Integrated Intervening Opportunities Model for Public Transit Trip Generation-Distribution: A Supply-dependent Approach

Mohsen Nazem, Ph.D. Candidate (Corresponding author)
Department of Mathematical and Industrial Engineering
École Polytechnique de Montréal, C.P. 6079, Station Centre-Ville
Montréal (Québec) CANADA H3C 3A7
Phone: +1 (514) 340-4711 ext. 4157
Fax: +1 (514) 340-5763
Email: mohsen.nazem@polymtl.ca

Martin Trépanier, Ph.D., P.Eng., Full Professor
Department of Mathematical and Industrial Engineering
École Polytechnique de Montréal, C.P. 6079, Station Centre-Ville
Montréal (Québec) CANADA H3C 3A7
Phone: +1 (514) 340-4711 ext.4911
Fax: +1 (514) 340-5763
Email: mtrepanier@polymtl.ca

Catherine Morency, Ph.D., P.Eng., Associate Professor
Department of Civil, Geological and Mining Engineering
École Polytechnique de Montréal, C.P. 6079, Station Centre-Ville
Montréal (Québec) CANADA H3C 3A7
Phone: +1 (514) 340-4711 ext.4502
Fax: +1 (514) 340-5763
Email: cmorency@polymtl.ca

Word count

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>5,168</td>
</tr>
<tr>
<td>Tables (4 X 250)</td>
<td>1,000</td>
</tr>
<tr>
<td>Figures (4 X 250)</td>
<td>1,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>7,168</strong></td>
</tr>
</tbody>
</table>

Transportation Research Board of the National Academies of Science
2013 Annual Meeting
Washington, DC

Date of submission: **July 31th, 2012**
ABSTRACT

In this study an Integrated Intervening Opportunities Model (IIOM) is developed for Public Transit (PT) trips. This model is a generation-distribution supply-dependent model, with single constraints only on trip production values for work and study PT 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, schools’ enrolment data, along with the geographical data of the GMA are used. The IIOM is a nonlinear model with sociodemographic, socioeconomic and PT supply characteristics, and also work and study spatial location attributes. Analysis of the modeling performance by means of several Goodness-of-Fit measures showed that the IIOM is well-behaved and more accurate than the classical Gravity Model (GM). Based on the explanatory variables used in the IIOM, 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 and socioeconomics characteristics, PT supply, etc. Also this study opens new opportunities for development of more accurate PT demand models with new emergent data such as smart card entries in the future.

Keywords: 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 (1-4). In GM, the number of trips between each Origin-Destination (OD) pair is based directly on the OD distance and trip production/attraction values, whereas the IOM considers the number of intervening opportunities as the main influencing factor. As it seems that both of these factors influence the trip distribution, some researchers developed unified hybrid models (5-7). In past studies, especially in the case of hybrid models, trip distribution is calibrated usually just for one purpose; work or study, and thus, the advantages and disadvantages of them are not well known.

In this study, we aim to calibrate an Integrated Intervening Opportunities Model (IIOM) for work and study Public Transit (PT) trips. This model considers the sociodemographic and socioeconomic attributes as well as the PT supply characteristics of Municipal Sectors (MS) located in the Island of Montreal, Canada. In other words, 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 supply modifications on the PT trip pattern. Also as the required entry data for this model could be available for the future, it could be used for forecasting purposes.

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

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 (8). However compared to the GM, the IOM is behavioral based (9), less sensitive to the size and shape of study area (10) and also produces better results in cases where destinations which satisfy the trip purpose are not uniformly distributed, like discrete attraction points such as ones for shopping or study purposes (11).

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 (12). In this study, we use the OD matrix derived from the large 2008 OD survey of the Greater Montreal Area (GMA).

In our previous studies, the IOM is calibrated with two different approaches, called Basic IOM (BIOM) and Hierarchical IOM (HIOM). The results showed that IOM compared to the GM has much better performance, in terms of trip production and trip attraction satisfaction and other Goodness-of-Fit measures (13). This model can be used later in middle and long-term forecasting. The models are developed at the MS level, for PT trips done during normal weekday peak hours (6:00AM to 9:00AM), within the Island of Montreal. This area consists of 41 MS, with a population of nearly 1.9 millions, distributed in almost 500 km² (14). This study is limited to the Island of Montreal in order to deal with less calculation complexities, remembering that the methodology can be generalized on the whole GMA.

In contrast with the previous HIOM which has a hierarchical nature (13), in this study we develop and integrate the IIOM as one nonlinear equation. The total number of trips for each OD is defined by an integrated equation consisting of two distinct nonlinear formulations for work and study trips, and also two linear weighting coefficients (Equations 3-5). Afterwards the IIOM is calibrated to calculate the weighting coefficients and also the model parameters. A great advantage of the newly developed IIOM compared to the previous HIOM (13) is that all required entry data for the development of the new model comes from external and independent sources (Figure 2), which presents a great predictive capacity for the new IIOM. Also the sociodemographic, socioeconomic and PT supply variables allow study of the effect of changes in these variables on the PT trip pattern.

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 modeling development are described. In the next section, following a descriptive analysis of the data, the development of the IIOM is presented. Afterwards, the performance of the IIOM is studied and compared to the classical GM, by means of several Goodness-of-Fit measures. The next section presents spatial limitations of the IIOM, by means of a spatial residual errors analysis. In the conclusion, we discuss some interesting potentials of the IIOM 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, and among them the Gravity Model (GM) and the Intervening Opportunities Model (IOM) are the most common (1, 2).

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 (1-4, 15-17).
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 (18-20). The fundamental idea of this model is that generalized cost is not the only factor that affects destination choices. Contrariwise, this model considers the relative accessibility of opportunities that can satisfy the trip purpose is the main influencing factor. 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 anymore as it was used explicitly in gravity model, and it serves rather to find the ranking of destinations from a given origin point (1-3, 21). The use of this model in transportation planning is briefly presented in literature. In the 1980s, opportunity model was used for modeling during the Chicago Area Transportation Studies (22, 23). More recently, the IOM was used for simulating student flows and results confirmed such model has better performance than the GM (6). In order to consider both distance and intervening opportunities in a single trip distribution model, an integrated gravity-intervening opportunities model is also presented and tested in the literature (5, 7). In the current study we aim to develop and calibrate an IIOM for work and study trips and compare its performance to a classical GM.

For calibrating trip distribution models, reference OD matrices are generally required, which could be estimated by several ways; direct observation, synthesis and etc. (24). 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 (25).

After calibration of trip distribution models, we need some Goodness-of-Fit measures. Several measures are presented in the literature (10, 24, 26-30). We will present formulations of required Goodness-of-Fit 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 data sources and afterward, their preparation for the modeling.

GMA Origin-Destination survey

For almost forty years, the Greater Montreal Area (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 peoples movement habits" (25). 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. Demographic information such as dwelling location, household size, car ownership and class of income and age, gender and main occupation are also gathered. Each record presents an expansion factor that will be used to expand the dataset based on the collected 5% sample (25).

In this study, the OD survey is used for deriving the reference OD matrices for each trip purpose; the process will be explained in the "Data preparation" section.
Census of Canada
Census of Canada is a "unique undertaking on a vast scale" conducted every five years by Statistics Canada (14). It consists of collecting data from 31.6 million people and more than 13.5 million dwellings. In this study, we used the data coming from one of the most recent censuses that was conducted in 2006 (31), and derived the population per age group, number of opportunities for work trips and individual average income in each MS.

Schools' enrolment dataset
Number of opportunities for study trips is derived from the schools' enrolment dataset. This dataset lists all students and their related school and allows estimation of the capacity of educational institutions.

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 feeds allow PT agencies to publish their transit data and developers to write applications that consume that data in an interoperable way" (32). In this study we used the Island of Montreal GTFS obtained from Société de Transport de Montréal (STM), which is the public transit agency in the Island of Montreal.

The GTFS is used for several purposes such as characterizing the PT Level of Service (PTLOS) in each MS, and also calculating the PT trip duration for each OD pair. The PTLOS is represented with two variables, which are sensitive to changes in the PT service characteristics, such as the headway, etc.: • Total number of Passage-Stops (transit vehicle passing a stop) per 24 hours in each MS and; • Spatial density of stops in each MS.

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 travel times, and to obtain finally a PT travel time matrix between all 41 MS centroids in the Island of Montreal.

In an IOM, trip duration is used to calculate the spatially cumulative opportunities for each OD pair. The methodology will be described in more detail in the next sections.

The next section presents the data preparation, in order to prepare the required datasets for the modeling.

Data preparation
In this section, we discuss the data preparation for calibrating the IIOM. Based on the presented data sources, the creation of reference OD matrices, number of opportunities in each MS, number of spatially cumulative opportunities and the total number of Passage-Stops per 24 hours in each MS is described below.

Reference OD matrices
Two OD matrices, one for work and the other one for study trips are created from the 2008 OD survey data. Each record in the OD survey presents the complete characteristics of an individual movement. First
we exclude all return-home and off-peak trips to obtain the required data. Afterwards we calculate the $T_{ij}$ values for each purpose by summing up the expansion factor values for all concerned OD pairs. This will result in two distinct datasets, one for work and one for study trips. As the OD survey is limited to a 5% household sample, some OD pairs will have zero trips for some trip purposes.

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. After calculating the number of trips for each MS, we can obtain trip production and attraction values for each MS and trip purpose. A sample of deriving trip number ($T_{ij}$), production ($E_{i}$) and attraction ($A_{j}$) values for work trips is presented in Table 1.

**Number of work and study 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 fulfill that activity type, we need distinct datasets for different trip purposes.

Number of work opportunities in each MS is derived from the 2006 census of Canada data, and number of study opportunities except universities is derived by summing up the estimated capacities of educational institutions in each MS. The spatial distribution of work and study opportunities confirms that the distributions of opportunities for different purposes are not the same, as well as the necessity to consider two distinct purposes for the application of the IOM.

**Number of spatially cumulative opportunities for work and study trips**

For the development of an IOM, we need the number of intervening opportunities for each OD pair. As in the study area the PT fare remains constant, we only consider PT trip durations calculated via GTFS, in order to calculate the generalized cost for each OD pair. Based on this hypothesis, for each purpose, the number of intervening opportunities between zone $i$ 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. For each purpose, this number can be calculated by executing the following steps.

**Step 1:** From each origin MS, enumerate all destination MSs based on the increasing PT trip times.

**Step 2:** Sum up all opportunities in each destination MS in order to find the total number of opportunities in each MS.

**Step 3:** Calculate the number of spatially cumulative opportunities *including* each destination MS, based on the ranked MS.

**Step 4:** Calculate the number of spatially cumulative opportunities *excluding* each destination MS, based on the ranked MS.
Table 1 - Deriving the reference OD matrices and trip production and attraction values from the OD survey

<table>
<thead>
<tr>
<th>ID</th>
<th>Origin MS</th>
<th>Destination MS</th>
<th>Hour</th>
<th>Purpose</th>
<th>Expansion factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>688</td>
<td>106</td>
<td>106</td>
<td>7:40 AM</td>
<td>Work</td>
<td>27.46</td>
</tr>
<tr>
<td>321889</td>
<td>120</td>
<td>110</td>
<td>7:30 AM</td>
<td>Study</td>
<td>17.08</td>
</tr>
</tbody>
</table>

Calculating reference $T_{ij}$ values

<table>
<thead>
<tr>
<th>Origin MS</th>
<th>Destination MS</th>
<th>No. of work trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>104</td>
<td>101</td>
<td>2,404</td>
</tr>
<tr>
<td>104</td>
<td>105</td>
<td>801</td>
</tr>
<tr>
<td>109</td>
<td>101</td>
<td>1,411</td>
</tr>
</tbody>
</table>

Calculating reference $E_i$ and $A_j$ values

<table>
<thead>
<tr>
<th>MS</th>
<th>Work trip production</th>
<th>Work trip attraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>1,070</td>
<td>61,399</td>
</tr>
<tr>
<td>108</td>
<td>12,028</td>
<td>6,201</td>
</tr>
<tr>
<td>118</td>
<td>5,834</td>
<td>1,544</td>
</tr>
</tbody>
</table>

In order to have an idea about the data flow in this study, Figure 2 shows main data sources and two levels of data processing for developing the IIOM. In next section, the development of the IIOM is described.

DEVELOPMENT OF THE INTEGRATED GENERATION-DISTRIBUTION MODEL

In this section we develop the IIOM, which is an integrated generation-distribution supply-dependent model. In other words, using variables describing the PT supply in each MS, the model calculates the OD matrices as well as trip production and attraction at the MS level.
### Table 2 - Sample of the number of spatially cumulative opportunities for work trips

<table>
<thead>
<tr>
<th>i</th>
<th>j</th>
<th>PT trip time in ascending order</th>
<th>No. of work opportunities in MS j</th>
<th>No. of spatially cumulative work opportunities between i and j, including j</th>
<th>No. of spatially cumulative work opportunities between i and j, excluding j</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>101101</td>
<td>0</td>
<td>182,215</td>
<td>182,215</td>
<td>0</td>
</tr>
<tr>
<td>102</td>
<td>101102</td>
<td>13</td>
<td>106,045</td>
<td>288,260</td>
<td>182,215</td>
</tr>
<tr>
<td>106</td>
<td>101106</td>
<td>15</td>
<td>43,994</td>
<td>332,254</td>
<td>288,260</td>
</tr>
<tr>
<td>107</td>
<td>101107</td>
<td>19</td>
<td>48,878</td>
<td>381,132</td>
<td>332,254</td>
</tr>
<tr>
<td>105</td>
<td>101105</td>
<td>26</td>
<td>65,779</td>
<td>446,911</td>
<td>381,132</td>
</tr>
<tr>
<td></td>
<td>138101138</td>
<td>74</td>
<td>3,165</td>
<td>1,095,350</td>
<td>1,092,185</td>
</tr>
<tr>
<td></td>
<td>137101137</td>
<td>75</td>
<td>9,780</td>
<td>1,105,130</td>
<td>1,095,350</td>
</tr>
<tr>
<td></td>
<td>140101140</td>
<td>85</td>
<td>4,485</td>
<td>1,109,615</td>
<td>1,105,130</td>
</tr>
<tr>
<td></td>
<td>134101134</td>
<td>96</td>
<td>1,835</td>
<td>1,111,450</td>
<td>1,109,615</td>
</tr>
<tr>
<td>141</td>
<td>101141</td>
<td>102</td>
<td>1,165</td>
<td>1,112,615</td>
<td>1,111,450</td>
</tr>
<tr>
<td></td>
<td>141141141</td>
<td>0</td>
<td>1,165</td>
<td>1,165</td>
<td>0</td>
</tr>
<tr>
<td>140</td>
<td>141140</td>
<td>30</td>
<td>4,485</td>
<td>5,650</td>
<td>1,165</td>
</tr>
<tr>
<td>139</td>
<td>141139</td>
<td>49</td>
<td>4,360</td>
<td>10,010</td>
<td>5,650</td>
</tr>
<tr>
<td>137</td>
<td>141137</td>
<td>53</td>
<td>9,780</td>
<td>19,790</td>
<td>10,010</td>
</tr>
<tr>
<td>131</td>
<td>141131</td>
<td>55</td>
<td>33,140</td>
<td>52,930</td>
<td>19,790</td>
</tr>
<tr>
<td>141</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>141117</td>
<td>141</td>
<td>27,245</td>
<td>1,049,580</td>
<td>1,022,335</td>
</tr>
<tr>
<td>116</td>
<td>141116</td>
<td>151</td>
<td>28,665</td>
<td>1,078,245</td>
<td>1,049,580</td>
</tr>
<tr>
<td>113</td>
<td>141113</td>
<td>153</td>
<td>12,135</td>
<td>1,090,380</td>
<td>1,078,245</td>
</tr>
<tr>
<td>114</td>
<td>141114</td>
<td>158</td>
<td>15,275</td>
<td>1,105,655</td>
<td>1,090,380</td>
</tr>
<tr>
<td>115</td>
<td>141115</td>
<td>160</td>
<td>6,960</td>
<td>1,112,615</td>
<td>1,105,655</td>
</tr>
</tbody>
</table>

### Descriptive analysis of the data

In this study, we use the 2008 Montreal OD survey dataset for all work and study trips done by PT during the weekday morning peak hours (6:00AM to 9:00AM), within the Island of Montreal. After applying the expansion factors, we obtain 278,005 PT trips for calibrating the model. These AM peak hour PT trips within the Island of Montreal represent 67.5% of all PT trips in the GMA, and more globally 37.6% of all 24 hours PT trips in the GMA. The distribution of all PT AM peak trips based on purpose is 57.1% for
work, 36.7% for study and 6.2% for other purposes. In this study, we aim to model work and study trips which gather almost 94% of all PT AM peak trips.

Development of the IIOM

The IOM in its general form is presented as (1):

\[ T_{ij} = \alpha E_i \left( e^{-P \cdot O_j} - e^{-P \cdot O_{j-1}} \right), \forall i, j \]  
(Equation 1)

- \( T_{ij} \): Number of trips from \( i \) to \( j \)
- \( E_i \): Trip production at \( i \)
- \( P \): Probability of choosing a potential opportunity
- \( O_j \): Number of spatially cumulative opportunities between \( i \) and \( j \), including \( j \)
- \( O_{j-1} \): Number of spatially cumulative opportunities between \( i \) and \( j \), excluding \( j \)
- \( \alpha \): Adjustment coefficient

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

\[ \sum_{i=1}^{j} T_{ii} = E_i, \forall i \]  
(Equation 2)

- \( T_{ij} \): Number of trips from \( i \) to \( j \)
- \( E_i \): Trip production at \( i \)

If we substitute the \( T_{ij} \) in Equation 2 with the general form of \( T_{ij} \) in Equation 1, we derive the single-constrained IOM as:

\[ T_{ij} = \beta E_i \frac{e^{-P \cdot O_j} - e^{-P \cdot O_{j-1}}}{1 - e^{-P \cdot O_J}}, \forall i, j \]  
(Equation 3)

- \( T_{ij} \): Number of trips from \( i \) to \( j \)
- \( E_i \): Trip production at \( i \)
- \( P \): Probability of choosing a potential opportunity
- \( O_j \): Number of spatially cumulative opportunities between \( i \) and \( j \), including \( j \)
- \( O_{j-1} \): Number of spatially cumulative opportunities between \( i \) and \( j \), excluding \( j \)
- \( O_J \): Total number of opportunities
- \( \beta \): Adjustment coefficient
Trip generation, $E_i$, in Equation 3 is set as a linear function combining sociodemographic, socioeconomic and PTLOS variables. In this study, we consider $E_i$ for work trips as:

$$E_{i}^{\text{Work}} = \alpha_{1i}^{\text{Work}} \cdot P_{1i} + \alpha_{2i}^{\text{Work}} \cdot P_{2i} + \alpha_{3i}^{\text{Work}} \cdot P_{3i} + \beta_{1i}^{\text{Work}} \cdot \text{LOS}_{1i} + \beta_{2i}^{\text{Work}} \cdot \text{LOS}_{2i} + \gamma_{i}^{\text{Work}} \cdot \text{INC}_{i}$$

(Equation 4)

- $E_{i}^{\text{Work}}$: Work trip production at $i$
- $P_{1i}$: Population group age between 0-19 years old at $i$
- $P_{2i}$: Population group age between 20-64 years old at $i$
- $P_{3i}$: Population group age 65 years old and older at $i$
- $\text{LOS}_{1i}$: Total number of PT Passage-Stops per 24 hours at $i$
- $\text{LOS}_{2i}$: Spatial density of PT Stops at $i$ (Number of Stops at $i$ / Area of $i$)
- $\text{INC}_{i}$: Mean income per person at $i$
- $\alpha_{1i}^{\text{Work}}, \alpha_{2i}^{\text{Work}}, \alpha_{3i}^{\text{Work}}, \beta_{1i}^{\text{Work}}, \beta_{2i}^{\text{Work}}, \gamma_{i}^{\text{Work}}$: Model parameters for work trips

Trip production values for study trips is also presented with a similar equation. Afterwards we suppose a linear weighted formulation for $T_{ij}$ as follows:
\[ T_{ij} = k_W T_{ij}^{\text{Work}} + k_S T_{ij}^{\text{Study}} \quad \forall i, j \] (Equation 5)

- \( T_{ij} \): Total number of work and study trips from \( i \) to \( j \)
- \( k_W \): Weighting coefficient for work trips
- \( k_S \): Weighting coefficient for study trips

Regarding the IIOM, the calibration is done with an integrated dataset of work and study trips in the statistical software, STATA®, via a nonlinear optimization procedure. Results which are presented in Table 3 show that all model parameters are statistically significant with acceptable confidence intervals.

### Table 3 - IIOM calibrated parameters

<table>
<thead>
<tr>
<th>IIOM Parameters</th>
<th>Work</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k )</td>
<td>0.573979</td>
<td>0.0933546</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>-0.4265714</td>
<td>1.067614</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0.4838109</td>
<td>-0.4852591</td>
</tr>
<tr>
<td>( \alpha_3 )</td>
<td>-0.075471</td>
<td>2.001871</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.1788247</td>
<td>-0.2829593</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-41.75296</td>
<td>88.87629</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>-0.0061696</td>
<td>-0.0721553</td>
</tr>
<tr>
<td>( P )</td>
<td>0.000151%</td>
<td>0.022300%</td>
</tr>
</tbody>
</table>

Observed values of \( T_{ij} \) versus the estimated values can be presented as follows.

\[ T_{ij} = 0.9992 T_{ij}^* \quad \text{with} \quad R^2 = 83.67\% \] (Equation 6)

- \( T_{ij} \): Observed number of work and study trips from \( i \) to \( j \)
- \( T_{ij}^* \): Estimated number of work and study trips from \( i \) to \( j \)

### Modeling performance analysis

After presenting the coefficient of determination \( (R^2) \) for the IIOM, here we analyze the performance of the model by means of several Goodness-of-Fit measures. First we compare the real and estimated cumulative values of \( T_{ij} \) versus trip duration. Afterwards, we discuss the performance of the IIOM in reproducing the observed values of trip production and attraction, followed by a detailed analysis of the IIOM with different Goodness-of-Fit measures.

**Cumulative number of trips versus trip duration**

Figure 3 shows the observed and estimated cumulative number of trips versus trip durations. The figure confirms that IIOM and OD values have very similar curves for trip durations that are less than about 60 minutes. For trip durations more than 60 minutes, the IIOM slightly overestimates the number of trips.
The evaluation of the estimated numbers for trip production and attraction is of great interest for understanding the appropriateness of a trip distribution model. In this study we calibrated the IIOM with single constraints on trip production values. These values are reported from the IIOM calibration:

\[ E_i = 0.9684E_i^* \quad \text{with} \quad R^2 = 94.23\% \quad \text{(Equation 7)} \]

\[ A_j = 0.9695A_j^* \quad \text{with} \quad R^2 = 92.52\% \quad \text{(Equation 8)} \]

\( E_i \): Observed value of trip production at \( i \)
\( E_i^* \): Estimated value of trip production at \( i \)
\( A_j \): Observed value of trip attraction at \( j \)
\( A_j^* \): Estimated value of trip attraction at \( j \)

Equations 7 and 8 confirm the great appropriateness of the IIOM for reproducing the trip production and attraction values.
Goodness-of-Fit measures
In this section we present several Goodness-of-Fit measures that compare entries in the observed and estimated matrices. In order to have an idea about the presented values for the IIOM, we present also the Goodness-of-Fit measures of the GM calibrated with the same dataset in our previous studies (13).

Mean trip duration error
Mean trip duration error is the difference between the mean duration estimated by the model and the mean trip duration based on the OD survey. In other words, trip duration for each OD is calculated from the GTFS data and then the error is calculated by means of observed and estimated $T_{ij}$. The error values reported in Table 4 show that the IIOM is more accurate than the GM.

Coefficient of determination ($R^2$)
Although some studies showed that in some cases the coefficient of determination or $R^2$ 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 (24, 26, 27, 33). The values of $R^2$ reported in Table 4 show that IIOM is more accurate than the GM.

PHI statistics (PHI)
PHI statistics (PHI) are presented as follows (27, 34):

$$PHI = \sum_{ij}(\max\{1, T_{ij}\} \ln \begin{vmatrix} \max\{T_{ij}\} \\ \max\{T_{ij}\} \end{vmatrix})$$  (Equation 9)

$T_{ij}$: Observed number of trips from $i$ to $j$

$T_{ij}^*$: Estimated number of trips from $i$ to $j$

Larger values of PHI present poorer model fits. The reported values in Table 4 show that the IIOM compared to the GM, yields smaller values for PHI statistics, which means it has a better performance.

Mean Absolute Error (MAE) and Normalized MAE (NMAE)
MAE and NMAE are defined as (27):

$$MAE = \frac{\sum_{ij}|T_{ij}-T_{ij}^*|}{N}$$  (Equation 10)

$$NMAE = \frac{MAE}{T/N}$$  (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

As for PHI, larger values of MAE and NMAE represent less accurate model fits. These measures also show that IIOM is more accurate than the GM.
Table 4 - Goodness-of-fit measures

<table>
<thead>
<tr>
<th>Performance analysis measures</th>
<th>IIOM</th>
<th>GM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean trip duration error</td>
<td>2.47%</td>
<td>4.15%</td>
</tr>
<tr>
<td>$T_{ij} = k \cdot T_{ij}^*$ (Value of $k$)</td>
<td>0.9992</td>
<td>1.0018</td>
</tr>
<tr>
<td>$R^2$</td>
<td>83.67%</td>
<td>78.73%</td>
</tr>
<tr>
<td>PHI</td>
<td>92,634.7</td>
<td>94,503.7</td>
</tr>
<tr>
<td>MAE</td>
<td>144.58</td>
<td>153.32</td>
</tr>
<tr>
<td>NMAE</td>
<td>0.448</td>
<td>0.474</td>
</tr>
<tr>
<td>DI (PME)</td>
<td>23.57</td>
<td>23.73</td>
</tr>
<tr>
<td>RMSE</td>
<td>242.9</td>
<td>250.2</td>
</tr>
<tr>
<td>Trip production measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E_i = k \cdot E_i^*$ (Value of $k$)</td>
<td>0.9684</td>
<td>0.8996</td>
</tr>
<tr>
<td>$R^2$</td>
<td>94.23%</td>
<td>99.67%</td>
</tr>
<tr>
<td>Trip attraction measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_j = k \cdot A_j^*$ (Value of $k$)</td>
<td>0.9596</td>
<td>1.0401</td>
</tr>
<tr>
<td>$R^2$</td>
<td>92.52%</td>
<td>93.43%</td>
</tr>
</tbody>
</table>

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 (28, 29):

$$DI = \frac{50}{T} \sum_{ij} |T_{ij} - T_{ij}^*|$$  \hspace{1cm} (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 4 shows that the IIOM behaves slightly better than the GM.

Root Mean Square Error (RMSE)

RMSE is defined as (24):

$$RMSE = \sqrt{\frac{\sum_{ij}(T_{ij} - T_{ij}^*)^2}{N}}$$  \hspace{1cm} (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 4 confirms that the IIOM behaves better than the GM, based on RMSE values.
Trip production and trip attraction measures

The coefficients of determination values presented for trip production and trip attraction in Table 4 show that the IIOM is more efficient for reproducing trip production values, whereas regarding trip attraction values, both models behave almost the same way.

Based on the presented Goodness-of-Fit measures, the IIOM is more accurate than the classical GM. Moreover, the integrated supply-dependent formulation developed in this study allows policymakers and analysts to study potential changes in PT trip pattern due to modifications in demography, economy, job and study spatial location and also the PT supply.

In the next section, we will study the limitations of the IIOM from a spatial point of view, in order to understand weaknesses and strengths of the IIOM.

Spatial limitations of the IIOM

In this section, we discuss the limitations of the IIOM from a spatial point of view. Figure 4 shows the spatial distribution of overestimation and underestimation residual errors for the IIOM, by means of desire lines plotted between MS for errors between 50% and 100%. The following findings are highlighted.

- The IIOM underestimates $T_{ij}$ more for OD pairs in which origin and destination are far away. We see also that most of this type of error occurs between non-central and central areas.

- Most of overestimation errors reported from the IIOM occur for close-distance OD pairs. We can explain the occurrence of these types of error in IIOM by the fact that the IOM family assigns trips to nearer potential opportunities. A possible solution to prevent this type of error might be the development of a more accurate method to define intervening opportunities for each trip purpose, as in this study we used only PT trip duration for calculating the number of intervening opportunities.

We performed also the same spatial residual error analysis for the GM calibrated in our previous studies (13), and found out random underestimation and overestimation errors, for both far and close-distance OD pairs.

CONCLUSIONS AND FUTURE RESEARCH

In this study we developed an integrated supply-dependent generation-distribution model for PT trips. This new model called IIOM is calibrated for study and work purpose trips, with single constraints on trip production values for PT morning peak hours trips in the Island of Montreal. The analysis of modeling performance by means of several Goodness-of-Fit measures showed that the IIOM is well-behaved, and more accurate than the classic GM which was presented in our previous studies (13).

We studied also limitations of the IIOM from a spatial point of view. In other words, by comparing observed and estimated trip numbers on an underestimation-overestimation basis, spatial weaknesses of the IIOM are known. This could help us to find improvement strategies in terms of explanatory variable choice or even model formulation modifications, which presents a part of our ongoing research.

The new formulation of the IIOM represents great potential for policy-makers or PT analysts, due to the sociodemographic, socioeconomic and PT supply variables:
Figure 4 - Spatial residual errors reported from the IIOM
As all the required data for calibrating the IIOM comes from external and independent sources, the IIOM could be used for the sake of PT trip generation and distribution forecasting for the future. More precisely, for calibrating the IIOM, we need reference OD matrices, number of potential work and study opportunities, OD trip durations, sociodemographic, socioeconomic, and PT supply characteristics at the MS level.

- The presence of sociodemographic, socioeconomic and especially PT supply characteristics in the IIOM 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 behaviors in a PT network (35). Regarding the PT supply characteristics, the developed IIOM could be used as a tool to forecast the PT trip pattern due to modifications in the PT network. As in the IIOM, two LOS variables of "total number of PT Passage-Stops per 24 hours at MS i" and "spatial density of PT Stops at MS i" are used, the analyst could verify the effect of different types of PT network modifications on the PT trip pattern.

Regarding the required trip duration values for calibrating trip distribution models, we suppose intuitively that IIOM compared to GM needs less accurate data. As a matter of fact in IIOM, trip duration is used only to rank zones according to distance, but in GM, trip duration values are used explicitly in the model formulation. 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 can 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 the smart card data. At the first level, we will define MS trip production \( E_j \) 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 IOM 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 to use the additional data derived from smart card validations in the GMA.

Another research topic is the development of the IIOM at a grid level. As the present study is done at the MS level, a grid study could present interesting results, in order to compare the sensitivity of PT trip distribution models to the study level.

We conclude that the use of the new developed IIOM is advantageous due to its behavioral and supply-dependent bases. This paper presents the first step in using the IIOM for PT trip analysis, planning and forecasting in the GMA, and evidently will be followed in the future by the development of more accurate models by means of continuous smart card data.

**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 that provided the Greater Montreal Origin-Destination survey and also the GTFS data. Special thanks to François Godefroy, Audrey Godin, Pierre-Léo Mongeon-Bourbonnais, Félix Pépin, 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


