REVISITING THE DESTINATION RANKING PROCEDURE IN DEVELOPMENT OF AN INTERVENING OPPORTUNITIES MODEL FOR PUBLIC TRANSIT TRIP DISTRIBUTION

Mohsen Nazem, Ph.D. Candidate (mohsen.nazem@polymtl.ca) [1,2,3]

Martin Trépanier, Ph.D., P.Eng., Full Professor (mtrepanier@polymtl.ca) [1,2,3,*]

Catherine Morency, Ph.D., P.Eng., Associate Professor (cmorency@polymtl.ca) [1,2,3]

1. Polytechnique Montréal (www.polymtl.ca)

C.P. 6079, Station Centre-Ville, Montreal (QC) H3C 3A7, Canada.

2. CIRRELT (www.cirrelt.ca)

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

3. MOBILITÉ Chair (www.polymtl.ca/mobilite)

Research Chair on Evaluation and Implementation of Sustainability in Transportation

* *Corresponding author. Phone: +1 (514) 340-4711 ext.4911 Fax: +1 (514) 340-5763*

ABSTRACT

In this study an Enhanced Intervening Opportunities Model (EIOM) is developed for Public Transit (PT) trips. This model is a distribution supply-dependent model, with single constraints only on trip production for work trips done during morning peak hours (6:00AM to 9:00AM) within the Island of Montreal, Canada. Different datasets including the 2008 Origin-Destination (OD) survey of the Greater Montreal Area (GMA), 2006 Census of Canada, GTFS network data, along with the geographical data of the study area are used. The EIOM is a nonlinear model composed of sociodemographic, PT supply and work location attributes. An enhanced destination ranking procedure is used for calculating the number of spatially cumulative opportunities, which is the basic variable of the EIOM. For comparison, a Basic Intervening Opportunities Model (BIOM) is developed by using the basic destination ranking procedure. In fact, the main difference between the EIOM and the BIOM is in the destination ranking procedure; the EIOM considers the maximization of a utility function composed of PT Level Of Service (LOS) and number of opportunities at the destination, along with the OD trip

duration, whereas the BIOM is based on a destination ranking derived only from OD trip durations. The performance of the EIOM is analyzed by means of several global and Goodness-of-Fit measures. Results confirm that the EIOM is well-behaved and more accurate than the BIOM. Based on the explanatory variables that are used in the EIOM, and also the enhanced destination ranking procedure, this study presents a new tool for PT analysts, planners and policy-makers to study the potential changes in PT trip patterns, due to changes in sociodemographic characteristics, PT supply, etc. Also this study opens new opportunities for the development of more accurate PT demand models with new emergent data such as smart card entries.
Keywords: Destination ranking, Intervening opportunities model, Public transit planning, Supply-dependent model, Trip distribution.

INTRODUCTION

Trip distribution is the second step in the classical sequential four step models for aggregate transport planning. Several families of trip distribution models such as Gravity Model (GM) and Intervening Opportunities Model (IOM) are presented in the literature (Bonnel, 2004; Hensher & Button, 2008; Ortuzar & Willumsen, 1994; Wilson, 1970). In GM, the number of trips between each Origin-Destination (OD) pair is based directly on the OD distance (or travel time) and trip production/attraction values, whereas the IOM considers the number of intervening opportunities as the main influencing factor. As it seems that both these factors influence the trip distribution, unified hybrid models are also developed and calibrated (Almeida & Goncalves, 2001; Goncalves & Ulysséa Neto, 1993; Wills, 1986).
Based on past studies, each family of distribution models has its own advantages and disadvantages compared to other models. At the same time, the calibration of each model requires different data and calibration procedures. The IOM has been used less than the GM due to complexities in terms of the calibration procedure (Goncalves & De Cursi, 2001). However compared to the GM, the IOM is behavioural based (Kermanshah, 2004), less sensitive to the size and shape of the study area (Chow, Zhao, Li, & Li, 2005) and also produces better results in cases where destinations which satisfy the trip purpose are not uniformly distributed (Veenstra, Thomas, & Tutert, 2010). In our previous studies, the IOM was calibrated with several different approaches for the Island of Montreal, Canada, and the results showed that IOM has much better performance than GM, in terms of trip production and trip attraction and other Goodness-of-Fit measures (Nazem, Trépanier, & Morency, 2012, 2013).
For the calibration of most distribution models, we need an OD reference matrix. Several approaches are presented in the literature, such as asking passengers to fill out questionnaires on board of transit vehicles, estimating the number of passengers from counting the boarding and alighting passengers at stops, or from smart card validation data (Feng & Li, 2004). In this study, we use an OD matrix obtained from 2008 OD survey held in the Greater Montreal Area (GMA); this survey reaches some 5% of the residing households.
This study area that is presented in Figure 1 consists of 41 Municipal Sector (MS), with a population of nearly 1.9 millions, distributed over almost 500 km^2 (StatCan, 2011b). The present study is limited to the Island of Montreal and PT work trips in order to deal with less

calculation complexities, remembering that the methodology can be generalized to the whole GMA.

As for the IOM (Equation 5) the number of intervening opportunities represents the most influencing factor, the procedure of its calculation becomes essential. For the calibration of an IOM, from each origin zone we need to rank all destination zones based on their relative attractiveness. Until now, most studies consider trip time or distance for representing the attractiveness of each destination zone (Nazem, et al., 2012, 2013). But it seems that travel time or distance is not the only influencing factor for public transit trips; other factors like Public Transit (PT) Level Of Service (LOS) and number of potential opportunities at the destination zone can also be relevant. Thus, we propose a new ranking procedure based on OD trip duration, PT LOS and number of potential opportunities at the destination zone.

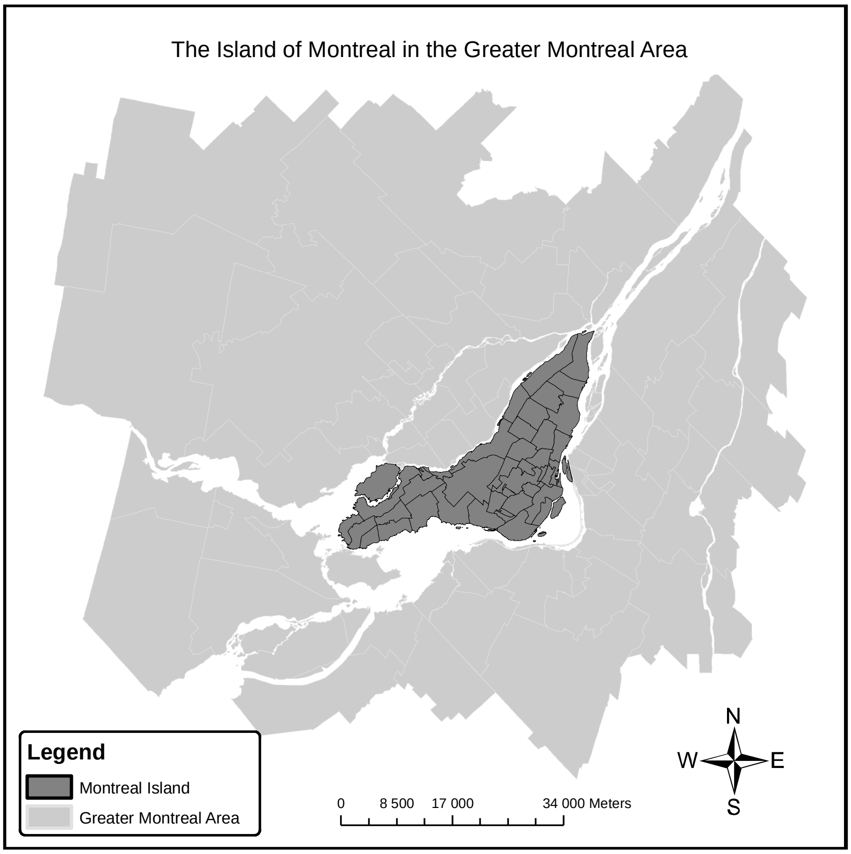

Figure 1 - Spatial distribution of the Municipal Sectors (MS) in the Island of Montreal, Canada

In this study, we aim to calibrate an Enhanced Intervening Opportunities Model (EIOM) with a new destination ranking procedure, called “enhanced destination ranking procedure”, based on the economic consumer theory. The EIOM will be calibrated and validated for work PT trips during average weekday morning peak hours (6:00AM - 9:00AM) within the Island of Montreal. This model considers sociodemographic attributes as well as PT supply characteristics of Municipal Sectors (MS) located in the study area. The Transportation Analysis Zones (TAZ) in this study are the MS presented in Figure 1. The advantage of this model is that it allows policy-makers to study the effects of PT supply modifications on the trip pattern. Also as the required entry data for this model could be available for the future, it could be used for forecasting purposes.

In contrast with the Basic IOM (BIOM) which considers only trip duration for ranking destination zones (Nazem, et al., 2012), in this study we develop a utility function based on OD trip duration, PT LOS and number of opportunities at the destination (Equation 7). We aim to calculate the coefficients of the utility function by means of a conditional logit model for the alternative specific variables, based on the observations of the 2008 GMA OD survey. Afterwards based on this utility function, we rank all destination zones from each origin zone and calculate the number of intervening opportunities for each OD pair. Then the EIOM (Equation 5) is calibrated based on the new number of intervening opportunities and its performance is studied.

The structure of the paper is as follows. First we present a literature review to learn more about advances in trip distribution models; then the datasets and their preparation for modelling development are described. In the next section, following a descriptive analysis of the data, the development of the BIOM and the EIOM is presented. Afterwards, the performance of the EIOM is studied and compared to the BIOM, by means of several global and Goodness-of-Fit measures. The next section presents spatial limitations of the models by means of a spatial residual errors analysis. In the conclusion, we discuss some interesting potentials of the EIOM for analysts and policy-makers, and also present some of our ongoing and future research topics.

LITERATURE REVIEW

Several families of models are used for estimating trip distribution in its general aggregate form. Among them, the Gravity Model (GM) and the Intervening Opportunities Model (IOM) are the most common (Bonnel, 2004; Ortuzar & Willumsen, 1994).

The GM for trip distribution inspired by Newton's law uses an impedance function, which is generally represented by a generalized cost. This model and its applications are well presented in the literature (Bonnel, 2004; De Grange, Troncoso, Ibeas, & González, 2009; Hensher & Button, 2008; Ortuzar & Willumsen, 1994; Rajesson, 2009; Thamizh Arasan, Wermuth, & Srinivas, 1996; Wilson, 1970).

The main idea of the opportunity model came from some theoretical concepts that relate the mobility, the migration distances and the spatial locations of services; the theory of this model in its present form was developed later (Schneider, 1959; Stouffer, 1940, 1960). The fundamental idea of this model is that distance is not the main factor that affects destination choices. Contrariwise, this model considers the main influencing factor as the relative accessibility of the opportunities that can satisfy each trip purpose. This model assumes that an individual chooses the closest destination location that gives him the opportunity to meet his needs. Distance or more widely, generalized cost, is not a continuous variable as it was used explicitly in gravity model, and it serves rather to find the ranking of destinations from a given origin point (Bonnel, 2004; Hensher, 1977; Hensher & Button, 2008; Ortuzar & Willumsen, 1994). The use of this model in transportation planning is briefly presented in literature. In the 1980s, opportunity model was used for modelling during the Chicago Area Transportation Studies (Eash, 1983, 1984). More recently, the IOM was used for simulating student flows and results confirmed such model has better performance than the GM (Almeida & Goncalves, 2001). In order to consider both distance and intervening opportunities in a single trip distribution model, a hybrid gravity-intervening opportunities model is also presented and tested in the literature (Goncalves & Ulysséa Neto, 1993; Wills, 1986). Also a recent study shows that a destination choice model based on the utility maximization principle behaves better than the GM for reproducing an observed OD matrix (Mishra, Wang, Zhu, Moeckel, & Mahaparta, 2013). An Intergraded IOM (IIOM) which integrates trip generation and trip distribution in a supply-dependant model is calibrated on PT trips within the Island of Montreal, and the results showed that the IIOM behaves much better than the GM in order to reproduce an observed OD matrix (Nazem, et al., 2013). In the current study we aim to develop and calibrate an EIOM, based on an enhanced destination ranking procedure, and compare its performance to the BIOM which used the basic destination ranking procedure, based only on trip durations.

For calibrating trip distribution models, reference OD matrices are generally required, which could be estimated by several methods; direct observation, synthesis and etc. (Yaldi, Taylor, & Yue, 2011). In this study, we obtain reference OD matrices by processing the data collected during a large-scale OD survey held in 2008 in the GMA (Mobilité des Personnes, 2010).

After calibration of trip distribution models, we need some global and Goodness-of-Fit measures for analyzing the model performances. Several measures are presented in the literature (Akwawua & Pooler, 2001; Black, 1991; Chow, et al., 2005; Evans & Pooler, 1987; Hu & Pooler, 2002; Smith & Hutchinson, 1981; Yaldi, et al., 2011). We will present formulations of required measures and other related works in the following sections.

DATA FOR ANALYSIS

In this section we introduce the datasets used in this study. First we present the reference datasets and afterward, their preparation for the modelling. Figure 2 shows a data flow diagram of all reference and prepared datasets.

Reference datasets

Reference datasets represent the raw data used for modelling. In this study, we used the 2008 GMA OD survey, the 2006 census of Canada, the General Transit Feed Specification (GTFS) files, along with the geographical data of the study area, at the MS level.

GMA Origin-Destination survey

For almost forty years, the GMA authorities have been conducting telephone OD travel surveys approximately every five years. This data includes rich information regarding all trips made by every person in a 5% sample of residing households, which makes the OD survey "the primary source of information on people movement habits" (*Mobilité des Personnes*, 2010). Precise spatiotemporal details are collected on all-purpose and all-mode trips. In this study, we used the data coming from the most recent OD survey that was conducted in 2008. In 2008, the sample contains almost 319,900 trips. Sociodemographic information such as dwelling location, household size, car ownership and class of income and age, gender and main occupation are also gathered. For each record an expansion factor is calculated that will be used to expand the dataset to all population based on the collected 5%

sample (*Mobilité des Personnes*, 2010). The OD survey is used for deriving the reference OD matrices for work trips, and also for the sake of trip production/attraction performance analysis.

Census of Canada

Census of Canada is a "unique undertaking on a vast scale" conducted every five years by Statistics Canada (StatCan, 2011b). It consists of collecting data from almost 31.6 million people and more than 13.5 million dwellings. Since the travel data is calibrated by using the corresponding population, we used the data from the 2006 census, for deriving the population per age group and also the number of opportunities for work trips at the MS level (StatCan, 2011a).

General Transit Feed Specifications (GTFS)

Based on the definition given by Google®, "the GTFS defines a common format for PT schedules and associated geographic information. GTFS allow PT agencies to publish their transit data and developers to write applications that consume that data in an interoperable way" (Google, 2012). In this study we used the GTFS data of the study area that is obtained from Société de Transport de Montréal (STM), the PT agency of the Island of Montreal.
The GTFS is used for several purposes such as characterizing the PT Level of Service in each MS, and also calculating the PT trip duration for each OD pair. The PT LOS is represented with the total number of Passage-Stops (transit vehicle passing a stop) per 24 hours in each MS, which is a variable sensitive to changes in PT supply.
For calculating PT trip durations for each OD pair, as this study is aggregate at the MS level, we consider the geographic centroid of each MS as its spatial delegate. Morning peak hour week schedules from the Montreal GTFS data were used to get the shortest routes and related trip durations, and to obtain finally a PT trip duration matrix between all 41 MS centroids in the Island of Montreal.
The next section presents the data preparation.

Data preparation

In this section, we discuss the data preparation for calibrating the models. Based on the presented data sources in Figure 2, we obtain a reference OD matrix, number of opportunities in each MS, number of spatially cumulative opportunities for each OD pair for the BIOM and the EIOM, and the total number of Passage-Stops per 24 hours in each MS.

Reference OD matrix

The OD matrix for work trips is derived from the 2008 OD survey data. Each record in the OD survey presents complete characteristics of a person’s trips. First we exclude all return-home and off-peak trips to obtain the required data. Afterwards we calculate the number of work trips for each OD pair by summing up the expansion factors for all concerned OD records in the survey. As the OD survey is limited to a 5% household sample, some OD pairs will have no trips, meaning that there is no data in the survey for those OD pairs.

This kind of OD reference dataset is usually presented in a matrix form, but in this study we turned columns into rows to obtain a table form (sample presented in Table 2). Afterwards, we can obtain trip production and attraction for each MS, which is used for analysing the modelling performance.

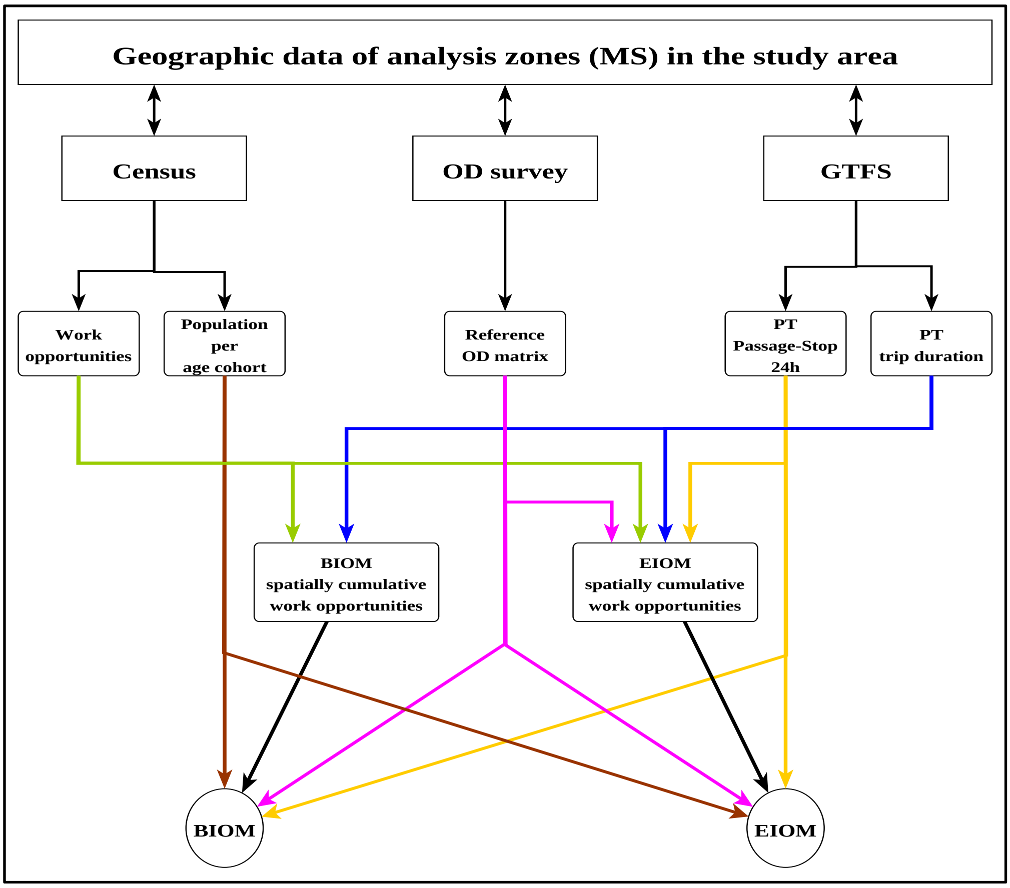

Figure 2 - Data flow diagram for the development of the BIOM and the EIOM

Number of opportunities in each MS

The development of trip distribution models needs the number of opportunities in each MS. As number of opportunities that can satisfy each trip purpose depends on activity locations that can fulfil that activity type, we need a dataset that presents the number of work opportunities at the MS level. This is derived from the 2006 census of Canada data.

Number of spatially cumulative opportunities in each MS

For the development of an IOM, we need the number of intervening opportunities for each OD pair. The number of intervening opportunities between zones *i* and *j* is the sum of all opportunities located *between i* and *j*. This number is called spatially cumulative opportunities, and will be used in the calibration of the IOM (Equation 5). In this study, the number of spatially cumulative opportunities is calculated in two different ways, which will result in two different models called BIOM and EIOM:

- BIOM: The calculation of spatially cumulative opportunities is done by a *basic destination ranking procedure* based on OD trip durations.
- EIOM: The calculation of spatially cumulative opportunities is done by an *enhanced destination ranking procedure* based on the maximization of a utility function composed of OD trip duration and PT LOS and number of potential opportunities at each destination. In contrast with the BIOM where only trip duration was used for ranking all possible destinations from each given origin, in the EIOM, values of the utility function for each OD pair are used for ranking.

Figure 2 shows a data flow diagram for the development of both BIOM and EIOM. In next section, the development of the models is described.

DEVELOPMENT OF THE MODELS

First we present the general form of the IOM, followed by the calibration of the BIOM and the EIOM and the analysis of their performance. The last part of this section presents a spatial analysis of the models in order to understand their limitations.
All models are calibrated on the 2008 GMA OD survey dataset for work trips done by PT during the weekday morning peak hours within the Island of Montreal. After applying the expansion factors, we obtain almost 158,790 PT trips, which represents 57.1% of all PT AM peak trips in the study area.

General form of the IOM

The IOM in its general form is presented as (Bonnel, 2004):

$$\boldsymbol{T_{ij} = k.E_i.(e^{-P*O_{j-1}} - e^{-P*O_j})} \quad , \forall\ \boldsymbol{i, j} \qquad \textbf{(Equation 1)}$$

Tij: Number of trips from i to j
Ei: Trip production at i
P: Probability of choosing a potential opportunity
Oj: Number of spatially cumulative opportunities between i and j, including j
Oj-1: Number of spatially cumulative opportunities between i and j, excluding j
k: Adjustment coefficient

We suppose a single-constrained model on trip production, presented as:

$$\boldsymbol{\sum_{l=1}^{J} T_{il} = E_i} \quad , \forall\ \mathbf{i} \qquad \textbf{(Equation 2)}$$

Tij: Number of trips from i to j
Ei: Trip production at i

If we substitute the T_{ij} in Equation 2 with the general form of T_{ij} in Equation 1, the single-constrained IOM is derived as:

$$\mathbf{T_{ij} = k.E_i.\frac{e^{-P*O_{j-1}} - e^{-P*O_j}}{1 - e^{-P.O_J}}} \quad , \forall\ \mathbf{i, j} \qquad \textbf{(Equation 3)}$$

Tij: Number of trips from i to j
Ei: Trip production at i
P: Probability of choosing a potential opportunity
Oj: Number of spatially cumulative opportunities between i and j, including j
Oj-1: Number of spatially cumulative opportunities between i and j, excluding j
O_J: Total number of opportunities
k: Adjustment coefficient

Trip generation , E_i in Equation 3 is set as a linear function combining sociodemographic and PT LOS characteristics. In this study, we consider E_i as:

$$\mathbf{E_i = a_1.P_{1i} + a_2.P_{2i} + a_3.P_{3i} + b.LOS_i} \quad , \forall \mathbf{i} \qquad \textbf{(Equation 4)}$$

E_i: Trip production at i
P_{1i}: Population age between 0-19 years old at i
P_{2i}: Population age between 20-64 years old at i
P_{3i}: Population age 65 years old and older at i
LOS_i: Total number of PT Passage-Stops per 24 hours at i
a_1, a_2, a_3, b: Model parameters

By combining Equations 3 and 4, the general form of the IOM is derived as:

$$\mathbf{T_{ij} = k.(a_1.P_{1i} + a_2.P_{2i} + a_3.P_{3i} + b.LOS_i).\frac{e^{-P*O_{j-1}} - e^{-P*O_j}}{1 - e^{-P.O_J}}} \quad , \forall \mathbf{i, j} \qquad \textbf{(Equation 5)}$$

Tij: Number of trips from i to j
k: Adjustment coefficient
P_{1i}: Population age between 0-19 years old at i
P_{2i}: Population age between 20-64 years old at i
P_{3i}: Population age 65 years old and older at i
LOS_i: Total number of PT Passage-Stops per 24 hours at i
a_1, a_2, a_3, b: Model parameters
P: Probability of choosing a potential opportunity
Oj: Number of spatially cumulative opportunities between i and j, including j
Oj-1: Number of spatially cumulative opportunities between i and j, excluding j
O_J: Total number of opportunities

Both BIOM and EIOM have the same form (as presented in Equation 5); the only difference between them is the way that the destination ranking procedure is applied. Supposing k=1, in next sections we describe the development of the BIOM and the EIOM for work trips.

BIOM

The Basic Integrated Opportunity Model is based on the formulation presented in Equation 5. As the BIOM in this study is calibrated on work trips, we need the number of work opportunities in each MS, which is derived from the 2006 census of Canada data. Regarding the intervening opportunities which are presented in Equation 5 by the number of spatially cumulative opportunities, the basic destination ranking procedure is applied. To do so, after calculating PT trip durations for all OD pairs via GTFS, we suppose that the number of intervening opportunities for each OD pair is the sum of all opportunities *located spatially between the origin and the destination.* This number can be calculated by executing the following steps:

- Step 1: Sum up all opportunities in each destination MS in order to find the total number of opportunities in each MS.
- Step 2: From each origin MS, rank all destination MS based on the increasing PT trip duration.
- Step 3: Calculate the number of spatially cumulative opportunities including each destination MS, based on the ranked MS in step 2.
- Step 4: Calculate the number of spatially cumulative opportunities excluding each destination MS, based on the ranked MS in step 2.

Table 1 shows the form of the derived dataset for the origin MS 101.

By integrating the spatially cumulative opportunities data, OD trip matrix, and PT LOS and sociodemographic data, we derive all data that is required for calibrating the BIOM. Afterwards the calibration is done by the statistical software, STATA®, via a nonlinear optimization procedure; results presented in Table 6.

Observed values of T_{ij} versus the estimated values can be presented as:

$$\boldsymbol{T_{ij} = 0.8662T^*_{ij}} \quad \boldsymbol{with} \quad \mathbf{R2 = 84.84\%} \qquad \text{(Equation 6)}$$

T_{ij} *: Observed number of work trips from i to j*
T^*_{ij} *: Estimated number of work trips from i to j*

Table 1 - Sample of spatially cumulative opportunities for work trips (Basic destination ranking procedure)

Origin	Destination	OD index	PT trip duration in ascending order (RANKING)	Number of opportunities at each destination	Number of spatially cumulative opportunities between i and j, including j	Number of spatially cumulative opportunities between i and j, excluding j
i	*j*	*ij*	*d_i_j*	*opp_j*	*o_j*	*o_j-1*
...						
	101	101101	0	182,215	182,215	0
	102	101102	13	106,045	288,260	182,215
	106	101106	15	43,994	332,254	288,260
	107	101107	19	48,878	381,132	332,254
	105	101105	26	65,779	446,911	381,132
101	...					
	138	101138	74	3,165	1,095,350	1,092,185
	137	101137	75	9,780	1,105,130	1,095,350
	140	101140	85	4,485	1,109,615	1,105,130
	134	101134	96	1,835	1,111,450	1,109,615
	141	101141	102	1,165	**1,112,615**	1,111,450
...						

EIOM

For calibrating the Enhanced Integrated Opportunity Model (Equation 5), we use the same dataset for sociodemographic and PT LOS attributes; the only difference between the EIOM and the BIOM is the destination ranking procedure. This will affect the number of spatially cumulative opportunities.

Enhanced destination ranking procedure

From each origin zone we aim to rank all possible destinations based on their attractiveness. The attractiveness of each destination, *j*, from each origin, *i*, is represented by a utility function, U_{ij}, composed of several alternative specific variables:

$$\mathbf{U_i^j} = \boldsymbol{u_1.d_{ij} + u_2.LOS_j + u_3.OPP_j} \quad , \forall\ \mathbf{i, j} \qquad \textbf{(Equation 7)}$$

U_i^j: *Utility of choosing destination zone, j, from origin zone, i*
d_{ij}: *PT trip duration between i and j (min.)*
LOS_j: *Total number of PT Passage-Stops per 24 hours at j*
OPP_j: *Total number of work opportunities at j*
u_1, u_2, u_3: *Model parameters*

As the study area is composed of 41 MS, we can suppose 41 different alternative destination zones from each given origin zone. Based on this hypothesis, we make a choice set of 41 MS for each declared OD trip in the 2008 OD survey. Table 2 shows a sample of data derived from the 2008 GMA OD survey. For creating the choice set, each line of Table 2 will be expanded to 41 lines, each one representing one of the 41 possible alternative destination MS in the study area. Table 3 shows a sample of the created choice set. In this table, the column *altj* shows the 41 alternative MS destinations, and *choice* column shows the chosen alternative. The values in the *t_ij_work* column represent the frequency of each choice in the dataset, and the last three columns represent the values of the alternative specific variables in Equation 6.
We calibrated a conditional logit model by the statistical software, STATA® on the dataset in Table 3 to calculate the coefficients of the utility function presented in Equation 7. Results that are presented in Table 4 show that all variables are statistically significant with 95% confidence intervals.
Now we can calculate the utility value for each OD pair and also the probability of choosing each possible alternative destination from each given origin:

$$\boldsymbol{P_i^j} = \frac{e^{U_i^j}}{\sum_{k=101}^{141} e^{U_i^k}} \quad , \forall\ \mathbf{i, j} \qquad \textbf{(Equation 8)}$$

P_i^j: *Probability of choosing destination zone, j, from origin zone, i*
U_i^j: *Utility of choosing destination zone, j, from origin zone, i*

The enhanced ranking procedure is based on the probabilities calculated by Equation 8. Table 5 shows a sample of the new destination ranks from the origin zone 102. The fact that the trip duration values, column *d_i_j* are not anymore in the ascending order confirms the difference between the basic and the enhanced ranking procedures.

Table 2 - Sample of OD trip numbers derived from the 2008 GMA OD survey

Index used for modelling	Origin	Destination	Work OD trips
n	*i*	*j*	*t_ij_work*
1	101	101	388
2	101	102	102
3	101	106	81
4	101	107	165
5	101	105	49
…	…	...	...

Table 3 - Sample of the data layout for the conditional logit model (Enhanced destination ranking procedure)

Index used for modelling	Origin	List of 41 alternative destinations	Chosen alternative	Frequency of each choice	PT trip duration between origin and each alternative destination	PT Pass-Stop 24 hours at each alternative destination	Number of opportunities at each alternative destination
n	*i*	*altj*	*choice*	*t_ij_work*	*d_i_altj*	*los_altj*	*opp_altj*
…							
		101	0	102	0	28802	182215
		102	1	102	13	24003	106045
		103	0	102	30	23859	22814.5836
		104	0	102	27	30880	19892.1996
		105	0	102	26	42064	65778.599
2	101	…					
		136	0	102	67	15993	8290
		137	0	102	75	3546	9780
		138	0	102	74	6455	3165
		139	0	102	57	3582	4360
		140	0	102	85	2932	4485
		141	0	102	102	129	1165
…							

Table 4 - Calibrated coefficients of the utility function, Equation 7 (Enhanced destination ranking procedure)

Conditional logit regression		*Prob>chi2 = 0.000*			*Pseudo R2 =35.31%*	
Model parameter	Coefficient	Std. Err.	z	P>|z|	[95% Conf. Interval]	
u_1	**-0.0289716**	.0001372	-211.19	0.000	-.0292405	-.0287027
u_2	**0.000011**	3.68e-08	545.85	0.000	.00002	.0000202
u_3	**0.0000201**	1.99e-07	55.34	0.000	.0000106	.0000114

Table 5 - Sample of new destination ranks based on probabilities (Enhanced destination ranking procedure)

Origin	Destination	PT trip duration between origin and each destination	Probability of choosing j, when leaving from i
i	*j*	*d_i_j*	*p_i_j*
...			
	101	14	48.88%
	102	0	15.05%
	119	53	5.34%
	105	31	3.33%
	107	20	3.14%
102	…		
	138	80	0.15%
	135	79	0.14%
	140	90	0.11%
	134	102	0.08%
	141	107	0.06%
...			

After calculating the probabilities of choosing each destination zone from each origin zone, we perform the following steps in order to calculate the number of spatially cumulative opportunities, based on the enhanced ranking procedure:

- Step 1: Sum up all opportunities in each destination MS in order to find the total number of opportunities in each MS.
- Step 2: From each origin MS, rank all destination MS based on the increasing probability calculated from Equation 8 (Sample presented in Table 5).
- Step 3: Calculate the number of spatially cumulative opportunities including each destination MS, based on the ranked MS in step 2.
- Step 4: Calculate the number of spatially cumulative opportunities excluding each destination MS, based on the ranked MS in step 2.

This will result in a dataset similar to Table 1, but derived from the enhanced ranking procedure. By integrating this dataset with the OD trip matrix, PT LOS and sociodemographic data we derive all required data for calibrating the EIOM, remembering that the EIOM is also based on the formulation presented in Equation 5. The calibration is done by the statistical software, STATA® via a nonlinear optimization procedure. Results show that all variables are statistically significant with acceptable confidence intervals (Table 6).

Observed values of Tij versus the estimated values can be presented as follows:

$$\boldsymbol{T_{ij} = 0.8923T^*_{ij} \quad with \quad \mathrm{R2} = 88.47\%} \qquad \textbf{(Equation 9)}$$

T_{ij} : Observed number of work trips from i to j
T^*_{ij} : Estimated number of work trips from i to j

In the next section, we compare the performance of the BIOM and the EIOM, in order to confirm the appropriateness of the enhanced destination ranking procedure.

Table 6 - Calibrated parameters of the BIOM and the EIOM (Equation 5)

Model parameter	EIOM						BIOM
	Coefficient	Std. Err.	t	P>|t|	[95% Conf. Interval]		Coefficient
$\mathbf{a_1}$	**-0.1641485**	0.0340167	-4.83	0.000	-0.2309518	-0.0973452	**-0.1985162**
$\mathbf{a_2}$	**0.176492**	0.0095483	18.48	0.000	0.1577408	0.1952432	**0.2303826**
$\mathbf{a_3}$	**-0.0477782**	0.0412879	-1.16	0.248	-0.1288609	0.0333045	**0.0218092**
$\mathbf{b}$	**0.0288631**	0.0100224	2.88	0.004	0.0091808	0.0485455	**0.0203322**
$\boldsymbol{P}$	**0.000244%**	1.18E-07	20.74	0.000	0.00000221	0.00000267	**0.000144%**

Modelling performance analysis

In this section, an analysis of the models' performance is presented by means of several global and Goodness-of-Fit measures. Almost all measures confirm that the EIOM compared to the BIOM behaves much better in reproducing the observed OD matrix derived from the 2008 GMA OD survey.

Cumulative number of trips versus trip duration

Figure 3 shows the observed and estimated cumulative number of trips versus trip duration for the BIOM and the EIOM. The figure confirms that both models reproduce similar curves to the observed curve. Also we see that globally the EIOM compared to the BIOM yields closer values to the observed cumulative number of trips.

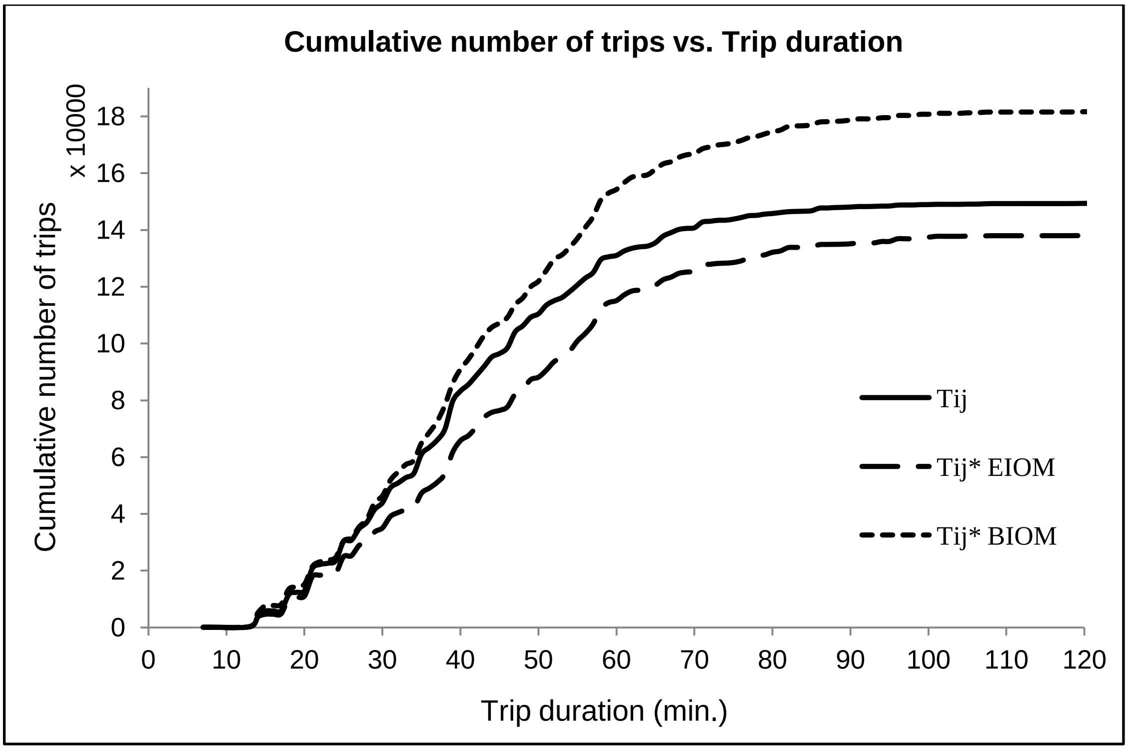

Figure 3 - Cumulative number of trips vs. Trip duration for BIOM and EIOM

Trip production (E_i) and trip attraction (A_j) modelling performance

Analyzing the performance of trip distribution models in reproducing trip production and attraction values is of great interest. Trip production and attraction measures that are presented in Table 7 confirms that the EIOM behaves much better than the BIOM.

Goodness-of-Fit measures

Several Goodness-of-Fit measures that compare entries cell-by-cell in the observed and estimated matrices are presented in Table 7.

Mean trip duration error

Mean trip duration error is the difference between the mean trip duration estimated by the model and the mean trip duration derived from the OD survey. Results that are reported in Table 7 confirm that based on this measure, both BIOM and EIOM behave good.

Coefficient of determination (R2)

Although some studies showed that in some cases the coefficient of determination *(R2)* may yield artificially high values in Goodness-of-Fit applications, we present it as a traditional measure, because it is one of the most cited measures in the literature (Black, 1991;

Knudsen & Fotheringham, 1986; Smith & Hutchinson, 1981; Yaldi, et al., 2011). The values of *R2* reported in Table 7 show that the EIOM is more accurate than the BIOM.

Mean Absolute Error (MAE) and Normalized MAE (NMAE)

MAE and NMAE are defined as (Smith & Hutchinson, 1981):

$$MAE = \frac{\sum_{i,j}|T_{ij} - T_{ij}^*|}{N} \qquad \text{(Equation 10)}$$

$$NMAE = \frac{MAE}{T/N} \qquad \text{(Equation 11)}$$

T_{ij}: Observed number of trips from i to j
T_{ij}^*: Estimated number of trips from i to j
T : Total number of trips derived from the OD survey
N : Number of estimated OD pairs

Table 7 - Goodness-of-fit measures

Performance analysis measures		BIOM	EIOM
Global measures	Mean trip duration (37.5 min observed mean trip duration)	35.7 min	39.4 min
	T_{ij}=k. T_{ij}* (Value of k)	0.8662	**0.8923**
	R2	84.88%	**88.47%**
	MAE	120.17	**97.36**
	NMAE	0.37	**0.30**
	DI (PME)	24.93	**20.32**
	RMSE	208.08	**190.80**
Trip production measures	E_i=k.E_i* (Value of k)	1.1898	**0.8805**
	R2	**95.13%**	94.43%
	DI (PME)	12.58	**8.98**
Trip attraction measures	A_j=k.A_j* (Value of k)	0.9133	**0.9706**
	R2	91.88%	**94.85%**
	DI (PME)	19.60	**12.41**

Larger values of the MAE and NMAE represent less accurate model fits. These measures also show that the EIOM is more accurate than the BIOM.

Dissimilarity Index (DI) or Percentage Misallocated Error (PME)

DI or PME which shows the percentage of the flows that are allocated to wrong cells in the matrix is defined as (Evans & Pooler, 1987; Hu & Pooler, 2002):

$$DI = \frac{50}{T}\sum_{i,j}|T_{ij} - T_{ij}^*| \qquad \text{(Equation 12)}$$

T_{ij}: Observed number of trips from i to j
T_{ij}^*: Estimated number of trips from i to j
T : Total number of trips derived from the OD survey

Larger values of DI show larger dissimilarities between the estimated and the observed OD survey matrices. Table 7 shows that based on this measure, the EIOM behaves much better than the BIOM.

Root Mean Square Error (RMSE)

RMSE is defined as (Yaldi, et al., 2011):

$$RMSE = \sqrt{\frac{\sum_{i,j}(T_{ij} - T^*_{ij})^2}{N}} \qquad \text{(Equation 13)}$$

T_{ij}: Observed number of trips from i to j
T^*_{ij}: Estimated number of trips from i to j
N : Number of estimated OD pairs

Table 7 confirms that the EIOM behaves better than the BIOM, based on RMSE values.
Based on all Goodness-of-Fit measures that are presented in Table 7, the EIOM is more accurate than the BIOM. Moreover, the EIOM allows policy-makers and analysts to study potential changes in PT trip distribution pattern and also destination attractiveness due to modifications in demography, job spatial location and also PT supply.
In the next section, we study limitations of the BIOM and the EIOM from a spatial point of view, to understand their weaknesses and strengths.

Spatial limitations of the models

The spatial limitations of the BIOM and the EIOM are presented by analyzing the spatial distribution of underestimation and overestimation residual errors.
Regarding the underestimation residuals with both models, few OD pairs yield spatial errors, and also all these spatial errors have a random nature.
Concerning the overestimation errors, Figures 4a and 4b present desire lines that are plotted between zones with overestimation errors of almost 100% or more. These figures show less residual errors reported by the EIOM compared to the BIOM, that confirms a much better performance for the EIOM. Moreover, in contrast with the random nature of the errors reported by the BIOM, the overestimation spatial errors of the EIOM are more concentrated in a big destination MS. This leads us to suppose that probably changing the zoning system, which is a part of our ongoing research to make more homogeneous zones can resolve this problem.

CONCLUSION AND FUTURE RESEARCH

In this study we developed an Enhanced Intervening Opportunities Model (EIOM) for PT trip distribution. The main difference between this model and the Basic Intervening Opportunities Model (BIOM) is in the destination ranking procedure that is used for the model calibration. Both models are calibrated on work trips, with single constraints on trip production for PT morning peak hours trips within the Island of Montreal, Canada. The flow of data that is presented in Figure 2 shows different data sources and the way that each required dataset is obtained from the raw data. The analysis of modelling performance by means of several global and Goodness-of-Fit measures confirmed that the EIOM is well-behaved and more accurate than the BIOM.
We studied also limitations of the EIOM from a spatial point of view. In other words, by comparing observed and estimated number of trips on an underestimation-overestimation basis, and spatial weaknesses and strengths of the models are discovered. This could help us to find improvement strategies in terms of explanatory variable choice, zoning system and model formulation modifications, which presents a part of our ongoing and future research.

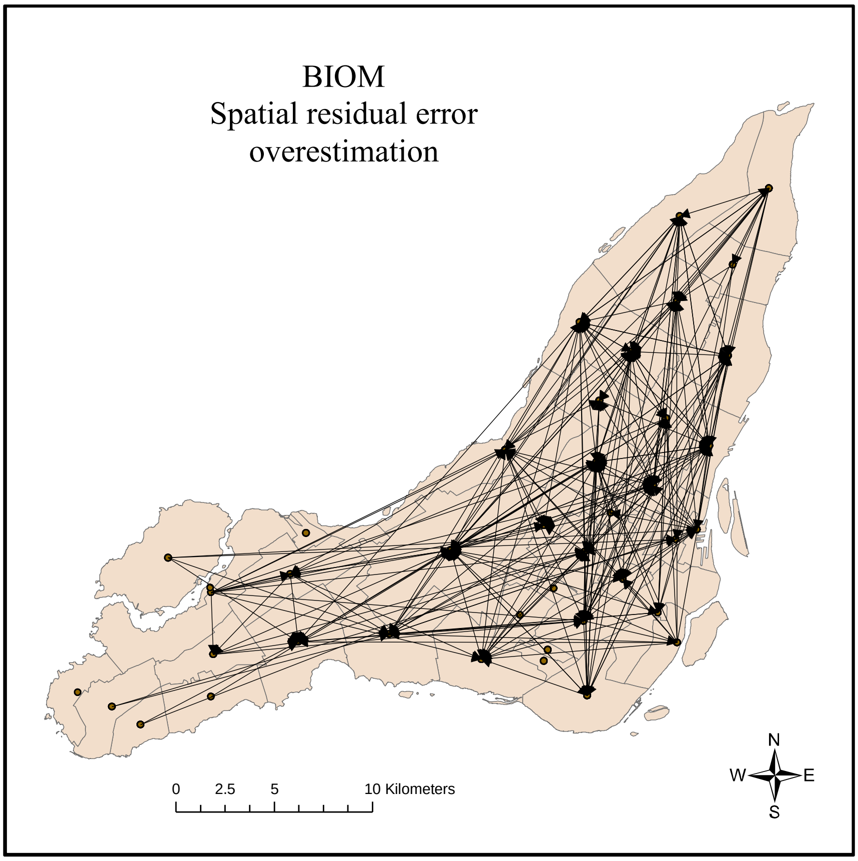

Figure 4a - Spatial residual errors reported from the BIOM

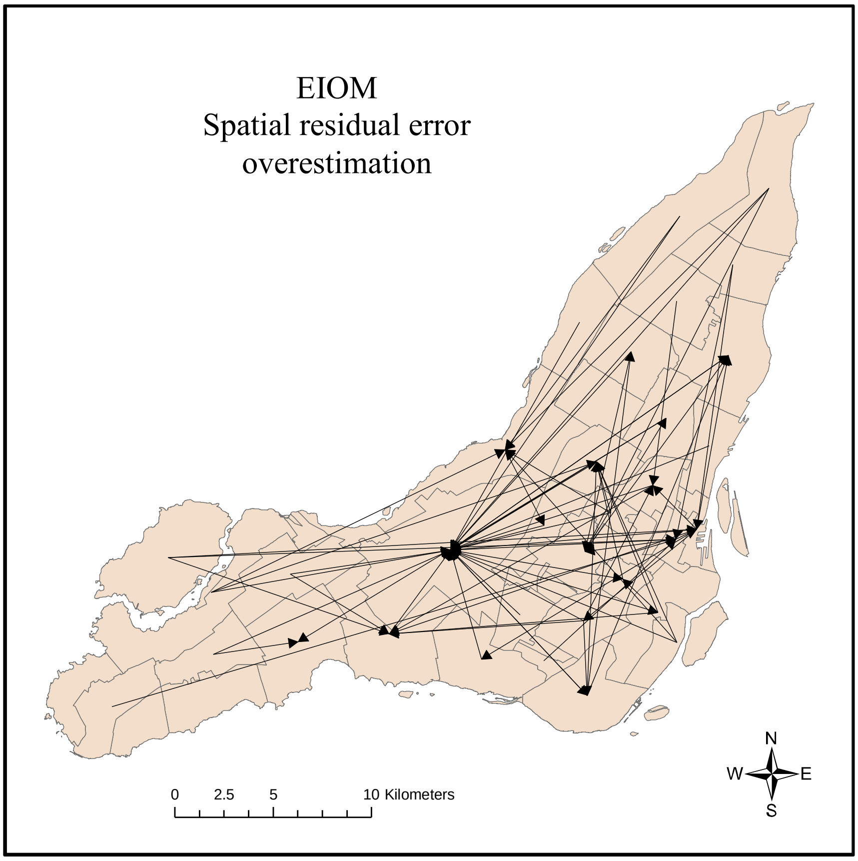

Figure 4b - Spatial residual errors reported from the EIOM

The presented formulation of the EIOM (Equation 5) and the enhanced destination ranking procedure represent great potentials for policy-makers and PT analysts, due to the sociodemographic and PT supply variables. Also, as all the required data for calibrating the EIOM comes from external and independent sources, this model could be used for the sake of PT trip generation and distribution forecasting for the future. Also the presence of sociodemographic and PT supply characteristics in the EIOM allows us to study the effect of potential changes of these variables on PT trip pattern. As we discussed in our previous studies, different sociodemographic groups have different behaviours in a PT network (Nazem, Trépanier, & Morency, 2011), and this is the main reason for considering different sociodemographic groups in the present study.

In contrast with the basic procedure, the enhanced destination ranking procedure considers trip purpose in ranking alternative destinations from a given origin. This results different destination ranking lists for different trip purposes, which is more realistic and reasonable.

Regarding trip duration values, the fact that they are used only in the destination ranking procedure, and not explicitly in the model as it was used in the GM, makes the BIOM and the EIOM less dependent to exact trip duration values. In contrast, it is of a great interest to study the sensitivity of the IIOM to trip duration values, because a minor change in the trip duration might change the ranking of zones, which will result in changes in the number of intervening opportunities.

For further research, we propose a bi-level optimization approach, by using data from an automated fare collection system. At the first level, we will define MS trip production based on sociodemographic, socioeconomic and PT LOS variables, and then we will calibrate it with data derived from smart card validations. The second level is dedicated to the calibration of an EIOM by means of the OD survey data and trip production values calculated at the first level. This new model could be interesting, because it allows us the usage of the detailed data derived from smart card validations. Also it allows the integration of two data sources of OD survey and smart card validation data in the calibration of a single model.

Another research topic is the development of the EIOM at a grid level. This could present interesting results for analyzing the sensitivity of PT trip distribution models to the study level. Moreover, we could probably ameliorate spatial errors experienced in the present study at the MS level.

We conclude that the EIOM is advantageous due to its behavioural and supply-dependent bases. Also the enhanced destination ranking procedure might be used for behavioural analysis, in terms of choosing alternative destinations based on their PT LOS, number of opportunities and OD trip duration values.

ACKNOWLEDGMENTS

The authors acknowledge the support of Natural Sciences and Engineering Research Council of Canada (NSERC) and of the Agence Métropolitaine de Transport (AMT). We thank the survey consortium which provided the Greater Montreal Origin-Destination survey and the GTFS data. Special thanks to François Godefroy, Audrey Godin, Pierre-Léo Mongeon-Bourbonnais, Éric Poliquin, Hubert Verreault, and other members of the Mobilité chair at École Polytechnique de Montréal for their help to obtain and prepare different required datasets.

REFERENCES

Akwawua, S., & Pooler, J. (2001). The Development of an Intervening Opportunities Model with Spatial Dominance Effects. Journal of Geographical Systems, 3, 69-86.

Almeida, L. M. W., & Goncalves, M. B. (2001). A Methodology to Incorporate Behavioral Aspects in Trip Distribution Models with an Application to Estimate Student Flow. Environment and Planning A, 33, 1125-1138.

Black, W. R. (1991). A Note on the Use of Correlation Coefficients for Assessing Goodness-of-fit in Spatial Interaction Models. [Article]. Transportation, 18(3), 199-206.

Bonnel, P. (2004). Prévoir la Demande de Transport. Paris, France: Presses Ponts et Chaussées.

Chow, L.-F., Zhao, F., Li, M.-T., & Li, S.-C. (2005). Development and Evaluation of Aggregate Destination Choice Models for Trip Distribution in Florida. Transportation Research Record(1931), 18-27.

De Grange, L., Troncoso, R., Ibeas, A., & González, F. (2009). Gravity Model Estimation with Proxy Variables and the Impact of Endogeneity on Transportation Planning. Transportation Research Part A, 43, 105-116.

Eash, R. (1983). Several More Improvements in Understanding, Calibrating, and Applying the Opportunity Model Chicago, IL: Chicago Area Transportation Study.

Eash, R. (1984). Development of a Doubly Constrained Intervening Opportunity Model for Trip Distribution. Chicago, IL: Chicago Area Transportation Study.

Evans, N. J., & Pooler, J. (1987). Distance Deterrence Effects in Constrained Spatial Interaction Models of Interprovincial Migration. Canadian Journal of Regional Science, Autumn 1987, 259-279.

Feng, S., & Li, X. (2004). Research on Public Transit Passenger OD Matrix Estimation. Paper presented at the 8th International Conference on Applications of Advanced Technologies in Transportaion Engineering.

Goncalves, M. B., & De Cursi, J. E. S. (2001). Parameter Estimation in a Trip Distribution Model by Random Perturbation of a Descent Method. Transportation Research Part B, 35, 137-161.

Goncalves, M. B., & Ulysséa Neto, I. (1993). The Development of a new Gravity - Opportunity Model for Trip Distribution. Environment and Planning A, 25(6), 817-826.

Google. (2012). What is GTFS? Retrieved July 7, 2012, from https://developers.google.com/transit/gtfs/

Hensher, D. A. (1977). Urbain Transport Economics. Cambridge: Cambridge University Press.

Hensher, D. A., & Button, K. J. (2008). Handbook of Transport Modelling. Amsterdam, The Netherlands: Elsevier.

Hu, P., & Pooler, J. (2002). An Empirical Test of the Competing Destinations Model. Journal of Geographical Systems, 4, 301-323.

Kermanshah, M. (2004). Notes de Cours "Prévision de la Demande en Transport". Sharif University of Technology.

Knudsen, D. C., & Fotheringham, A. S. (1986). Matrix Comparison, Goodness-of-Fit and Spatial Interaction Modeling. International Regional Science Review, 10(2), 127-147.

Mishra, S., Wang, Y., Zhu, X., Moeckel, R., & Mahaparta, S. (2013). Comparison between Gravity and Destination Choice Models for Trip Distribution in Maryland. Paper presented at the TRB of the National Acdemies of Science Annual Meeting.

Mobilité des Personnes. (2010). Montreal, QC: Secrétariat à l'Enquête Origine-Destination (OD), Agence Métropolitaine de Transport (AMT).

Nazem, M., Trépanier, M., & Morency, C. (2011). Demographic Analysis of Route Choice for Public Transit. Transportation Research Record, 2(2217), 71-78.

Nazem, M., Trépanier, M., & Morency, C. (2012). Hierarchical Intervening Opportunities Model for Public Transit Trip Distribution. Paper presented at the Conference on Advanced Systems for Public Transit (CASPT).

Nazem, M., Trépanier, M., & Morency, C. (2013). Integrated Intervening Opportunities Model for Public Transit Trip Generation-Distribution: A Supply-dependent Approach. Paper presented at the TRB of the National Academies of Science Annual Meeting.

Ortuzar, J. d. D., & Willumsen, L. G. (1994). Modelling Transport. Chichester, England: John Wiley & Sons Ltd.

Rajesson, F. (2009). Modèle Hybride d'Estimation de la Demande de Transport Collectif. École Polytechnique de Montréal, Montréal, Québec.

Schneider, M. (1959). Gravity Models and Trip Distribution Theory. Regional Science Association, 1, 51-56.

Smith, D. P., & Hutchinson, B. G. (1981). Goodness-of-Fit Statistics for Trip Distribution Models. Transportation Research, 15A, 295-303.

StatCan. (2011a). 2006 Census of Population, Quebec - Cat. No. 94-581-X2006002. http://ivt.crepuq.qc.ca/ [Accessed October 11, 2011]

StatCan. (2011b). Statistics Canada. www.statcan.gc.ca

Stouffer, S. A. (1940). Intervening Opportunities: A Theory Relating Mobility and Distance. American Sociological Review, 5(6), 845-867.

Stouffer, S. A. (1960). Intervening Opportunities and Competing Migrants. Journal of Regional Science, 2, 1-26.

Thamizh Arasan, V., Wermuth, M., & Srinivas, B. S. (1996). Modelling of stratified urban trip distribution. Journal of Transportation Engineering, 122(5), 342-349.

Veenstra, S. A., Thomas, T., & Tutert, S. I. A. (2010). Trip distribution for limited destinations: A case study for grocery shopping trips in the Netherlands. Paper presented at the 89th Annual Meeting of the Transportation Research Board.

Wills, M. J. (1986). A flexible gravity-opportunities model for trip distribution. Transportation Research Part B: Methodological, 20(2), 89-111.

Wilson, A. G. (1970). Advances and Problems in Distribution Modelling. Transportation Research, 4, 1-18.

Yaldi, G., Taylor, M. A. P., & Yue, W. L. (2011). Forecasting Origin-Destination Matrices by Using Neural Network Approach: A Comparison of Testing Performance between Back Propagation, Variable Learning Rate and Levenberg-Marquardt Algorithms. Paper presented at the Australasian Transport Research Forum 2011 Proceedings, Adelaide, Australia.