USING GPS TECHNOLOGIES TO COLLECT MULTIDAY TRAVEL DATA

Govardhan R. Muthyalagari
Department of Civil and Environmental Engineering
University of South Florida, ENB 118
4202 East Fowler Avenue
Tampa, FL 33620-5350
Ph: 813-974-8727
Fax: 813-974-2957
Email: muthyala@eng.usf.edu

Abhishankar Parashar
Department of Civil and Environmental Engineering
University of South Florida, ENB 118
4202 East Fowler Avenue
Tampa, FL 33620-5350
Ph: 813-974-8727
Fax: 813-974-2957
Email: parashar@eng.usf.edu

Ram M. Pendyala*
Department of Civil and Environmental Engineering
University of South Florida, ENB 118
4202 East Fowler Avenue
Tampa, FL 33620-5350
Ph: 813-974-1084
Fax: 813-974-2957
Email: pendyala@eng.usf.edu

*Corresponding author

December 2000

ABSTRACT

GPS-based data collection methods offer considerable potential for collecting multiple days of travel information without placing undue burden on travel survey respondents. Multiday travel data are valuable for analyzing day-to-day variability in travel behavior, which is an extremely important aspect of travel behavior from a modeling, data collection, and policy analysis perspective. However, there is very limited knowledge regarding the application of GPS-based methods to collect multiday travel information. Utilizing data from the Lexington, Kentucky GPS experiment, this paper aims to examine the potential of GPS to measure travel characteristics over multiple days. A major finding of this paper is that the GPS-based experiment provides indications of within-person day-to-day variability in travel consistent with that reported in the literature. Thus, it appears that GPS-based technologies offer considerable potential for collecting multiday travel data. Some of the differences that were found may partially be attributed to socio-economic differences across samples and partially to the greater ability of GPS-based technologies to capture short and infrequent trips consistently across multiple days. Future research efforts should attempt to determine the extent to which each of these factors accounts for the differences. Overall, the paper shows that GPS-based technologies offer considerable potential for collecting multiday travel data.

Keywords: multiday travel data, GPS technology, travel characteristics, inter-person variability, intra-person variability

INTRODUCTION

The collection of multiday travel data has been an issue of much interest to travel behavior researchers for several decades. More often than not, travel demand models are estimated on travel survey data sets that are collected for a one-day travel period. Researchers and planners have questioned whether a one-day data set captures the full range of travel that is undertaken by an individual as there are many activities (say, personal business, shopping, social visits, medical/dental, etc.) that are not necessarily done on a daily basis.

Despite the recognition of the importance of multiday travel data, many planners have shied away from conducting multiday personal travel surveys because of the potential ill-effects of survey fatigue and low response rates that are typically attributed to longer survey durations. Then, it may be conjectured that an alternative travel data collection methodology that does not impose undue burden on survey respondents would offer great benefits for planners interested in collecting multiday travel data.

The Federal Highway Administration (FHWA) recently conducted a pilot experiment in which one vehicle in each of 100 households in the Lexington, KY area was fit with a Global Positioning System (GPS) unit and a small hand-held computer. These devices afforded the opportunity to track the movements of individual vehicles over a target period of six days, but for which some were tracked as along as eight or nine days, while minimizing respondent burden. The GPS-based data collection methodology also provides the added benefit of capturing information on short and infrequent trips, route choice, and speed that is typically more burdensome to collect in a traditional personal recall survey.

This paper focuses on examining the potential use of GPS-based technologies to collect multiday travel data with a view to minimizing respondent burden. The specific objectives of this research are to:

- examine measures of day-to-day variability in travel behavior as obtained from a GPS-based travel data set, and

- compare indicators of day-to-day variability obtained from the GPS-data set against those reported in the literature to assess the potential of GPS-based survey methodologies for collecting multiday travel data.

The remainder of this paper is organized as follows. The next section provides a detailed review of research on the importance of multiday data in travel demand analysis. This review is followed by a brief description of the Lexington, Kentucky GPS travel data collection experiment. The fourth section provides an exploratory comparison of measures of travel characteristics across days of the week. The fifth section presents an analysis of the degree of repetition seen in the GPS survey sample with respect to various travel characteristics. In the sixth section, within person variability as measured by the GPS survey is quantified and then compared against that reported in the literature. Finally, conclusions regarding the potential of GPS based technologies for multiday travel data collection and directions for further research are provided in the final section.

IMPORTANCE OF MULTIDAY TRAVEL DATA

Over the past several decades, increasing attention has been paid to the collection and use of multiday travel survey data to capture the variation in travel behavior. However, a review of the literature and planning practice indicates that the predominant procedure for the conduct of household travel surveys has been the traditional "one-day travel diary" approach. Early examples of multiday surveys include the 1971 household travel survey in Uppsala, Sweden that collected travel information for a 35-day period [Hanson and Huff, 1982] and the 1973 seven day activity survey conducted in Reading, England [Pas, 1986]. Since then, many multiday travel surveys have been conducted around the world with a view to obtaining multiple days of travel information.

An excellent discussion on the importance of collecting multiday information is provided by Jones and Clarke [1988]. They note that as the emphasis of transportation planning shifted from capacity expansion to travel demand management, some of the issues facing transportation planners could not be addressed by one-day data (regardless of the sample size) because “by their nature, they are questions about variations in behavior over time”. They point out that multiday data provides information about the distribution of the frequency of participation, in addition to the mean participation rate. Jones and Clarke [1988] indicate that information on the frequency of participation allows the planner to gauge the exposure of different demographic and travel segments to various policy scenarios (say, over the course of a week or month). In essence, the measurement of travel behavior over multiple days has important policy implications in the current planning context.

Even when policy considerations do not necessarily require the use of multiday travel survey data, efficiency considerations may call for the collection and use of such data. Pas [1986] shows that substantial economies can be achieved with respect to survey cost and model parameter efficiency if a multiday sample, as opposed to a one-day sample, is employed in travel demand analysis. Koppelman and Pas [1984] report on an examination of linear regression models of trip generation estimated using multiday data. They find that the estimators of multiday models are more efficient than those of the single day model. For example, using five-day data from the 1973 Reading, England survey, they find that the standard errors of parameter estimates are reduced by 31 percent through the use of a five day sample rather than a single day sample.

Hanson and Huff conducted several studies using multiday travel data from the 1971 Uppsala (Sweden) household travel survey [Hanson and Huff, 1982, 1986, 1988a, 1988b; Huff and Hanson, 1986, 1990]. The Uppsala survey obtained information on all out-of-home travel-activity behavior using self-administered travel diaries for a 35-day period. Hanson and Huff used a representative sample of 149 individuals who completed diaries for the entire 35-day period to examine the day-to-day and week-to-week variability in travel behavior. They note that some behaviors, when examined in a disjointed framework (say, a work trip examined in isolation of the overall daily activity-travel pattern), are repeated on a day-to-day or week-to-week basis. However, when the overall daily activity-travel pattern is examined in its entirety, they find that a one-day pattern is not representative of a persons travel behavior.

Much of the work reported by Pas and his colleagues in a series of papers [Pas, 1986, 1987, 1988; Koppelman and Pas, 1984; Pas and Koppelman, 1987] utilizes seven-day activity data collected in 1973 in Reading, England. More recently, Pas and Sundar [1995] extended the work done previously by examining day-to-day variability for different travel indicators and across household members using three-day travel diary data collected in 1989 in Seattle. All of their work shows that there is substantial intra-person multiday variability in travel behavior. For example, Pas [1987] found that about 50 percent of the total variability in trip-making in the Reading data set could be attributed to intra-personal day-to-day variability in trip generation. Pas and Koppelman [1987], in an extension of this work, found that the level of intra-personal variability varies significantly across demographic segments. For example, they found that females exhibit higher levels of intra-personal variability than males, possibly due to the roles traditionally played by females in households.

Several other studies have examined variability in travel characteristics across multiple days. Examples include Kitamura and van der Hoorn [1987] and Hirsh, et. al. [1986]. A series of papers by Mahmassani and his colleagues [Mahmassani, 1997; Mahmassani and Chang, 1985, 1986; Mahmassani and Stephan, 1988; Mahmassani and Herman, 1990] study the day-to-day departure time choice dynamics under congested network conditions. Their work and that of Mannering [1989] show that both departure time choice and route choice show variability across days, but that variability in departure time choice is greater than that of route choice.

In summary, a vast body of knowledge and evidence supports the notion that there is considerable information and insights to be gained by collecting multiday travel data. Considering that the collection of multiday travel information may be difficult both from a survey administration and a respondent burden standpoint, the potential of using GPS-based technologies to collect such information is appealing. It is in this context that this paper examines the potential of using GPS-based technologies to collect multiday travel data.

LEXINGTON GPS TRAVEL SURVEY

This section provides a brief overview of the GPS-based travel survey experiment that was conducted by Federal Highway Administration (FHWA) in Lexington, Kentucky area of the United States. Much of the material in this chapter is derived from a report prepared by Battelle [1997] that includes detailed descriptions of the experiment and the databases assembled.

The GPS-based travel survey experiment required the participation and cooperation of a Metropolitan Planning Organization (MPO). The selected MPO would need to have an up-to-date and accurate digital map of the test area and its highway network, with a minimum accuracy satisfying Federal National Map Accuracy Standards. In addition, the participating MPO would have to be able to provide about 250-300 hours of total staff support during the field experiment. Following a nationwide solicitation and selection process, the Lexington, Kentucky area MPO was chosen as the host MPO for the field test. The Lexington area MPO is the principal planning agency for the two-county area of Fayette and Jessamine in Central Kentucky. The two counties together encompass an area of approximately 461 square miles with a total population of approximately 350,000.

Considering the specialized nature of the GPS-based field experiment, it was not possible to utilize a pure random sample for the survey. A special recruitment process was used to obtain a sample of 100 participating households for the experiment. A sample of nearly 2,000 listed telephone numbers from households in Fayette and Jessamine counties (weighted by population) was purchased from a commercial source. The participating households were recruited using a stratified sampling plan based on demographic factors. The demographic factors considered included gender, age, and presence of children in the household. Licensed drivers under the age of 18 were not permitted to participate as the principal eligible driver within a household. In addition, principal eligible drivers recruited for the field test were required to drive at least three days a week. This was done to ensure that a reasonably large amount of data would be collected despite the small sample size and short duration of the experiment.

The recruiting process started by mailing out pre-solicitation letters without any prior contact with the households. Once the letters were mailed, telephone contacts began with the households. If the household was responsive to the initial call, they were asked to participate in a brief screening survey to determine their eligibility for the field test. Following the screening interview, a pre-usage interview was conducted to ensure that the household would be able to use the field equipment without any problems.

A pre-solicitation letter was sent to a total of about 1,300 households with listed telephone numbers. Once the telephone interviewers determined that there was an eligible driver and vehicle in the household, 67 percent of those eligible consented to participate in the field test. At the end of the recruitment process, a total of 100 households were successfully recruited for participation in the field test.

The field equipment consisted of the following individual components:

- Hand-held computer: The hand-held computer was a Sony MagicLink PIC-2000 personal digital assistant (PDA), with a backlit touch screen user interface.

- GPS Receiver: The GPS receiver was a Garmin TracPak-30 equipped with a magnetic roof mount or a suction cup device for mounting inside the windshield of a car.

- PTS Software: The Personal Travel Survey (PTS) software was a user-friendly interface software that controlled the recording of GPS data and allowed respondents to enter trip information.

- Accessories: A SRAM PCMCIA memory card containing the PTS application software and upto 2 MB of memory for data collection was provided. In addition, connecting cables that enabled communication among the devices were also furnished to respondents.

The Personal Travel Survey (PTS) software had two primary functions. First, it allowed respondents to easily enter trip information and second, it facilitated recording of GPS positional data for each trip. It consisted primarily of three interfaces – administrative, GPS, and respondent.

The administrative interface allowed the field test administrator to set the operational parameters of the data collection devices and personalize the respondent interface for each household. Household users were not provided access to this interface. The GPS interface displays raw GPS output as it is being written to memory in the computer. This interface provides the mechanism by which the GPS receiver data stream is written and stored in memory. This interface also does not have any user control.

The respondent interface is the major feature of the PTS software that respondents control and utilize throughout the field experiment. The interface allows drivers to use a touch-screen to initiate a trip, identify the driver and his or her activity (trip purpose), identify passengers and their activities (trip purposes), and end a trip. Time, speed, and positional data are recorded directly through the GPS receiver.

In general, both the hardware devices and software interfaces performed well in the field. There were very few instances of hardware or software malfunction that called for corrective action. Further details about the hardware and software performance can be found in Battelle [1997].

OVERVIEW OF SURVEY SAMPLE

The field experiment yielded a total of four databases containing personal travel data and vehicle movement profiles. The four databases are as follows:

- PDA Data: This data set consisted of information and statistics derived directly from the data recorded in the MagicLink Personal Digital Assistant (PDA). Information in this data set included trip start and end times, trip occupancy, and trip purpose. All information, except trip start and end times, was input into the device by individual respondents. Each record in the data set represents a unique trip.

- GPS Data: This data set consisted of information and statistics derived directly from the data recorded through the GPS receiver. Information in this data set included trip start and end times, positional data (latitude and longitude), speed, and calculated trip distance. Positional data was recorded every 2-3 seconds.

- Match Data: This data set consisted of information derived from post-processing the GPS data in conjunction with the GIS-based travel network. Information in this data set included trip start and end times, network link identification, highway functional class, and trip distances.

- Recall Data: This data set consisted of information and statistics derived directly from the post-usage interviews of the Lexington field test respondents that were conducted by telephone following the field experiment. This data set consists of self-reported information for one day of travel (corresponding to a day on which the respondent used the GPS and PDA devices). Information collected on a recall basis for trips undertaken in a 24-hour period included trip start and end times, trip distance, destination, mode, and purpose.

In addition to the detailed travel-related information, data was also available for selected socio-economic variables. Information on household size, number of children under 16 years, vehicle ownership, number of licensed drivers, age and gender of drivers, and household income were also available in the data sets. However, it was found that there was some missing data for these variables thus limiting the extent to which these variables could be used for analysis.

As this paper focuses on the collection and analysis of multiday travel data using GPS technologies, the one-day recall data set derived from the post-usage telephone interview is not used within the scope of this study. A comparison between the GPS and recall trip data sets (documented in a previous paper; see Yalamanchili, et. al., 1999) showed that the GPS data set successfully recorded the travel undertaken by the principal eligible driver over the course of a day. In fact, the comparison showed that the GPS data set recorded short and linked trips that were not recorded in the recall trip data set. As such, for the sample of principal eligible drivers, it appears that the GPS data set is a complete record of their travel. The question being addressed in this paper is whether GPS-based data sets also offer robust multiday travel information with minimal increase in respondent burden.

A total of 100 households participated in the Lexington, KY GPS experiment. Data collection devices were installed in one vehicle for each household. Any licensed driver in the household was permitted and encouraged to use the GPS-fitted vehicle. Within the context of this paper, one driver in each household was identified as a primary driver. The driver who provided travel information for the most number of days in each household was chosen as the primary driver. If two household members provided travel information for the same number of days, then the driver that made the larger number of trips using the GPS-fitted vehicle was designated as the primary driver for each household. Thus, there are 100 primary drivers in the final sample used for analysis in this paper. It is to be noted that this sample may differ from that used in other publications (Zhou and Golledge, 2000; Battelle, 1997; and Murakami and Wagner, 1999) and this is likely to contribute to differences in analysis results. In this paper, the desire to maximize the amount of travel information motivated the use of number of days and trip frequency as the criteria for identifying primary drivers. Information on socio-economic and demographic characteristics of the 100 primary drivers used in this paper is shown in Table 1.

For the sample for which information is available, it is found that their characteristics are consistent with expectations. The average household size is 2.94 persons per household with more than 50 percent of the households having 3 or more persons. 28 households reported incomes less than $35,000 per year. The average car ownership is 2.17 vehicles per household with most of the households having 2 or more cars which is consistent with the household size distribution. There is an even split in the sample between males and females. Battelle (1997) reports that most of the drivers are well educated with about 40 percent having a college education. Only 10 primary drivers are less than 26 years of age, while 12 drivers are greater than 65 years of age. 40 percent of the households reported having at least one child less than 16 years of age. Based on these characteristics, one can see that the sample is a rather selective one with households exhibiting mature lifecycle stages, greater household sizes, and higher car ownership levels.

A note is due here regarding the biased nature of the sample. Within the context of this paper, it was desired to examine the potential of using GPS technologies to collect multiday travel data while minimizing respondent burden. The paper does not aim to calculate and provide measures of day-to-day variability in travel behavior that are applicable to the population as a whole. Therefore, the sample is neither weighted nor expanded to correct for its biased nature. Also, in this paper, measures of day-to-day variability obtained from the GPS data set are compared against those reported by Pas and Sundar (1995). They used a raw unweighted and unexpanded sample of 142 drivers drawn from the Seattle area in their analysis. Thus, for comparative purposes, it was felt that the analysis in this paper should also be performed on the raw unweighted and unexpanded sample. By no means should the travel characteristics or measures of day-to-day variability reported in this paper be considered representative of the general population of the Lexington, KY area.

COMPARISON OF AVERAGE TRAVEL CHARACTERISTICS ACROSS DAYS

The first analysis performed in this paper is an exploratory comparison of sample-wide average travel characteristics across days of the week. The survey entailed drivers using the GPS equipment and recording travel information for about a six-day period, with several households providing additional days of information and several providing a little less than six days of usable information. For comparison purposes, all seven days of the week were considered in the context of this paper.

Tables 2 and 3 provide comparisons of average travel characteristics across the days of the week for the sample of primary drivers. While Table 2 focuses on a comparison across weekdays, Table 3 focuses on a comparison between the weekend-days. In the tables, the sample size varies from one day to the next because the comparison is not being done for exactly the same sample of individuals. There are individuals in one day’s sample who are not in another day’s sample. The tables provide sample sizes corresponding to each day and to each measure of travel considered in the comparison. Each sample size corresponds to the number of drivers for whom complete information was available for the particular travel day and travel measure.

The first three variables in the tables represent trip frequencies. Mid-day non-work trips refers to trips made for non-work purposes during the period of 10 AM to 3 PM (trip must begin and end in this period). The next two variables examine daily travel times reported by individuals in the PDA device and as recorded by the GPS device. In general, the PDA device travel time is found to consistently exceed the GPS recorded travel time. This is presumably because the PDA device would start recording time information prior to the GPS receiver. Whereas the PDA device would start recording time as soon as the respondent hit “start trip”, the GPS receiver would start recording time a little bit later after it identified a starting position. Thus, this difference may be largely attributed to hardware issues. While the PDA device time shows total travel time to be a little more than one hour per day, the GPS device time shows total travel time to be a little less than one hour. Regardless of the measure used, their averages are quite similar across the days of the week.

The two variables measuring travel distance are obtained through the information recorded in the GPS device. One measure is directly obtained from the GPS positional information while the other is obtained by plotting the information within a GIS. The final three variables considered in the tables examine variability in three temporal events, the first time of departure from home, the final arrival at home (for the day), and the final departure from work (for workers in the sample).

In general, it is found that Thursday is different from other days (see Table 2). The number of trips, travel times, and travel distances are all considerably larger on Thursday than on other days. Whether this is a simple data issue or a true behavioral phenomenon should be investigated further. Also, it is interesting to note that Friday depicts the lowest trip rates but reasonably comparable travel times and distances indicating that Friday’s trips may be of longer duration and length. On average, the first time of departure from home for this sample is at about 10:45 AM with Tuesday showing the greatest deviation. On average, final arrival at home (that is, the time at which an individual returns home for the day) is a little after 6:30 PM. Final departure times from work (for the workers in the sample) are found to have an average around 4:15 PM.

In Table 3, it is found that the two weekend days are quite similar to one another. A comparison of weekday averages (from Table 2) and weekend averages is also provided through the last two columns of this table. Substantial differences between weekdays and weekends are found in total trips, first departure time from home, and final arrival at home. All other measures including non-work trips, mid-day non-work trips, and travel times and distances show considerable similarity between weekdays and weekends. As this is not a pure commuter sample, this result is generally consistent with expectations.

Overall, it is found that total trip frequency on weekends is less than that on weekdays, presumably because of the absence of work trips. The first time of departure from home on weekends is a little after 12 noon as opposed to the average 10:25 AM departure time found on weekdays. With respect to final arrival at home, the weekend average is about 5:45 PM as opposed to the weekday average of about 6:30 PM. Once again, these findings are consistent with expectations about differences between weekday and weekend travel.

In summary, the aggregate sample-wide comparison of averages presented in this section shows considerable similarities in travel characteristics across days of week, particularly if one were to control for weekday vs. weekend differences. These aggregate-level findings (of similarity across days) are generally consistent with those reported in the literature. The widespread adoption of one-day travel data collection procedures may, in part, be attributed to the appearance of uniformity in travel measures across multiple days. However, as has been documented in the literature, the aggregate and average nature of the analysis presented in this section masks some of the more disaggregate variability present in the data set. The next two sections focus on a more disaggregate analysis of day-to-day variability to see if GPS technologies provide indications consistent with the literature when one examines multiday data at the level of the individual traveler.

ANALYSIS OF REPETITION IN BEHAVIOR ACROSS MULTIPLE DAYS

This section uses the GPS dataset to examine the degree of repetition in travel characteristics across multiple days. Table 4 provides the analysis for various multiday samples while controlling for day-of-week (by including weekday samples only). The table considers various multiday samples separately to account for the fact that some individuals provided fewer days of data than others. For each sample, the percent of individuals who exhibited the same value of a characteristic over all days, all but one day, and all but two days is provided. For travel time and distance measures, the percent of individuals who exhibited values of characteristics within ±20% of their median value is identified. For example, suppose an individual's median travel time over a five day period is 60 minutes. If the person had daily travel times that all lie between 48 minutes and 72 minutes, then the person is considered for inclusion in the first column (% same on all days). The median was used instead of the mean to eliminate the effect of outlier values. With respect to departure times and arrival times, the time of day was converted to a continuous clock time (e.g., 6 AM is equal to 360 min starting at 12 midnight) for ease of calculation of median and the ±20% range.

In general, Table 4 shows that there is considerable variability in travel behavior across multiple weekdays. The percent of individuals in each sample who exhibit the same characteristic across all days (the first column in each table) is extremely small, with many entries showing zero. The table also shows that the extent of variability increases as the period of observation increases. This finding is consistent with expectations; as one observes an individual for a longer duration, greater differences will surface. In general, the table shows that the greatest stability occurs in regard to the last three variables that represent temporal events – first departure from home, final arrival at home, and final departure from work. The figures in these columns are a function of the range of time considered (here, ±20% of the median is considered) and should be interpreted with caution. When ±10% of the median value was chosen as the defining range, it was found that the percent of individuals who exhibited times within that range over multiple days dropped dramatically. In comparing across characteristics and using "±20% of the median" as the defining range for time and distance variables, it appears that mid-day non-work trips and time of arrival at home show the greatest level of similarity across multiple days.

The descriptive analysis presented in Table 4 provides a clear indication that, even though sample-wide day-to-day variability of average travel characteristics (as seen in Tables 2 and 3) does not appear substantial, there is considerable variation in travel characteristics from one day to the next when travelers are examined at a disaggregate level. It was therefore considered appropriate to measure the extent of intra-personal variability present in the data set and compare it against that reported in the literature to see whether GPS technologies are able to provide robust multiday travel data.

MEASURING VARIABILITY IN TRAVEL CHARACTERISTICS

Koppelman and Pas (1984), Pas (1987), and Pas and Sundar (1995) provide a framework for measuring and quantifying variability in travel characteristics. In their work, an important distinction is made between inter-personal variability and intra-personal variability. Together, these two measures of variability account for the total variability in a travel survey data set. Figure 1 shows the framework as adopted from their work. Inter-personal variability refers to differences in behavior among different individuals on the same or different days. Behavioral differences among persons may be explained partially by differences in the characteristics of individuals. By incorporating such characteristics into a model, one can account for systematic differences in behavior among individuals. The portion of inter-personal variability that can be explained systematically through differences in socio-economic characteristics is referred to as explained variability. The remainder is referred to as unexplained variability.

Similarly, intra-personal variability may also be considered to have two components. The first component is called systematic day-of-week variability. This refers to the portion of intra-personal variability that may be attributed to systematic day-of-week effects. Intra-personal variability that can not be explained by day-of-week effects is random and is referred to as residual intra-personal variability. Within the context of the GPS data set being used in this study, it is not possible to uniquely quantify and measure the two components of intra-personal variability. To do so, one would require several weeks of travel survey information in order to determine the systematic day-of-week component of intra-personal variability (e.g., how does one Monday differ from other Mondays?). As such, in this paper, only the total intra-personal variability is measured and reported. Likewise, no models are estimated to explain differences in travel characteristics between persons (to determine explained variability) and only the total inter-personal variability is quantified.

The methodological approached adopted in this paper follows that presented in Pas (1987) and Pas and Sundar (1995). The total variability in various measures of travel is split into its two components which are represented by appropriate sums of squares. In this representation, the total variability is represented by the total sum of squares (TSS), as follows:

$$TSS = \sum_i \sum_j (t_{ij} - \bar{t})^2 \tag{1}$$

where t_{ij} = number of trips made by person i on day j

$\bar{t}$ = overall sample mean number of trips made per person per day

The total sum of squares may be considered to be consisting of a between-person sum of squares and a within-person sum of squares. The between-person sum of squares is representative of the inter-personal variability while the within-person sum of squares is representative of the intra-personal variability in the data set. The between-person sum of squares (BPSS) is given by:

$$BPSS = \sum_i J_i (\bar{t}_i - \bar{t})^2 \tag{2}$$

where J_i = the number of days for which individual i reported travel information

$\bar{t}_i$ = mean number of trips made per day by person i.

Within-person sum of squares (WPSS) is given by:

$$WPSS = \sum_i \sum_j (t_{ij} - \bar{t}_i)^2 \tag{3}$$

Also, we have:

$$\bar{t}_i = \frac{\sum_j t_{ij}}{J_i} \quad \text{and} \quad \bar{t} = \frac{\sum_i J_i \bar{t}_i}{\sum_i J_i} = \frac{\sum_i \sum_j t_{ij}}{\sum_i J_i}$$

It can be readily seen that:

$$TSS = BPSS + WPSS \tag{4}$$

Therefore, the ratio WPSS/TSS provides a measure of the proportion of total variability in travel that may be attributed to within-person variability. Similarly, BPSS/TSS provides a measure of the proportion of total variability in travel that may be attributed to between-person variability. It should be noted that, even though the above formulation was provided in terms of trip frequency, the discussion and formulas can be extended in a straightforward manner to other measures of travel such as travel time, distance, etc.

Results of the analysis of intra-personal variability for ten different travel characteristics are provided in Figures 2 and 3. These figures show the extent of intra-personal variability (as a percentage of the total variability in the data set) for different sample considerations. The 81 individuals who provided at least three weekdays of information (see Table 4) are considered for analysis of intra-personal variability. In the first type of analysis, the first 3 to 5 days of complete information are included regardless of the day-of-week. In the second type of analysis, the first 3 to 5 weekdays of complete information are included. Thus, a comparison between the first two types of analyses provides an indication of the impact of day-of-week effect on measures of intra-personal variability. Finally, the third type of analysis includes only three weekdays of travel information for the 81 individuals. A comparison between the second and third types of analyses provides an indication of the impact of survey duration on measures of intra-personal variability. The reason for doing these comparisons is to determine whether GPS-based data sets offer plausible results that are consistent with expectations. In this particular context, one would expect the following two results:

- Intra-personal variability is lower when one controls for day-of-week effects.
- Intra-personal variability is lower in data sets derived from shorter survey durations.

If the GPS-based data set offers indications consistent with these expectations, it would probably mean that GPS is able to offer robust multiday travel information. On the other hand, if the results obtained are contrary to these expectations, then additional investigations would have to be undertaken to identify the reasons behind the contradictory findings.

There are some clear messages that are provided by Figures 2 and 3. First, the degree of intra-person variability is influenced by day-of-week effects. In comparing the intra-personal variability found in the 3-5 day sample with that found in the 3-5 weekday sample, it is found that the extent of intra-personal variability reduces for virtually all travel measures. The second major finding is that the duration for which travel information is measured impacts the degree of intra-personal variability. A comparison of the 3-5 weekday sample against the 3-weekday sample shows that the extent of intra-personal variability is smaller for the 3-weekday sample. This shows that the degree of intra-personal variability increases with increasing duration of observation. Both of these findings are very consistent with expectations and corroborate earlier research reported in the literature (Pas, 1987; Pas and Sundar, 1995). For example, Pas (1987) reports that, in a five-day activity data set from Reading, England, about 50 percent of the total variability in trip making may be attributed to intra-personal variability. On the other hand, Pas and Sundar (1995) find the corresponding percentage to be only 38 percent when considering a three-day sample. Thus, they too postulate that the extent of intra-personal variability increases as the duration of observation increases. As the results are found to be consistent with the expectations noted earlier, one can say that GPS-based technologies are indeed able to offer robust multiday travel datasets.

There are two travel characteristics for which measures of intra-personal variability are directly comparable to results reported by Pas (1987) and Pas and Sundar (1995). Pas and Sundar (1995) utilize travel survey data collected over three weekdays for a raw sample of 142 persons from the Puget Sound area in Washington. As such, their results are most comparable to the third sample considered in this paper (3-weekday sample). The two measures are total trips and total travel time. With respect to total trips, the percent of total variability that is attributable to intra-personal variability for the particular sample in this analysis is about 49 percent. The corresponding percentage reported in Pas and Sundar (1995) is 38 percent. With respect to total travel time, the percent of total variability that is attributable to intra-personal variability for the

3-weekday sample in this analysis is about 52 percent. The corresponding percentage reported in Pas and Sundar (1995) is 42 percent.

Thus, there is a consistent difference of about 10 percentage points between the measures of intra-person variability obtained from the GPS data set and that reported in the literature. This finding is also quite consistent with expectations. In a previous paper (Yalamanchili, et al., 1999), it was found that the GPS based data set obtained in this experiment captured short, infrequent, and linked trips that tended to be missed in the recall (traditional telephone interview) data set. As traditional recall data sets tend to capture the more regular, frequent, and major trips undertaken by individuals, intra-person variability tends to be lower in such data sets. Thus, based on the findings in this paper, one may conjecture that GPS based technologies are able to offer a more comprehensive picture of multiday travel behavior.

However, it would be premature to draw strong conclusions at this point. In comparing the results from this experiment with those reported in Pas and Sundar (1995), no consideration has been given to differences in socio-economic, demographic, commute, and spatial characteristics that may exist between the two samples. Even though both samples are raw and unweighted, it is likely that they differ with respect to various characteristics. At least a portion of the greater variability in travel across days reported in the GPS data may be attributed to such differences.

DISCUSSION AND CONCLUSIONS

In this paper, the potential of using GPS technologies to collect multiday travel data is explored by analyzing a GPS-based travel data set collected from a sample of 100 households in the Lexington, Kentucky area. One vehicle in each of the 100 households was fitted with a GPS device and a personal digital assistant (PDA) device in order to track and record movements of the vehicle over a period of several days. Most households provided at least three days of complete and usable information, thus providing a valuable data set for examining the potential of adopting GPS technologies in larger scale travel data collection efforts. The analysis in this paper focused on trips reported by primary drivers and did not include secondary and tertiary drivers who may have also used the GPS-fitted vehicle for trip making purposes.

The first analysis performed for this paper focused on comparing aggregate sample-wide averages across the days of the week. The comparison was done across five weekdays and between the two weekend days. In general, it was found that sample-wide averages show considerable stability across multiple days.. However as expected, some differences in trip making patterns were found between weekdays and weekend days with regard to trip frequencies and first time of departure from home. Interestingly, travel distances and total travel time were quite similar between weekdays and weekend days. In general, it was found that the GPS data set yielded trip frequencies slightly higher than the typical average frequencies reported in NPTS and other traditional travel surveys. This is presumably because the GPS travel data collection method is better able to capture short, infrequent, and linked trips (Yalamanchili, et. al., 1999) that tend to be missed in traditional recall surveys These findings were consistent with expectations and showed that GPS technologies are capable of offering plausible measures of travel over multiple days.

The paper next focused on a disaggregate analysis of the GPS-based multiday travel data set. First, an analysis was done to assess the degree of repetition of behavior in the sample. The sample was divided into groups based on the total number of days for which they provided complete usable travel information. Then, the percents of each sample group that repeated behavior on all reported days, all but one reported days, and all but two reported days were determined. In general, it was found that the percentage of individuals who repeated their behavior on all days is extremely small regardless of the travel characteristic being examined. Consistent with research reported in the literature, this analysis showed that there is considerable within-person variability in behavior across multiple days. Once again, the consistency of this finding showed that GPS technologies are capable of serving as useful tools for collecting multiple days of travel information.

In an effort to compare the degree of intra-personal variability between the GPS-based data set and that reported in the literature, the methodology adopted by Pas (1987) and Pas and Sundar

(1995) was applied to the Lexington GPS data set. In this approach, the total variability in the data set consists of two components – inter-personal variability and intra-personal variability. By representing total variability by the total sum of squares, inter-personal variability by the between-person sum of squares, and intra-personal variability by within-person sum of squares, the percent of total variability that may be attributed to intra-personal variability can be determined. This analysis was done for several different travel characteristics including trip frequencies, travel times, travel distances, and selected departure/arrival times.

The analysis showed that the extent of intra-personal variability measured in a data set is sensitive to two dimensions. First, it is sensitive to the day of the week effect. If one controls for the day of the week when analyzing variability across multiple days, then the degree of intra-personal variability measured is lower. Second, it is sensitive to the period or duration for which travel is observed. The longer the period of observation, the greater the degree of intra-personal variability. Both of these findings are consistent with the literature, thus showing that GPS-based travel data sets are capable of offering plausible and reasonable inferences on multiday travel characteristics.

In comparing the estimates of intra-personal variability in travel offered by the GPS-based data set with those reported in the literature by Pas (1987) and Pas and Sundar (1995), it is found that the estimates obtained from the GPS data set in this study are consistently greater than those reported in the literature by nearly 10 percentage points. Further research is needed to clearly identify the sources of this difference; however, several hypotheses may be put forth here. First, there may be socio-economic differences between the samples under consideration that are partially contributing to this difference. The GPS-based data set used in this study is a small sample data set, includes only those trips undertaken by the primary driver using the GPS-fitted vehicle, and includes a self-selected sample of households that has larger household sizes, vehicle ownership, income, and education levels. The extent to which socio-economic differences and sample composition contribute to differences in measures of intra-personal variability merit further investigation.

A second major hypothesis, and more noteworthy in the context of the growing interest in GPS-based data collection methods, is that GPS technologies are able to better capture variability in travel behavior across multiple days. There are two potential inter-related reasons for this. First, as demonstrated by Yalamanchili, et. al. (1999), the GPS-based data set captures (to a greater extent) the short and infrequent trips that may not be obtained in a traditional travel diary survey. As these short and infrequent trips are quite different from one day to the next, they contribute to a greater degree of intra-personal variability than if they were omitted from the analysis. Second, in a multiday travel diary survey (especially those exceeding two days), there is likely to be diary fatigue (Pendyala and Pas, 2000). Even if a respondent provides very detailed information (including that on short and infrequent trips) on one or two days, it is likely that the quality of the information will deteriorate on subsequent days. Usually diary fatigue manifests itself in the form of unreported trips (usually the short and infrequent trips go unreported) and missing data items. Then, it is likely that a multiday GPS travel data set that includes detailed information on all trips across all days will provide a greater (and potentially more accurate) measure of intra-personal variability than a multiday travel diary data set where some infrequent trips go unreported in latter days of the survey.

Overall, it appears that GPS based travel data sets are able to:

- better capture short and infrequent trips (Yalamanchili, et. al., 1999),
- provide accurate temporal (time-of-day) information without round-off (Murakami and Wagner, 1999),
- provide multiple days of travel information with plausible measures of intra-person variability (this paper), and
- offer detailed route choice, spatial location, and travel itinerary information not available in other travel survey data sets.

More importantly, GPS-based data collection methods accomplish the above without placing undue or additional burden on the respondent. It is no surprise that GPS-based data collection methods are gaining increasing popularity for use in travel surveys. Several major ongoing travel surveys include GPS-based data collection components.

Within the scope of this paper, the full potential of GPS based travel data sets was not exploited. Detailed route choice, spatial location, and travel itinerary data present in the data set was not used in the analysis of this paper. As the findings of this paper seem to indicate that GPS-based data sets serve as robust sources of multiday travel information, further research should utilize GPS-based data sets to examine variability in route choice, spatial location, and action space across multiple days. Such studies would be unique as it is very difficult to perform such analyses using traditional travel diary survey data sets.

ACKNOWLEDGMENT

The authors are extremely grateful to Elaine Murakami who provided the Lexington GPS data set and valuable advice and comments throughout the course of this research. Funding provided by Federal Highway Administration is gratefully acknowledged. Sridhar Sunkanapalli assisted in the data preparation tasks.

REFERENCES

Battelle Transportation Division, 1997. Global Positioning Systems for Personal Travel Surveys. Lexington Area Travel Data Collection Test. Final Report to Office of Highway Information Management, Federal Highway Administration, Washington, D.C.

Hanson, S., Huff, J.O., 1988a. Systematic variability in repetitious travel. *Transportation* 15, pp. 111-135.

Hanson, S., Huff, J.O., 1988b. Repetition and day-to-day variability in individual travel patterns: Implications for classification. In: Golledge, R. and H. Timmermans (eds) *Behavioral Modelling in Geography and Planning*, Croom Helm, New York.

Hanson, S., Huff, J.O., 1986. Classification issues in the analysis of complex travel behavior. *Transportation* 13, pp. 271-293.

Hanson, S., Huff, J.O., 1982. Assessing day-to-day variability in complex travel patterns. *Transportation Research Record* 891, pp. 18-24.

Hirsh, M., Prashker, J.N., Ben-Akiva, M., 1986. Dynamic model of weekly activity pattern. *Transportation Science* 20(1), pp. 24-36.

Huff, J.O., Hanson, S., 1990. Measurement of habitual behavior: Examining systematic variability in repetitive travel. In: Jones, P. (ed) *Developments in Dynamic and Activity-Based Approaches to Travel Analysis*. Gower Publishing Co., Aldershot, England, pp. 229-249.

Huff, J.O., Hanson, S. 1986. Repetition and variability in urban travel. *Geographical Analysis* 18(2), pp. 97-113.

Jones, P., Clarke, M., 1988. The significance and measurement of variability in travel behaviour. *Transportation* 15, pp. 65-87.

Kitamura, R., van der Hoorn, T., 1987. Regularity and irreversibility of weekly travel behavior. *Transportation* 14, pp. 227-251.

Koppelman, F.S., Pas, E.I., 1984. Estimation of disaggregate regression models of person trip generation with multiday data. In: Volmuller, J. and R. Hamerslag (eds) *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*, VNU Science Press, Utrecht, The Netherlands, pp. 513-529.

Mahmassani, H., Chang, G.L., 1985 Dynamic aspects of departure-time choice behavior in a commuting system: Theoretical framework and experimental analysis. *Transportation Research Record* 1037, pp. 88-101.

Mahmassani, H., Chang, G.L., 1986. Experiments with departure time choice dynamics of urban commuters. *Transportation Research* 20B, pp. 297-320.

Mahmassani, H., Stephan, D., 1988. Experimental investigation of route and departure time dynamics of urban commuters. *Transportation Research Record* 1203, pp. 69-84.

Mahmassani, H., Herman, R., 1990. Interactive experiments for the study of tripmaker behavior dynamics in congested commuting systems. In: Jones, P. (ed) *Developments in Dynamic and Activity-Based Approaches to Travel Analysis*. Gower Publishing Co., Aldershot, England, pp. 272-297.

Mannering, F.L., 1989. Poisson analysis of commuter flexibility in changing routes and departure times. *Transportation Research* 23B(1), pp. 53-60.

Murakami, E., Wagner, D.P., 1999. Can Using Global Positioning System (GPS) Improve Trip Reporting? *Transportation Research C*, Vol. 7, pp. 149-165.

Pas, E.I., 1988. Weekly travel-activity behavior. *Transportation* 15, pp. 89-109.

Pas, E.I., 1987. Intra-personal variability and model goodness-of-fit. *Transportation Research* 21A(6), pp. 431-438.

Pas, E.I., 1986. Multiday samples, parameter estimation precision, and data collection costs for least squares regression trip-generation models. *Environment and Planning A* 18, pp. 73-87.

Pas, E.I., Koppelman, F.S., 1987. An examination of the determinants of day-to-day variability in individuals' urban travel behavior. *Transportation* 14, pp. 3-20.

Pas, E.I., Sundar, S., 1995. Intra-personal variability in daily urban travel behavior: Some additional evidence. *Transportation* 22, pp. 135-150.

Pendyala, R.M., Pas, E.I., 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 E-Circular Number E-C008, Transportation Research Board, National Research Council, Washington, D.C., pp. II-B/1 – II-B/22.

Yalamanchili, L., Pendyala, R.M., Prabaharan, N., Chakravarthy, P., 1999. Analysis of Activity Chaining Using Lexington, Kentucky GPS Data. *Transportation Research Record* 1660, Transportation Research Board, National Research Council, Washington, D.C., pp. 58-65.

Zhou, J., Golledge, R., 2000. An Analysis of Household Travel Behavior Based on GPS. CD-ROM Proceedings of the 9th International Association for Travel Behavior Research Conference, Gold Coast, Australia, July 2-7.

Table 1. Descriptive Characteristics of Survey Sample

Characteristic	Value (N=100)
Mean household size	2.94
1 person hhld	13
2 person hhld	35
3 person hhld	15
4+ person hhld	37
Income Level	
Low (< $35K)	28
Medium ($35K - $75K)	49
High (> $75K)	17
Average Vehicle Ownership	2.17
1 car hhld	21
2 car hhld	47
3 car hhld	27
4+ car hhld	5
Gender	
Male	44
Female	42
Average Age (in years)	46.3
Less than 26 years	10
26 to 45 years	32
46 to 55 years	23
56 to 65 years	9
Greater than 65 years	12
Number of Children <16 years	0.68
0 child hhld	60
1 child hhld	21
2 child hhld	11
3 child hhld	7
4+ child hhld	1

Table 2. Comparison of Average Travel Characteristics Across Weekdays

Characteristic	Monday		Tuesday		Wednesday		Thursday		Friday		Wtd Avg
	Average	N	Average	N	Average	N	Average	N	Average	N	
Total trips	5.2	70	5.4	73	5.5	75	6.3	72	4.6	66	5.4
Non-work trips	4.6	70	4.4	73	4.4	75	5.2	72	3.8	66	4.5
Mid-day non-work trips	1.4	70	1.7	73	1.3	75	1.8	72	1.1	66	1.5
PDA travel time(min)	67.0	70	71.0	71	70.5	76	89.1	72	72.3	69	74.0
GPS travel time(min)	52.8	49	69.0	54	55.3	59	68.2	58	58.0	56	60.8
GPS travel distance (miles)	20.0	49	28.0	54	19.9	59	40.4	58	23.4	56	26.5
GIS travel distance (miles)	20.0	49	26.1	54	20.1	59	29.1	58	23.4	55	23.8
First departure from home	10:31	70	9:36	73	10:52	76	10:20	72	10:47	69	10:25
Final arrival at home	18:50	70	18:22	73	18:46	76	18:20	72	18:25	69	18:32
Final departure time from work	15:49	28	15:22	39	16:47	40	17:20	37	15:23	32	16:11

Table 3. Comparison of Average Travel Characteristics Across Weekend Days

Characteristic	Saturday		Sunday		Wtd Avg	Weekday Wtd Avg (from Table 2)
	Average	N	Average	N		
Total trips	4.7	58	4.3	52	4.5	5.4
Non-work trips	4.3	58	4.2	52	4.2	4.5
Mid-day non-work trips	1.3	58	1.5	52	1.4	1.5
PDA travel time (min)	68.7	59	65.3	53	67.1	74.0
GPS travel time (min)	53.7	46	46.0	39	50.1	60.8
GPS travel distance (miles)	26.9	46	26.2	39	26.5	26.5
GIS travel distance (miles)	27.0	45	23.0	38	25.2	23.8
First departure from home	11:46	59	12:25	50	12:04	10:25
Final arrival at home	17:38	59	17:57	53	17:47	18:32
Final departure time from work	16:27	17	17:20	8	16:44	16:11

Table 4. Analysis of Repetition of Behavior in Weekday Samples

Characteristic	% Same on all days	% Same on all but 1 day	% Same on all but 2 days	Total
3-WEEKDAY SAMPLE (N=25)				
Total trips	8.0	40.0	--	48.0
Non-work trips	12.0	40.0	--	52.0
Mid-day non-work trips	32.0	44.0	--	76.0
GPS travel time (±20% of median)	4.0	4.0	--	8.0
PDA travel time (±20% of median)	4.0	32.0	--	36.0
GPS travel distance (±20% of median)	0.0	24.0	--	24.0
GIS travel distance (±20% of median)	28.0	60.0	--	88.0
First departure from home (±20% of median)	44.0	40.0	--	84.0
Final arrival at home (±20% of median)	72.0	24.0	--	96.0
Final work departure time (±20% of median)	0.0	12.0	--	12.0
4-WEEKDAY SAMPLE (N=32)				
Total trips	3.1	6.3	40.6	50.0
Non-work trips	3.1	18.8	43.8	63.6
Mid-day non-work trips	21.9	25.0	31.3	88.2
GPS travel time (±20% of median)	0.0	0.0	37.5	37.5
PDA travel time (±20% of median)	3.1	21.9	43.8	68.8
GPS travel distance (±20% of median)	0.0	21.9	43.8	65.7
GIS travel distance (±20% of median)	0.0	0.0	43.8	43.8
First departure from home (±20% of median)	34.4	37.5	9.4	81.3
Final arrival at home (±20% of median)	59.4	40.6	0.0	100.0
Final work departure time (±20% of median)	0.0	9.4	18.8	28.2
5-WEEKDAY SAMPLE (N=24)				
Total trips	0.0	4.1	16.7	20.8
Non-work trips	0.0	4.1	16.7	20.8
Mid-day non-work trips	12.5	16.6	25.0	54.1
GPS travel time (±20% of median)	0.0	0.0	8.3	8.3
PDA travel time (±20% of median)	0.0	12.5	33.3	45.8
GPS travel distance (±20% of median)	0.0	12.5	12.5	25.0
GIS travel distance (±20% of median)	0.0	16.6	8.3	24.9
First departure from home (±20% of median)	8.3	41.6	33.3	83.2
Final arrival at home (±20% of median)	50.0	37.5	8.3	95.8
Final work departure time (±20% of median)	0.0	0.0	17.7	17.7

Note: For the 3-weekday sample, the last column is not applicable (no repetition).

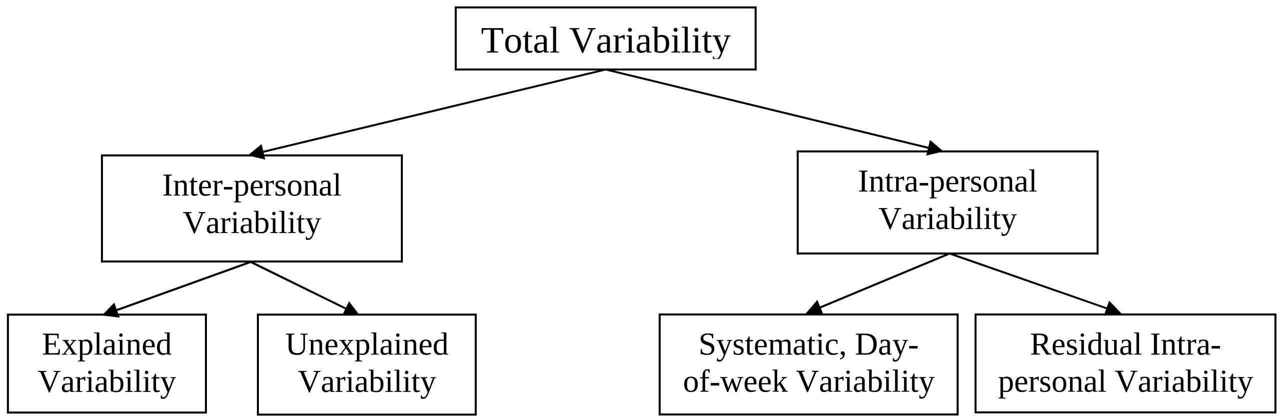

Figure 1. Methodological Framework for Measuring Variability in Travel (Source: Pas, 1987)

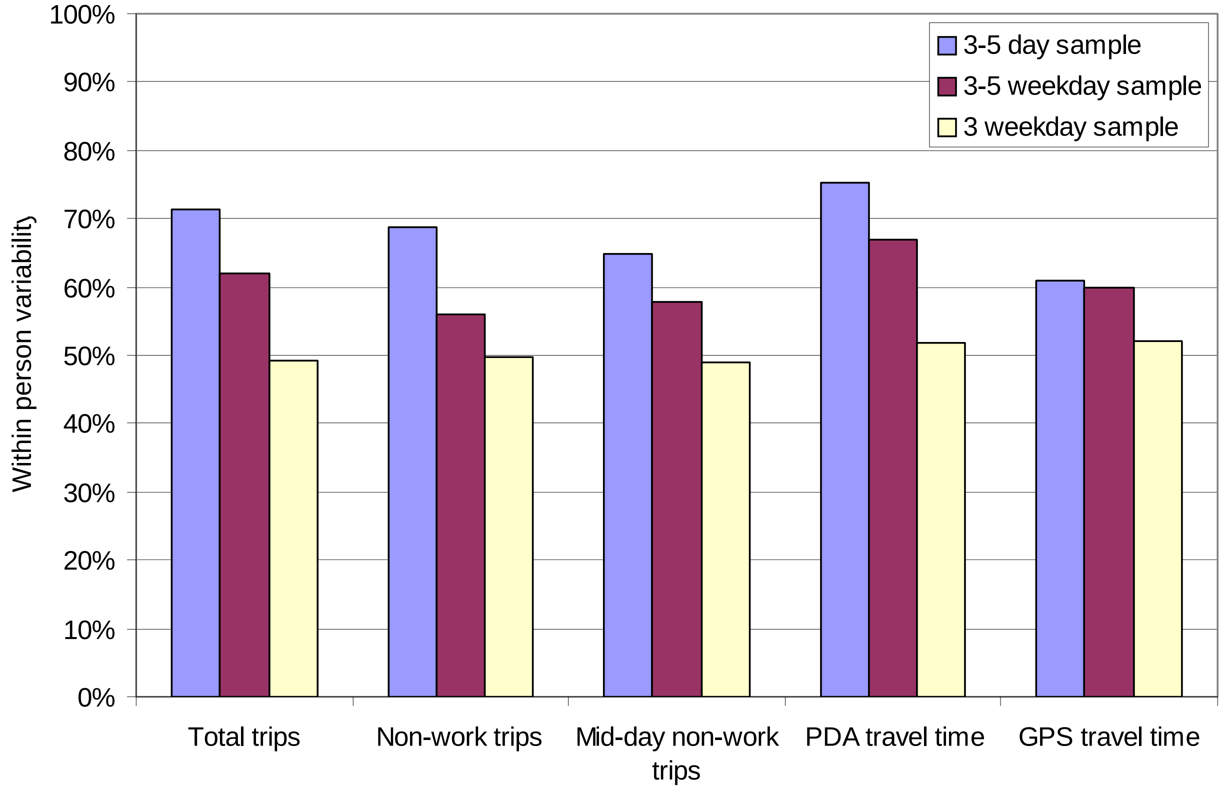

Figure 2. Intra-personal Variability for Trip Frequencies and Travel Times

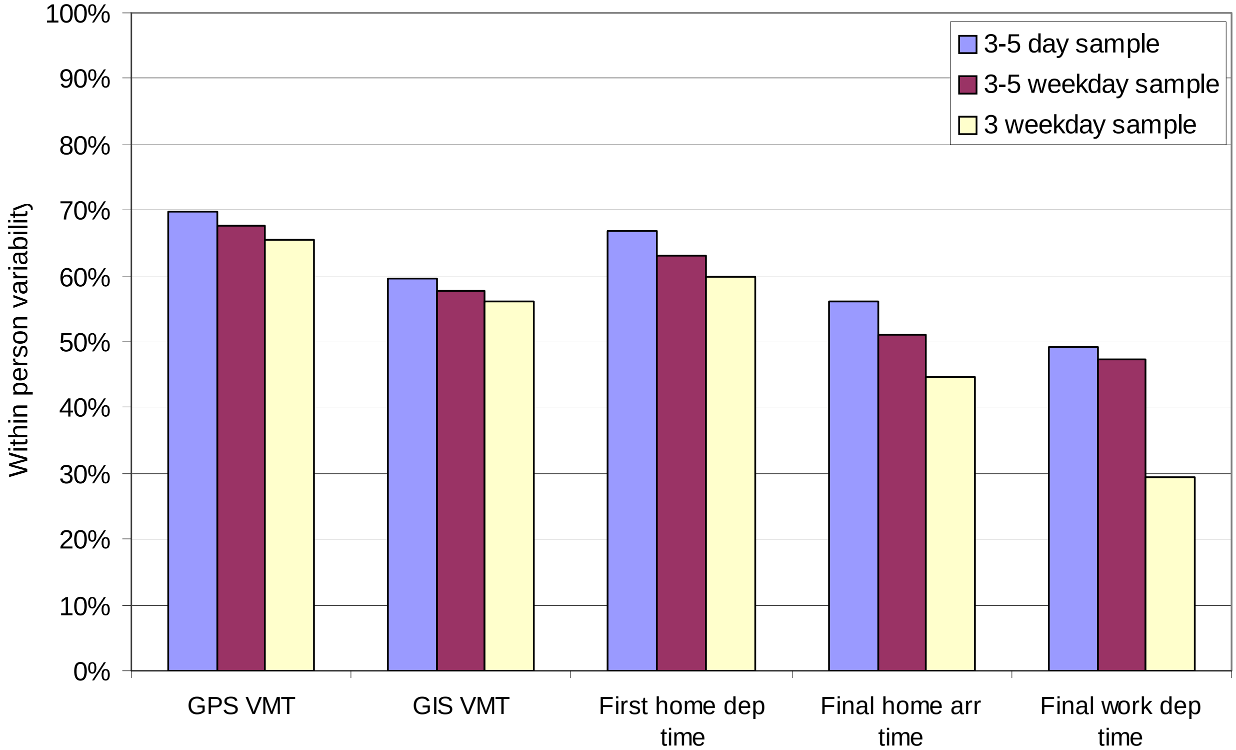

Figure 3. Intra-personal Variability for VMT and Departure/Arrival Times