $i$-th person’s pattern is closest to all other activity-travel patterns. Thus, we obtain the “standard” activity-travel pattern as $S_i$ and population-wide single-day variability of person $n$ as:

$$\text{PIV}(n) = L(S_n, S_i)$$  

(1)
where \( i = \arg \min_{n=1,\ldots,N} \sum_{n'=1}^{N} L(S_n, S_{n'}) \) \hspace{1cm} (2)

Obviously, we always have \( \text{PIV}(i) = 0 \).

### 2.3.2 Multi-day intrapersonal variability (MIV)

Different from data with single-day activity-travel patterns, a multi-day dataset would provide more information on travelers’ 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 in different days. Similar as PIV, we 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 in all days as the multi-day intrapersonal variability. With this definition, we would be able to measure how driver’s behavior varies from the 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 \( S_{n,m} \) in the matrix denotes the single-day activity of person \( n \) on day \( m \). Obviously, we have \( N \times M \) days of data and we use \( W \) to denote this number.

For each person \( n \), we can define the multi-day intrapersonal variability as follows. Since we have the multi-day activity-travel pattern data \( S_{n,m}, m = 1,2,\ldots,M \), we can calculate the Levenshtein distance \( L(S_{n,m}, S_{n,m'}) \) between day \( m \) and day \( m' \) for this person \( n \). Similar as in PIV, we let \( I(n) \) denote the index of the standard activity-travel pattern for this person if the pattern of person \( n \) on day \( I(n) \) is closest to activity-travel patterns of person \( n \) on all other days. Then, we define multi-day intrapersonal variability for person \( n \) as:

\[
\text{MIV}(n) = \sum_{m=1}^{M} L(S_{n,m}, S_{n,I(n)})
\]

\[
\text{where } I(n) = \arg \min_{m=1,\ldots,M} \sum_{m'=1}^{M} L(S_{n,m}, S_{n,m'})
\]

### 2.3.3 Multi-day intrapersonal distance variability

In addition to the intrapersonal variability of activity travel pattern, travel distance (or travel time given constant vehicle speed) has been identified as one of the most important metrics and the traveler’s daily travel distance also contributes to the total variability of travel behavior (Pas and Sundar, 1995; Schlich and Axhausen, 2003; Stopher et al., 2008; Chikaraishi et al., 2011). 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 to help limiting the sample pool. Thus, we can set a range of travel distances that may occur for next consecutive days when single-day travel distance is given.
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 discarded.

In Figure 1, suppose we have average travel distance \( d_{\text{avg}} \) of multi-days, we can estimate the maximum travel distance \( d_{\text{max}} \) and the minimum travel distance \( d_{\text{min}} \) with the trend line equations as follows:

\[
d_{\text{max}} = 1.5197d_{\text{avg}} + 13.646 \\
d_{\text{min}} = 0.6573d_{\text{avg}} - 6.0886
\]

The \( R \)-squared 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.

### 3 Connection between multi-day intrapersonal variability and single-day population-wide interpersonal variability

The connection between intrapersonal variability and interpersonal variability is helpful, when we estimate driver’s behavior and generate multi-day samples from single-day travel data. We argue that the variability distribution in single-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 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 multiply \( \frac{1}{360} \) as the adjusted constant coefficient for PIV and \( \frac{1}{1080} \) for MIV in order to normalize the variability value. 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 compares 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 a person has high MIV, his/her activity-travel pattern will vary a lot from day to day. On the other hand, PIV gives us the difference between various people comparing to the standard activity-travel pattern. According to our definition, the standard activity-travel pattern is some sort of “average” among all people. The standard pattern is the one that is 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. It is very likely that our “standard” is actually a work-only 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.
Figure 3: Standard activity-travel pattern of Person No.23 on a single day

3.1 Low MIV

For people with low intrapersonal variability in Figure 4, they follow similar activity-travel patterns 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 the horizontal 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.

3.2 Medium MIV

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
Figure 4: Samples of person with low MIV
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.

3.3 High MIV

Persons in Figure 6, 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 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.

3.4 Remarks

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 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.

We make some remarks:
Low MIV: Typical full-time workers will have very low MIV, since they spend most of their time working. They also have small PIV comparing to the “standard” pattern of working activity type. Note that the “standard” pattern could belong to a person with higher intrapersonal variability. People can have distinct activities on multi-days, while only one day data is included in the dataset. Therefore, it is possible that the work-only pattern is chosen into the single-day dataset, although the person is not a typical full-time worker and happened to have went working on that specific day.

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

High MIV: 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.

4 Sampling Procedure

Our sampling method is designed to generate multi-day activity-travel pattern data from single-day activity-travel pattern data. The main 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.
4.1 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 (MacQueen et al., 1967). 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 a total number of $N$ travelers, there are $N$ attributes which are the SAM scores against all patterns, including itself. We can also apply $K$-medoids algorithm (Kaufman, 1987), 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.

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,
Table 4: Transition probability matrix $\Psi$

<table>
<thead>
<tr>
<th>Cluster</th>
<th>$c_{n,m-1} = 1$</th>
<th>$c_{n,m-1} = 2$</th>
<th>$c_{n,m-1} = 3$</th>
<th>$c_{n,m-1} = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_{n,m} = 1$</td>
<td>0.24</td>
<td>0.155</td>
<td>0</td>
<td>0.101</td>
</tr>
<tr>
<td>$c_{n,m} = 2$</td>
<td>0.56</td>
<td>0.619</td>
<td>0.455</td>
<td>0.144</td>
</tr>
<tr>
<td>$c_{n,m} = 3$</td>
<td>0</td>
<td>0.083</td>
<td>0.273</td>
<td>0.058</td>
</tr>
<tr>
<td>$c_{n,m} = 4$</td>
<td>0.2</td>
<td>0.143</td>
<td>0.273</td>
<td>0.696</td>
</tr>
</tbody>
</table>

Cluster MIV distribution is applied during sampling instead of cluster PIV distribution.

4.2 Transition probability

With clustering 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 more 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 categorize 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.

After creating $K$ clusters of travel activities for an $M$-day-$N$-person data set, we let $c_{n,m} = k$ if the $m$-th day’s activity pattern of person $n$ falls in the $k$-th cluster. We also define a $K \times K$ matrix $Q$ with each element $Q_{k,k'}$ denoting the total number of day-person pairs such that $c_{n,m} = k$ and $c_{n,m-1} = k'$ for all $n = 1, ..., N$ and $m = 1, ..., M$, which counts the number of cases transitioning from cluster $k'$ in one day to cluster $k$ in the next day. Then, we can obtain the $K \times K$ transition probability matrix $\Psi$ with each element being:

$$\Psi_{k,k'} = \frac{Q_{k,k'}}{\sum_{k_1=1}^{K} Q_{k_1,k'}}$$

for all $k, k' = 1, ..., K$.

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 3. Here, for element $\Psi_{ij}$ in the matrix, it means that the probability of transferring from the cluster $j$ of the previous day ($c_{n,m-1} = j$) to the cluster $i$ of the current day ($c_{n,m} = i$) is $\Psi_{ij}$ and we use $m$ to denote the current day index and $n$ for the current person.

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 Sampling method

Based on the former definitions, we introduce our method for sampling activity-travel patterns as follows.

Suppose we have single-day activity-travel pattern sequence data for \( N \) persons as \( P_{\text{single-day}} = [S_1, S_2, ..., S_N] \). We calculate the variability \( v_{n,n'} = L(S_n, S_{n'}) \) between all possible pairs of activity-travel pattern sequences \( S_n \) and \( S_{n'} \); thus generate the \( N \times N \) variability matrix \( V \). Then, this variability matrix \( V \) is the input as the cost matrix for \( K \)-medoids 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 most time on out of home. Thus, we can divide all activity-travel pattern vector \( S_i \) into \( K \) different clusters, obtaining a \( K \)-clustered result matrix \( C \). Since we only have one day data (\( M = 1 \)), our cluster result matrix \( C = (c_{n,1} : n = 1, ..., N) \) is actually a vector. Thus, when we have \( c_{n,1} = k \), we know the activity sequence \( S_n \) falls into the \( k \)-th cluster.

We can then generate a \( K \times K \) transition probability matrix \( \Psi \) 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 a summation of counted values from the whole population.

With the cumulative distribution of MIV for cluster \( k \), we can randomly generate the intrapersonal variability \( \text{MIV}(n) \) for each person \( n \) by the inverse of the cumulative distribution function (cdf).

We only need to generate \( M - 1 \) days of activity-travel pattern \( \tilde{S}_{n,m}, m = 2, 3, ..., M \), since we can use the original single-day data as the sample of the first day in our \( M \)-day sample. Thus, we can construct an \( M \)-day sample \( \tilde{S}_{n,m}, m = 1, 2, 3, ..., M \) for each person \( n = 1, 2, ..., N \). Since we have the clustering result \( c_{n,1} \) for \( \tilde{S}_{n,1} \), we can generate all \( c_{n,m}, m = 2, ..., M \) based on the transition probability \( \Psi \) and the former day’s clustering result \( c_{n,m-1} \) for the same person. With the clustering result \( c_{n,m} \) for multi-day sample and intrapersonal variability \( \text{MIV}(n) \) for person \( n \), we can set limits the sample pool from original data. Only activity-travel patterns that fall in the \( c_{n,m} \)-th cluster with MIV smaller than \( \text{MIV}(n) \) 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 \( \tilde{S}_{n,m} \) for the \( n \)-th person on the \( m \)-th day. After repeating 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.

4.4 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 $K$-medoids 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.

We also define a single-day trivial method in this paper as duplicate one-day’s travel activities for multi-days to generate multi-day data. In that case, people’s trips will always be the same on each day.

5 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.
5.1 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 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.

5.2 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).

Suppose that we have the original $M$-day activity-travel pattern data $P_n = P_{multi-day,n} = [S_{n,1}, S_{n,2}, ..., S_{n,M}]$ for person $n$ and sample $M$-day activity-travel pattern data $\tilde{P}_n = \tilde{P}_{multi-day,n'} = [\tilde{S}_{n',1}, \tilde{S}_{n',2}, ..., \tilde{S}_{n',M}]$ for person $n'$. We can generate all possible permutations of $\tilde{P}_n$ with a total number of $M!$. We let $\text{perm} (\tilde{P}_n')$ denote the collection of all permutations of elements in $\tilde{P}_n'$, and $\text{perm}_i (\tilde{P}_n')$ denote the $i$-th permutation, for any person $n' = 1, ..., N$. Thus, we can define $VD(\cdot, \cdot)$ as the summation of Levenshtein distance between each activity-travel pattern pair, and $L(\cdot, \cdot)$ denotes the Levenshtein distance. Then, we can easily define the MIV error (MIVE) between $P_n$ and $\tilde{P}_n'$ as follows:

$$MIVE(P_n, \tilde{P}_n') = \min_{1 \leq i \leq M!} VD(P_n, \text{perm}_i (\tilde{P}_n'))$$

where $VD(P_n, \tilde{P}_n') = \sum_{m=1}^{M} L(S_{n,m}, \tilde{S}_{n',m})$

Suppose we have an $M$-day activity-travel pattern dataset $P_{multi-day,n}$ of $N$ people as well as the corresponding multi-day sample $\tilde{P}_{multi-day,n}$. 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 $E$, where each element $e_{n,n'} = MIVE(P_n, \tilde{P}_{n'})$ denoting the MIV error between person $n$ and person $n'$. 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 $n$ to sample person $n'$ is $e_{n,n'}$. Thus, we can define a variable $x_{n,n'} = 1$ to denote original person $n$ matching sample person $n'$, otherwise $x_{n,n'} = 0$.

$$\min \sum_{n=1}^{N} \sum_{n'=1}^{N} e_{n,n'} x_{n,n'}$$

$$\sum_{n=1}^{N} x_{n,n'} = 1 \quad \forall n'$$

$$\sum_{n'=1}^{N} x_{n,n'} = 1 \quad \forall n$$

$$x_{n,n'} \in \{0, 1\} \quad \forall n, n'$$

We can solve this problem with an optimization solver such as CPLEX, and obtain 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 Figures 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 Figures 9a and 9b with person ordered by MIV.
We can see that the MIV error is smaller with our sampling method for most people with higher MIV during their travels, since the single-day 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 single-day trivial method since their activity patterns are similar each day, and our proposed sampling method may overestimate MIV due to random selection of single-day trips.

5.3 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.

6 Case Study: Multi-Day Battery Electric Vehicle (BEV) Feasibility Assessment

We apply the proposed sampling of multi-day activity-travel patterns to feasibility assessment for Battery Electric Vehicles (BEVs). For BEVs, the biggest challenge is whether a traveler’s intended travel activities can be served by a BEV, since BEVs have a shorter range and longer charging times. Although the coming generation of luxury EVs 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 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).

However, 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 objective of this case study is to identify the market potential of BEVs based on multi-day activity-travel patterns, incorporating day-to-day intrapersonal variability in activity-travel patterns.

6.1 Empirical results

The 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 \( SOC^i_{t} \), 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 \( \Delta t \) 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:

\[
\text{Traveling: } SOC^i_{t_1 + k\Delta t} = SOC^i_{t_1} + \frac{k - n}{n}d, \quad k = \{1, 2, ..., n\},
\]

\[
\text{Charging: } SOC^i_{t_1 + k\Delta t} = \min(SOC^i_{t_1} + Ck\Delta t, R), \quad k = \{1, 2, ..., n\}.
\]

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 persons 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 using single-day trivial method.

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-travel patterns for each person, while we will use the \( l^1 \)-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 single-day 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 single-day trivial method.
Figure 11: Example SOC curves based on different samples
We compare our method with the trivial method in Figure 12 based on the relative SOC error, which is defined as the absolute SOC error divided by MIV. Our method is able to reduce the SOC error comparing to the 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.

There could be different standards while considering if a user is available for BEV, and we define availability as all trips (or a proportion of all trips) are finished 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 \textit{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 single-day trivial method as well as the original multi-day data are shown in the following Figure 13 where the vertical axis shows the percentage value of the SOC positive rate and
the horizontal axis shows 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 single-day 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 single-day 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 set 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.

7 Conclusion

In this paper, 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 a single-day trivial method since both methods help to bind 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.

8 Acknowledgment

This work was, in part, supported by Region II University Transportation Research Center (UTRC). Authors are grateful for their generous support. Authors would 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.

References


Appendix

A Summary of previous studies on intrapersonal variabilities

Here is a summary of the previous studies on intrapersonal variabilities in Table 4.
Table 5: Results of various measure of day-to-day intrapersonal variability in previous studies

<table>
<thead>
<tr>
<th>Paper</th>
<th>Data Source</th>
<th>Intrapersonal variability standard</th>
<th>Large proportion of total variance</th>
<th>Major findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pas and Koppelman (1986)</td>
<td>Reading, England (1973)</td>
<td>daily trip frequency</td>
<td>Yes</td>
<td>sociodemographic groups dependent</td>
</tr>
<tr>
<td>Pas (1988)</td>
<td>Employed people in Reading, England (1973)</td>
<td>day of week, daily pattern type</td>
<td>Yes</td>
<td>day-of-week independent for daily pattern, sociodemographic characteristic important for weekly pattern</td>
</tr>
<tr>
<td>Pas and Sundar (1995)</td>
<td>North King County, Washington (1989)</td>
<td>trip frequency, daily travel time, trip chaining</td>
<td>Yes</td>
<td>similar magnitude under different standards</td>
</tr>
<tr>
<td>Schlich and Axhausen (2003)</td>
<td>Mobidrive in Germany (1999)</td>
<td>trip based method, time budgets-based method</td>
<td>Not mentioned</td>
<td>trip-based better, complexity needs consideration, additional proofs for previous works</td>
</tr>
<tr>
<td>Susilo and Kitamura (2005)</td>
<td>Mobidrive in Germany (1999)</td>
<td>action space(the second moment of the out-of-home activity locations)</td>
<td>No</td>
<td>workers and students more stable</td>
</tr>
<tr>
<td>Elango et al. (2008)</td>
<td>Commute Atlanta study (2004)</td>
<td>delta trips</td>
<td>Not mentioned</td>
<td>significant demographic variables effects, day-of-week effects, less seasonal effects</td>
</tr>
<tr>
<td>Buliung et al. (2008)</td>
<td>Toronto Travel Activity Panel Survey</td>
<td>minimum convex polygon (MCP) metric(smallest convex polygon containing all activity locations within a respondents activity-travel pattern)</td>
<td>Not mentioned</td>
<td>spatial variety in travel behaviors, typically conduct activities at repeated locations, human spatial behavior sensitive to temporal scale of analysis</td>
</tr>
<tr>
<td>Study</td>
<td>Methodology</td>
<td>Data Collection</td>
<td>Analysis</td>
<td>Findings</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Stopher et al. (2008)</td>
<td>28-day GPS survey in Sydney</td>
<td>daily travel distance, daily number of trips, average travel time per trip, etc.</td>
<td>Yes</td>
<td>1 or 2 days of data overestimate variances, stable at 18 or 19 days</td>
</tr>
<tr>
<td>Kang and Scott (2010)</td>
<td>Toronto Travel-Activity Panel Survey with GIS toolkit (2003)</td>
<td>activity time-use patterns</td>
<td>Yes for weekdays, No for weekends</td>
<td>joint activities (interact with other household members) have higher proportion of intra-var</td>
</tr>
<tr>
<td>Chikaraishi et al. (2011)</td>
<td>German Mobility Panel survey data (1999-2008)</td>
<td>travel time expenditure</td>
<td>Yes</td>
<td>situational attributes dependent, longer observation time is important</td>
</tr>
<tr>
<td>Susilo and Axhausen (2014)</td>
<td>Mobidrive in Germany (1999)</td>
<td>repetitiveness of activity travel patterns</td>
<td>Not mentioned</td>
<td>individual’s behavioral choices dependent, activity type affects repetition pattern</td>
</tr>
<tr>
<td>Moiseeva et al. (2014)</td>
<td>GPS data in Eindhoven, Netherlands (2010)</td>
<td>activity travel pattern</td>
<td>No</td>
<td>sociodemographic characteristics dependent, intra-variability reflect environment learning speed</td>
</tr>
</tbody>
</table>