International Conference on Travel Survey Methods

Investigating the capacity of continuous household travel surveys in capturing the temporal rhythms of travel demand

Wafic El-Assi[a], Dr. Eric J. Miller[a], Dr. Catherine Morency[b], Dr. Khandker Nurul Habib[a]*

[a]*Department of Civil Engineering, University of Toronto, Canada*
[b]*Department of Civil, Geological and Mining Engineering, École Polytechnique de Montréal, Canada*

Abstract

Continuous household travel surveys have been identified as a potential replacement for traditional one-off cross-sectional surveys. The main claimed advantage of continuous surveys is the availability of data over a continuous spectrum of time, thus allowing for the investigation of the temporal variation of trip behavior. This claim is put to the test by estimating mixed effects models on the individual, household spatial, trip and modal level using the Montreal Continuous Survey (2009-2012) data. The findings conclude that the temporal variability in trip behavior is only observed when modelling on the spatial and modal level.

"Keywords: continuous surveys, mixed effects model, variance partition coefficient analysis;"

1. Introduction

Household travel surveys are fundamental for the understanding of the socio-economic factors underlying travel behavior. In many regions around the world, travel survey data are used almost entirely for their richness, depicting fluctuations in travel patterns and household socio-demographics by calculating basic statistical measures such as trip activity means and standard deviations (Ampt & Ortuzar, 2011). Other regions collect such data for the training and development of sophisticated policy-oriented travel demand models.

The Montreal metropolitan agency has been conducting large cross-sectional household travel surveys every 5 years since the 1970s (Habib & El-Assi, 2015). The surveys are relatively large with a sampling rate of approximately 5%. A cross-sectional survey is defined as a survey executed at a point in time and conducted on a one-off basis. Large regional household travel surveys, while typically conducted over weeks or months, are still considered cross-sectional, as the data are pooled to represent a “typical day” (Verreault & Morency, 2011). The Transportation Tomorrow Survey (TTS), conducted in the Greater Toronto and Hamilton Area (GTHA) and repeated every 5 years, also falls under this category.

In 2009, right after Montreal’s most recent OD survey, the agency launched an experimental continuous/repeated cross-sectional survey (Tremblay, 2014). In other words, households were sampled every day with each household only to be surveyed once. No repeated observations were recorded for any of the households. The survey ended at the end of the year 2012, a few months before the start of Montreal's next major OD survey in 2013.

Like other metropolitan areas, Montreal has been facing increasing challenges in the conduct of its typical household travel survey in relation to declining response rates, the incompleteness of sampling frame, inability to monitor changes, etc (Tremblay, 2014). As many other regions, Montreal relies on its large-scale household travel surveys to support decision making regarding transportation investments (subway extension for instance) and being able to measure the changes in behaviours after important changes in transportation supply is of great importance. It is the aim of the region to build on lessons learned from previous household travel survey, and incorporate the necessary changes pertained to sampling frame and survey design to improve data quality and answer previously neglected questions such as seasonality of behavior.

This study investigates the temporal variability of travel behavior by estimating a set of mixed effects models on the individual, household and spatial level using the Montreal Continuous OD survey, and conducting a variance partition coefficient (VPC) analysis on the ensuing results. The study concludes that seasonality, or more generally the temporal variability, of travel behavior can only be captured on the spatial (e.g. regional) level if using continuous survey data. The paper is organized as follows: A literature review is first presented on the use of continuous surveys and on complex models used to depict the temporality of travel behavior from survey data. Then the data and its source (Montreal Continuous OD survey) are briefly described. After that, the data preparation steps preceding the modelling exercise are listed. Next, the methodology behind the modelling exercise is explained, followed by the results. Finally, the paper objectives and results are summarized in the conclusion along with limitations and future research.

2. Literature Review

A continuous survey has many advantages over its cross-sectional counterpart. In a full on-going continuous survey, data are collected for an entire weekday, every day of the week, 52 weeks a year (Ortuzar, et al., 2010). Such effort should ideally be kept going for several years. This data, collected over a large period of time, can potentially be used to observe temporal trends in travel patterns and behavior at an aggregate level in the survey area. However, the specific time period where the survey is undertaken may be subject to unpredictable events (Stopher & Greaves, 2007). For example, it is expected that a regional continuous survey should be capable of depicting the 2008 recession by observing a decline in the number of freight deliveries over that time period. Furthermore, a similar survey should also be able to capture the effect of the current decrease in fuel prices on mode choice. In this way, the global evolution of mobility behavior over time can be captured.

In addition, a cross-sectional survey does not allow for the comparison between short term and long-term trends, as data are only collected at a point in time (Kish, 1965). From a modelling perspective, the continuous nature of the data may permit the use of more sophisticated models, such as mixed effects models, to investigate dynamics and adaptation of travel behavior. This is in contrast to traditional practice where multivariable regression models are the base statistical tool to predict trip generation and changes in travel behavior as a function of land use attributes and socio-demographic factors (Stopher & Greaves, 2007). Although a powerful statistical tool, linear regression is governed by a number of limiting factors; namely, the independence of observations. Moreover, a multivariable regression model in its most basic form only accounts for a single level outcome (whether it be an individual, household or zone), rather than attributing behavior to additional macro level causes (DiPrete & Grusky, 1990). Several countries and metropolitan regions around the world have conducted continuous household travel surveys (Ampt & Ortuzar, 2011). Examples of such surveys include the National Travel Survey in Britain, the eldest ongoing continuous household travel since 1988, and the famous Sydney Household Travel Survey (Ampt & Ortuzar, 2011).

Nevertheless, it is very difficult to find literature on the statistical modelling tools and techniques used with continuous data to depict transportation behavior.

On the other hand, several types of research have opted to use (repeated) cross-sectional and panel datasets to depict behavior. In numerous cases, the statistical model of choice was a mixed effects model[†]. DiPrete and David Grusky (1990) developed a multilevel model for the analysis of trends within repeated cross-sectional samples. The proposed model is first-order auto regressive at the macro-level equation (highest level in the multilevel model), defined to be a time variable such as a year. Such a custom model allows for time series analysis by serially correlating the errors of the upper-level equation (DiPrete & Grusky, 1990).

A less programming intensive approach was presented by Lipps and Kunart (2005), where they used four cross-sectional data sets of the National Travel Survey (NTS) conducted in (West) Germany in 1976, 1982, 1989 and 2002 to build a hierarchical linear model (Lipps & Kunert, 2005). The dependent variable of the model was the logarithmic transformation of the daily travel distance covered by survey respondents. Travel distance was regressed against a series of socio-demographic and land use variables, such as employment, number of cars available, population size and household size. The structure of the model was setup so as to have individuals nested within households nested within zones. The study showed that, by estimating a sequentially pooled (over the various time periods) mixed effects model, the total variance of daily travel distance decreased. This may indicate that at least the population surveyed, is slowly developing increasingly homogeneous behavior over time. Further, the authors also show by calculating the variance partition coefficients of every level within the hierarchy, that over 90% of the variation in travel distance may be attributed to variations between individuals within households, and variations between households. The study, although unique, does not correct for the differences in sampling frame and methods adopted across the four surveys, which may lead to biased estimates over time (Ampt & Stopher, 2006).

A final great piece of work was completed by Goulias (2002). Goulias used a panel dataset, the Puget Sound Transportation Panel (PTSP), conducted in California to estimate a set of four correlated activity based multilevel models (Goulias, 2002). The four multilevel models investigated individual choices in time allocation to maintenance, subsistence, leisure and travel time. A three-level nested hierarchy was exploited with occasions of measurements as the lowest level, individuals as the second and households as the third. The joint and multivariate correlation structure of the dependent variables, along with the flexibility offered via the use of mixed effects models, allowed for the investigation of three key factors: the behavioural context of individuals, heterogeneity of behavior and longitudinal variation of time allocation. The author’s key finding is that the household level variance was more than one-third of that of the individual level and thus was considered significance. Further, the author also concluded that clear evidence exists of non-linear dynamic behavior in time-allocation. None of the above have combined the flexibility of mixed effects models with continuous travel survey data.

2.1 Travel survey description

At the end of the 2008 cross-sectional OD survey conducted in Montreal, the metropolitan agency decided to move into an experimental ongoing survey. Already at that time, partners were looking for ways to monitor the evolution of behaviors and be able to report on the relation between travel conditions and choices, and changes in transportation supply (opening of a new bridge or subway station for instance), at different time lags. The objective was to gather continuous data to assess its potential contribution to the understanding of

---

[†] A mixed effect model is also known in literature as a multilevel model, hierarchical linear model, and random intercept/coefficient model (Rabe-Hesketh & Skrondal, 2012)

changes in travel behavior.

Data were collected on a continuous basis using non-repeated sampling from January 2009 to December 2012. Interviews were conducted using the same CATI tool that was used during the previous large-scale survey. This experimentation was also the testbed for new questions and was used as a preparation for the 2013 survey.

On a typical week, some 250-400 households were surveyed, amounting to 14,400 households in the first year of the survey, and 16,000 to 16,700 for the other three years. The data were collected from 8 regions in the Montreal Metropolitan Area. No official results were published after the conduct of the survey. Some analyses were reported, namely the study of changes in cycling levels, but there was no systematic modelling of travel behavior using the data.

2.2 Data cleaning and preparation

After completion of the survey, the data was validated to limit the presence of erroneous records in the data. The resulting dataset contains all trips, their related attributes as well as data on individuals and households. While some variables were readily available for modelling, others required preprocessing such as trip chain identification and duration estimation. Other databases were fused with the survey dataset; namely data from Environment Canada on daily weather conditions from the international airport sensor (snow, rain, average temperature), and fuel price from the Régie de l'énergie du Québec.

Overall, the total number of individuals surveyed was 152,157. The dataset was then prepared for modelling. Holidays were removed from the dataset so as to capture trip distance variation on an "average" workday. The total number of records removed was 958. After that, records with missing values were deleted, bringing down the total number of surveyed individuals to 148,992. Respondents who answered a survey question by "I refuse to answer" or by "I don't know", or records with missing values were also removed.

2.2.1 Data preparation for individual level modelling

It was noticed that approximately 17% of the remaining respondents reported zero trips on the day they were surveyed. Therefore, to avoid floored residuals, individuals who didn't conduct any trip, or conducted a trip of less than 0.5 km in distance, were removed. This provides for a more homogeneous group for analysis. The final dataset has 88,156 individual records.

2.2.2 Data preparation for household level modelling

The total trip distance per household was calculated from the survey dataset. Household level attributes were also aggregated accordingly. Further, households with a total of zero trips were not included in the analysis. Indeed, every row in the resulting dataset constituted a household. The final dataset has 42,895 household records.

2.2.3 Data preparation for spatial level modelling

The total number of trips per spatial unit (region or municipal sector) per time period were aggregated. No individuals were excluded.

2.2.4 Data preparation for trip level modelling

The survey dataset was converted from wide to long format. That is, every row was a trip, rather than an individual. The grouping factors (random effects) considered were the mode, spatial unit (either region or municipal sector) and time period.

2.2.5 Data preparation for modal level modelling

Trip level data were aggregated by mode. Modal, spatial and temporal random factors were included. A separate model was estimated for each combination of random factors.

2.2.6 Data preparation for active modes modelling

A subset of the dataset that included trips conducted by walking or biking was used for modelling. A random effects model was then estimated with the log of trip distance as the dependent variable for a set of models, and the log of a number of trips by mode for another set of models. The grouping factors (random effects) considered were mode, region and different time periods. The only region was considered for spatial units due to the small sample size of trips conducted by active modes of transport.

A description of the available variables in the dataset may be seen in table 1 below. Table 2 provides summary statistics.

Table 1. Definition of variables in dataset

	Variable	Definition	Variable Type
Trip Attributes	tripdu	Total duration of a respondents' trip chain	Continuous
	tripd	Total distance of a respondents' trip chain	Continuous
	tripr	Total number of trips in a respondent's trip chain	Count
Person Attributes	age	Age in years	Continuous
	driv_lic	Driving license = 1 if respondent carries a driving license; = 2 otherwise	Binary
	gender	Gender of respondent = 1 if male; = 2 if female	Binary
	occ_status	Occupation Status of respondent; 1 = Full time worker; 2 = Part time worker; 3 = Student; 4 = retired; 5 = work at home	Categorical
Household Attributes	nb_child	Number of children under 16 years of age in a household	Continuous
	hhsize	Number of persons in a household	Continuous
	carown	Number of cars owned by household	Continuous
	income	Household income 1= 0$ - 20 000$; 2 = 20 000$ - 40 000$; 3 = 40 000$ - 60 000$; 4 = 60 000$ - 80 000$; 5 = 80 000$ - 100 000$; 6 = 100 000$+	Categorical
Other Variables	rainday	Rainday = 1 if it rained on the day of the survey; = 0 otherwise	Binary
	snowday	Snowday = 1 if it snowed on the day of the survey; = 0 otherwise	Binary
	Population	Number of residents per municipal sector	Continuous
	fuelprice	Fuel price on the day of the survey	Continuous
Spatio-Temporal Variables	region	Region address of Household: based on 8 large regions	Categorical
	Municipal Sector	108 zones representing municipalities or districts	Categorical
	Year	Year of survey	Categorical
	Season	Season of survey	Categorical
	month	Month in year of survey	Categorical
	no week	Week in year of survey	Categorical

dow	Day of week in year of survey	Categorical
Season by Year	Season of year (e.g. fall 2011, winter 2011, spring 2011, summer 2011, fall 2012, etc.)	Categorical
Month by Year	Month of year (e.g. Jan 2011, Feb 2011, March 2011, etc.)	Categorical

Table 2. Summary of Descriptive Statistics

	Variable	Unit	Min	Max	Range	Median	Mean	Std Dev
Trip Attributes	tripdu	min	0	1410	1410	445	365	269.57
	tripd	Km	0	354	354	10	18	20.78
	tripr	N/A	0	22	17	2	2.4	1.69
Person Attributes	age	years	0	99	99	42	39.96	22.31
	driv_lic	N/A	1	2	1	N/A	1.29	N/A
	gender	N/A	1	2	1	N/A	1.52	N/A
Household Attributes	nb_child	N/A	0	8	8	N/A	N/A	N/A
	hhsize	N/A	1	21	20	3	3.07	1.4
	carown	N/A	0	14	14	2	1.61	1.03
Other Variables	rainday	N/A	0	1	1	N/A	0.37	N/A
	snowday	N/A	0	1	1	N/A	0.13	N/A
	fuel.price	$	81.5	146.9	65.4	115.5	118.03	16.69
	Population	N/A	962	126600	125638	55530	55370	30846

3. Methodology

Six groups of models were estimated:

- A group of mixed effects models with individual level observations; the chosen dependent variable was the logarithmic transformation of travel distance
- A group of mixed effects models with household level observations; the chosen dependent variable was the logarithmic transformation of travel distance
- A group of mixed effects models with regional level observations; the chosen dependent variable was numbers of trips generated per region per temporal variable
- A group of mixed effects models with trip level observations; the chosen dependent variable was the logarithmic transformation of travel distance
- A group of mixed effects models with aggregated modal level observations; the chosen dependent variable was the logarithmic transformation of travel distance
- A group of mixed effects models for aggregated walking and cycling trips; the chosen dependent variable was the logarithmic transformation of travel distance for one set of models, and the logarithmic transformation of trip counts for another set

Within every group, various temporal variables were tested as random effects to investigate their contribution to the total variance of the dependent variable.

The mixed effects models attempt to answer the main research questions investigated in this paper. That is, is the time period component of a mixed effect model estimated using a continuous survey imperative to the understanding of the factors affecting the overall variation in total trip distance or trip generation, and the use of a hierarchical approach to model travel beavior.

All models were estimated in **R** using the lme4 package.

3.1 Linear mixed effects model

Referring back to the design of the Montreal continuous survey, respondents were interviewed within households randomly sampled from regions at different time points. Therefore, it is logical to assume that the collected data has an inherent nested structure. The appropriate methodology to analyze hierarchically nested data is by using a mixed effects model (Rabe-Hesketh & Skrondal, 2012). A mixed effects model attempts to describe the contextual effect of the data while accounting for the variation in the dependent variable originated from multiple levels (Goulias, 2002). Further, a mixed effects model handles random effects. That includes the grouping of observations under higher levels (or clusters) such as the grouping of individuals under households. The act of clustering observations within groups leads to correlated error terms. Treating clustering as a nuisance, as in simple regression, causes biased estimates of parameter standard errors (Garson, 2013). This can lead to mistakes in interpreting the significance of coefficients. Figure 1 shows the nested hierarchy of the survey data.

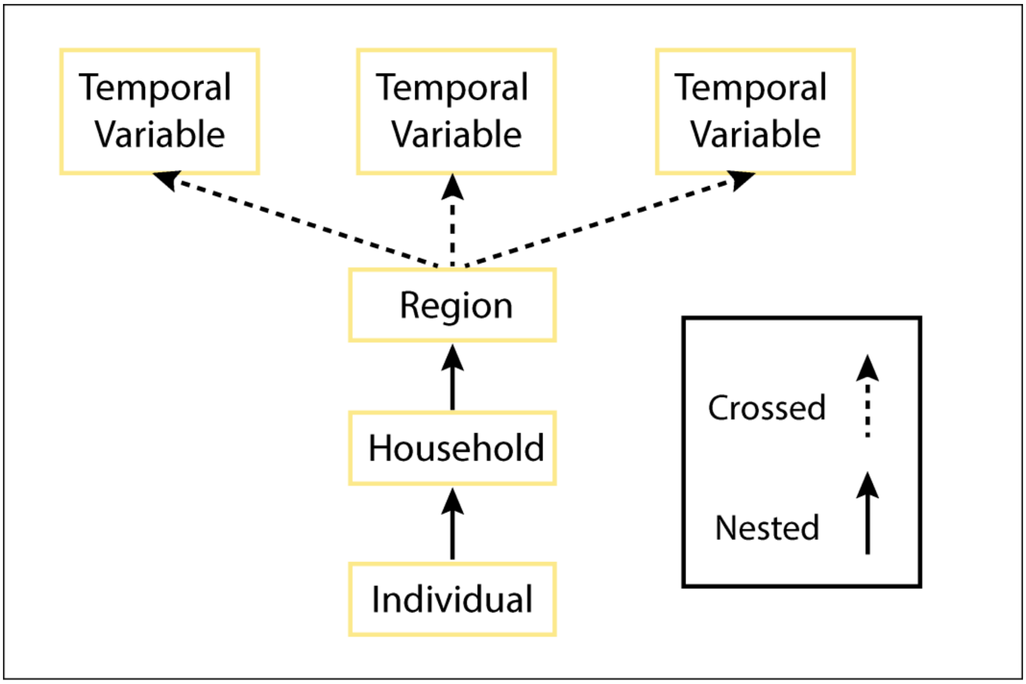

Fig. 1. Nested hierarchy of mixed effects model

The figure shows individual respondents nested in households and households nested in their respective regions, as expected. Usually, individuals belong to a single household and households can only be located in one spatial area. On the other hand, the figure shows regions crossed with time periods. This is because data were collected from all regions at continuous time points. In other words, no region belongs to a single time point only, rather the survey design ensured a distributed sampling effort. It is important to recognize the cross-classified structure of the model, for applying a model with nested regions in time points can seriously bias standard errors of parameters and variance component estimates - an important factor in this paper (Garson, 2013).

To understand a mixed effects model with a structure similar to that displayed in figure 1, it is convenient to first start with a simple two-level hierarchy. A two level mixed effects model maybe expressed in the following form (Scott, et al., 2013):

$$y_{ij} = \beta_0 + u_j + e_{ij}, \quad i = 1, \dots, N, \quad j = 1, \dots, J$$

where y_{ij} is an n x 1 vector of random variables representing the observed value for individual i nested in household (group) j. The term u_j is called the group random effects. It is an independent error term (or group effect) assumed to follow a normal distribution of mean 0 and variance σ^2. The individual residuals e_{ij} also represents an independent error term assumed to follow a normal distribution of mean 0 and variance σ^2. Adding explanatory variables is fairly simple:

$$y_{ij} = \beta_0 + \beta x_{ij} + u_j + e_{ij}, \quad i = 1, \dots, N, \quad j = 1, \dots, J$$

Where β is an n x q matrix of regressors; it represents the coefficient for x_{ij}[‡]. The two-level notation can be expanded to form a three level mixed effects model, where individuals *i* are nested in household *j* and region *k*:

$$y_{ijk} = \beta_0 + u_j + u_{jk} + e_{ijk},$$
$$i = 1, \dots, N, \quad j = 1, \dots, J, \quad k = 1, \dots, k$$

Here, u_{jk} is the effect of household *j* nested in region *k*. It is also an independent error term assumed to follow a normal distribution of mean 0 and variance σ^2. To represent the crossed effects, the model notation maybe denoted as follows:

$$y_{ijkt} = \beta_0 + u_j + u_{jk} + u_{jt} + e_{ijkt},$$
$$i = 1, \dots, N, \quad j = 1, \dots, J, \quad k = 1, \dots, k,, \quad t = 1, \dots, t$$

The above notation adds another random effect u_{jt}. The subscript of the effect implies the nesting of household *j* in time periods *t*[§]. The absence of the subscript k in the additional random term denotes that the effects u_{jk} and u_{jt} are crossed. That is, region *k* is crossed with time period *t* (Scott, et al., 2013). It is important to note that effects u_{jk} and u_{jt} are no longer independent of each other, rather they have a bivariate normal distribution with zero means and an unstructured 2x2 covariance matrix.

3.2 Variance partition coefficient analysis

[‡] The fixed effects factor was removed from subsequent notation as the focus is more on the correct representation of the random effects

Following specification and model estimation, the VPCs of each grouping level were calculated using the following formula (Rabe-Hesketh & Skrondal, 2012):

$$\frac{\sigma_u^2}{\sigma_T^2}\text{, where } \sigma_T^2 = \sigma_j^2 + \sigma_{jk}^2 + \sigma_{jt}^2 + \sigma_{ijkt}^2$$

The VPC ranges from 0 to 1. If the VPC of a level is 0, no between group differences exist. If the VPC is equal to 1, no within group differences exist (Fiona, 2008). The VPC measures the proportion of total variance in the dependent variable that is due to the differences between groups. For a simple mixed effects model, the VPC is equal to the intra-class correlation (Fiona, 2008). The intra-class correlation is the correlation between the selected dependent variable of two individuals from the same group (e.g. household).

To illustrate with an arbitrary example, a VPC of 0.2 for time periods implies that 20% of the variation in the dependent variable is between time periods and 80% is within. The intra-class correlation is also equal to 20%.

In this investigation, different time periods (months, seasons, and years) will be tested for significance, and their VPC, along with that of other groups, will be calculated accordingly. This exercise is essential to understand the reasons behind the variation in trip behavior in general. A log likelihood ratio test will be used to identify the significance of the grouping factors (Rabe-Hesketh & Skrondal, 2012).

4. Empirical Results

4.1 Individual level mixed effects model

For the individual level mixed effects model, the effect of clustering was taken into account by nesting individuals in households, and households in regions. Regions were crossed with various time periods.

Table 3 lists the parameter estimates, t-statistics and confidence intervals. An ANOVA comparison showed that the season model was the best. Thus, only the fixed effects of the season model are presented. For the income variable, income category 1 (0 – $20,000) was used as a base. Similarly, the “other” work category was set as the base for the variable *occupation status*.

Table 3. Seasonal Individual Level Mixed Effects Model

Variable	Description	Estimate	Std. Error	t-value
(Intercept)	N/A	2.033	0.111	18.311
income2	$20,000 - $40,000	0.133	0.013	10.222
income3	$40,000 - $60,000	0.241	0.014	17.851
income4	$60,000 - $80,000	0.304	0.014	21.03
income5	$80,000 - $100,000	0.359	0.017	21.687
income6	$100,000+	0.41	0.015	26.967
Age	N/A	0.006	0	18.075
Female	N/A	-0.098	0.006	-16.846
occ_status1	Full Time	0.449	0.019	24.033
occ_status2	Part Time	0.185	0.022	8.281

occ_status3	Student	-0.142	0.021	-6.803
occ_status5	Work at Home	-0.212	0.022	-9.828
occ_status6	Retired	-0.09	0.025	-3.533
hhsize	Household Size	-0.004	0.003	-1.323

Almost all parameters were estimated with the expected signs and were statistically significant at the 95% confidence interval, with the exception of household size. Interestingly, household size was a significant variable until the addition of the region random effect. Thus, it may be that the effect of household size is region dependent, with individuals living further away from the downtown core may be travelling longer distances on a daily basis and vice versa[**].

The income variable, as in all income categories compared to the base category, was statistically significant and positively correlated with total trip distance travelled. This is in line with transportation literature (Meyer & Miller, 2001). Further, women seem to prefer travelling shorter travel distances as the women variable proved to be negatively correlated with the dependent variable (taking men travellers as a base). This may be because women tend to work closer to home (Hanson & Johnston, 2013). An interaction variable consisting of gender and occupation status was tested to determine if this behavior may vary across different employment conditions. Nevertheless, the results proved insignificant and the interaction term was removed from the model.

A significant positive relationship between age and total trip distance was also identified. This is a reasonable conclusion as with age comes more household responsibility, resulting in longer distance travel. Further, full time and part time workers showed a positive correlation with total trip distance as compared to the “other” work category. The survey did not ask whether the individual was unemployed, rather it included an “other” category. Thus, full and part time workers may well travel more than other non-workers for commuting and other activities. On the other hand, Individuals who work at home alongside retirees and students may choose to travel on shorter trips for leisure, maintenance and subsistence activities (Goulias, 2002). Overall, a working individual (or even a retired man or women) may have a larger spending capacity and thus justifying the feasibility of longer trip making.

4.1.1 Variance partition coefficient analysis

Table 4. Individual level VPC analysis

Time Period	Cluster	Variance	VPC
Season	Time Period	0.00036	0.04%
	Region	0.09186	9.54%
	Household	0.22563	23.42%
	Residual	0.645399	67.00%
Season by Year	Time Period	0.000842	0.09%

[**] The model was re-estimated while eliminating the hhsize variable. All parameter estimates were more or less identical. An ANOVA test was conducted to determine whether the model is better off without hhsize. Nevertheless, the null hypothesis could not be rejected and it was decided to keep the variable.

	Region	0.09183	9.53%
	Household	0.22514	23.37%
	Residual	0.6454	67.00%
Month	Time Period	0.000369	0.04%
	Region	0.091807	9.53%
	Household	0.22555	23.42%
	Residual	0.64541	67.01%
Month by Year	Time Period	0.00103	0.11%
	Region	0.091856	9.54%
	Household	0.224984	23.36%
	Residual	0.64536	67.00%
Year	Time Period	0.00038	0.04%
	Region	0.0918	9.53%
	Household	0.225812	23.44%
	Residual	0.645279	66.99%

Five mixed effects models were estimated using the same previously described variables, but while varying the time period component (table 4). That is, mixed model 1 was assigned “Season” as its time variable, mixed model 2 was assigned “Season by Year”, mixed model 3 was assigned “Month” as its time variable, etc... The objective of including a time period as a random effect is to understand the total variation in the dependent variable – total trip distance travelled in this case – attributed to a specific temporal variable. If the random effect is significant, and the VPC is substantive, then it is safe to say that continuous surveys are more effective than cross-sectional surveys in the sense that the variation of trip behavior over time may be observed.

Interestingly, all time period random effects were proven to be statistically significant via a chi-squared test. Nevertheless, the VPC of every single time period is below 1%. This means that only a small fraction (<1%) of the variance of the total trip distance travelled may be explained by varying time periods. Most of the variation in the total trip distance covered was explained by the differences between individuals (~67%), followed by the variation between households (~23%), and that between regions (~9.5%). The relatively large VPC for households and regions gives support for active research areas in transportation planning that tackle household interactions (e.g. who gets the car?) (Roorda, et al., 2009). Further, the fact that more than 30% of the variation in trip behavior is explained by the different random effects implies that the data exhibit some degree of clustering (Rabe-Hesketh & Skrondal, 2012).

4.2 Household level mixed effects model

Similar to the individual-level model, the effect of clustering was taken into account by nesting households in regions. The regions were also crossed with time periods to assess the temporal variability of travel distance on the household level.

Table 5 lists the fixed effects chosen along with their parameter estimates, t-statistics and confidence intervals. All variables were shown to be significant with the expected signs. The results show that travel distance is positively correlated with increasing income, car ownership, and household size.

Table 5. Seasonal household level mixed effects model

Variable	Description	Estimate	Std. Error	t-value
(Intercept)	N/A	1.910648	0.0991	19.28
income2	$20,000 - $40,000	0.320781	0.014214	22.57
income3	$40,000 - $60,000	0.538065	0.015092	35.65
income4	$60,000 - $80,000	0.651359	0.016648	39.13
income5	$80,000 - $100,000	0.743763	0.019633	37.88
income6	$100,000+	0.771728	0.018289	42.2
carown	Car ownership	0.191677	0.005612	34.15
hhsize	Household Size	0.243418	0.003801	64.05

Five temporal variables were assessed for significance: month, season, year, month by year and season by year. All temporal variables were shown to be significant via an ANOVA test. Contrary to the individual level analysis, the VPCs were calculated for both region and municipal sector spatial units.

The variance contribution of temporal variables to the total variance in total travel distance on the household level was less than 1%. Thus, although statistically, a temporal variation exists, the magnitude of that significance is negligible. Table 6 provides a summary of the variation contribution of the different temporal variables on the dependent variable, alongside the other random effects. It is evident from the table that approximately 90% of the trip variation in the dependent variable is due to the variation within regions/municipal sectors and between households, with the remaining 9%-to-10% is attributed to differences between regions/municipal sectors. Minor differences in VPC were reported between the group of models estimated with a region grouping variable versus municipal sector.

Table 6. Household level VPC analysis

		Region		Municipal Sector	
Time Period	Cluster	Variance	VPC	Variance	VPC
Season	Time Period	0.000321	0.04%	0.0003247	0.04%
	Spatial Unit	0.076363	8.68%	0.0886567	10.21%
	Residual	0.8029	91.28%	0.7791814	89.75%
Season by Year	Time Period	0.00223	0.25%	0.002249	0.26%
	Spatial Unit	0.07707	8.75%	0.089473	10.29%
	Residual	0.80116	90.99%	0.777406	89.45%
Month	Time Period	0.0003724	0.04%	0.0004068	0.05%
	Spatial Unit	0.0764017	8.69%	0.0886413	10.21%
	Residual	0.8027961	91.27%	0.7790549	89.74%
Month by Year	Time Period	0.002449	0.28%	0.002513	0.29%
	Spatial Unit	0.077042	8.75%	0.089387	10.29%

	Residual	0.800837	90.97%	0.777058	89.42%
Year	Time Period	0.002017	0.23%	0.001882	0.22%
	Spatial Unit	0.07745	8.79%	0.089629	10.31%
	Residual	0.801711	90.98%	0.777955	89.48%

4.3 Spatial Level Mixed Effects Model

Unlike individuals and households, a continuous survey dataset is likely to exhibit repeated observations at a spatial unit recorded over time. This is especially true if the spatial unit is large (such as a region). In other words, the continuous dataset on the region level exhibits a panel like structure. Therefore, a more interesting modelling exercise is to investigate the contribution of various temporal variables to the variation in the total variance of the dependent variable on different spatial units.

A series of random-intercept only models (with the exception of including the logarithm of population for all municipal sector models) were estimated on the region and municipal sector level, with data aggregated temporally over seven different time periods: year, day of week, day by year, month, month by year, season and season by year. All of the temporal random effects mentioned were found to be significant at the 95% confidence interval. Table 7 provides a summary of the VPC results.

Table 7. Spatial level VPC analysis

		Region		Municipal Sector	
Time Period	Cluster	Variance	VPC	Variance	VPC
Season	Time Period	0.017787	2.89%	0.01531	25.65%
	Spatial Unit	0.596044	96.92%	0.02824	47.32%
	Residual	0.001127	0.18%	0.01613	27.03%
Season by Year	Time Period	0.025871	4.09%	0.02635	19.63%
	Spatial Unit	0.599652	94.79%	0.03013	22.44%
	Residual	0.007113	1.12%	0.07778	57.93%
Month	Time Period	0.046105	7.15%	0.05675	35.37%
	Spatial Unit	0.592891	91.99%	0.03032	18.90%
	Residual	0.005521	0.86%	0.07339	45.74%
Month by Year	Time Period	0.04018	6.04%	0.0497	15.67%
	Spatial Unit	0.60056	90.25%	0.02905	9.16%
	Residual	0.02469	3.71%	0.23835	75.17%
Day of Week	Time Period	0.003888	0.65%	0.002984	4.87%
	Spatial Unit	0.594501	99.09%	0.027841	45.48%
	Residual	0.001575	0.26%	0.030392	49.65%
Day of Week by Year	Time Period	0.009915	1.61%	0.007831	4.71%
	Spatial Unit	0.599095	97.15%	0.029191	17.57%

	Residual	0.00764	1.24%	0.129124	77.72%
Year	Time Period	0.0066726	1.11%	0.003186	7.24%
	Spatial Unit	0.5945349	98.79%	0.029365	66.72%
	Residual	0.0005903	0.10%	0.011461	26.04%

Approximately 90% to 98% of regional level trip generation variance may be attributed to between region differences. Nevertheless, the VPC analysis for the region level model provided unique insights on the effect of various temporal variables, and between and within region differences, on trip generation. It can be observed that, on the year level, approximately 1% of the variation in trip generation is to be attributed to between year differences. Almost all of the remaining variance is attributed to between region differences. Therefore, it can be concluded that no major changes in trip generation occurred over the 4-year period of the continuous survey due to differences in years. The between season and between month VPCs were larger at approximately 3% and 6%, respectively. That is, the variation in a trip generation on the region level is increasingly explained by more disaggregate time units. This may be attributed to the fact that seasons and months may differ significantly from one another affecting travel behavior (potential reasons: weather changes, school year, vacations calendar, etc..) as opposed to a homogeneous set of years. Nevertheless, the trend does not follow for the between day VPC as weekday day-to-day trip generation may not exhibit significant differences.

On the other hand, the VPC analysis for the municipal sector model yielded much larger time period coefficients with 7% for between year, 25.7% for between season and 35.4% for between month variation. The results indicate that a larger proportion of trip generation behavior can be explained when modelling on a more disaggregate spatial scale. One potential reason may be due to the land use and built environment differences that can be observed when comparing smaller geographic units as opposed to larger ones, influencing the mode of trips selected and the number of trips generated by residing populations. Moreover, a significant proportion of the variance in a trip generation is explained by the within municipal sector differences (differences in the trip generation between households and individuals for example) with the VPC ranging from 26% to approximately 46%. This is much larger than what can be observed on the region level for, while regional residents may exhibit behavioral differences, the average regional trip generation is potentially more or less the same.

4.4 Individual Trip Level Mixed Effects Model

The mixed effects modelling exercise was then extended to investigate the relationship between the logarithm of trip distance and a number of explanatory variables for every individual trip captured by the Montreal Continuous Survey. That is, the multiple trips conducted by an individual were modelled independently. Trips were nested in modes, and modes were crossed with regions and time periods. Ideally, the clustering effect of every individual person should be taken into account. However, adding such a random effect will multiply the complexity of the model leading to a failure in conversion.

Table 8 lists the fixed effects chosen along with their parameter estimates, t-statistics and confidence intervals. The seasonal model coefficients were chosen to remain consistent with the table results displayed on the individual level model. The parameter estimates as a result of varying the time component were very similar when compared to the seasonal model. All variables were shown to be significant with the expected signs. Household size was again an exception in this modelling exercise as the parameter estimate produced a negative sign. This may be because as the household size increases, individual trip distance per person may decrease as the chore of travelling is distributed across the many residents of the household. Aside from household size, trip distance increases with income and age, while women seem to travel on shorter trips than their male counterparts.

Table 8. Seasonal trip level mixed effects model

Variable	Description	Estimate	Std. Error	t-value
(Intercept)	N/A	1.8724	0.249	7.51
income2	$20,000 - $40,000	0.0382	0.01	3.72
income3	$40,000 - $60,000	0.1076	0.011	10.23
income4	$60,000 - $80,000	0.1355	0.011	12.25
income5	$80,000 - $100,000	0.168	0.012	13.56
income6	$100,000+	0.2054	0.011	17.93
Age	N/A	0.0038	0.0003	13.4
Female	N/A	-0.087	0.006	-15.67
occ_status1	Full Time	0.4584	0.011	40.52
occ_status2	Part Time	0.2789	0.015	18.01
occ_status3	Student	0.1655	0.017	9.55
occ_status5	Work at Home	0.1803	0.019	9.33
occ_status6	Retired	0.1571	0.019	8.11
hhsize	Household Size	-0.006	0.002	-2.53

4.4.1 Variance partition coefficient analysis

Five temporal variables were assessed for significance: month, season, year, month by year and season by year. Also, the analysis was conducted for both regions and municipal sectors. All temporal variables were shown to be significant via an ANOVA test, with the exception of the year variable. This indicates that no significant change has happened in the variation of the trip distance between years.

The variance contribution of temporal variables to the total variance in the trip distance was less than 1%. The results are similar to that of the individual level and household level modelling exercises and are expected since every trip is observed once (no repeated observation per trip). Table 9 provides a summary of the variance contribution of the different temporal variables on the dependent variable, alongside the other random effects. It is evident from the table that approximately 60% of the trip variation in the dependent variable is due to the variation between trips and within modes. Also, approximately 35% of the variation in trip distance is attributed to between mode differences, indicating that the choice of travel mode is quite significant to understanding travel behavior. Finally, about 3% of the variation in trip distance is attributed to between-region differences, while between municipal sector differences explain about 7% of that variation. This is in line with the results of previous modelling exercises in this paper.

Table 9. Trip level VPC analysis

	Region	Municipal Sector

Time Period	Cluster	Variance	VPC	Variance	VPC
Season	Time Period	0.000686	0.05%	0.000666	0.05%
	Spatial Unit	0.04173	3.34%	0.088759	7.03%
	Mode	0.449462	35.92%	0.435268	34.50%
	Residual	0.759269	60.72%	0.737024	58.41%
Season by Year	Time Period	0.000692	0.06%	0.000635	0.05%
	Spatial Unit	0.04182	3.34%	0.08886	7.05%
	Mode	0.448743	35.89%	0.434725	34.47%
	Residual	0.759136	60.71%	0.736926	58.43%
Month	Time Period	0.000696	0.06%	0.000656	0.05%
	Spatial Unit	0.041707	3.33%	0.088687	7.03%
	Mode	0.449362	35.92%	0.435061	34.49%
	Residual	0.759163	60.69%	0.736939	58.42%
Month by Year	Time Period	0.00093	0.07%	0.000862	0.07%
	Spatial Unit	0.041823	3.34%	0.088941	7.05%
	Mode	0.448701	35.89%	0.434487	34.46%
	Residual	0.758894	60.69%	0.736689	58.42%
Year	Time Period	9.26E-06	0.00%	0	0.00%
	Spatial Unit	4.17E-02	3.34%	0.08873	7.04%
	Mode	4.48E-01	35.83%	0.43368	34.42%
	Residual	7.60E-01	60.83%	0.73752	58.54%

4.5 Modal level mixed effects model

Contrary to individual trips, a continuous survey dataset is likely to exhibit repeated observations on the modal level. Therefore, in an attempt to investigate the temporal variation in travel behavior, a series of random-intercept only models were estimated at the modal level. That is, data were aggregated by mode, alongside the commonly used spatial and temporal variables. The modelling exercise was carried out for both regions and municipal sectors, with data aggregated temporally over five different time periods: season, season by year, month, month by year, and year. The dependent variable was chosen to be the logarithm of trip distance.

All of the temporal random effects mentioned were found to be significant at the 95% confidence interval, with the exception of the year random effect in the regional modelling exercise. Table 10 provides a summary of the variance partition coefficient results.

Table 10. Modal level VPC analysis

		Region		Municipal Sector	
Time Period	Cluster	Variance	VPC	Variance	VPC
Season	Time Period	0.03797	0.92%	0.01429	0.39%
	Spatial Unit	0.90423	21.92%	0.38175	10.30%

	Mode	2.41213	58.48%	2.36938	63.92%
	Residual	0.77061	18.68%	0.94114	25.39%
Season by Year	Time Period	0.02567	0.66%	0.01752	0.57%
	Spatial Unit	0.80538	20.76%	0.32297	10.55%
	Mode	2.3217	59.85%	1.87548	61.26%
	Residual	0.72643	18.73%	0.8632	28.19%
Month	Time Period	0.0416	1.03%	0.03959	1.23%
	Spatial Unit	0.8283	20.56%	0.33714	10.48%
	Mode	2.392	59.37%	1.98853	61.79%
	Residual	0.7672	19.04%	0.85298	26.50%
Month by Year	Time Period	0.03473	0.94%	0.02623	1.08%
	Spatial Unit	0.67824	18.43%	0.25832	10.63%
	Mode	2.19074	59.52%	1.29124	53.11%
	Residual	0.77673	21.10%	0.85533	35.18%
Year	Time Period	0.00228	0.06%	0.00553	0.15%
	Spatial Unit	0.88543	24.84%	0.39501	10.84%
	Mode	2.19569	61.60%	2.36861	65.00%
	Residual	0.48128	13.50%	0.87503	24.01%

The hypothesis in this paper has been that if a particular variable, such as mode or region/municipal sector, exhibited repeated observations, then the magnitude of the temporal VPC is likely to be significant. That is, a sizable proportion of the total variance of the dependent variable is explained by the temporal random effect. Nevertheless, the results showed that, at least for the modal level modelling exercise, the temporal random effect explains very little (less than 1%) of the total trip distance variance. There may be two main reasons for such a conclusion. The first is that the variance contribution of the temporal variables is overshadowed by the between-mode differences. Indeed, the between-mode differences are attributed between 58% and 65% of the overall variation in the trip distance by mode. The other reason may be that travel behavior over the selected time periods is homogeneous. That is, individuals, travel the same distance by mode every month, season or year. Intuitively, this explanation may stand for auto and transit users but is rather difficult to justify for active modes such as walking and cycling. The next section is devoted to investigating whether temporal variation in trip behavior may be observed for active modes. Here, active modes are defined as either walking or cycling trips (Mahmoud, et al., 2015).

Aside from the between mode differences, the within mode differences were attributed between 13% to 26% of the total variation in modal travel distance. In addition, the between region/municipal sector differences were attributed anywhere between 10% and 25% of the total variation.

It is important to note that the analysis in this section was repeated for (the logarithm of) trip counts by mode as a dependent variable to validate the results. However, the aforementioned conclusions were largely similar.

4.6 Active mode level mixed effects model

After aggregating trips by mode, a subset of the dataset that includes trips conducted by walking or biking was taken out and used for modelling. A random effects model was then estimated with the log of trip distance as the dependent variable for a set of models, and the log of trip counts by mode for another set of models. The grouping factors (random effects) considered were mode, region and different time periods (season, year, month, season by year, month by year). Only the region grouping factor was considered as the number of trips, or total trip distance covered, by active modes would have been too thinly distributed across different municipal sectors for analysis purposes. Table 11 summarizes the obtained VPC results.

Table 11. Active mode VPC analysis

		Trip Distance		Trip Rates	
Time Period	Cluster	Variance	VPC	Variance	VPC
Season	Time Period	0.4605	23.96%	0.4056	14.50%
	Region	0.4846	25.21%	0.5183	18.53%
	Mode	0.3465	18.03%	1.3849	49.50%
	Residual	0.6304	32.80%	0.489	17.48%
Season by Year	Time Period	0.2917	18.01%	0.2028	9.39%
	Region	0.5671	35.01%	0.5315	24.62%
	Mode	0.2513	15.51%	1.1135	51.57%
	Residual	0.5097	31.47%	0.3114	14.42%
Month	Time Period	0.3356	18.34%	0.2897	11.63%
	Region	0.56	30.61%	0.5372	21.57%
	Mode	0.323	17.65%	1.2561	50.43%
	Residual	0.611	33.40%	0.4078	16.37%
Month by Year	Time Period	0.1418	10.21%	0.1075	6.47%
	Region	0.475	34.21%	0.4352	26.19%
	Mode	0.1346	9.70%	0.8035	48.36%
	Residual	0.6369	45.88%	0.3153	18.98%
Year	Time Period	0.02391	2.87%	0.00993	0.62%
	Region	0.5986	71.87%	0.60677	38.18%
	Mode	0.11165	13.40%	0.89716	56.45%
	Residual	0.09879	11.86%	0.07549	4.75%

Interestingly, the VPCs of the different temporal variables were significant at the 95% confidence interval (with the exception of the year variable for the trip counts model) and ranged from 1% to 25%. That is, approximately 1% to 25% of the variation in travel behavior, whether it is trip distance or a number of trips, is attributed to between time period differences (e.g. season to season). This means that, in the case of active modes, the temporal nature of the data are heteroscedastic. This is in contrast to the conclusion of the previous section, where the temporal component of the estimated models proved negligible in explaining the variation in travel behavior. Here lies the advantage of continuous surveys, as their continuous data elements can be leveraged to conduct time series analysis and identify temporal trends for various policy purposes.

In the case of total travel distance covered, anywhere from 25% to 71% of the total variance may be attributed to between-region differences. Further, modal differences still played a role in explaining the dependent variable variance (10% to 18%). On the other hand, in the case of trip rates, or count of trips by mode, between mode differences played a bigger role in explaining the variance of the dependent variable. That may be because, while the trip distances covered by walking and cycling can likely be very similar, the number of trips by each mode can vary significantly. It is also possible that such trips are under-reported in the Montreal Continuous Survey. The same set of models were estimated for the remaining dataset that included all other modes with the exception of walking and biking. The variance component in travel behavior attributed to the temporal component of the model was below 1%.

5. Conclusion

The VPC analysis conducted suggests that only a very small percentage of the total variation in trip distance travelled by individuals and/or households in a typical weekday can be attributed to the variation between time periods. This begs the question of whether continuous surveys are any more advantageous than large one-off cross-sectional surveys, the dominant practice in most major Canadian cities (Habib & El-Assi, 2015), for investigating temporal differences in trip behavior on the household or individual level. However, the continuous nature of the data may allow for time series analysis of trip behavior at a more aggregate level such as zones/municipal sectors, modes or regions due to the presence of repeated observations. Further, continuous surveys can also be used for conducting before and after studies, an area that has yet to be investigated.

The study, however, is not without its limitations. For instance, to develop a more elaborate understanding of the Montreal metropolitan area trip behavior, it is imperative to also investigate the subset of the population that did not conduct any trips on the day of the survey. Further, random coefficients were not introduced as part of the modelling structure. Such variables can alter the variance of the dependent variable, thus affecting the calculated VPCs. Moreover, repeated observations are likely to be available for the different modes provided in the survey. As such, a VPC analysis can be conducted to investigate the variation attributed to modal differences in explaining travel behavior.

Acknowledgements

The authors would like to thank the AMT (Metropolitan Transportation Agency) for providing access to the data (continuous survey) for research purpose, as well as to Hubert Verreault (Polytechnique Montreal) for his contributions in data processing.

References

Ampt, E. & Ortuzar, J., 2011. On Best Practice in Continuous Large-Scale Mobility Surveys. *Transportation Reviews,* 24(3), pp. 337-363.

Ampt, E. & Stopher, P., 2006. Mixed Methods Data Collection in Travel Surveys: Challenges and Opportunities. *Presented at the 28th Australian Transport Research Forum.*

Borgoni, R., Ewert, U. & Furnkranz-Prskawetz, A., 2002. *How Important are Household Demographic Characteristics to Explain Private Car Use Patters? A Multilevel Approach to Austrian Data,* s.l.: Max Plank Institute for Demographic Research Working Paper.

Data Management Group, 2013. *Transportation Tomorrow Survey.* [Online] Available at: http://www.dmg.utoronto.ca/transportationtomorrowsurvey/ [Accessed 2015].

DiPrete, T. & Grusky, D., 1990. The Multilevel Analysis of Trends with Repeated Cross-Sectional Data. *American Sociological Association,* Volume 20, pp. 337-368.

Fiona, S., 2008. Module 5: Introduction to Multilevel Modelling Concepts. In: *Learning Environment for Multilevel Methodology and Applications.* s.l.:University of Bristol, Centre for Multilevel Modelling.

Fiona, S., 2011. Module 6: Regression Models for Binary Response Concepts. In: *Learning environment for multilevel methodology and applications.* s.l.:University of Bristol, Centre for Multilevel Modelling.

Garson, G., 2013. *Hierarchical Linear Modelling - Guide and Applications.* 1st ed. Thousands Oak, California: SAGE Publications Inc..

Goulias, K., 2002. Multilevel Analysis of Daily Time Use and Time Allocation to Activity Types Accounting For Complex Covariance Structures Using Correlated Random Effects. *Transportation,* Volume 29, pp. 31-48.

Habib, K. & El-Assi, W., 2015. How Large is too Large? The Issue of Sample Size Requirements of Regional Household Travel Surveys, the Case of the Transportation Tomorrow Survey in the Greater Toronto and Hamilton Area. *Presented at the 95th Annual Meeting of The Transportation Research Board.*

Hanson, S. & Johnston, I., 2013. Gender Differences in Work Trip Length: Explanations and Implications. *Urban Geography,* 6(3), pp. 193-219.

Kish, L., 1965. *Survey Sampling.* New York: John Wiley & Sons Inc..

Lipps, O. & Kunert, U., 2005. Measuring and Explaining The Increase in Travel Distance: A Multilevel Analysis Using Repeated Cross-Sectional Travel Surveys. *DIW-Diskussionspapiere,* Volume 492.

Mahmoud, M., El-Assi, W., Habib, K. & Shalaby, A., 2015. How Active Modes Compete With Motorized Modes in High Density Areas: A Case Study of Downtown Toronto. *Canadian Transportation Research Forum.*

Meyer, M. & Miller, E., 2001. *Urban Transportation Planning.* Second ed. New York: McGraw Hill.

Miller, M. et al., 2012. *Changing Practices in Data Collection on the Movement of People,* Quebec: Lee-Gosselin Associates Limited.

Ortuzar, J., Armoogum, J., Madre, J. & Potier, F., 2010. *Continuous Mobility Surveys: The State of Practice,* s.l.: Transport Reviews.

Peachman, J. & Battellino, H., 2007. The Joys and Tribulations of A Continuous Survey. *International Conference On Transportation Survey Quality and Innovation.*

Rabe-Hesketh, S. & Skrondal, A., 2008. *Multilevel and longitudinal modeling using Stata.* s.l.:STATA press.

Rabe-Hesketh, S. & Skrondal, A., 2012. *Multilevel and Longitudinal Modelling Using Stata.* Texas: Stata Press.

Rashad, M., Srikuhenthiran, S., Habib, K. & Miller, E., 2015. *Current State of Smart Phone Survey Methods,* Toronto: University of Toronto Transportation Research Institute.

Roorda, M., Carrasco, J. & Miller, E., 2009. An integrated model of vehicle transactions, activity scheduling and mode choice. *Transportation Research Part B: Methodological,* 43(2), pp. 217-229.

Scott, M., Shrout, P. & Weinberg, S., 2013. Multilevel Modelling Notation-Establishing Commonalities. In: *The SAGE Handbook of Multilevel Modelling.* London: SAGE Publications ltd.

Stopher, P. & Greaves, S., 2007. Household Travel Surveys: Where Are We Going?. *Transportation Research Part A,* Volume 41, pp. 367-381.

Tremblay, P., 2014. *An overview of OD-Surveys in Quebec.* [Online]
Available at: http://uttri.utoronto.ca/files/2014/10/5-Tremblay-An-Overview-of-OD-Surveys-in-Quebec.pdf
[Accessed 2015].