WEIGHTING METHOD OF HOUSEHOLD TRAVEL SURVEY TO ALLEVIATE DISCREPANCIES IN TRANSPORT PLANNING MODELS

Professor Robert CHAPLEAU, ing., Ph.D.
Department of Civil, Geological and Mining Engineering
École Polytechnique de Montréal, Université de Montréal
Director of MADITUC group

Daniel PICHÉ, ing., M.Sc.A.
Research Associate, MADITUC group
Department of Civil, Geological and Mining Engineering
École Polytechnique de Montréal, Université de Montréal

ABSTRACT

The Montreal Region has maintained the tradition over the last 40 years of conducting large scale (5% sample of the household universe of the Greater Montreal Area) so-called Origin-Destination travel surveys. Every 5 years, about 65,000 households are subjected to a CATI (Computer Assisted Telephone Interview) method in which every trip made by each member of the household is characterized in terms of origin, destination, purpose, departure time, travel modes, bus routes taken, car ownership, status, etc… The information is interactively validated, and trip ends, junction points (subway or railway stations), bridges, freeways, monuments, trip generators and residences are geo-coded. This type of survey is usually fundamental for a large set of transit and roadway simulation projects.
In the past, several weighting factors were independently determined for household, persons and trips, thus creating diverse inconsistencies. The available reference databases consist of total population demographic distribution by 5-year cohorts and enumeration of private households, for about 100 municipal sectors.
To acknowledge the fact that personal travel behaviour is conditioned by household structure, age, and residence location, and to investigate possible biases in the observed sample, the weighting method utilizes an iterative proportional fitting algorithm (Furness type growth factor method) to respect the household composition distribution and a simplified 5-behavior-typical-cohort demographic distribution. The weights are also constrained to some specific variation, and the method application results in about 7,000 different weighting factors preserving the coherence and consistence between household and population surveyed travel behaviour. Impacts from different weighting strategies are compared.

Keywords: household travel survey, Origin and Destination, population behaviour, weighting method, iterative proportional fitting, Montreal case

INTRODUCTION

Montreal household travel survey context

Household travel surveys (HTSs) are one of the most and common ways to collect fundamental information for planning and simulating transit and road networks. In the Greater Montréal Area (GMA), Origin-Destination travel surveys began in the 1970's. Since then, large scale HTSs (a sample of around 5%, or nearly 65 000 households) have been conducted every five years, with each subsequent HTS bearing witness to methodological and technological improvements (Trépanier et al., 2008). In 1998 the CATI (Computer-Assisted Telephone Interview) method was incorporated into the HTS. In addition to guiding the interviewers during the interview process, the survey software interactively gather, validate information (e.g. transit route taken, age and driver's license) and geocodes, with complete GIS-T reference, all locations related (residences, trips ends, junction points - subway and railway stations-, trip generators, bridges and freeways). More information on the Montréal OD travel survey software is documented in Chapleau et al. (2001). The most recent HTS was conducted in autumn 2008 and covered an area of over 8 120 square kilometers. From 220 000 phone calls and 66 100 completed interviews, a database of 66 100 households, 157 000 persons and 323 000 trips were assembled.

The processing of HTS data is often perceived as usual activity with less interest. However, processing activities, more precisely data weighting, are key steps to achieve high quality data (Sammer, 2006). Traditionally, Montreal travel survey units - households and persons/trips - are weighted independently without consideration of the inherent relationship between each other. Trip weights are estimates based on a person's demographic attributes without consideration of the household to which the person belongs. The reference population for HTS weighting comes from the Canadian census, which, like HTS is conducted every five years (but with a two years interval). Total population demographic distribution by 5-year cohorts and enumeration of private households are specific available data. Yet it is known that travel behavior is influenced by decisions and constraints imposed by the household. Furthermore, household activities are influenced by location and the age of its members. Key facts about demography and mobility confirm that personal travel behaviour is conditioned by household structure, age, and residential location. Different geo-demographic demonstrations relying on totally disaggregate analyses of HTS data have revealed consistent spatial phenomena over time, such as demography and travel behaviours (Chapleau and Morency, 2004). In Figure 1 below, certain spatial trends may be observed:

- A proportional relationship between a household's distance from the CBD and its size. In general, one-person households live in the central districts whilst family-households prefer suburban areas ;
- An inversely proportional relationship between distance from CBD and the transit share; transit share is higher in the central districts while it is basically insignificant in suburban areas.

> "The simultaneous examination of demographic and mobility indicators furthers the rendering of urban interactions. The major attributes retained for this analysis are synthesised in the following table (Table 1). Hence, Chernoff faces are used to compare the features of each region for the 1987 and 1998 Origin-Destination surveys (10-years evolution). This table also describes the features used in the Chernoff faces to illustrate the extent of each mobility or demographic attribute. The statistical technique allows easy detection of the similarities and differences between regions and time periods. It also permits the five detailed comparisons between specific features. Figure 2 presents the corresponding Chernoff faces. The heavy trends are rapidly detected." (Chapleau and Morency, 2004)

For data dissemination purposes, the GMA is traditionally separated in eight geopolitical significant regions like those in Figure 2. Theses analytical regions will be used throughout this paper as spatial units for the purpose of presenting analysis results.

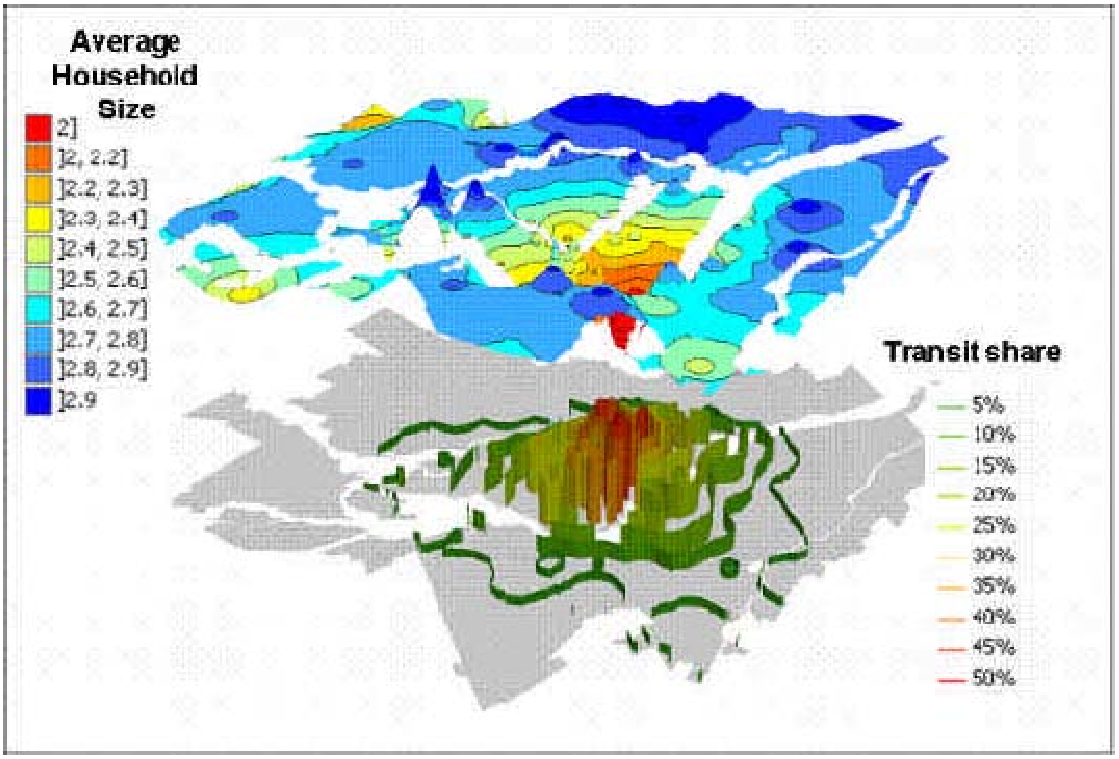


Figure 1- 3D visualisation of spatial processes: household size and transit share [Chapleau and Morency, 2004]

Table 1- Features of the Chernoff faces: 17 attributes associated with the eight analysis region of the GMA (comparable scale for the two surveys) [Chapleau and Morency, 2004]

Facial Feature	Variable	Minimum	Maximum	Average 87	Average 98
Face width	people / household	1.59	3.04	**2.55**	**2.42**
Ear level	cars / household	0.44	1.68	**1.05**	**1.16**
Half-face height	Households / km^2	58	2 936	**336**	**391**
Eccentricity of upper face	People / km^2	176	6 090	**855**	**947**
Eccentricity of lower face	% 1-person households	11.2%	60.6%	**25.3%**	**29.5%**
Length of nose	% 0 car households	5.1%	61.4%	**25.9%**	**22.8%**
Position of center of mouth	% zero-trippers	19.5%	23.8%	**22.7%**	**21.3%**
Curvature of mouth	% 00-15 years old	6.4%	27.3%	**19.6%**	**20.0%**
Length of mouth	% 65 years old and +	5.4%	14.4%	**9.8%**	**10.8%**
Height of center eyes	Men / Women ratio	0.89	1.18	**0.93**	**0.97**
Separation of eyes	% Trips to CBD	2.6%	76.9%	**8.9%**	**8.2%**
Radius of eyes	Transit share (%)	2.3%	54.9%	**25.0%**	**17.2%**
Eccentricity of eyes	Motorised trips / person	1.30	2.03	**1.71**	**1.89**
Half-length of eye	Km / household	16.06	77.57	**39.43**	**46.62**
Position of pupils	Average household income	36 176 $	54 147 $	**44 634 $**	**44 997 $**
Height of eyebrows	% Single-detached house	0.2%	75.9%	**27.7%**	**30.2%**
Angle of eyebrows	% dwelling owners	10.3%	79.3%	**44.6%**	**48.0%**

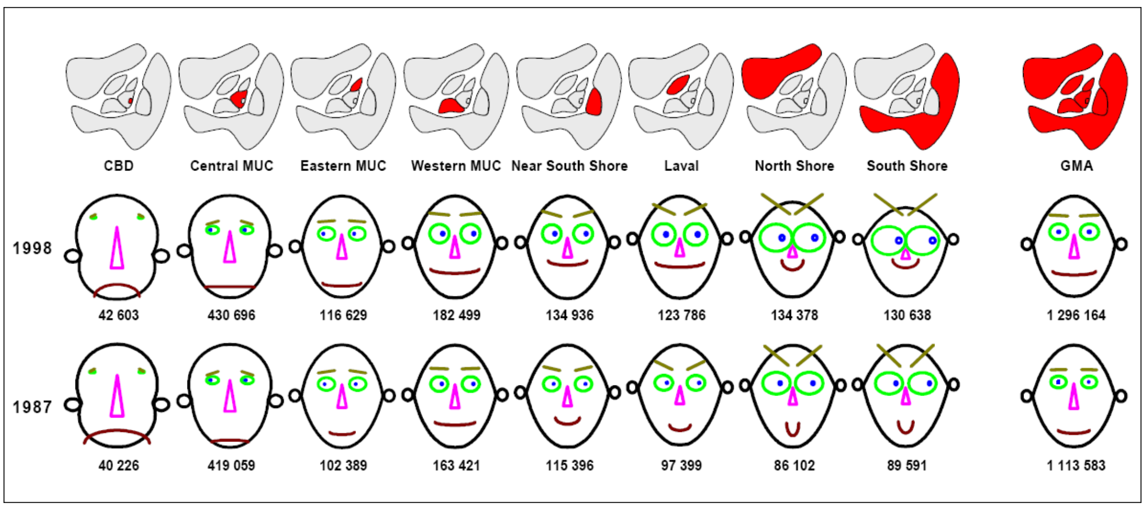


Figure 2- Chernoff faces for the eight analytical regions of the GMA - data from the 1987 and 1998 Origin Destination surveys [in Chapleau and Morency, 2004]

Conceptual Overview

From acknowledged geodemographic confirmations, a new method of determining the weights given to travel survey observation units was explored for the 2008 Montreal HTS, with the aim of preserving consistency between *household* and *person* units. Figure 3 conceptually presents an overview of HTS concepts.

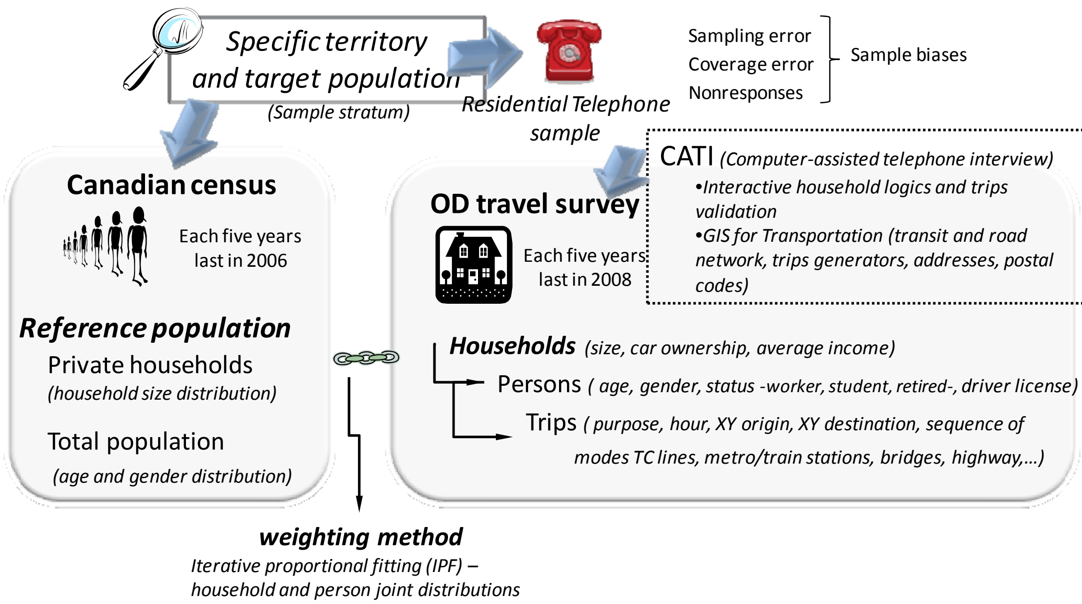


Figure 3- Conceptual schema and objects of the Households Travel Survey background

To achieve HTS data calibration, an iterative proportional fitting (IPF) procedure is proposed in this paper. First presented for estimating contingency table from known marginal distributions (Deming and Stephan, 1940), IPF is now used in diverse application domains (Norman, 1999). In transportation engineering, the IPF occurs during distribution of the origin-destination matrix as part of the standard transport modelling process (Ortuzar and Willumsen, 2006). In travel survey weighting, the IPF procedure is usually applied individually to household and to person level (Stopher et al., 2006). IPF has also found use in techniques developed for the generation of synthetic populations as part of the microsimulation of travel behaviour (Beckman et al., 1996). Unfortunately, the IPF procedure used in creation of synthesis population reconstructs joint distributions at household or person-level attributes without guarantee for consistency with the other observation unit. More recently, a number of studies in population synthesis note this problem (Guo and Baht, 2007; Arentze et al., 2007; Ye et al., 2009). Along these lines, an IPF heuristic so-called Iterative Proportional Updating (IPU), and similar at the one uses in this paper, is proposed by Ye et al. (2009). The aim of IPU is to generate household synthetic population with simultaneous consideration of household and persons attributes.

This paper is structured as follows: The first section documents the 2008 Montreal HTS sample composition and biases by a spatial comparison with reference population data. Secondly, the Iterative Proportional Fitting procedure and alternative weighting scenarios are described. This section also offers an overview of weighting statistics and visualizations. Finally, a discussion of the weighting scenarios and future analyses is presented.

2008 OD TRAVEL SURVEY SAMPLE DIAGNOSTIC AND BIAS

Over a period of 72 days (September 2 to December 19, 2008), nearly 220,000 phone calls were necessary to produce 66,100 completed and validated weekday interviews. Figure 4 illustrates the number of households surveyed each day of the HTS, as well as the percentages represented by region. Under ideal circumstances these values would remain constant throughout the survey. Figure 4 highlights the non-uniformity of the sample rate achieved: temporally, and spatially. During the first month we are able to observe a steady rise in the number of completed interviews, whilst the team members honed their interviewing skills and telephone manner and the lack of qualified interviewers was redressed. During the last week of the interview period, previsions in the survey design ensured that the team direct its attention to those sectors of which an insufficient number of households have been contacted. By the end of the survey period, that minimum number of households is surpassed for each sector of the central Montreal Urban Community (MUC) region.

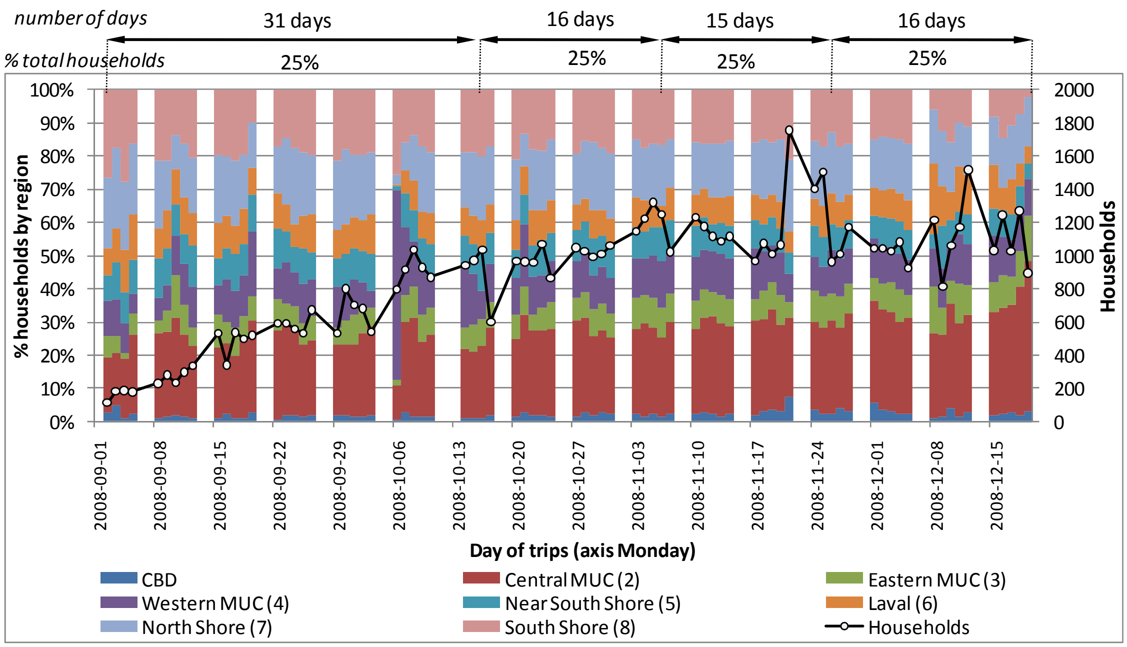


Figure 4 – Week day variability of sample and completed households proportion by region (2008 travel survey)

The refusal rate was about 40.4% (based on eligible households), a significant increase since the first computer assisted and monitored OD travel survey in 1998. A classification of the interviews processed during the last three OD travel surveys is shown in Table 2. It shows the rising inherent difficulties to establish contact with households. This phenomenon is consistent with general tendencies observed in telephone surveys (Stopher, 2007; Tuckel, 2001). Another element to be considered in the evaluation of refusal rates is the interviewers' experience. Analysis of the performance of the interviewers from the 1998 HTS found that the level of experience of the interviewers correlated with the refusal rates (Chapleau, 2003). Over the course of the surveys, as the interviewers gained experience they obtained a lower rate of refusal.

In addition, a significant increase in the number of calls in 2008 that was "screened-out". The only change in methodology for the 2008 survey that may help to explain this occurrence is the phone number screening system undertaken by the firm hired to complete the interviews. Phone numbers deemed ineligible by the screening procedure were not contacted by interviewer and validity of screened out units has not been confirmed by human.

Table 2- State of Montreal 1998, 2003 and 2008 OD survey sample (weekday)

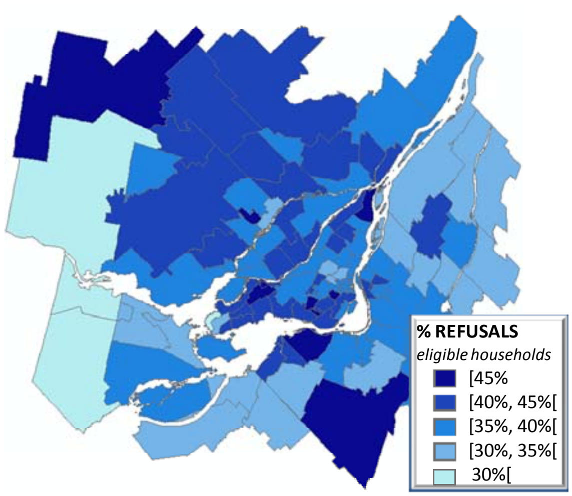
	1998		2003		2008	
Total units used in sample (phone numbers)	**110 938**		**168 958**		**221 202**	
Eligible units	**90 763**	81.8%	**131 235**	77.7%	**137 108**	62.0%
Completed households	*65 227*	*71.9%*	*70 418*	*53.7%*	*66 100*	*48.2%*
Refusals	*19 273*	*21.2%*	*48 910*	*37.3%*	*55 445*	*40.4%*
Breakoffs/Incompletes (1)	*6 263*	*6.9%*	*11 907*	*9.1%*	*15 563*	*11.4%*
Eligibility unknown (2)	**13 437**	12.1%	**21 528**	12.7%	**43 566**	19.7%
Screened out units (3)	**6 738**	6.1%	**16 194**	9.6%	**40 528**	18.3%

(1) Respondent not available, Appointment, Missed appointment, language barrier, breakoff

(2) Answering machine, No answer, Line busy

(3) Non-residential, Outside sample territory, disconnect number, fax number, permanent disability

With regards to the localisation of refused and incomplete calls, a correlation has been shown to exist between the rate of these calls and household's socio-demographic attributes. The main spoken language and the average income are two example characteristics, and indeed for the 1998 HTS (Chapleau, 2003) it was observed that refusal and incomplete call rates correlated negatively with the French-speaking population and correlated positively with the English-speaking population and the more wealthy sectors. Consistent with this observation is Figure 5, which conveys (at left) higher refusal rates in the sectors recognised as majority Anglophone (e.g. those in the Montreal West Island). Figure 5 also conveys (at right) a higher rate of incomplete calls mainly on the Island of Montreal, an area known for its multi-ethnicity and hence a slight predisposition for language difficulties.



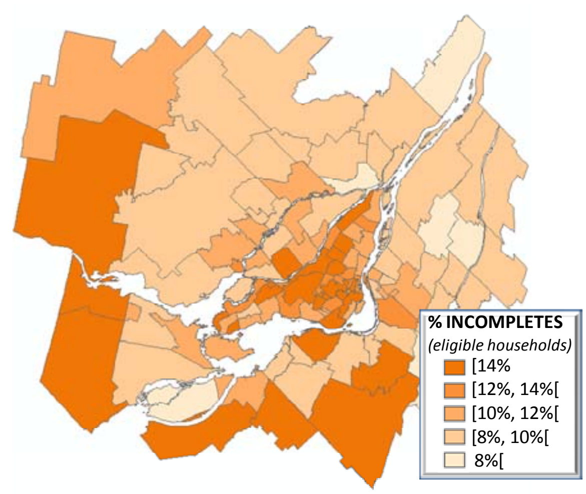


Figure 5 – Spatial distribution of refusals and incompletes rates (2008 travel survey)

Household sample biases

In this section, sample biases are considered by performing a spatial comparison of the population reconstructed from survey data versus the reference population. The spatial distribution of households according to their size is depicted in Figure 6, with classification by households of 1, 2, 3 and 4 or more persons. As is often observed in telephone survey, 1-person households are under-represented. For the 2008 Survey, although the overall sample of households constitutes 4.16% of the entire territory, only 3.55% of 1-person households were contacted.

Another observation that can be made with reference to Figure 6 is the consistent underestimation across all household categories in specific sectors. This underestimation is a consequence of a municipal confusion in the sample units provided by the survey sampler. This resulted in a reduced number of households surveyed from those sectors, since households believed to be located in the particular sectors were actually located elsewhere. This type of sampling error remains difficult to correct when discovered post-survey.

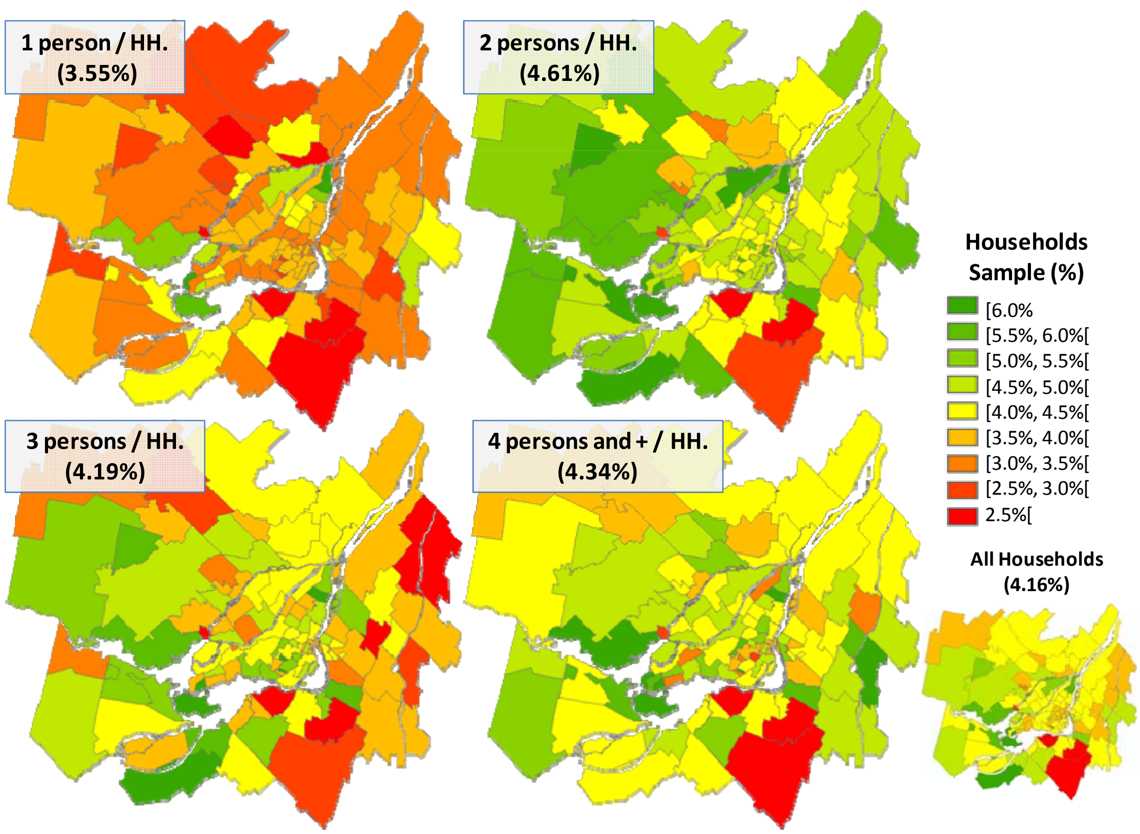


Figure 6 – Sample households by municipality sectors with classification by persons/HH.

A different perspective is given by assessing the combined population of the surveyed households relative to the total reference population of the HTS. To more accurately assess what biases pertaining to individuals are inherent at the scale of the household, every individual who is surveyed is given a weighted score. This weight measures the representativeness of a given individual's household as a function of its size and sector. Figure 7 captures one such population comparison, spatially distributed, and highlights an underestimation of children and young adults (0-39 y.o.) compared to older adults (40+ y.o.). More precisely, a clear underrated of persons aged 25-39 y.o. can be seen across the territory, while persons aged between 15 and 24 are mostly underestimated in the central region (Downtown and surrounding areas). In Figure 8, a more detailed comparison is made between the demographic structures of the population reconstructed from HTS data (colored histogram) and the reference population (hollow histogram), with differentiation of gender and segmentation by cohorts of five years. The structure of the entire survey territory and of each great analysis region is presented side-by-side to illustrate the influence of the location of a place of residence.

This is the first time in the history of the Montreal HTS that young adult part of the population finds itself underestimated to such an extent. Amongst the hypotheses that may account for this bias is an increased prevalence of households foregoing fixed-line telephones in favor of their cell phones. These households were not part of the telephone sample supplied by the provider. Further discussion of the cell-phone only household and other related telephone survey phenomena may be found in other works (Keeter and Kennedy, 2006 ; Stopher and Greaves, 2007).

An annual survey conducted since 1998 by Statistics Canada (Statistics Canada, 2009) adds weight to this hypothesis. The Residential Telephone Service Survey aims to record telephone penetration rates across the country. Available statistics are given per province, with coverage of rural areas. In December 2008, 65.5% of households in the Province of Quebec used at least one cell phone. An estimated 6.7% used only cell phones as their form of communication, compared with 2.2% in 2003. Of most relevance to the current discussion is that the rate of young — i.e. comprised solely of 18-34 year olds — cell-phone only households approaches 25%. Moreover, Statistics Canada (2006) reveals that households in urbanised regions such as Montreal area are most likely to be those with cell phones only. In 2005, this household proportion was around 6.4% in GMA but in 2008 had rise to 8.1%.

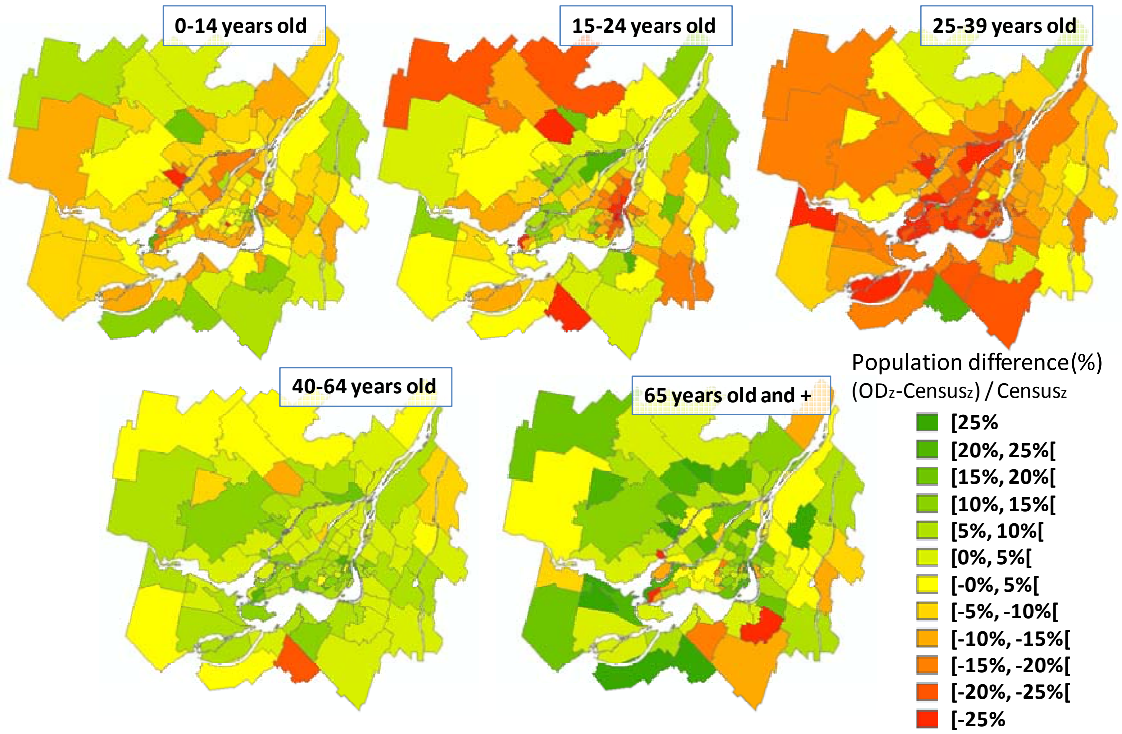


Figure 7 – Spatial distribution (by age group) of population difference (%) between census and OD travel survey population weighted by household size stratification

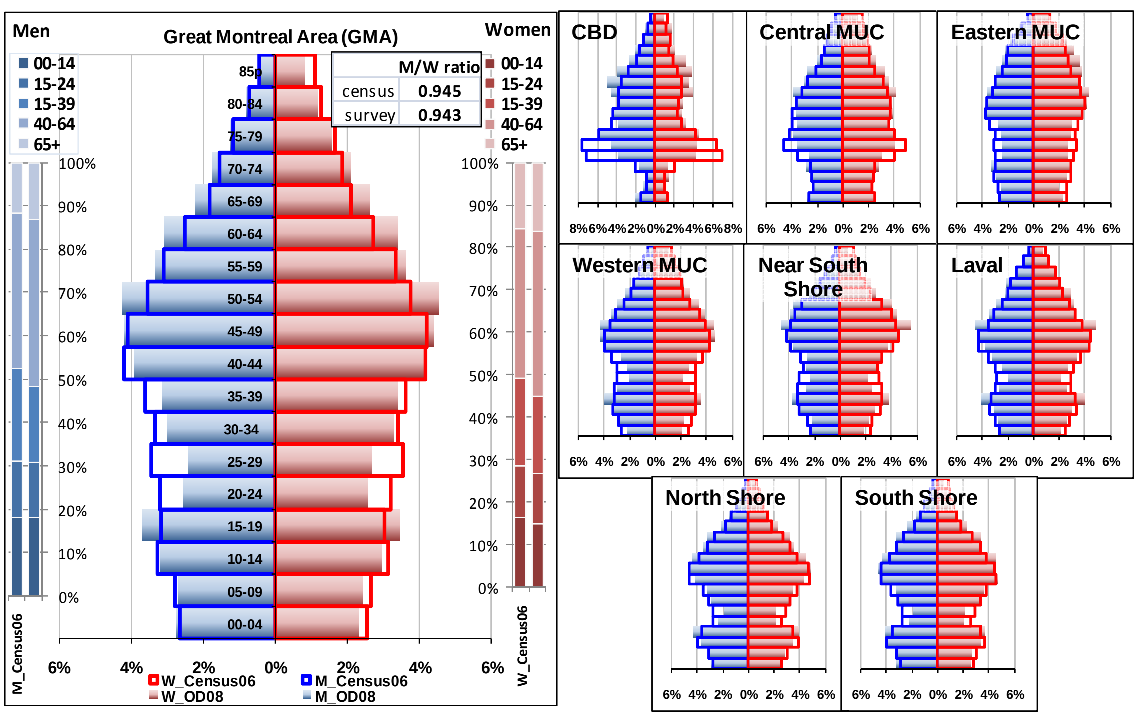


Figure 8 – Demographic profile by five years cohort of person sample and comparison with census population (2008 OD travel survey person sample and 2006 Canadian census)

IPF PROCEDURE AS TRIPS WEIGHTING METHOD

The aim of the weighting method investigated for the 2008 travel survey is to lessen the discrepancies between the survey sample and the reference population whilst maintaining consistency between the weights assigned to the household and person survey units. To achieve this ambition, a calibration of the weights is performed by an iterative proportional fitting (IPF) procedure that recreates distributions of household and person units simultaneously. Ultimately, trips performed by a household are given the same weighting factor as the household.

Population reference data

The available reference databases (2006 Canadian census) provide information on total population demographic distribution by 5-year cohorts and enumeration of private households by number of occupants (1, 2, 3, 4-5 and 6+ persons). However, a problem is encountered when attempting to unify these two census populations. The Canadian census contains data on the distribution of private households by size, but the demographic data (age and gender distribution) is presented only for the total population, which includes persons in private and collective households (collective household refers to residents occupying lodging or rooming houses, hotels, motels, hospitals, staff residences, jails, missions, and so on..). A difference of 1.5%, or nearly 60 000 persons, exists between the two population universes for the GMA. Historically, Montreal HTSs have always produced a reconstructed population without concern for this difference. To preserve comparability of the 2008 reference population with those of earlier HTSs, the demographic distribution of persons in collective households is thus considered with an identical demographic distribution to those of private households.

Weighting algorithm

The applied weighting method uses an iterative proportional fitting algorithm based on the classic Furness type growth factor method (equations 1 to 3). This iterative algorithm performs a calibration on margins by repeated application of growth factors to household weights. This is done in such a way as to simultaneously reproduce the distribution of households across household size and distribution of persons by age group. At each iteration, the algorithm alternates between adjusting a row (person-level) and adjusting a column (household-level) until a solution satisfies a convergence constraint (after a respectable number of iterations). Minimisation of squared error (MSE) was the convergence constraint chosen for our analysis, but this constraint can be expressed in a multitude of fashions.

A precise reconstruction of household distribution is possible upon the termination of the algorithm with a set of proportional adjustment values for each household. The adjustment procedure is applied for each of the 107 spatial units sampled during the HTS. These units correspond to the 107 municipality sectors, the preferred units for disseminating HTS results in the GMA community.

Furness equations

$$T'_{ij} = T_{ij} a_i b_j \qquad (1)$$

$$a_i = \frac{O_i}{\sum_i T_{ij}} \; \forall i \qquad (2)$$

$$b_j = \frac{D_j}{\sum_j T_{ij}} \; \forall j \qquad (3)$$

Table 3- Initial weighted population matrix and Furness procedure schematisation

Census private HH.	Households sample	Initial HH. Weigth (mean)	Persons /HH.	A 0-14	B 15-24	C 25-39	D 40-64	E 65+	Population Private households (census)
516 919	18 244	28.33	1						
540 374	24 162	22.36	2			Tij			Dj
260 908	10 228	25.51	3						
334 124	13 590	24.59	4+						
			Population Travel survey	631 338	460 644	691 430	1 486 919	578 825	3 849 155
			Total Population (Census)			Oi			3 939 819
			Difference	-41 462	-35 896	-135 634	82 660	39 669	-90 664
			Difference (%)	↓ -6.16%	↓ -7.23%	↓ -16.40%	↑ 5.89%	↑ 7.36%	↓ -2.30%

The application of this procedure relies upon a computational processing of the surveyed households. Table 4 is a tabular representation of the data being processed by the procedure for a specific sector. Each household is handled individually and categorised against a typology of households defined by the household's structure (number and age of residents). The typology selected distinguishes a total of 393 categories of household across the GMA for the 2008 HTS, based on a division of all interviewed households according to size (1 person, 2 persons, 3 persons and 4+ persons) and age group (A: 0-14 y.o., B: 15-24 y.o., C: 25-39 y.o., D: 40-64 y.o., E: 65+ y.o.). Such a division ensures a sufficient number of observations per cell for the 107 weighted sector and prevents empty cells. To illustrate the labelling of a given household, "1A1C" denotes a household of two persons and composed of one person from group A (i.e. between 0 and 14 y.o.) and another person from group C (between 25 and 39 y.o.). The number of distinct categories of household represented per sector averages between 20 (sector 141, Senneville) and 117 (sector 118, Montreal North). Ultimately all of the households and the persons from the households associated with a given household category and sector receive the same weighting factor. Application of the IPF algorithm for the 2008 Montreal HTS results in approximately 7000 different weighting factors. Finally, a proportional person's weight adjustment is done to recreate the census total population universe.

Table 4 – Example of individual household treatment for factorisation by IPF (Example of sector 121-Outremont)

HH_ID	HH_SECTOR	HH CATEGORIE	HH_SIZE	HH_1p	HH_2p	HH_3p	HH_4pmore	age group A	age group B	age group C	age group D	age group E	INITAL weight	FEXP iteration k (k=1)	FEXP iteration k+1
10317	121	1E	1	1	0	0	0	0	0	0	0	1	28.45	28.45	26.73
10557	121	1D	1	1	0	0	0	0	0	0	1	0	28.45	28.45	26.91
10903	121	3A2D	5	0	0	0	1	3	0	0	2	0	22.08	22.08	15.73
11137	121	5A2C	7	0	0	0	1	5	0	2	0	0	22.08	22.08	24.43
12999	121	1C	1	1	0	0	0	0	0	1	0	0	28.45	28.45	41.79
13589	121	1C2E	3	0	0	1	0	0	0	1	0	2	23.39	23.39	24.11
							…								
14561	121	2E	2	0	1	0	0	0	0	0	0	2	21.19	21.19	18.69
15891	121	1E	1	1	0	0	0	0	0	0	0	1	28.45	28.45	26.73
16234	121	1D	1	1	0	0	0	0	0	0	1	0	28.45	28.45	26.91
16585	121	2A2D	4	0	0	0	1	2	0	0	2	0	22.08	22.08	15.73
16849	121	3A1B1C	5	0	0	0	1	3	1	1	0	0	22.08	22.08	33.24
18456	121	2D	2	0	1	0	0	0	0	0	2	0	21.19	21.19	18.82
Total weigthed HHs (HTS)				3 528	3 009	1 333	1 855	4 504	3 049	4 306	7 688	3 408	*Total weigthed persons (HTS)*		
Total private HHs (census)				3 528	3 009	1 333	1 855	4 490	3 043	4 297	7 668	3 399	*Total population (census)*		
			a_i	1.000	1.000	1.000	1.000	1.003	1.002	1.002	1.003	1.003	**b_i**		
								0.32%	0.20%	0.22%	0.25%	0.25%	*Population difference (%)*		
											789		***MSE***		

Initial weights

To initialize the algorithm a first solution is required, which is obtained by reconstructing the population with a weighting factor that is the ratio of the number of private households by size group from census population to the total number of surveyed households. These weights reproduce the distribution of private households as a function of their size, and a population divided into five pre-determined age groups. This constitutes the initialization step.

Weight trimming

The unconstrained application of the iterative proportional fitting (IPF) method outlined above may produce weights for some sectors exceeding an acceptable value in relation to sampling rates. For instance, some weights reach in excess of 300 while a sample size of around 4% would suggest an approximate weight to be around 25. To restrain weight variability a constraint is imposed upon them. Two scenarios are explored, where the weights are kept within a plus/minus value of the initial weight:

- FEXP25: IPF method with weights limited within initial weight ±25%;
- FEXP50: IPF method with weights limited within initial weight ±50%.

Evaluation of weighting scenario

In this section, different examples of weighting statistics to evaluate scenarios are summarising. First, for each scenario, the population reconstruction matrix of the GMA is presented (Table 5 and Table 7), as well as the difference between the solution offered by the particular scenario and the initial weights solution (Table 6 and Table 8).

Table 5 – Population matrix result of scenario FEXP_25 (trimming at ±25% of initial weight)

Census private HH.	Households sample	Initial HH. weigth	Persons /HH.	A 0-14	B 15-24	C 25-39	D 40-64	E 65+	Population Private households
516 919	18 244	28.33	**1**	0	12 100	85 171	214 396	205 250	516 917
540 374	24 162	22.36	**2**	20 822	64 939	222 282	497 464	275 268	1 080 775
260 908	10 228	25.51	**3**	132 481	127 062	194 979	291 604	36 637	782 762
334 124	13 590	24.59	**4+**	500 084	262 856	285 762	402 741	21 961	1 473 404
			Population Travel survey	653 387	466 957	788 194	1 406 204	539 116	**3 853 859**
			Total Population (Census)	672 800	496 540	827 064	1 404 259	539 156	**3 939 819**
			Difference	-19 413	-29 583	-38 870	1 945	-40	**-85 960**
			Difference (%)	-2.89%	-5.96%	-4.70%	0.14%	-0.01%	**-2.18%**

Table 6 – Population cells adjustment (scenario FEXP_25)

Persons /HH.	A 0-14	B 15-24	C 25-39	D 40-64	E 65+	Population Private households
1		23.4%	23.2%	-5.2%	-3.1%	0.0%
2	17.5%	20.1%	20.6%	-5.3%	-8.1%	0.0%
3	4.8%	0.8%	9.9%	-5.8%	-15.6%	0.0%
4+	2.6%	-2.9%	9.6%	-5.5%	-9.3%	0.3%
Population Travel survey	3.5%	1.4%	14.0%	-5.4%	-6.9%	0.1%

Table 7 – Population matrix result of scenario FEXP_50 (trimming at ±50% of initial weight)

Census private HH.	Households sample	Initial HH. weigth	Persons /HH.	A 0-14	B 15-24	C 25-39	D 40-64	E 65+	Population Private households
516 919	18 244	28.33	**1**	0	13 434	91 995	208 494	203 007	516 931
540 374	24 162	22.36	**2**	22 597	71 932	230 832	486 875	268 549	1 080 784
260 908	10 228	25.51	**3**	136 331	126 918	200 302	284 324	34 910	782 784
334 124	13 590	24.59	**4+**	499 717	268 016	284 490	404 376	23 197	1 479 796
			Population Travel survey	658 644	480 299	807 619	1 384 070	529 664	**3 860 296**
			Total Population (Census)	672 800	496 540	827 064	1 404 259	539 156	**3 939 819**
			Difference	-14 156	-16 241	-19 445	-20 189	-9 492	**-79 523**
			Difference (%)	-2.10%	-3.27%	-2.35%	-1.44%	-1.76%	**-2.02%**

Table 8 – Population cells adjustment (scenario FEXP_50)

Persons /HH.	A 0-14	B 15-24	C 25-39	D 40-64	E 65+	Population Private households
1		37.0%	33.0%	-7.8%	-4.1%	0.0%
2	27.5%	33.0%	25.3%	-7.3%	-10.3%	0.0%
3	7.9%	0.7%	12.9%	-8.1%	-19.5%	0.0%
4+	2.6%	-1.0%	9.1%	-5.1%	-4.2%	0.4%
Population Travel survey	4.3%	4.3%	16.8%	-6.9%	-8.5%	0.3%

Figure 9 compares distributions of household's weight estimated in scenario FEXP50 and initial IPF weights. Weighted person shares are illustrated by household size where age group are marked (Figure 10 and Figure 11). Theses figures help to distinguish household's category and person's group adjustments and the impact of lessening weight constraints on weight variability.

Figure 12 shows the resultant population following adjustment via the weighting method (scenario FEXP50). This map can be juxtaposed with Figure 7, which shows the same population derived from the sample but before application of IPF correction. Note the more uniform adherence to the reference population by the reconstructed population. A difficulty in reproducing a handful of groups still remains, such as young adults (15-24 y.o.) in the Downtown district and certain groups from sectors with sampling errors or a low population. More detailed demographic (5 y.o. cohorts) profile reconstructed by IPF procedure is shown at Figure 13.

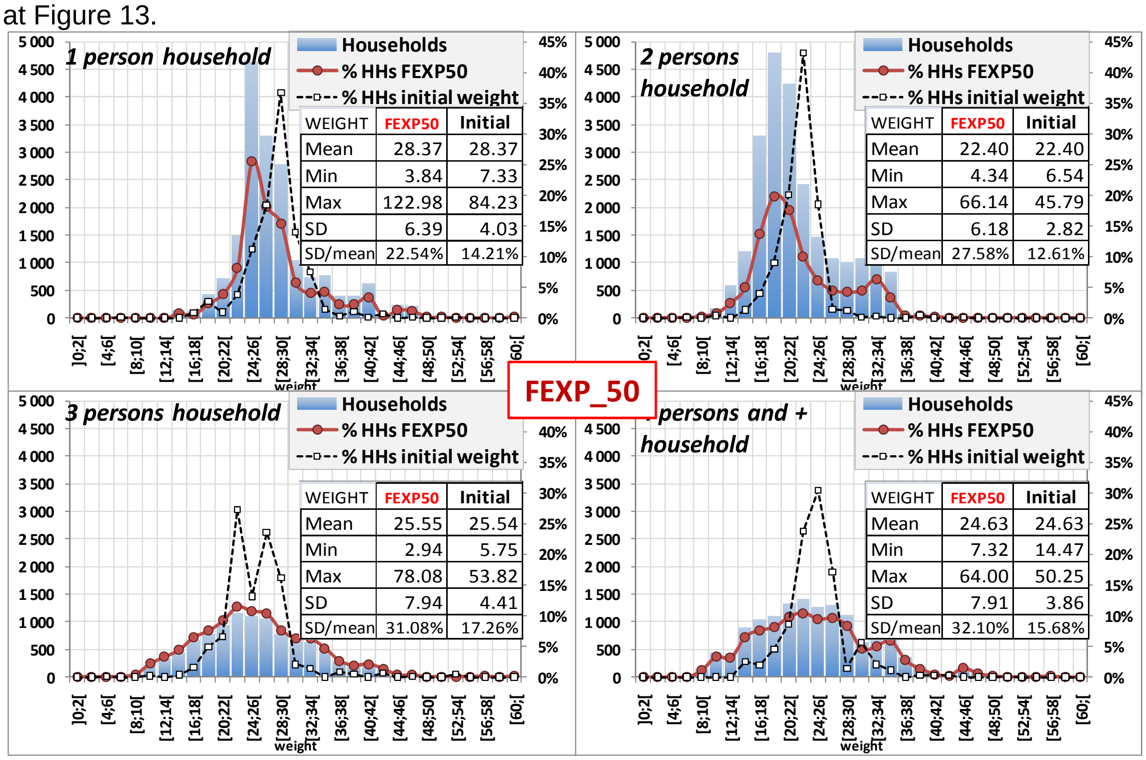


Figure 9 - Distribution of household's weight by HH size (scenario FEXP50) and variability statistics

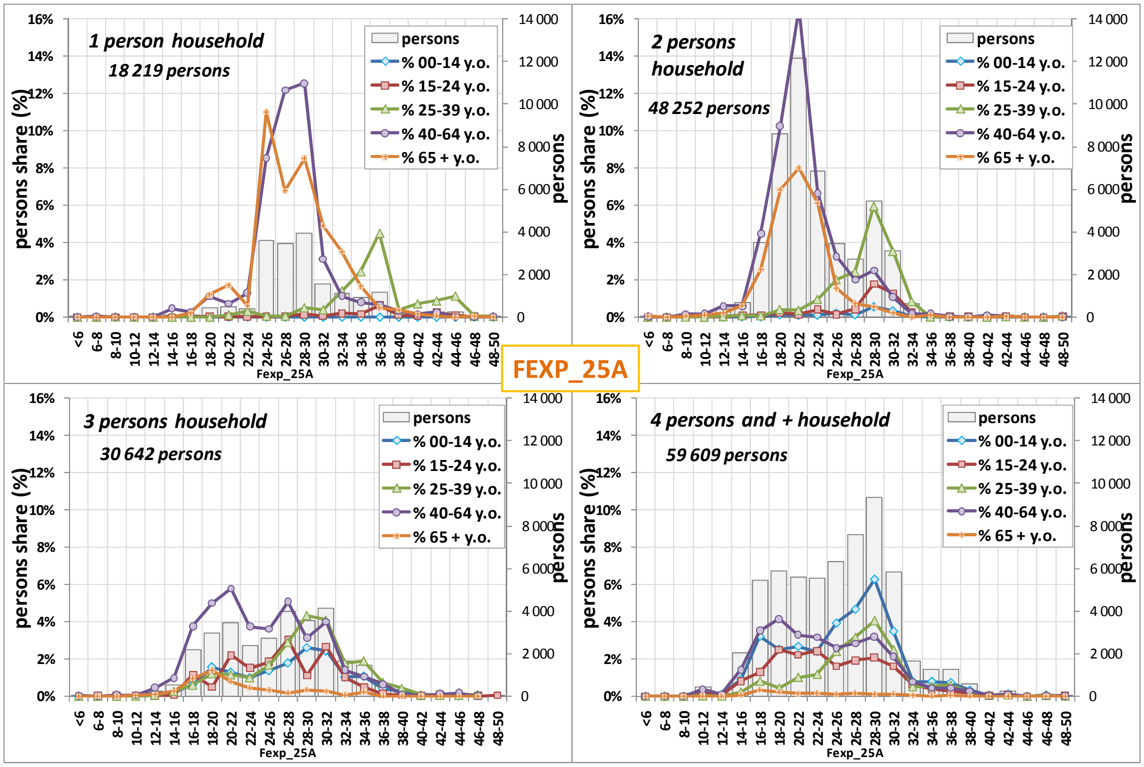


Figure 10- Distribution of person's weight categorized by household size and group age (scenario FEXP25_A)

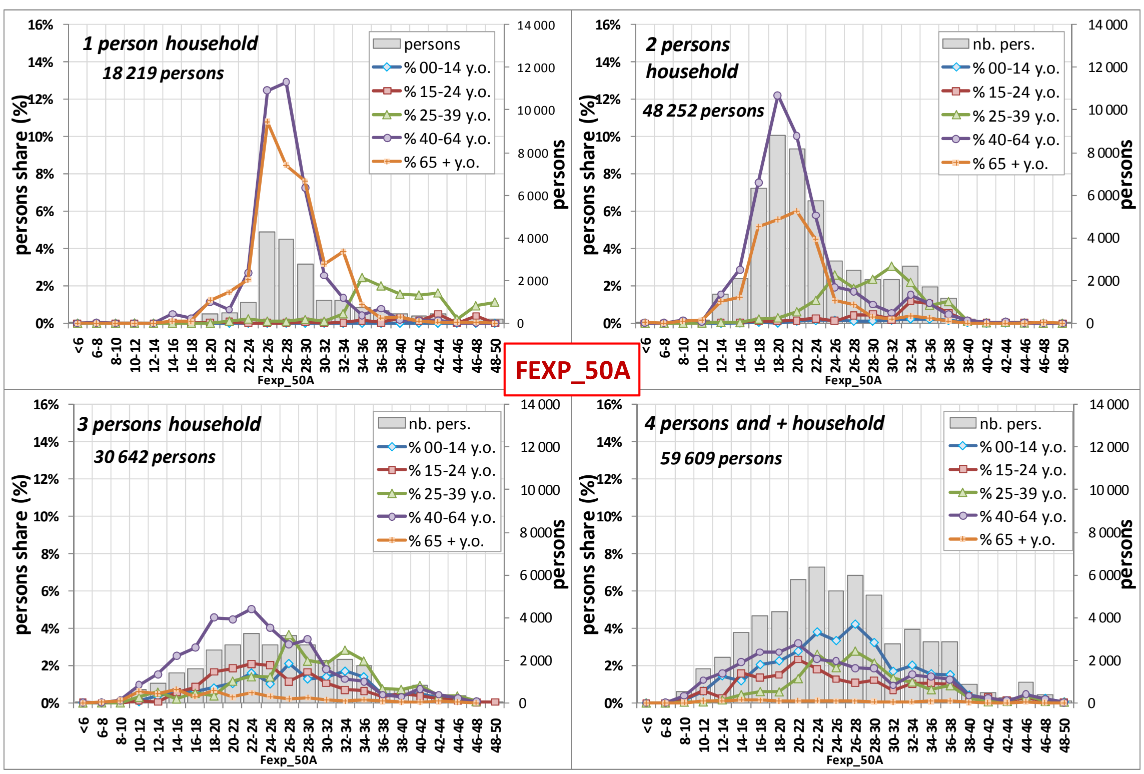


Figure 11- Distribution of person's weight categorized by household size and group age (scenario FEXP50_A)

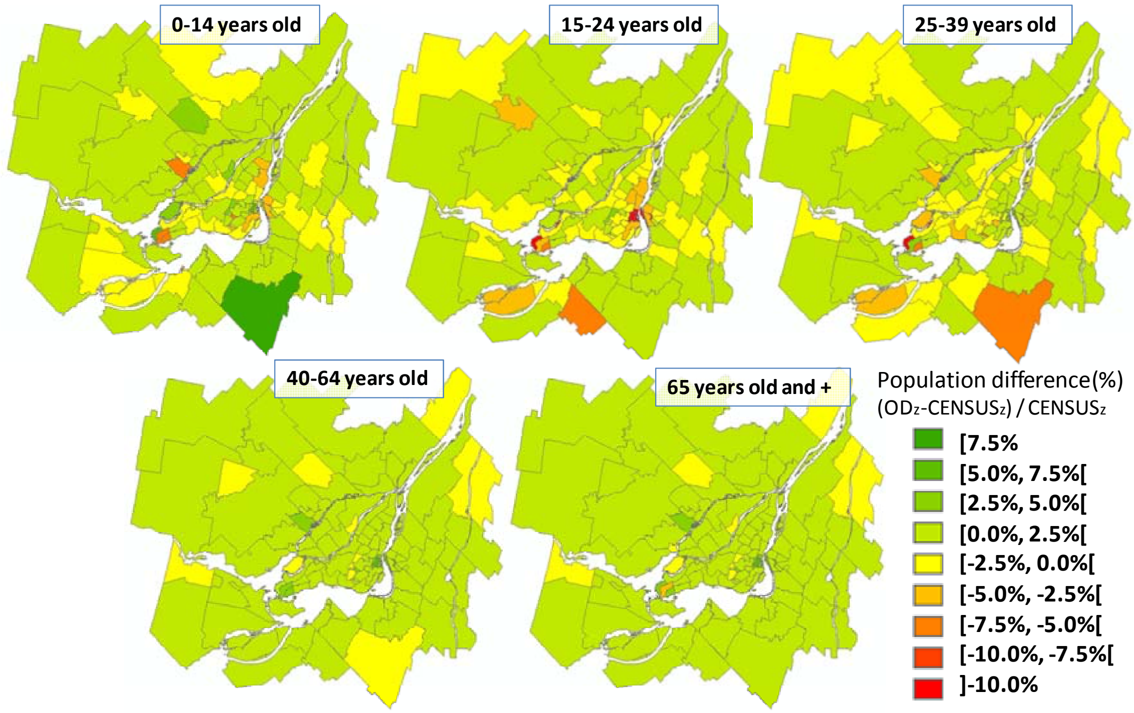
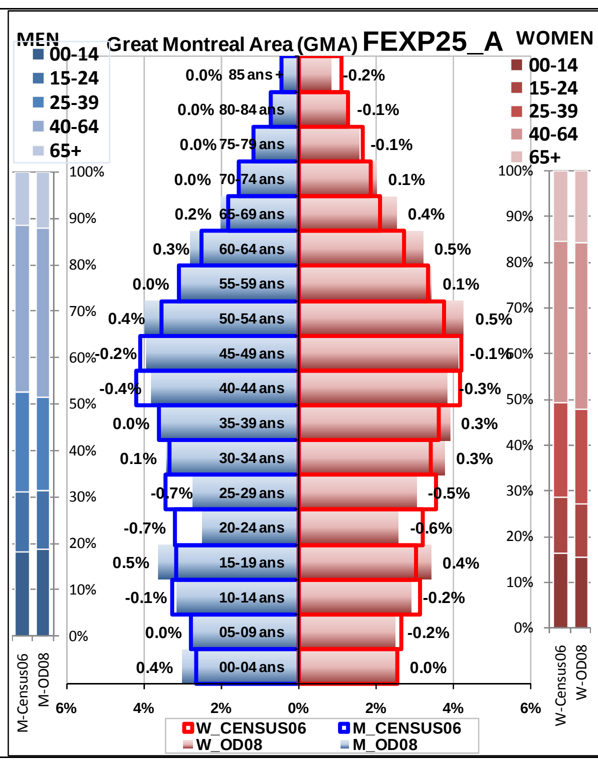


Figure 12 – Spatial distribution (by age group) of population difference (%) between census and OD travel survey population weighted by scenario FEXP_50A



Greater Montreal Area

FEXP_ 25A

age groups	Census persons	OD persons	Diff.	Diff. %
00-14 y.o.	672 800	666 950	-5 850	*-0.9%*
15-24 y.o.	496 540	477 231	-19 309	*-3.9%*
25-39 y.o.	827 064	806 294	-20 770	*-2.5%*
40-64 y.o.	1 404 259	1 437 246	32 987	*2.3%*
65 y.o. and +	539 156	552 102	12 946	*2.4%*
Total pop.	**3 939 819**	**3 939 823**	**4**	***0.0%***

FEXP_50A

age groups	Census persons	OD persons	Diff.	Diff. %
00-14 y.o.	672 800	671 190	-1 610	*-0.2%*
15-24 y.o.	496 540	490 114	-6 426	*-1.3%*
25-39 y.o.	827 064	824 950	-2 114	*-0.3%*
40-64 y.o.	1 404 259	1 412 167	7 908	*0.6%*
65 y.o. and +	539 156	541 460	2 304	*0.4%*
Total pop.	**3 939 819**	**3 939 881**	**62**	***0.0%***

Figure 13 – Demographic profile by five years cohort of weighted persons and comparison with census population (OD survey weighted persons and 2006 Canadian census)

At the level of aggregation of geopolitical significant regions, weighting impacts on general indicators are documented below in Table 9 at household, person and trip-level. Differences between indicators of the FEXP50 scenario and the initial survey sample are evaluated. In short, it can be seen that the weighting procedure reveals change to mobility indicators aside from bringing about a population that is slightly younger (mean age decreases of 1.24) and with fractionally altered modal patterns (transit share increases of 1.19). Greater indicators changes are perceived in central regions (CBD and Central MUC).

Table 9 – General indicator differences between FEXP_50 weighting strategy and survey sample

Highlights - OD travel survey 2008 - (Difference between weighting scenario FEXP50 - survey sample)	1- CBD	2- CENTRAL MUC	3-EAST MUC	4- WEST MUC	5-Near South shore	6-Laval	7-North shore	8-South shore	GMA OD08 (survey sample)	GMA OD08 (FEXP50)	GMA OD08 (FEXP50 - survey sample)
HOUSEHOLDS											
Cars/Household	-0.061	-0.034	-0.001	-0.067	-0.038	-0.035	-0.053	-0.062	1.373	1.310	-0.063
Persons/Household	-0.076	-0.090	0.036	-0.053	-0.027	-0.052	-0.045	-0.062	2.399	2.333	-0.066
% 1 person HH	6.61%	5.44%	0.81%	4.55%	4.18%	3.82%	4.05%	4.11%	26.71%	31.45%	4.74%
% 4 persons and + HH	-0.24%	-1.13%	1.07%	-0.24%	0.20%	-1.22%	-0.63%	-1.47%	21.13%	20.19%	-0.94%
% no car household	3.50%	2.99%	0.36%	1.74%	1.07%	0.85%	0.74%	1.04%	18.55%	20.86%	2.31%
% household: < 20 000$ / year	2.04%	1.53%	0.08%	1.11%	0.58%	0.50%	0.42%	1.31%	15.25%	16.48%	1.23%
% household: 20 000$ -60 000$/ year	1.33%	1.07%	0.20%	0.85%	0.91%	0.59%	-0.67%	-0.25%	45.80%	46.27%	0.47%
% household.: 60 000$ -100 000$/ y.	-1.13%	-1.29%	-0.31%	-0.59%	-0.72%	-0.51%	0.06%	-0.54%	24.77%	23.87%	-0.90%
% household: > 100 000$ / year	-2.23%	-1.31%	0.03%	-1.38%	-0.77%	-0.58%	0.19%	-0.52%	14.19%	13.38%	-0.81%
PERSONS											
Car/person	-0.021	0.002	-0.008	-0.015	-0.009	-0.001	-0.009	-0.008	0.572	0.561	-0.011
Ratio Men/Women	0.011	-0.019	-0.009	-0.014	-0.011	-0.005	-0.013	-0.010	0.943	0.929	-0.014
mean age	-2.89	-1.08	-2.22	-1.47	-1.70	-1.24	-1.31	-0.67	40.05	38.81	-1.24
% 00-15 years old	-0.67%	0.10%	1.93%	1.44%	1.67%	1.24%	1.38%	0.55%	17.56%	18.40%	0.84%
% 65 years old and +	-3.39%	-0.71%	-2.87%	-1.03%	-1.39%	-1.30%	-1.04%	-0.25%	14.72%	13.73%	-0.99%
% workers	1.64%	0.72%	0.88%	0.56%	0.30%	0.88%	0.88%	0.57%	35.93%	36.63%	0.70%
% students	2.21%	0.86%	1.35%	0.51%	1.62%	0.31%	1.00%	0.52%	18.68%	19.51%	0.82%
% others	-2.21%	-0.29%	-1.17%	-0.71%	-0.87%	-0.59%	-0.61%	-0.23%	22.92%	22.38%	-0.54%
% no trip	-1.64%	-1.29%	-1.06%	-0.36%	-1.05%	-0.59%	-1.28%	-0.86%	22.47%	21.48%	-0.99%
% car owners (16 y.o. and +)	-2.39%	-0.87%	0.32%	0.05%	-0.01%	0.54%	0.36%	-0.16%	63.70%	62.89%	-0.81%
% car licence (16 y.o. and +)	-0.44%	-0.72%	0.31%	-0.19%	-0.14%	0.18%	-0.04%	-0.45%	80.50%	79.85%	-0.65%
% Transit users	2.04%	1.67%	0.97%	0.44%	0.67%	0.15%	0.27%	0.34%	16.28%	17.46%	1.18%
TRIPS											
trips/person	0.046	0.039	0.034	0.013	0.043	0.024	0.046	0.031	2.036	2.072	0.036
motorised trips/person	-0.018	0.025	0.029	0.010	0.032	0.023	0.035	0.019	1.771	1.791	0.020
car trips/person	-0.058	-0.012	0.008	-0.004	0.009	0.016	0.024	0.014	1.320	1.314	-0.006
Transit trips/person	0.041	0.037	0.020	0.010	0.012	0.001	0.004	0.006	0.296	0.320	0.025
Work trips/person	0.014	0.008	0.008	0.005	0.004	0.009	0.010	0.006	0.390	0.397	0.007
Study trips/person	0.027	0.010	0.015	0.006	0.018	0.003	0.012	0.006	0.200	0.209	0.009
leisure trips/person	0.012	0.005	-0.001	-0.002	0.000	-0.002	0.001	0.000	0.139	0.140	0.002
shopping trips/person	-0.007	0.000	-0.008	-0.004	-0.006	-0.002	0.000	0.000	0.159	0.158	-0.001
others trips/person	-0.024	-0.001	0.005	0.004	0.009	0.009	0.005	0.005	0.247	0.250	0.004
CDB trips/person	0.050	0.009	0.001	-0.001	0.002	0.002	0.003	0.003	0.149	0.159	0.011
% Transit trips/ motorised trips	3.89%	1.69%	0.75%	0.45%	0.35%	-0.09%	0.16%	0.26%	16.70%	17.89%	1.19%
% Home based trips	0.21%	-0.03%	-0.07%	-0.11%	-0.09%	-0.23%	-0.22%	-0.17%	46.83%	46.69%	-0.14%
% back home trips	0.10%	0.00%	-0.01%	-0.10%	-0.07%	-0.19%	-0.17%	-0.14%	46.32%	46.22%	-0.10%
% intermediate trips	-0.31%	0.03%	0.09%	0.21%	0.15%	0.42%	0.39%	0.31%	6.85%	7.09%	0.24%
% AM peak trips (6:00 to 8:59)	0.06%	-0.05%	0.73%	0.39%	0.62%	0.39%	0.38%	0.17%	24.81%	25.04%	0.23%
% PM peaks trips (15:30 to 18:29)	0.11%	-0.03%	0.32%	0.20%	0.22%	0.35%	0.33%	0.16%	28.58%	28.75%	0.18%
trip distance (euclidean)	-0.02	-0.03	-0.03	-0.11	-0.09	-0.03	-0.36	-0.33	8.018	7.747	-0.272

Weight adjustments of persons to preserve comparability

As stated earlier, the proposed weighting method enables an exact reconstruction of the household universe with proper distribution of household size and demographic structure. However, when regrouped into 5-year cohorts as per Figure 13, the weighting method fails to sufficiently redress the marked underestimation of certain groups of the population, in particular young adults between 20 and 39. By relaxing the main constraint of the IPF algorithm, a more faithful reconstruction of the aforementioned population groups may be performed, but at the cost of bringing about an explosion in value of the weights. To ensure a level of comparability across the series of surveys, an adjustment of the weights on people is applied. This adjustment enables the reconstruction of the reference population, in age cohorts of 5 or 10 years, of the 107 municipal sectors. It should be noted that this procedure generates a distinct weight for each member of the household, instead of the IPF method that generates a unique weight for the household and its occupants.

The adjusted weighting factor of a person (FACPER) belonging to household **M** is determined by the product of the weighting factor of her household (FEXP) and the adjustment factor **Faj** acting for residence **S**, age cohort **C** and gender **G** of the person.

$$FACPER_{M,C,G} = FEXP_M * Faj_{S,C,G}$$

The adjustment factor corresponds to the ratio between the reference population (census 2006) and the population reconstructed by the household weighting factor for the sector, cohort and gender concerned.

$$Faj_{S,C,G} = \frac{Census\ Total\ population\ _{S,C,G}}{\sum_{S,C,G} FEXP_M}$$

The age groups used in the establishment of the adjustment factors are 5-year or 10-year cohorts for persons less than 70, and a single group for persons aged 70 and older. To assist age stratification choice, Figure 14 shows the census distribution of the residents of all 107 sectors (grouped by analysis region) with separation into gender and 5-year cohorts. Based on person enumeration per cohort and the assumption of a perfect sample expected, it is possible to determine the probability that a certain population groups have been inadequately sampled. For a five percent sample size target and an expected number of at least thirty observations, a theoretical minimum of six-hundred persons as population is thus required by gender-cohort to ensure a good level of representativeness. In Figure 14, these groups not satisfying this requirement are marked in red. For those municipal sectors where the number of observations taken was insufficient (31 out of 107), the age cohort size has been enlarged to 10 years.

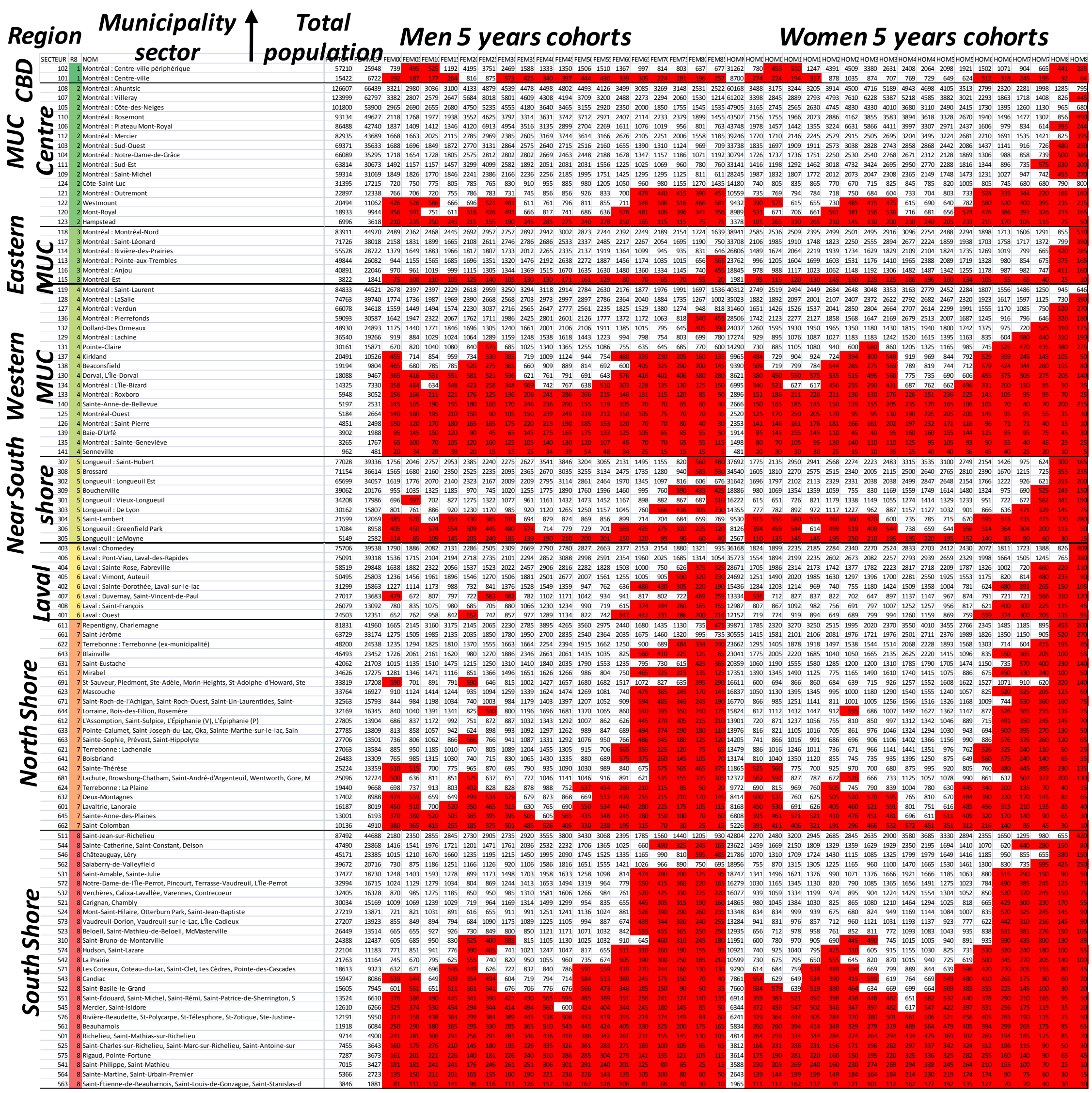


Figure 14- Visualisation of person 5 years cohorts with insufficient population for each municipality sector regrouped by analytical region

CONCLUSIONS

Household travel survey is becoming more and more perilous when population behavior representativeness has to be confirmed. In particular, household samples obtained by a randomly selected residential phone list are subjected to under-representation of several categories of peoples. Singles, seniors, young adults, affluent people are not easily reached or ready to participate to a travel survey very often oriented to promote public transport.

Notwithstanding these difficulties, the paper has presented the application of an iterative proportional fitting method to ensure proper similar weights to persons and trips belonging to the same household. It ensures coherence and consistency in the enumeration of people and trips. Based on well-known travel behavior categories of households (composition of persons of different age categories), several types of visualization have demonstrated where and in which spatio-demographic areas the travel behavior estimates are statistically sounded or not. Multiple other reasons independent of the survey original sample have also to be examined: seasonality, adequacy of reference data, etc.

In context, this project constitutes the first step to a more serious investigation of special biases due to factors such as: language, ethnicity, gender, income, mode insecurity, student population, cell telephone, weather, etc....

ACKNOWLEDGMENTS

The authors wish to acknowledge the transport authorities by whom the large-scale surveys, mainly Household Origin-Destination surveys, are conducted. Those surveys authorize the continuation of a travel behaviour observational and analytical culture in the GMA: STM, RTL, STL, AMT and MTQ.

REFERENCES

Arentze T., H.J.P. Timmermans, and F. Hofman (2007). Creating Synthetic Household Populations: Problem and Approach. Transportation Research Record: Journal of the Transportation Research Board, 2014, 2007, pp. 85-91.

Beckman, R.J., K.A. Baggerly and M.D. McKay (1996). Creating Synthetic Baseline Populations. Transportation Research A, Vol. 30, Number 6, pp. 415-429.

Chapleau, R., Allard B., Trépanier, M., Morency, C. (2001). Origin-Destination Travel Survey Software: Cognitive and Technological Assistance, World Conference on Transportation Research, July 2001, Séoul, Korea, CD-ROM, 20 pages.

Chapleau, R. (2003). Measuring the internal quality of a CATI travel household survey. In: Stopher, Peter, Jones, Peter (2003). Transport Survey Quality and Innovation, Kruger, South Africa, Pergamon, pages 69-87.

Chapleau, R., Morency, C. (2004). Modelling urban Mobility with Geodemographics, 10th World Conference on Transportation Research, Istanbul, Turquie.

Deming, W.E. and Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Annals of Mathematical Statistics,11, 427-444.

Guo, J.Y. and Bhat, C.R. (2007). Population Synthesis for Microsimulating Travel Behavior, Transportation Research Record 2014 (2007), pp. 92–101.

Keeter, S., C. Kennedy (2006). The Cell Phone Challenge to Survey Research. The Pew Research Center.

Norman, P. (1999) Putting Iterative Proportional Fitting on the Researcher's Desk. Working Paper. School of Geography, University of Leeds.

Ortúzar, J. and Willumsen, L. G. (2006). Modelling Transport, John Wiley and Sons,Inc., West Sussex, England.

Sammer, G. (2006). Processing, analysis, and archiving of travel survey data. In: Travel surveys methods: quality and future directions (Stopher, P. R. and Stecher, C., ed.), 1st edition 2006, pp239-269. Elsevier, Oxford.

Stopher, P. R., Wilmot, C.G., Stecher, C., Alsnih, R. (2006). Household travel surveys: proposed standards and guidelines. In: Travel surveys methods: quality and future directions (Stopher, P. R. and Stecher, C., ed.), 1st edition 2006, pp19-74. Elsevier, Oxford.

Stopher, P. R., Greaves, S. P. (2007). Household travel surveys: Where are we going? Transportation Research Part A: Policy and Practice, 41(5), 367-381.

Statistics Canada (2009). Residential Telephone Service Survey 2008. In: the Daily (Statistics Canada),http://www.statcan.gc.ca/daily-quotidien/090615/dq090615c-eng.htm

Statistics Canada (2006). Residential Telephone Service Survey 2005. In: the Daily (Statistics Canada), http://www.statcan.gc.ca/daily-quotidien/060405/dq060405b-eng.htm

Trépanier, M., Chapleau, R., Morency, C. (2008). Tools and Methods for a Transportation Household Survey. Urban and Regional Information Systems Association Journal, 20(1), p. 35-43.

Tuckel, P., O'Neill, H. (2001) The vanishing respondent in telephone surveys. Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001.

Ye, X., Konduri, K.C., Pendayala, R.M., Sana, B., Waddell, P. (2009). Methodology to match distributions of both household and person attributes in generation of synthetic populations. In: TRB 88th Annual Meeting Compendium of Papers.