Multi-day Activity-Travel Pattern Sampling Based on Single-Day Data

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

January 28, 2018

Abstract

Although it is important to consider multi-day activities in transportation planning, multi-day activity-travel data are expensive to acquire and therefore rarely available. In this study, we propose to generate multi-day activity-travel data through sampling from readily available single-day household travel survey data. A 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. Thus, interpersonal variability observed in cross-sectional single-day data of a group of people 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.

Keywords: Activity-Travel Patterns; Day-to-day variability; Interpersonal variability; Sampling Multiday Activity-Travel Patterns

[*]Corresponding author; `jeeeunka@buffalo.edu`

List of Figures

List of Tables

1 Introduction and background

Intrapersonal variability, also known as day-to-day variation, of activity-travel patterns is found to show strong repetitions, yet with considerable variations (Hanson & Huff 1981, 1988, Pas & Sundar 1995, Pendyala & Pas 2000, Chikaraishi et al. 2009, Chikaraishi, Fujiwara, Zhang & Axhausen 2011). Observations of day-to-day variation 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 & Pas 1984, Pendyala et al. 2001, Pas & Koppelman 1986, Pas & Sundar 1995, Susilo & Axhausen 2014). Variations of travel behavior have also been explained by day-of-week factors. In previous studies, Pas & 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-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 & Sundar (1995) extended trip generation rate day-to-day variation analysis with similar formulations of the total sum of squares in travel behaviors. 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 behaviors of individuals. Elango et al. (2007) introduced delta trips as the measurement of day-to-day trip making variability. Their experiment results showed that 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 1, we show whether previous works conclude that intrapersonal variability occupies a large proportion of the total individual travel variability. We 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 measurements in previous studies is given in Appendix A. However, how the new generations of travelers will change in terms of activity-travel pattern remains uncertain according to some recent works. Lyons (2015) focused on the interactions between motor age and digital age based on a socio-technical conceptualization of society, and they showed the future of transport is uncertain in the digital age. On the other hand, Garikapati et al. (2016) introduced the potential changes of activity pattern as well as time usage from a generation of people. According to their work, the activity pattern might change based on the age of travelers instead of the generation. They showed that millennial tend to mimic the activity-time use patterns of prior generations and fundamental shift in travel demand in the future should not be expected based on similar work (McDonald 2015) with earlier household travel survey data. Given diverse measurements and distinct numerical results in previous work, intrapersonal variability has been proved to be closely related to the variation of people's activity-travel patterns even if the effect of new generations of travelers remains uncertain.

Despite this evidence, 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% comparing to each other (Chen et al. 2016). In case of applications or case studies in reality, working with only single-day data could result in distinct conclusions.

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, industry, 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 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.

There is limited work focusing on the generation of multiple-day travel dataset. Recently, Medina (2016) presented two discrete choice models to generate multiple-day travel activity types based on the revealed sample trips according to the likeliness of the activity, using large AFC data and survey. We intend to generate multi-day data with consideration of day-to-day variation via sampling from a single-day dataset, which is easier to collect comparing to smart card data. *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 (day-of-year). 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 the 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 a 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 pattern on some days and work-and-shopping activity-travel pattern 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.

Table 1: Variability proportion and measurements in previous studies

Large proportion of total variance	Paper	Intrapersonal variability measurements			
		Trip frequency	Travel time	Actvity location	Activity-travel pattern
Yes	Pas & Koppelman (1986)	✓			
	Pas (1988)				✓
	Pas & Sundar (1995)	✓	✓		
	Kang & Scott (2010) (Weekdays)				✓
	Chikaraishi, Fujiwara, Zhang, Axhausen & Zumkeller (2011)		✓		
No	Susilo & Kitamura (2005)			✓	
	Chikaraishi, Fujiwara, Zhang, Axhausen & Zumkeller (2011)				
	Kang & Scott (2010) (Weekends)				✓
	Moiseeva et al. (2014)				✓
Not mentioned	Schlich & Axhausen (2003)	✓	✓		
	Elango et al. (2007)	✓			
	Buliung et al. (2008)			✓	
	Susilo & Axhausen (2014)				✓

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 considerations, we measure intrapersonal variability based on the similarity of activity cluster types. With 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 the distribution of day-to-day intrapersonal variability, our sampling method is based on estimated day-to-day based transition probabilities between clustered activity-travel-pattern types. Travel distances are also considered to provide more accurate samples. Our major contribution is the proposed sampling method since there is little previous work studying multi-day dataset sampling based on single-day datasets with consideration of intrapersonal variability. In empirical experiments, our generated sample shows a variability distribution that is similar to the true multi-day intrapersonal variability under individual-level day-to-day person-to-person matching.

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.

2 Proposed variability measurements

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

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 various previous studies (Schlich & Axhausen 2003, Susilo & Kitamura 2005, Susilo & Axhausen 2014). The data contains a continuous travel diary of six weeks, helping to find the behavior patterns of the respondents. The travel diary survey was conducted in two German cities of Karlsruhe and Halle both 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 given in Axhausen et al. (2002). Given that non habitual long distance trips could still be hard to capture with consideration of day-to-day patterns, we focus on the population group of people with mainly habitual trips in order to reduce sociodemographic impacts. Thus, we only include weekday trips from workers with vehicles in our sample, since weekday trips include more habitual activities and workers' activity patterns are easier to track according to the work of Pas (1988). Based on these considerations, we obtain a sample of workers with all vehicle trips, and each person's data contain 5-day travel information. We first excluded 4966 trips with missing information, and chose 7962 trips from households that own vehicles. Then, we considered 2887 vehicle trips from these travelers, and ensured that each person has at least five weekdays of data available starting from the first Tuesday. Thus, we have a subset of 1157 trips from 50 workers and 16 non workers. Our final sample data contains 927 daily trips from 50 workers, 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 & Huff 1981, Joh et al. 2001, Pas 1988, Recker et al. 1985, Susilo & 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. (2017) 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 work, we follow the representation seen in Wilson (1998), Allahviranloo et al. (2014), Xu et al. (2017) and Ebadi et al. (2017). We include 'Home', 'Work', 'Shopping', 'Leisure', 'School', 'Personal Business' and 'Other' as the 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. Each time slot of 6 min is labeled with one of the eight defined activity types. Thus, we achieve a daily vector of activity-travel pattern with 240 elements of activity types. However, there are only 32 night trips happened between 0:00 and 6:00, which is even less than the 57 trips happened between 6:00 and 7:00 among the total 927 trips. Thus, we consider 0:00 to 6:00 as the night time for most travelers and these night trips will be excluded from our analysis. We include only the 180 elements denoting the activities from 6:00 to 23:59 in order to exclude the night time with few activities.

A sample for the original data containing travel activity data in single day for one person is shown in Table 2. There are various attributes of the original data, including household no (hh_nr),

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

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, and *P1D1* is used to denote data of person 1 on day 1.

2.3 Measurements 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 operation 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 costs of insertion and deletion are set to 1 and the cost of substitution is 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)

We can define the population-wide interpersonal variability (PIV) for single-day data as the upper bound measurement of variability. 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 the 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 among all people.

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 the closest to all other activity-travel patterns. Thus, we obtain the "standard" activity-travel pattern as S_i and the population-wide single-day variability of person n as:

$$\text{PIV}(n) = L(S_n, S_i) \tag{1}$$

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

Obviously, we always have $\text{PIV}(i) = 0$. If ties are obtained while choosing standard pattern, we will randomly choose one of the likely activity-travel patterns since these patterns are likely to be similar.

2.3.2 Multi-day intrapersonal variability (MIV)

Different from data with single-day travel activities only, 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 on various days. Similar as PIV, we find the 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 are able to measure how a traveler's behaviors vary from his/her normal activity-travel pattern on different days.

Suppose we have a multi-day activity-travel pattern dataset $P_{\text{multi-day}}$ of N persons, and M-day data is recorded for each person. We can reshape the 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 $S_{n,m}, m = 1, 2, ..., M$, we can calculate the Levenshtein distance $L(S_{n,m}, S_{n,m'})$ between day m and day m' for this person n. Similarly, 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 the closest to the activity-travel patterns of person n on all other days. Then, we define multi-day intrapersonal variability for person n as:

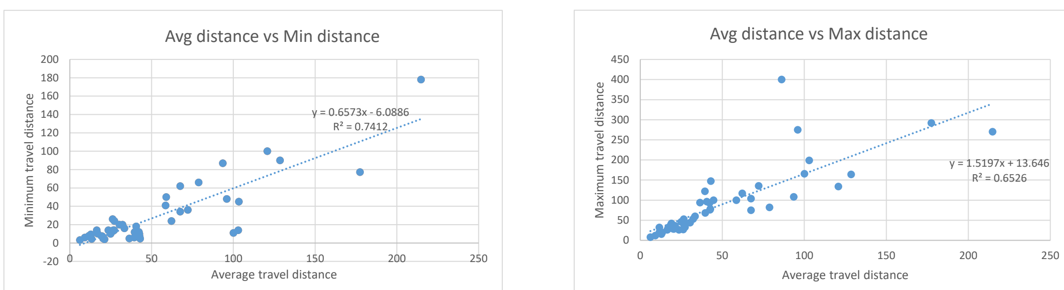
$$\text{MIV}(n) = \sum_{m=1}^{M} L(S_{n,m}, S_{n,I(n)}) \tag{3}$$

$$\text{where } I(n) = \arg \min_{m=1,\ldots,M} \sum_{m'=1}^{M} L(S_{n,m}, S_{n,m'}) \tag{4}$$

2.3.3 Multi-day intrapersonal distance variability

In addition to the intrapersonal variability of activity-travel patterns, 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 & Sundar 1995, Schlich & Axhausen 2003, Stopher et al. 2007, Chikaraishi, Fujiwara, Zhang, Axhausen & Zumkeller 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 observations. 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 a single-day travel distance is given.

We plot average travel distance, maximum travel distance as well as minimum travel distance of our testing 5-day dataset in the following Figure 1, and the trend line plots could reveal the relationship between different variables. Several outlier points in the original data have been discarded.



(a) Average vs maximum

(b) Average vs minimum

Figure 1: Multi-day travel distance variability

In Figure 1, 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:

$$d_{\text{max}} = 1.5197 d_{\text{avg}} + 13.646 \tag{5}$$

$$d_{\text{min}} = 0.6573 d_{\text{avg}} - 6.0886 \tag{6}$$

The R-squared values are both higher than 0.6, showing that our estimated equations fit the observed data well. The distance regression is to provide additional limits on travel distance. We assume that the travel distance observed on a single day will be the average travel distance of the traveler. With the regression, we can limit the maximum travel distance as well as the minimum travel distance of that traveler, and only the travel data satisfying the distance limits will be sampled as the multi-day data of that person. These can be applied as additional constraints during the later sampling process.

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

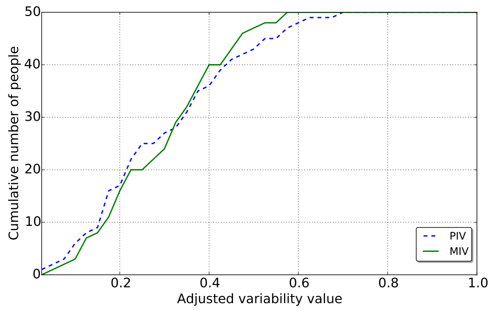
The connection between intrapersonal variability and interpersonal variability is helpful, when we estimate travelers' behaviors and generate multi-day samples from single-day travel data. We argue that the variability distribution in single-day PIV is similar to the variability distribution in MIV given the assumption that cross-sectional data for a group of people 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) in the above sections are plotted for the whole population in Figure 2a. Here, we use adjusted constant coefficient c to unitize PIV and MIV values so that they will fall in the range of $[0, 1]$. We multiply 1/360 as the adjusted constant coefficient for PIV and 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. We achieved the p-value of 0.9639 in the KS test, failing to reject the null hypothesis that the two samples are from the same distribution. In Figure 2, we also show the PIV/MIV comparison based on a subset of different numbers of workers randomly chosen from the original data. As we can see, the PIV curves are visually similar in all 5 cases, and KS tests all fail to reject the null hypothesis that the two samples are from the same distribution according to p-value.

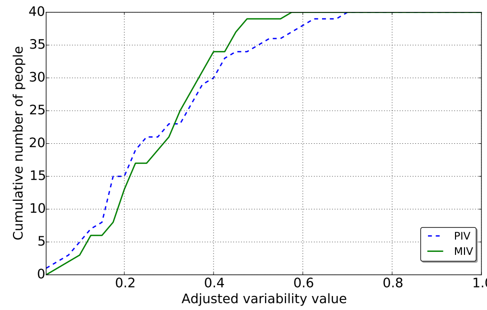
In addition, we show KS test results related to PIV comparison based on different sample sizes in Table 4. The number of workers in the sample is given in the first column, and all workers are randomly chosen from our 50 workers data set. In the second column, we present the p-values of KS tests between PIV and MIV for the same sample size. In the third column, we present the p-values of KS tests between PIV under different sizes and PIV under the case of $N = 50$. In the fourth column, we present the p-values of KS tests between PIV under different sizes and MIV under the case of $N = 50$. Thus, we can conclude that sample size will not affect the distribution of PIV since the $p-$values in the third column indicate that all PIV results are from the same distribution. According to the results from the fourth column, PIV values of different subsets are similar to MIV values from the whole population.

Table 4: $p-$value of KS test results based on PIV of various sample sizes

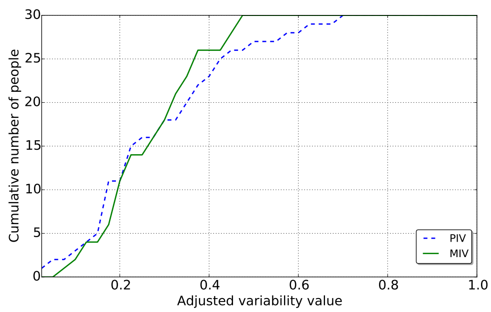
Sample size	vs Sample MIV	vs PIV(N=50)	vs MIV(N=50)
N=50	0.9639	0.999	0.9639
N=40	0.5726	0.999	0.6597
N=30	0.7989	0.999	0.8147
N=20	0.8186	0.969	0.9047



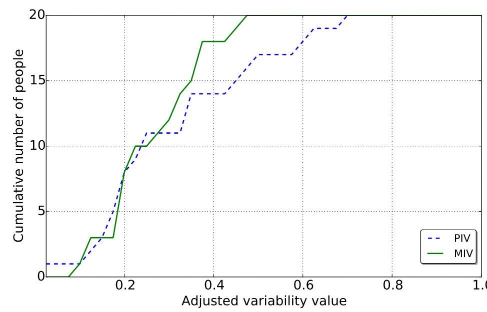
(a) Result of 50 workers, p-value= 0.9639



(b) Result of 40 workers, p-value= 0.5726



(c) Result of 30 workers, p-value= 0.7989



(d) Result of 20 workers, p-value= 0.8186

Figure 2: Comparison results based on sample size

The connection between the distributions of MIV and PIV can be attributed to each traveler's MIV contributing to PIV distribution. In order to explain this fact, we divide the population into three groups based on the values 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 indicates the variability between different days for each person. If MIV is low for a person, then his daily activity-travel patterns should be quite similar. However, if a person has a high MIV, his/her activity-travel pattern should 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* represents the "average" among all people since it is the closest one to other patterns. In other words, most activity-travel patterns should have similar 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. This person is a quite standard worker, and this is also the type of activity-travel pattern that appears the most in the data.

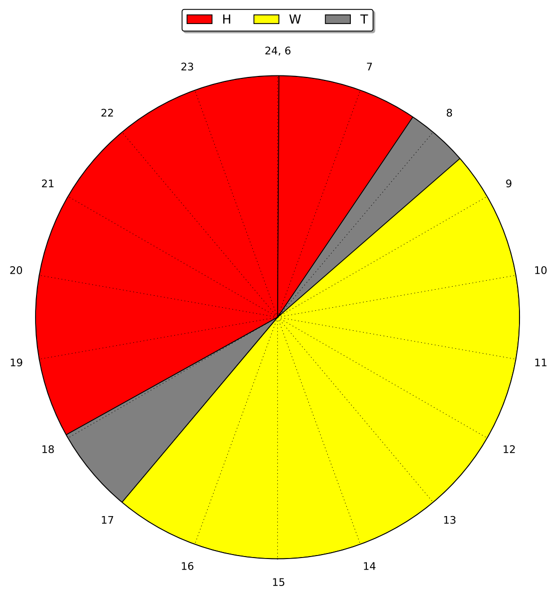


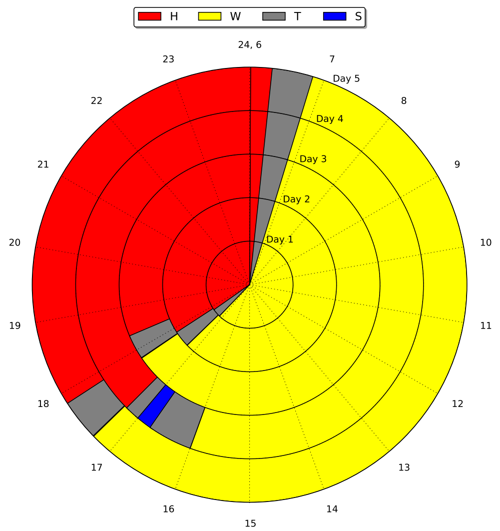
Figure 3: Standard activity-travel pattern of Person No.23 on a single day

3.1 Low MIV

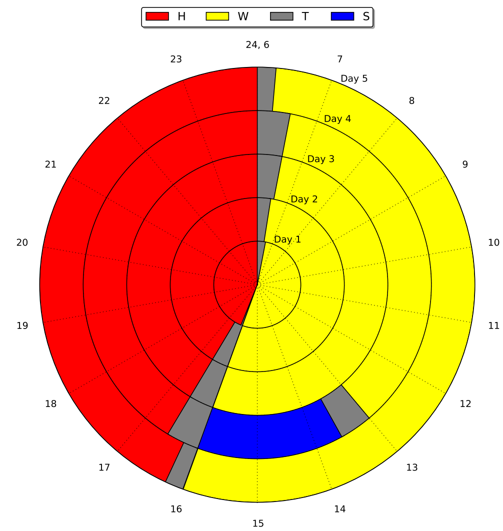
For people with low intrapersonal variability in Figure 4, they follow similar activity-travel patterns of a standard worker. They might have some difference in the time of arriving and departing from work places, which should slightly increase the variability between different people. They might go other places than work places, which will slightly increase both PIV and MIV. Thus, we can explain the feature in both cumulative curves on the left side of the horizontal axis. They have low MIV and PIV comparing to the standard pattern because they show standard working pattern. All people are workers 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 is positive for people in low MIV group since their activity patterns have little difference comparing to the standard pattern that affects both PIV and MIV.

3.2 Medium MIV

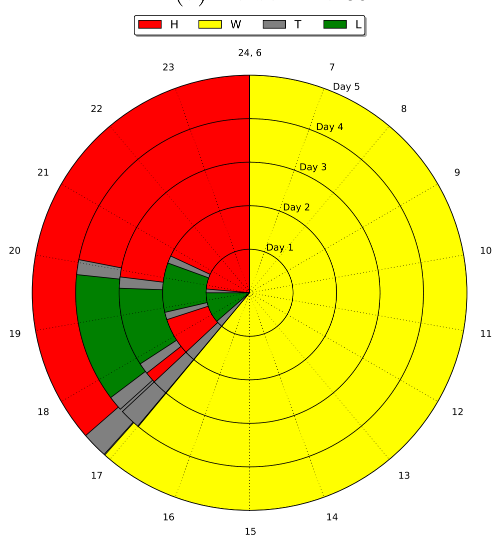
In addition to working, people also need to go for other activities like shopping and leisure. Thus, only a few of them have quite low MIV. However, for these people, their MIV would not be high since they only go to other places after work or on non-working days. We show two samples of medium MIV in Figure 5 where one person did not 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 patterns would be not significantly different from the "standard" since they spent most of the time on the activity type of working. For people with medium MIV, they can either have one day that is quite distinct from standard pattern or several days that are different from



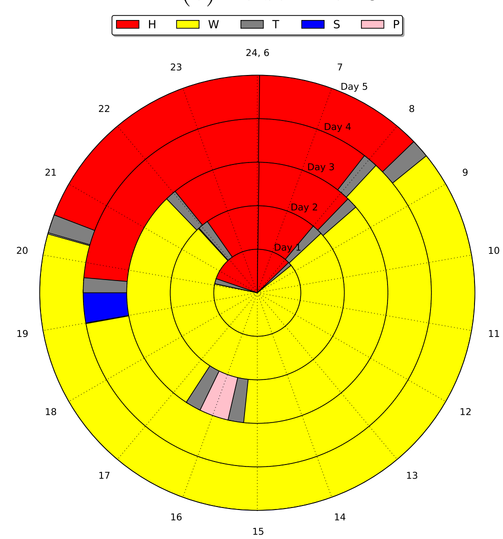
(a) Person No.39



(b) Person No.28

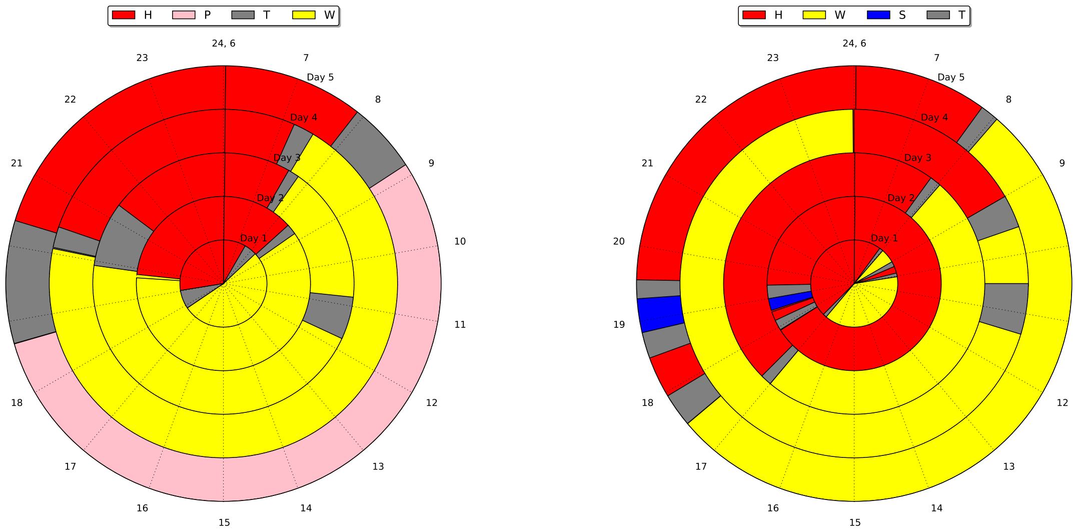


(c) Person No.8



(d) Person No.9

Figure 4: Samples of person with low MIV



(a) Person No.13

(b) Person No.31

Figure 5: Samples of person with medium MIV

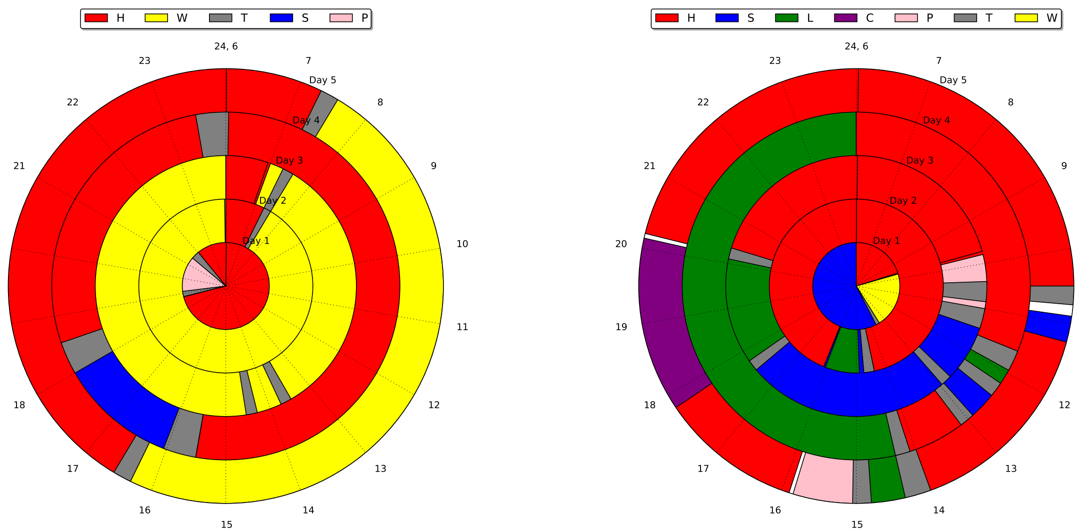
standard pattern. Although the difference falls in appropriate range, the correlation between PIV and MIV is still largely affected by random factors and the choice of standard pattern. Thus, we observe negative correlation among medium MIV group denoting the dependence between PIV and MIV.

3.3 High MIV

Persons in Figure 6 belong to high MIV group, and they have activity-travel patterns that are very different from PIV standard pattern on most days. Since one day out of the five days are recorded for each person in single-day data, there is a 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 validates the conclusion.

3.4 Remarks

In summary, MIV can be estimated based on PIV, and we should be able to see a good distribution of intrapersonal variability in single day data. The two distributions show similarity since they denote population wide information. Statistically, we have the correlation between adjusted PIV and MIV of 0.365 among the whole population, correlation of 0.388 among low MIV group, correlation



(a) Person No.32

(b) Person No.46

Figure 6: Samples of person with high MIV

of -0.129 among medium MIV group and correlation of 0.371 among high MIV group.

We make some remarks:

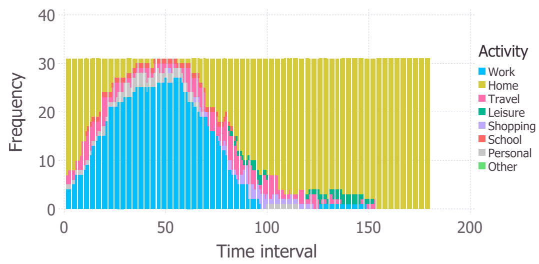
Low MIV: Typical full-time workers usually have very low MIV, since they spend most of their time working. They also have small PIV comparing to the "standard" pattern of working activity type. Note that the "standard" pattern could belong to a person with higher intrapersonal variability. People can 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 go working on that specific day.

Medium MIV: Most people 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.

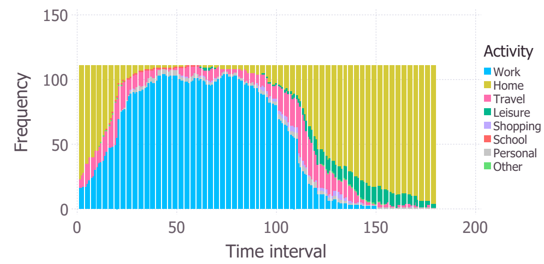
High MIV: People with high MIV are likely to have single-day activity-travel pattern that is very different from each other and PIV standard pattern, which could also lead to high PIV.

4 Sampling procedure

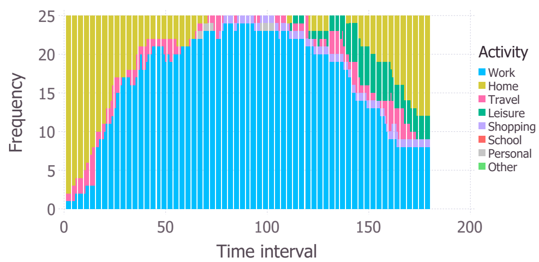
Our sampling method is designed to generate multi-day activity-travel patterns from single-day activity-travel patterns. The main idea is to pick different *single-day* activity-travel patterns from whole population based on the given personal single-day activity-travel patterns to construct reasonable personal *multi-day* activity-travel patterns.



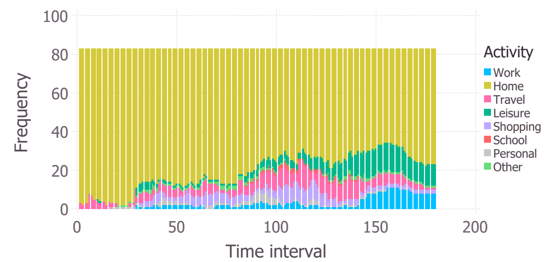
(a) Visualization for cluster 1



(b) Visualization for cluster 2



(c) Visualization for cluster 3



(d) Visualization for cluster 4

Figure 7: Activity-travel pattern visualization for all clusters

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). In the method, 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 (Kaufmann & Rousseeuw 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 results of the frequency of different activity types denoted by various 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 the longer work time. People in cluster 4 usually stay at 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 distributions for each group of people. Thus, we can obtain information about MIV distribution, if we know the cluster in which a person falls. We know cross-sectional data contains intrapersonal variability information, and adjusted PIV distribution is similar to adjusted MIV distribution. Thus, 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

Table 5: Transition probability matrix Ψ of the original dataset

Cluster	$c_{n,m-1} = 1$	$c_{n,m-1} = 2$	$c_{n,m-1} = 3$	$c_{n,m-1} = 4$
$c_{n,m} = 1$	0.24	0.155	0	0.101
$c_{n,m} = 2$	0.56	0.619	0.455	0.144
$c_{n,m} = 3$	0	0.083	0.273	0.058
$c_{n,m} = 4$	0.2	0.143	0.273	0.696

is available, clusters' MIV distribution can be applied during sampling instead of clusters' PIV distribution.

4.2 Transition probability

With clustering results, day-to-day intrapersonal activity pattern type transition probabilities can be calculated based on a multi-day dataset since day-to-day transferring information is given. If any form of multi-day data is not available, we also provide an estimation method to approximate the transition probability based on single-day dataset.

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 Ψ with each element being:

$$\Psi_{k,k'} = \frac{Q_{k,k'}}{\sum_{k_1=1}^{K} Q_{k_1,k'}} \tag{7}$$

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

The transition probability will give us the probability of transferring from one cluster to another cluster in general. The transition probability extracted from 5-day travel activity dataset of 50 employed people is shown in Table 5. Here, for element Ψ_{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 Ψ_{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. Given the fact that multi-day travel activity datasets might not be always available, we present an estimation of transition probability matrix based on single-day data as follows.

Suppose we have K clusters, we can reorder the clusters so that the cluster with more elements will have smaller index. Thus, cluster 1 will have more elements than other clusters. Given the ordered clusters $[C_1, ..., C_K]$, the number of elements in each cluster $[C_1^c, ..., C_K^c]$ as well as the distance between clusters $\iota(k, k')$. Here, we use average linkage $\iota(k, k') = \frac{1}{n_k n_{k'}} \sum_{n_k}^{i} \sum_{n_{k'}}^{j} L(x_{ki}, x_{k'j})$

as our distance measurement and $L(x_{ki}, x_{k'j})$ is the Levenshtein distance between two elements i and j from cluster k and k'.

Then we can first generate a $K \times K$ matrix Ψ' where $\Psi'_{k,k} = f(C_k^c)$ and $\Psi'_{k,k'} = \frac{\min_{i \neq k'} \iota(i,k')}{\iota(k,k')}$ to estimate how likely a cluster k will transfer to cluster k' based on the distance between these two clusters. Here, $f(\cdot)$ is an increasing function and can be flexibly defined. We assume that if there are more elements in a cluster, the cluster is more habitual in average and it is more likely to transfer to itself. Then, we can update the matrix Ψ' by linearly adjusting each column so that column summation is 1 to satisfy the axiomatic definition of probability.

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 the 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 Ψ 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 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 $\widetilde{S}_{n,m}, m = 2, 3, ..., M$, since we can use the original single-day data as the first day data in our M-day sample. Thus, we can construct an M-day sample $\widetilde{S}_{n,m}, m = 1, 2, 3, ..., M$ for each person $n = 1, 2, ..., N$. Since we have the clustering results $c_{n,1}$ for $\widetilde{S}_{n,1}$, we can generate all $c_{n,m}, m = 2, ..., M$ based on the transition probability Ψ and the former day's clustering result $c_{n,m-1}$ for the same person. With the clustering results $c_{n,m}$ for multi-day sample and intrapersonal variability MIV(n) for person n, we can generate a sample pool from original data. Only activity-travel patterns that fall in the $c_{n,m}$-th cluster 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 $\widetilde{S}_{n,m}$ 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 results based on transition probability as well as the 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 results and corresponding MIV cdf.

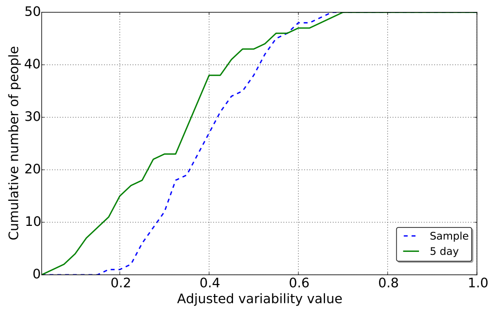
Step 6. Additional limits could be applied on activity-travel patterns in sample pool based on the original single-day travel distance and constraint function of $f(\cdot)$ and $g(\cdot)$ for the maximum and minimum travel distance, respectively.

Step 7. Randomly choose a single-day activity-travel pattern for each person on each day until our multi-day dataset is fully constructed.

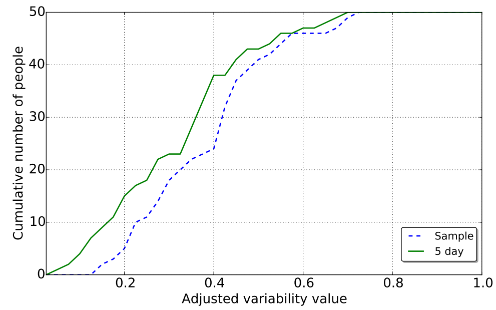
In this paper, we also define a *single-day trivial method* that duplicates single-day's travel activities for multi-days to generate multi-day data. In that case, people's trips will always be the same on each day.

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 patterns 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 match by day, 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



(a) MIV cumulative distribution with distance limit



(b) MIV cumulative distribution without distance limit

Figure 8: MIV results of our sample and original data

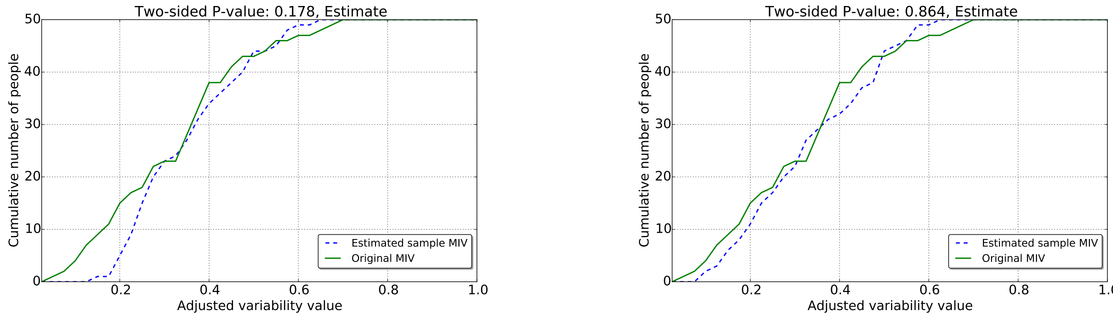
for the population instead of one individual person, we think the distribution of clusters is more important than the cluster order. So we provide two methods to validate the performance of our sampling method to show whether the generated 5-day activity-travel pattern represents the variability observed 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. The distribution shows how the MIV variability is distributed among the whole population based on original or sample data. This will provide us with the insight of the number of people with MIV variability that falls in some range. Given the fact that we are sampling randomly based on original data, the day-of-week is reordered. Since Pas (1988) mentioned that day-of-week is independent of daily travel activity type selection, the general property of our generated sample data should be similar to the original data. Thus, the general distribution of the whole population will be more meaningful and will be able to indicate the similarity to the original data. 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 0.1122 in KS test, 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 0.1777 in KS test, failing to reject the null hypothesis that the two data samples are from the same distribution. These statistics show that our sampling method performs well while generating samples from single-day travel dataset with consideration of intrapersonal variability.

We also verify the performance of our estimated transition probability matrix. For our current case of 4 clusters, we take the diagonal elements of Ψ' as $[4, 2, 0.5, 0.25]$ to consider different weights for different clusters, and take $1/240$ as the adjusted constant coefficient c for our estimated sample MIV. In the following Figure 9, we show the results for comparing 50 workers' original MIV and sample MIV for two randomly generated samples. We can see that both samples perform well in the KS test, showing that our estimated transition probability matrix can generate a reasonable good sample.



(a) Estimated sample 1 MIV comparison result (b) Estimated sample 2 MIV comparison result

Figure 9: MIV results based on estimated transition probability

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

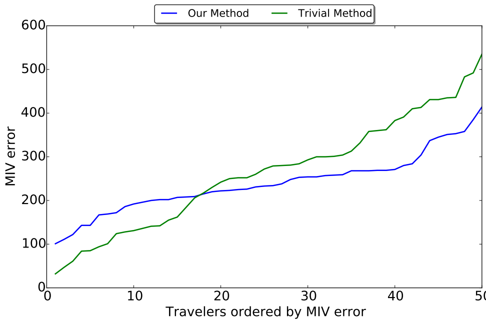
Since we want to compare the differences between original multi-day data and sampled multi-day data based on the intrapersonal variability, it is reasonable to define the difference between one person's original multi-day activity-travel pattern and another person's sample multi-day activity-travel pattern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

Suppose we have the original M-day activity-travel pattern data $P_n = P_{\text{multi-day},n} = [S_{n,1}, S_{n,2}, ..., S_{n,M}]$ for person n and sample M-day activity-travel pattern data $\widetilde{P}_n = \widetilde{P}_{\text{multi-day},n'} = [\widetilde{S}_{n',1}, \widetilde{S}_{n',2}, ..., \widetilde{S}_{n',M}]$ for person n'. We can generate all possible permutations of $\widetilde{P}_n$ with a total number of $M!$. We let $\texttt{perm}(\widetilde{P}_{n'})$ denote the collection of all permutations of elements in $\widetilde{P}_{n'}$, and $\texttt{perm}_i(\widetilde{P}_{n'})$ 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 $\widetilde{P}_{n'}$ as follows:

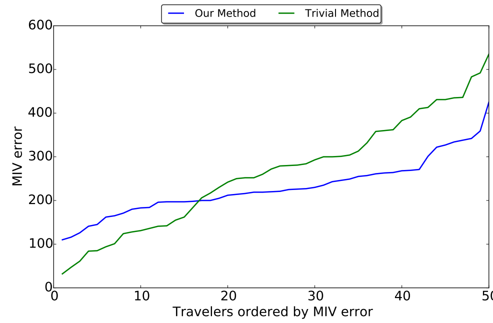
$$MIVE(P_n, \widetilde{P}_{n'}) = \min_{1 \leq i \leq M!} VD(P_n, \texttt{perm}_i(\widetilde{P}_{n'})) \tag{8}$$

$$\text{where } VD(P_n, \widetilde{P}_{n'}) = \sum_{m=1}^{M} L(S_{n,m}, \widetilde{S}_{n',m}) \tag{9}$$

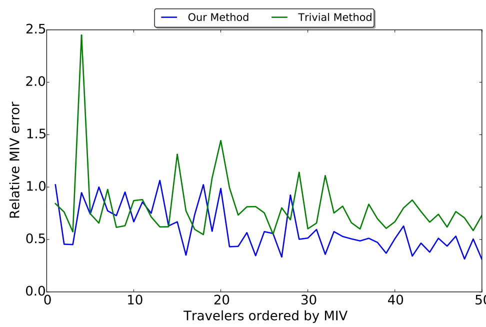
Suppose we have an M-day activity-travel pattern dataset $P_{\text{multi-day},n}$ of N people as well as the corresponding multi-day sample $\widetilde{P}_{\text{multi-day},n}$. Since our new validation method needs to find a day-to-day-person-to-person match with the least MIV error difference, we can first formulate a MIV error matrix E, where each element $e_{n,n'} = MIVE(P_n, \widetilde{P}_{n'})$ denotes 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



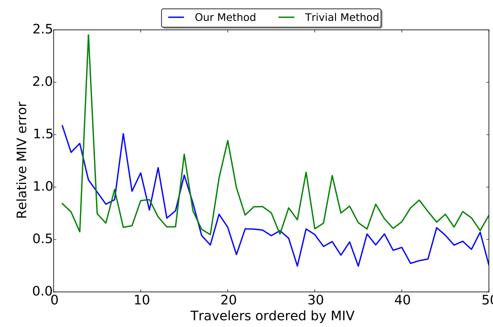
(a) Absolute error including distance limit



(b) Absolute error excluding distance limit



(c) Relative error including distance limit



(d) Relative error excluding distance limit

Figure 10: MIV error validation results

original person n to sample person n' is $e_{n,n'}$. Thus, we can define a variable $x_{n,n'} = 1$ to denote the person n matching with the person n', otherwise $x_{n,n'} = 0$.

$$\min \sum_{n=1}^{N} \sum_{n'=1}^{N} e_{n,n'} x_{n,n'} \tag{10}$$

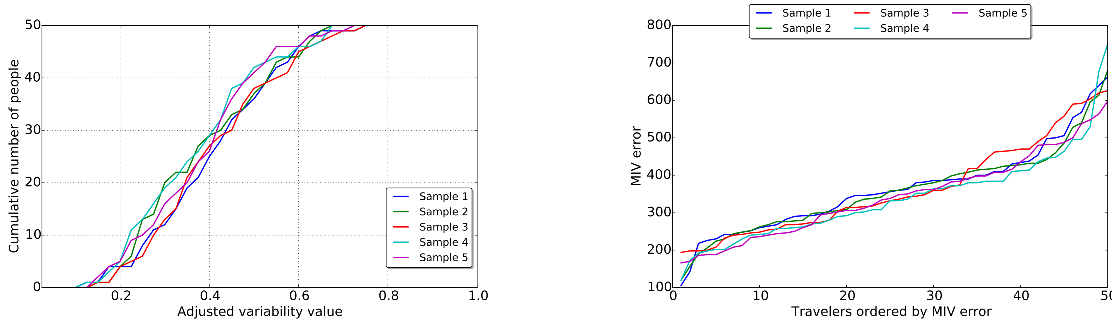
$$\sum_{n=1}^{N} x_{n,n'} = 1 \qquad \forall n' \tag{11}$$

$$\sum_{n'=1}^{N} x_{n,n'} = 1 \qquad \forall n \tag{12}$$

$$x_{n,n'} \in \{0,1\} \qquad \forall n, n' \tag{13}$$

We can solve this problem with an optimization solver such as CPLEX, and obtain the minimum MIV error match of sample and original data. The MIV comparison results are as follows in Figure 10. Here, we have 'Our Method' as our proposed sampling method as well as error considering population wide distribution and 'Trivial Method' as duplicated multi-day sampling method. In Figures 10a and 10b, we order the person by absolute MIV error while we show the trend of relative MIV error (absolute MIV error/MIV) in Figures 10a and 10b with person ordered by MIV.

We can see that the MIV error is smaller based on our sampling method for most people with



(a) MIV Cumulative distribution without distance limit

(b) Absolute error excluding distance limit

Figure 11: Multiple sample comparison results

higher MIV during their travels, since the trivial sampling method without considering day-to-day intrapersonal variability failed to estimate MIV well. For some people with low MIV, their travel activities can be well estimated by the trivial method since their activity patterns are similar in 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 samples 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 11, 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 differences between various samples are 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. In conclusion, our sampling method is able to provide random samples with stable MIV given same single-day travel activity data.

6 Conclusion & Future works

In this paper, we employ 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 and multi-day data. Given this evidence, we develop our sampling method to generate multi-day samples based on single-day travel data by clustering travelers into different groups. We also include the transition probability between cluster types as well as the daily travel distance to provide more accurate sampling results. An estimation method is developed to provide transition probability based on single-day data only. Although only a few longer non-habitual trips can be captured based on our current assumptions, our sampling method might still be applicable in charging infrastructure studies or other studies where mainly daily habitual trips of travelers are considered.

These generated multi-day activity-patterns can be used to 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 among whole population instead. Our multi-day sample data perform well comparing to the estimation results of single-day data although the multi-day sample data tend to overestimate the intrapersonal variability of some people. Our sampling method helps to extract more information of day-to-day intrapersonal variability contained in single-day cross sectional data.

Although trivial method could help generate a good sample of people with low MIV, we can not identify the group of low MIV people just by his/her single day activity-travel data. Social characteristic information could be helpful, and we might generalize our analysis by considering social characteristic information during clustering in future research. In addition, more accurate estimation of variability 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. Lastly, we believe our proposed method works for other types of data sets when the dissimilarity is defined for the specific information available from the data that are useful for one's focus. For example, the proposed comparison can be analyzed and sampling method can be applied to GPS-type trajectory data if (dis)similarity is defined accordingly (e.g., staying vs traveling, spatial similarity) depending on application.

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. We would improve our sampling method and include other trips of non-workers as well as trips during weekends to consider the effect of non-habitual trips if larger datasets are available in the future.

7 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

Allahviranloo, M., Regue, R. & Recker, W. (2014), Pattern clustering and activity inference, *in* 'Transportation Research Board 93rd Annual Meeting', number 14-1274.

Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfüser, G. & Haupt, T. (2002), 'Observing the rhythms of daily life: A six-week travel diary', *Transportation* **29**(2), 95–124.

Buliung, R. N., Roorda, M. J. & Remmel, T. K. (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., Ma, J., Susilo, Y., Liu, Y. & Wang, M. (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., Fujiwara, A., Zhang, J. & Axhausen, K. (2009), 'Exploring variation properties of departure time choice behavior by using multilevel analysis approach', *Transportation Research Record: Journal of the Transportation Research Board* (2134), 10–20.

Chikaraishi, M., Fujiwara, A., Zhang, J. & Axhausen, K. W. (2011), 'Identifying variations and co-variations in discrete choice models', *Transportation* **38**(6), 993–1016.

Chikaraishi, M., Fujiwara, A., Zhang, J., Axhausen, K. & Zumkeller, D. (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.

Ebadi, N., Kang, J. E. & Hasan, S. (2017), 'Constructing activitymobility trajectories of college students based on smart card transaction data', *International Journal of Transportation Science and Technology* **6**(4), 316 – 329.

Elango, V., Guensler, R. & Ogle, J. (2007), 'Day-to-day travel variability in the commute atlanta, georgia, study', *Transportation Research Record: Journal of the Transportation Research Board* (2014), 39–49.

Garikapati, V. M., Pendyala, R. M., Morris, E. A., Mokhtarian, P. L. & McDonald, N. (2016), 'Activity patterns, time use, and travel of millennials: a generation in transition?', *Transport Reviews* **36**(5), 558–584.

Hanson, S. & Huff, J. O. (1981), 'Assessing day-to-day variability in complex travel patterns', *Transportation Research Record* **891**, 18–24.

Hanson, S. & Huff, O. J. (1988), 'Systematic variability in repetitious travel', *Transportation* **15**(1-2), 111–135.

Joh, C.-H., Arentze, T. A. & Timmermans, H. J. (2001), 'Multidimensional sequence alignment methods for activity-travel pattern analysis: A comparison of dynamic programming and genetic algorithms', *Geographical Analysis* **33**(3), 247–270.

Joh, C.-H., Arentze, T., Hofman, F. & Timmermans, H. (2002), 'Activity pattern similarity: a multidimensional sequence alignment method', *Transportation Research Part B: Methodological* **36**(5), 385–403.

Kang, H. & Scott, D. M. (2010), 'Exploring day-to-day variability in time use for household members', *Transportation Research Part A: Policy and Practice* **44**(8), 609–619.

Kaufmann, L. & Rousseeuw, P. (1987), 'Clustering by means of medoids'.
URL: *https://www.researchgate.net/publication/243777819_Clustering_by_Means_of_Medoids*

Koppelman, F. & Pas, E. (1984), Estimation of disaggregate regression models of person trip generation with multiday data, *in* '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.

Levenshtein, V. I. (1966), Binary codes capable of correcting deletions, insertions, and reversals, *in* 'Soviet physics doklady', Vol. 10, pp. 707–710.

Lyons, G. (2015), 'Transport's digital age transition', *Journal of Transport and Land Use* **8**(2), 1–19.

MacQueen, J. et al. (1967), Some methods for classification and analysis of multivariate observations, *in* 'Proceedings of the fifth Berkeley symposium on mathematical statistics and probability', Vol. 1, Oakland, CA, USA., pp. 281–297.

McDonald, N. C. (2015), 'Are millennials really the "go-nowhere" generation?', *Journal of the American Planning Association* **81**(2), 90–103.

Medina, S. A. O. (2016), 'Inferring weekly primary activity patterns using public transport smart card data and a household travel survey', *Travel Behaviour and Society* .
URL: *https://doi.org/10.1016/j.tbs.2016.11.005*

Moiseeva, A., Timmermans, H., Choi, J. & Joh, C.-H. (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. & Koppelman, F. S. (1986), 'An examination of the determinants of day-to-day variability in individuals' urban travel behavior', *Transportation* **13**(2), 183–200.

Pas, E. I. & Sundar, S. (1995), 'Intrapersonal variability in daily urban travel behavior: some additional evidence', *Transportation* **22**(2), 135–150.

Pendyala, R., Parashar, A. & Muthyalagari, G. (2001), Measuring day-to day variability in travel characteristics using gps data, *in* '79th annual meeting of the Transportation Research Board'.

Pendyala, R. & Pas, E. (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* pp. II–B/1 – II–B/22.

Recker, W. W., McNally, M. G. & Root, G. S. (1985), 'Travel/activity analysis: pattern recognition, classification and interpretation', *Transportation Research Part A: General* **19**(4), 279–296.

Schlich, R. & Axhausen, K. W. (2003), 'Habitual travel behaviour: evidence from a six-week travel diary', *Transportation* **30**(1), 13–36.

Stopher, P., FitzGerald, C., Zhang, J. & Bretin, T. (2007), 'Analysis of 28-day global positioning system panel survey: Findings on variability of travel', *Transportation Research Record: Journal of the Transportation Research Board* (2014), 17–26.

Susilo, Y. & Kitamura, R. (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. & Axhausen, K. W. (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., Kang, J. E. & Chen, R. (2017), 'A random utility based estimation framework for the household activity pattern problem', *Transportation Research Procedia* **23**, 809–826.

Appendix

A Summary of previous studies on intrapersonal variabilities

Here is a summary of the previous studies on intrapersonal variabilities in Table 6.

Table 6: Results of various measure of day-to-day intrapersonal variability in previous studies

Paper	Data Source	Intrapersonal variability standard	Large proportion of total variance	Major findings
Pas & 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 & Sundar (1995)	North King County, Washington (1989)	trip frequency, daily travel time, trip chaining	Yes	similar magnitude under different standards
Schlich & 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 & Kitamura (2005)	Mobidrive in Germany (1999)	action space(the second moment of of the out-of-home activity locations)	No	workers and students more stable
Elango et al. (2007)	Commute Atlanta study (2004)	delta trips	Not mentioned	significant demographic variables effects, day-of-week effects, less seasonal effects
Buliung et al. (2008)	Toronto Travel Activity Panel Survey	minimum convex polygon (MCP) 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. (2007)	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 & Scott (2010)	Toronto Travel-Activity Panel Survey with GIS toolkit (2003)	activity time-use patterns	Yes for weekdays, No for weekends	joint activities(interact with other household members) have higher proportion of intra-var
Chikaraishi, Fujiwara, Zhang, Axhausen & Zumkeller (2011)	German Mobility Panel survey data (1999-2008)	travel time expenditure	Yes	situational attributes dependent, longer observation time is important
Susilo & Axhausen (2014)	Mobidrive in Germany (1999)	repetitiveness of activity-travel patterns	Not mentioned	individual's behavioral choices dependent, activity type affects repetition pattern
Moiseeva et al. (2014)	GPS data in Eindhoven, Netherlands (2010)	activity-travel pattern	No	sociodemographic characteristics dependent, intra-variability reflect environment learning speed