DIRECT SCHEDULE-BASED ASSIGNMENT OF SMART-CARD TRIPS TO A GTFS TRANSIT NETWORK

Tim Spurr, Agence métropolitaine de transport, tspurr@amt.qc.ca

Robert Chapleau, École Polytechnique de Montréal

Guillaume Bisaillon, École Polytechnique de Montréal

ABSTRACT

The availability of very large datasets and the evolution of transport modelling tools present an opportunity to develop detailed and precise models of transportation systems. This paper describes a methodology to construct a model of high spatiotemporal resolution that could provide new insights into the operational elements of a public transit system and the behavioural patterns of its users. The Montreal subway system (68 stations on 4 lines) is used as a case study. Seven days of entry-only smart card validation data are used to construct a complete weekday containing 883,599 person-trips. Since only the entry station of a given trip is indicated by the fare validation, the exit station must be derived using trip chain logic. An algorithm for performing this derivation and for evaluating its degree of certainty is described in detail. Next, the paper discusses the optimal coding of a simple transit network for the purpose of schedule based assignment based on GTFS information. In this simulation network, the movements of individual vehicles and people can be explicitly represented. The derived person-trips are simulated using two independent methods: the TRANSIMS 4 platform and an SQL-based "direct" assignment algorithm. In both cases, the results of a schedule-based simulation allow for the construction of dynamic load profiles of platforms and trains. The procedures described in the paper offer a means of using advanced simulation methods to generate coherent results from large, detailed and non-synthetic datasets. The described methods have potential applications for operational and strategic planning.

Keywords: public transit simulation, smart card data, GTFS, TRANSIMS

INTRODUCTION

The increasing prevalence of large and detailed data sets offers the possibility of modelling public transit systems in new ways. With respect to travel demand, smart-card data provide precise information on temporal variations and can register a complete population of travellers. On the supply side, standardized schedule formats designed for the diffusion of traveller information present a detailed portrayal of transit supply. In addition to these large quantities of information describing supply and demand, advanced simulation tools such as activity-based or agent-based microsimulation can model transportation systems with a high degree of spatiotemporal precision while preserving each modelled object and its attributes throughout the simulation process. The application of these three elements to the analysis of public transit systems may allow for better-informed operational and strategic planning decisions taken to address issues of resource allocation, service quality and passenger experience. This paper aims to demonstrate the feasibility of such an approach using the Montreal subway system as case study. The emphasis is placed on an assignment algorithm within TRANSIMS but a parallel method, called “direct assignment” and developed to validate the implicit application of correct TRANSIMS parameters, is discussed as well. This second method uses a strict database processing program to assign individual trips to specific vehicle runs.

The paper has the following structure. The next section attempts to position the paper generally within the context of established practice and previous research. The third section describes the data used in the detailed simulation of the Montreal subway system. The fourth section explains in detail a multiple-step methodology for converting smart card transactions into origin-destination pairs (trips), for coding a subway network suitable for dynamic disaggregate assignment and, finally, for executing the assignment algorithm itself. The fifth section presents some examples of the types of results that can be produced using this approach. The final section offers some concluding comments.

PUBLIC TRANSIT MODELS, BIG DATA AND MICROSIMULATION

Public transit models: established practice and emerging issues

The approach to planning, design and evaluation of large urban transport systems in Montreal, Toronto and other Canadian cities is somewhat unique in that it has been based for many years on the collection of travel demand data using large-sample household travel surveys conducted according to the CATI (Computer-Assisted Telephone Interview) method. The resulting datasets have been used as input for traditional four-step planning procedures as well as the totally disaggregate assignment of person-trips to a public transit network (Chapleau, 1992). The distinguishing feature of the totally disaggregate method is that it preserves each modelled object and its attributes throughout the simulation process thereby permitting a multi-dimensional analysis of the simulated system. Historically, the method has depended upon the existence of a detailed travel survey containing information on a large number of travellers (the most recent household travel survey in Montreal, conducted in 2008, generated a database containing just over 350,000 observed trips). Despite their size and scope, it is nonetheless possible that these surveys may not be ideally-suited to certain specific analyses. For example, it is pertinent to wonder whether or not the survey adequately represents a "typical average weekday", sufficiently captures variability in travel behaviour, and can be reliably used for analyses outside of peak travel periods.

Such preoccupations may be of increasing importance in the near future as telephone-based surveys become harder to carry out and as public transit enjoys growing popularity in many urban areas. Problems related to crowding and capacity as well as the construction of detailed and precise indicators of punctuality and other measures of performance are starting to receive more attention from transit modellers, planners and operators (for example (Zorn, Sall et al. 2012) and (Eom, Choi et al. 2012)). The growing availability of smart-card transaction data as well as detailed schedule information in standardized format hold significant potential for the examination of such problems and, more generally, could provide additional insight into the experience of public transit passengers. While comparable analyses are feasible using large-sample survey data (Chapleau 2002), new sources of voluminous information – sometimes referred to generically as "big data" – offer the possibility to investigate the public transit passenger experience at a microscopic level.

Travel demand modelling using "big data"

While the ultimate impact of the big data paradigm on transport planning, operations and research is still unclear, technological advancements in the domain of data collection certainly have the potential to significantly alter the standard practice in these domains. The applications and research issues associated with evolving information technology are summarily documented by Wolfson and Xu (2010), who perceive these ongoing developments as an opportunity:

> "In the envisioned environment, billions of sensors embedded in the infrastructure, in portable devices, and in vehicles will generate vast amounts of data whose interpretation could be exploited to spur the creation of innovative transportation services and policies."

Indeed, the evolution of technologies associated with data collection, data storage and data analysis has made possible highly detailed and precise representations of travel behaviour. Large samples of travel demand data in the form of smart card transactions have been employed to measure numerous travel phenomena associated with almost the entire population of transit system users, in contrast to the limited samples that are captured by conventional travel surveys (Chapleau, Trépanier et al. 2008). Although smart card transactions provide only partial information on travel patterns, statistical techniques are being developed to construct reliable origin-destination matrices from this type of data (Trépanier, Tranchant et al. 2007; Zhao, Rahbee et al. 2007; Chu and Chapleau 2008; Farzin 2008; Munizaga and Palma 2012).

Another manifestation of the effect of information technologies on the operation of transport systems is the passenger's expectation of accurate and precise service information at any time anywhere. Transit operators have responded by developing various advanced traveler information systems (ATIS) applications, many of which rely on the General Transit Feed

Specification (GTFS) – a new standard for the design of planned service information system intended for public use (Google Developers 2012). The willingness of many operators to make these data public, often within the context of larger open-data initiatives, has led to detailed information on transit supply at the level of individual vehicle departures at each transit stop becoming widely available. GTFS has already been used as a basis for the coding of simulation transit networks, particularly when a schedule-based assignment procedure is adopted (Sokolov 2010; Scherr, Burton et al. 2011).

Microsimulation, activity-based models, and population synthesis

The progress of information technology has influenced the domain of travel behaviour modelling as well. As an emerging alternative to the classical four-step model of urban transportation planning, activity-based microsimulation (ABM) models aim to represent behavioural variability explicitly through the definition of numerous objects (people, households, activities, vehicles, signals etc.) programmed to interact with one another. Examples of activity-based simulation platforms include MATSim (http://www.matsim.org/) and TRANSIMS (http://code.google.com/p/transims/), both of which are open-source applications. These simulators are often integrated within a larger urban modeling framework that extends the activity-based approach beyond the transport system to encompass land use and regional economics. Example implementations include (Salvini and Miller 2003; Moeckel, Schwarze et al. 2007; Pendyala, Konduri et al. 2012). Because the raw input data are aggregate or represent a limited sample, population synthesis procedures are necessary to generate data for each modelled agent (Bhat and Guo 2007; Pritchard 2008; Müller and Axhausen 2011). The individual trips made by these agents are constructed using a daily activity synthesizer like DaySim (Bradley, Bowman et al. 2010) or OpenAmos (Pendyala, Konduri et al. 2012). Some notable applications of activity-based modelling of public transit systems include the analysis of a proposed light-rail project in Phoenix (Volosin, Paul et al. 2012), the simulation of public transit in the Greater Chicago area using GTFS (Sokolov 2010) and several MATSim-based deployments in various metropolitan areas (Erath, Fourie et al. 2012).

DATA

The methodology described in the next section is made possible by the existence of two large and detailed datasets representing supply and demand, respectively: first, the GTFS for the Montreal subway system operated by the Société de transport de Montréal (STM); and second, seven consecutive days of smart-card entry transactions for the same system. From a methodological perspective, the subway is a convenient test case because the locations of its entry points are fixed in space. Data of comparable reliability for the bus system are not yet available for Montreal. Although the Montreal subway network is not especially large, the dimensions of the data for the case study are nonetheless considerable, as shown in Table 1. Automatic fare validation data for a full week of subway operations are used to construct a complete dataset for a single day, namely Thursday October 21st 2010 during which 883,599 transactions were recorded. Validations for the following seven days are used to impute exit stations. The seven-day database contained roughly 4.8 million transactions. The Montreal subway network consists of four lines and 68 stations. The corresponding GTFS for a typical weekday contains 30,281 stop times and 1,581 train departures. The assignment of smart-card-derived trips to the GTFS network produces roughly 7 million trip segments, including transfer legs.

Table 1 – Dimensions of the data associated with the simulation of the Montreal subway

Data	**Dimensions**
Subway network	4 lines 66 km of in-service track 68 stations 146 platforms 470 turnstiles
GTFS – weekday	11 trip patterns 1,581 subway departures (line-direction) 32,500 platform stop times
Smart card validations	883,000 on 21 October 2010 4.8 million during the week of 21-28 Oct. 2010
Simulation	7 million trip segments (legs)

METHODOLOGY

The methodology described in this section is outlined in Figure 1. Summarily, the operation of the subway system is described by the two primary sources of input data: smart card transactions and the GTFS. The smart card transactions are converted into origin-destination trips using an algorithm based on trip-chain logic. Because the resulting trip table represents the behaviour of a complete population of traveling agents, the population synthesis procedure normally associated with ABM modelling can be by-passed. Meanwhile, the GTFS serves as a basis for the development of a complete schedule for the subway system. Geographic information systems (GIS) are used for network coding and the processing of simulation results. In order to determine passenger itineraries and estimate vehicle loads, an ABM model (in this case, TRANSIMS) is used to perform dynamic disaggregate trip assignment and microsimulation of an observed population of subway users. As a verification, a parallel “direct assignment” method is implemented in SQL to generate comparable disaggregate and dynamic results.

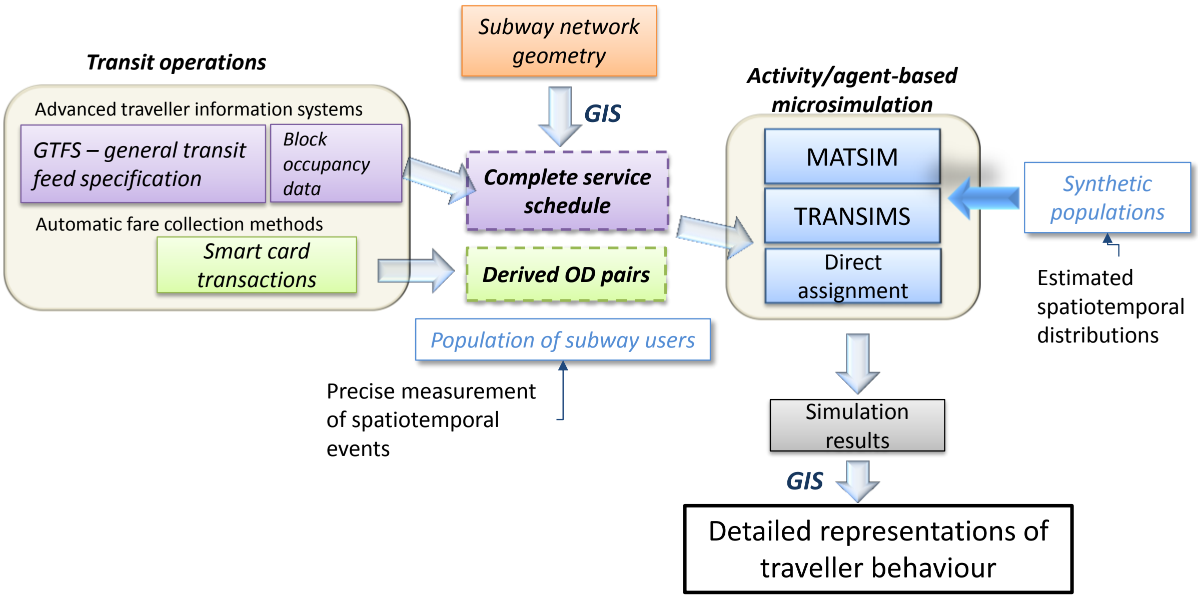

Figure 1 - A methodology for the microsimulation of smart-card derived subway trips

In Table 2, the methodology is separated into four major components: the definition of the transit network using geographic information systems (GIS); the conversion of entry-only smart card transactions into origin-destination pairs (trips); the construction of realistic subway schedules using GTFS and block occupancy data; and the trip assignment using TRANSIMS or SQL. Each component is associated with a specific methodological challenge. They are discussed in detail over the next four subsections.

Table 2 - Summary of the methodological challenges associated with the simulation process

Procedure	Model	Methodological challenge
Transit network definition	GIS	PLATFORM-level coding; network for representing the movements of vehicles and people
Travel demand	Smart card transactions	The imputation of trip destinations and classification of transactions according to destination certainty
Transit schedule construction	GTFS + subway block occupancy data	Determination of travel times and dwell times at high resolution (SECONDS)
Dynamic disaggregate trip assignment	TRANSIMS or SQL	Analysis of multiple objects: platforms, persons, vehicles, trips, legs

Transit network definition

The relative simplicity of Montreal subway system makes it an appropriate test case for the elaboration of network-coding methodology adapted to the assignment of smart-card derived trips. Figure 2 summarily illustrates the transformation of the basic network as typically represented to users to a simulation network constructed in a way such that all relevant system objects (lines, runs, platforms, station access and egress points, pedestrian tunnels etc.) are included. The network geometry is modified slightly to facilitate visualization of each object as well as the simulation results.

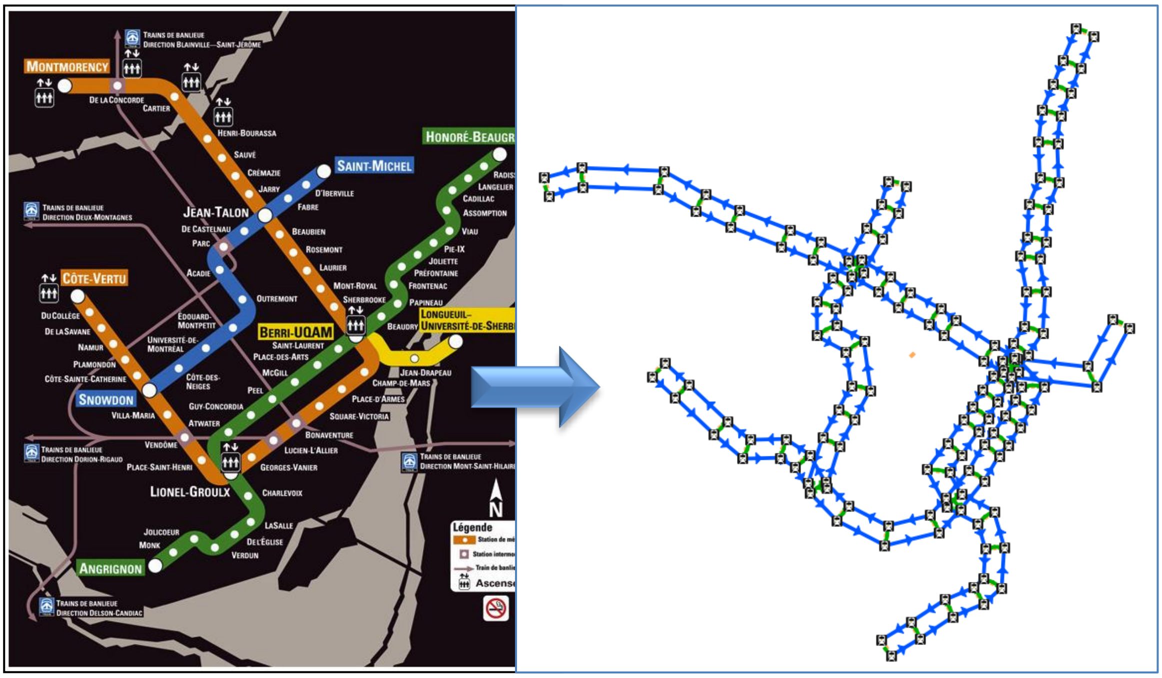

Figure 2 - The Montreal subway network as represented to the travelling public (left) and as coded for the assignment of smart card trips (right)

The methodological challenge presented by the definition of the transit network arises from the necessity of simulating the movements of vehicles and people simultaneously. Therefore, the coding of network links and nodes was performed following a strategic approach. The approach has two especially important features:

1. The coding of each station as a group of walkway-connected activity locations and transit stops representing, respectively, turnstiles where fare validations occur and particular station platforms;
2. A sequential numbering scheme for links, nodes and subway stops implicitly representing the eight line-directions of the subway network.

The result of this coding scheme is a hybrid transit-highway network. The presence of “local” connector links between non-intersecting “major” links is characteristic of networks used for automobile traffic assignment and these elements are used to model individual travellers when they move through the system on foot. The line and stop elements, however, are typical of transit assignment networks and serve to model individual transit vehicles. Each transit line is defined as an ordered sequence of stops (a trip pattern). Since the Montreal subway system consists of four bi-directional lines, the expected number of lines coded would be eight. The GTFS, however, reveals 11 distinct trip patterns including short-turns and movements to and from train storage locations for routine service-level adjustments.

Particular attention is paid to the numbering of links, nodes, stations and platforms, as shown in Figure 3. Each subway line is coded as a sequence of platforms identified by four digits.

The first digit is the line number, the next two digits comprise the station number and the last digit indicates the line direction. The station numbers are based on the planned stop sequence. As a result, the stop numbers for a given trip pattern are either ascending or descending. This numbering scheme is essential for the analysis of results and for the application of the SQL-based direct assignment algorithm.

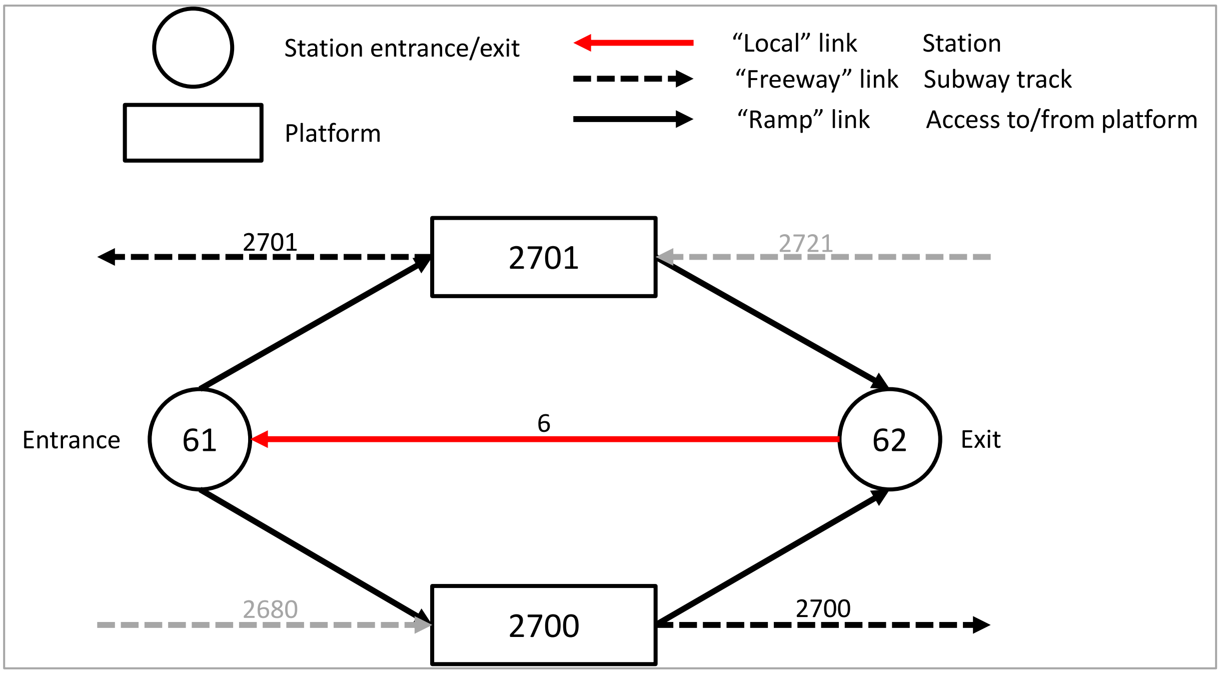

Figure 3 - Subway network objects and example numbering scheme for a single station

Conversion of Smart Card Transactions into O-D Trips

At this stage of the process, the methodological challenge associated with the representation of travel demand is to obtain origin-destination trips from smart card-transactions. To begin with, certain concepts must be clearly defined:

The **analysis day** is the 20-hour period from 5 a.m. on October 21st 2011 to 1 a.m. on the following day during which the Montreal subway system carries revenue passengers.

The **analysis week** is the seven days following the analysis day. The transactions recorded over the course of the week are used to complete the partial information available in the analysis day.

The term **smart card** in this paper refers to all fare media that can be used in the STM's automatic fare collection system. They include reloadable plastic cards containing chips or disposable, non-reloadable paper cards containing either chips or magnetized strips for short-term use. Note that it not possible to use the subway by paying cash. The purchased fare must be electronically loaded on one of these two media.

A **transaction** is the recorded validation of a fare on a specific card. The transaction has three essential attributes: first, the unique identifier of the fare medium (smart card or paper card) on which the fare information is stored; second, the time of day at which the validation occurred (precise to the nearest second); third, the unique identifier of the turnstile that performed the validation. This last attribute is used to determine the transaction station which serves as the trip origin.

A **trip** is defined as an entry station and an exit station visited sequentially by a single traveller. For the purposes of the present analysis, the entry station is considered the origin and exit station is the destination. Around half of all subway transactions were found to be transfers from other modes (bus or train). The true origins or destinations of these multimodal trips cannot be determined from the subway transaction data alone since users of the Montreal subway system validate their fare upon entry only. In the future, smart card transactions of travellers transferring to buses could assist in deriving a destination station. Currently, the surface network transaction data do not possess the necessary level of spatial precision.

The conversion of entry transactions into origin-destination trips requires an algorithm for imputing the most probable exit station given the available information. Initially, three generic transaction locations are defined:

1. The location (subway station) of the first transaction of the analysis day for a particular card *t* is called $s_{t,1}$.
2. The location of the last transaction during that day is called $s_{t,n}$.
3. The location of the first transaction following the transaction at $s_{t,n}$ is referred to as $s_{t,n+1}$.

Each derived trip is classified into one of seven types defined with reference to these three 1 locations. The following description of the seven trip types is summarized in Table 3

Type 1: Ok_loop

This type of trip is constructed from sequential transactions of card *t* at $s_{t,n}$ and $s_{t,n+1}$ where the location $s_{t,n+1}$ is the same as the location of the first transaction (i.e. $s_{t,n+1} = s_{t,1}$). An example of this type is a return-home trip at the end of the day.

Type 2: Ok_next

Type 2 encompasses sequential transactions of card *t* at $s_{t,n}$ and $s_{t,n+1}$ where the location of $s_{t,n+1}$ is not the same as the location of the first transaction and is not the same as the location of the last transaction (i.e. $s_{t,n+1} \neq s_{t,1}$ and $s_{t,n+1} \neq s_{t,n}$). A trip of this type would be made by someone who did not return home before the end of the analysis day.

Type 3: Ok_valid

These trips are defined by a sequence of two transactions which are neither the first nor the last transaction of the analysis day. The destinations of these intermediate trips can be inferred from the existence of subsequent transactions within the analysis day.

Type 4: Ok_firsttrip

This type refers to all trips from $s_{t,1}$ to a location confirmed by the existence of a subsequent transaction. This transaction type can be used to associate subways stations with types of activity by comparing its frequency with total number of transactions at each station. Stations where confirmed first transactions predominate are associated with residential activities, while stations where confirmed first transactions are rare are associated with work and study. In Figure 4, the former type of station is found toward the extremities of the network while the latter type is concentrated in the centre (downtown Montreal) as well at three stations serving the main campus of the University of Montréal.

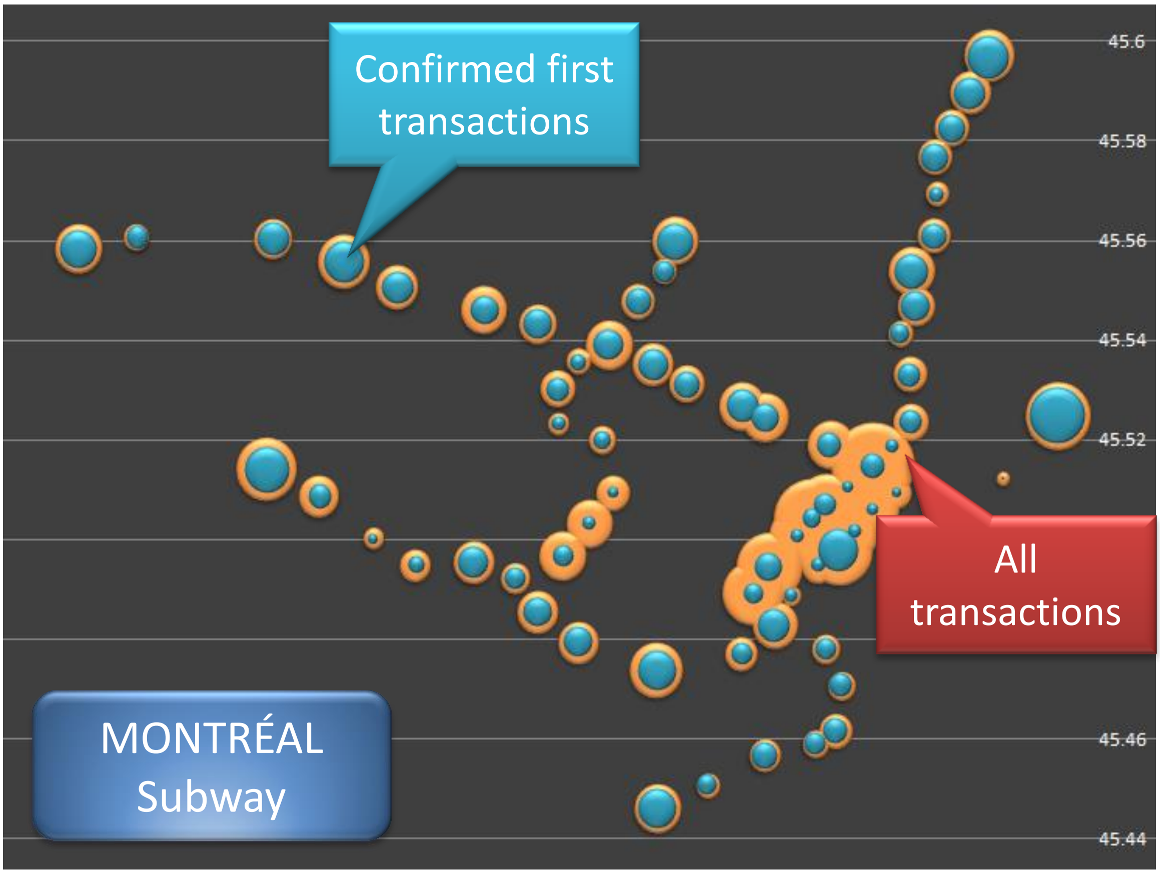

Figure 4 - Characterization of stations using confirmed trip chains

Type 5: Ok_prob, Type 6: Nb1_outbound and Type 7: Nb1_inbound

These types refer to smart cards having only one transaction within the analysis day and for which no subsequent transaction can be found at a location other than $s_{t,1}$. The underlying travel behaviour can be used to distinguish three distinct categories of traveller:

a) Travellers who rode the subway only once during the analysis week;
b) Travellers who rode the subway every day but only in one direction, using other modes of travel to complete their daily trip chain;
c) Travellers who rode the subway at least twice during the analysis day but whose movements cannot be traced because they use single-trip paper tickets instead of a smart card (multiple-trip paper tickets can be traced using the serial number of the paper card).

These categories of traveler all share an important characteristic: they are very unlikely to be using a monthly pass to access the system. In the first two cases, a monthly pass is not economical since the subway is used less than twice a day. Travellers in the third category cannot use a monthly pass because the pass can only be loaded onto a smart card. The destinations of type 6 and type 7 trips are therefore imputed through a process that matches outbound and inbound trips to form a complete round-trip. The matching procedure is partly justified by the symmetry of daily entry and exit volumes at subway stations which is apparent not only in the smart card database but also in the regional OD survey of 4% of households. The 2008 OD survey represents 807,737 subway trips (compared to the 883,599 subway trips represented by smart card transactions). Over an average 24-hour period, the station entry volumes display a very strong linear correlation with station exit volumes.

With this in mind, the first step in the matching process involves determining, for each transaction in the database not corresponding to any of the first four trip types, a most-probable destination given the origin, the transaction time and the confirmed destination distribution of travellers who use tickets rather than a monthly pass. This particular distribution is chosen because the travel behaviour of non-pass holders is expected to be different from that of pass-holders. Each trip made by a subway rider paying with individual tickets has a marginal cost equal to the cost of the ticket. Pass-holders, on the other hand, face no marginal cost per trip and are therefore more likely to make shorter intermediate trips. Overall, 130,708 individual trips representing 65,354 presumed round trips could be matched using this method. Within each pair, the trip with the earliest departure time is classified as Type 6 while the later trip is Type 7. The travellers making these trips fall into class c described above. Trips that cannot be paired are classified as Type 5 which represents the behaviour of travelers in categories a and b.

Table 3 - Trip typology based on smart card transactions

TYPE	CATEGORY	DESCRIPTION	TRIPS	%
7	Nb1_inbound	Transaction with no subsequent match – (probable inbound half of a round trip)	65,354	7.4%
6	Nb1_outbound	Transactions with no subsequent match (probable outbound half of a round trip)	65,354	7.4%
5	Ok_probable	Transaction with no subsequent match. Destination determined according to exit frequencies (random OD)	23,469	2.7%
4	Ok_firsttrip	First trip of the day with destination confirmed by at least one subsequent transaction (determined OD)	304,651	34.5%
3	Ok_valid	Intermediate trips with exit stations confirmed by the transaction chain (determined OD)	111,004	12.6%
2	Ok_next	Last transaction of the day where the first transaction of the next day is not in the original chain (determined OD)	85,801	9.7%
1	Ok_loop	Last transaction of the day where the first transaction of the next day occurs at the at the same station as the first transaction (determined OD)	227,966	25.8%
	TOTAL	**Transactions of the analysis day for which a destination can be determined.**	**883,599**	**100.0%**

For the entire analysis day, there were 154,177 transactions (17.5% of all transactions) without verifiable destinations. The destinations of these trips were imputed with a varying degree of certainty. Figure 5 displays the certainty of the trip imputation as a function of the system entry time. The destinations of trip types 1 to 4 are considered to be known "with certainty" since they all represent cases where a subsequent transaction exists at a location other than that of the first transaction. Trips without confirmed destinations are found in the three grey series. The series highlighted in red contains type 5 trips for which destinations are estimated using an observed spatiotemporal distribution of demand. The derived destinations of these trips are the most uncertain. The two other grey series are paired trips (types 6 and 7) which were derived using the same distribution as type 5 trips but constrained to the condition of symmetry between entry and exit volumes. The destinations of these trips are therefore estimated with greater confidence than the destinations of type 5 trips. Note that while the proportion of trips without verifiable destinations remains fairly constant (near 20%) throughout the analysis day, type 5 trips are almost non-existent before 9 a.m. meaning that the derived travel demand is represented with a maximal amount of certainty during the morning peak period.

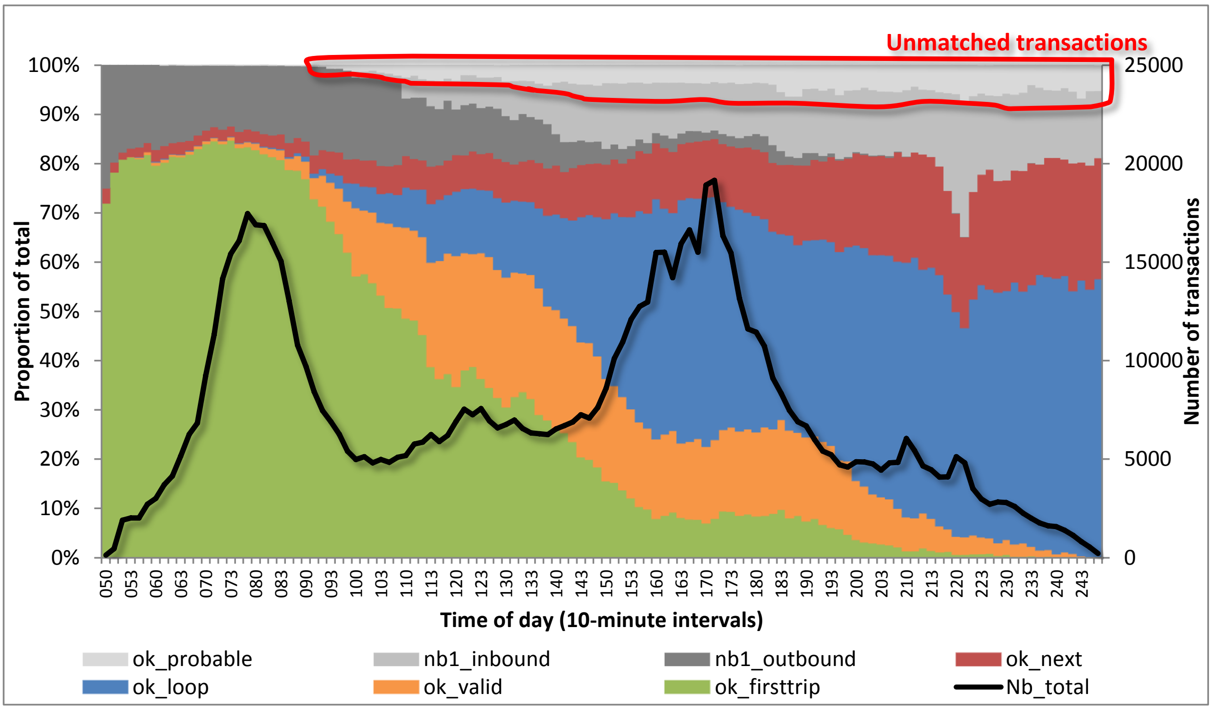

Figure 5 - Distribution of transaction times (entries) by trip type and time of day

Transit schedule construction

Parallel to the use of smart card transactions to represent travel demand, transit supply is represented using a GTFS stop time file for the Montreal subway system that indicates station arrival times for each planned subway run. It is far from certain, however, that the GTFS alone is suitable for the construction of a realistic representation of a transit schedule. Indeed, the Société de transport de Montréal (STM) who provides the data to the public through its website, says as much in the following disclaimer (http://www.stm.info/English/en-bref/a-developpeurs.htm):

> Metro schedules are provided by way of indication only and are essentially used to determine how long a trip will take. They cannot be used to develop an application for metro timetables, as they are only used to estimate the time needed to travel between two stations.

In order to construct a sufficiently accurate schedule, it is necessary to incorporate data that describes the movement of transit vehicles with greater precision. In the present case, subway block occupancy data for one a.m. peak period served this purpose. Using accurate interstation and platform length information in combination with the time and duration periods during which a train occupied each block, Blais (2007) generated average travel times and speeds between stations. Average station dwell times were estimated as well. These results are used to enrich the basic schedule information provided by the GTFS.

Dynamic disaggregate trip assignment

In this paper, the role of the network simulation is to enrich the existing data on supply and demand. The smart-card derived trip information includes the trip origin, destination and departure time but the itinerary and travel time of a given journey are unknown. The GTFS-based schedule describes the movements of in-service transit vehicles but not the number of passengers in each train. A schedule-based assignment of trips generates results that effectively represent the spatiotemporal variation of the transit system attributes while preserving each individual simulated object – a totally disaggregate approach.

Two methods are used to perform totally disaggregate schedule-based assignment. The first method, dubbed "direct assignment" involves an algorithm implemented using SQL. Because of the relative simplicity of the Montreal subway network, the number of plausible paths from one station to another is very limited. Consequently, the sequence used for each OD pair can be constructed before the execution of the assignment algorithm. For each trip, the assignment algorithm uses the predetermined stop, line and transfer sequence to find the subway train runs that minimize the wait and transfer times. The end result is a trip itinerary for each traveller composed of the specific subway train runs used to complete the journey.

The second solution is provided by the open-source TRANSIMS platform (http://code.google.com/p/transims/). The TRANSIMS modelling procedure involves a trip-by-trip schedule-based assignment followed by a microsimulation which considers both the individual train capacity as well as the time required for the loading and unloading of each train at each station. If the number of passengers waiting to board at a particular station exceeds the capacity of the train, the excess passengers remain on the platform and wait for the next train. The time a train spends in a station depends upon the boarding and alighting volumes. The microsimulation thereby includes certain operational characteristics that are not considered by the schedule-based path building algorithm. In addition, the microsimulation can produce second-by-second results that are easily adaptable to animated visualization tools.

In either case, the methodological challenge is to obtain wholly disaggregate simulation results, specifically output files that preserve all the attributes of the input objects, especially stations, vehicles and passengers. In general terms, this challenge is tackled by matching a complete description of a traveller's itinerary contained in a trip plan file with a complete description of vehicle movements. Spatiotemporal coincidences of vehicles and travellers represent boarding and alighting events occurring at specific locations (precise to the nearest meter) and at specific times (precise to the nearest second). A complete inventory of such events permits the generation of the results discussed in the next section.

RESULTS

This section of the paper demonstrates some of the results, potentially useful for strategic and operational planning, that can be obtained from the microsimulation based on large quantities of validated information in the form of smart-card derived trips and a modified

GTFS. The numerical results of a microsimulation can be compiled in multiple ways including animation of vehicle movements over the network (Spurr, Chapleau, & Piché, 2013), dynamic load profiles for specific trains and station platforms, a queuing analysis of platforms, an evaluation of individual passenger experience considering wait times and crowding and the measurement of congestion effects in a public transit network. Examples of a few of these results are presented below. While numerous experiments were conducted using a full-day simulation, for brevity's sake the results presented in this section are limited to a two hour simulation from 7:00 to 9:00. The simulation of the 157,665 trips corresponding to this time period generates a sufficient quantity of output for discussion purposes. Results for the time periods from 7:00 to 8:00 and after 9:00 are excluded since they represent transitory phenomena associated with the dynamic loading and unloading of the network.

The multidimensional structure of the results permits the construction of dynamic load profiles of station platforms and subway trains. In Figure 6, time is represented by five minute intervals between 8 a.m. and 9.am for a single line-direction (line 2 westbound). The matrix on the left contains the volume of passengers arriving on the platform on foot (either from the station entrance or from connecting subway lines) during each interval. The matrix on the right shows the number of people leaving each platform by subway. These representations of platform activity are indicative of the degree of crowding within stations. Platform 2601 is especially remarkable in this regard. In addition, the graphic provides some insight into the operational characteristics of individual stations along the line. A dramatic contrast is apparent between the stations upstream of 2601 where passenger boarding volumes are high and almost all of the subsequent stations where boarding volumes are often two orders of magnitude lower. Two exceptions, platforms 2521 and 2421, are at intermodal stations connecting the subway to regional bus service and commuter rail, respectively.

	PASSENGERS ARRIVING ON FOOT													PASSENGERS DEPARTING BY SUBWAY												
PLATFORM	**8:00**	**8:05**	**8:10**	**8:15**	**8:20**	**8:25**	**8:30**	**8:35**	**8:40**	**8:45**	**8:50**	**8:55**	**TOTAL**	**8:00**	**8:05**	**8:10**	**8:15**	**8:20**	**8:25**	**8:30**	**8:35**	**8:40**	**8:45**	**8:50**	**8:55**	**TOTAL**
2881	197	156	272	156	193	174	107	104	156	73	138	115	**1841**	136	248	268	130	283	161	0	161	173	0	159	0	**1719**
2861	68	61	136	178	58	45	45	53	84	95	41	27	**891**	27	103	65	190	84	39	52	55	74	0	153	0	**842**
2821	156	157	212	128	171	90	185	149	79	88	138	44	**1597**	329	41	254	194	97	213	174	0	190	117	0	201	**1810**
2801	280	305	206	288	191	318	183	236	205	175	138	150	**2675**	337	216	306	270	80	195	351	294	142	254	85	128	**2658**
2781	157	133	112	162	111	128	203	106	140	130	63	119	**1564**	221	162	118	119	120	118	94	219	153	97	152	83	**1656**
2761	124	130	141	120	98	85	123	89	88	81	70	60	**1209**	79	123	154	104	145	89	123	105	62	113	63	44	**1204**
2741	115	140	162	129	147	127	122	123	123	119	138	74	**1519**	141	140	155	126	130	93	125	166	130	74	194	64	**1538**
2721	262	249	321	413	326	355	325	403	317	311	248	264	**3794**	298	473	350	359	190	477	326	251	244	377	268	224	**3837**
2701	74	101	185	144	148	103	100	237	114	134	105	126	**1571**	62	86	136	207	82	139	123	201	173	90	142	95	**1536**
2681	108	68	71	137	25	132	100	64	90	84	47	74	**1000**	31	107	113	134	38	72	142	76	51	96	52	102	**1014**
2661	191	108	174	214	121	208	230	121	164	151	172	124	**1978**	160	96	190	203	137	197	200	170	177	108	162	106	**1906**
2641	111	105	120	139	152	113	127	134	175	130	124	122	**1552**	132	56	122	203	152	151	55	65	230	160	91	180	**1597**
2621	52	58	64	50	92	88	75	55	76	62	72	40	**784**	29	55	87	50	97	59	75	82	75	34	78	45	**766**
2601	597	547	608	520	520	555	680	501	412	470	514	451	**6375**	872	705	525	538	591	688	309	773	407	647	359	154	**6568**
2581	9	6	7	5	7	7	7	9	7	5	9	5	**83**	5	8	8	6	5	9	4	3	4	7	13	9	**81**
2561	13	9	7	5	7	8	5	5	11	8	5	12	**95**	13	10	6	6	6	9	5	4	8	11	7	5	**90**
2541	14	10	6	9	8	3	14	3	4	9	5	6	**91**	6	17	7	12	10	1	14	6	1	4	10	9	**97**
2521	86	102	71	54	67	58	71	60	75	43	49	109	**845**	85	92	108	37	64	62	75	60	66	63	57	50	**819**
2501	5	16	2	10	3	2	3	5	8	4	7	8	**73**	5	3	16	9	5	2	2	5	8	4	4	9	**72**
2481	31	38	29	27	15	9	6	9	11	17	9	4	**205**	32	40	33	22	23	16	4	8	10	13	15	6	**222**
2461	114	205	172	182	78	141	133	126	125	85	104	123	**1588**	194	194	112	269	105	149	15	181	62	62	196	122	**1661**
2441	19	34	19	11	17	20	10	16	13	16	21	24	**220**	11	29	36	12	14	16	15	12	13	19	23	19	**219**
2421	96	69	121	78	48	139	50	50	82	64	66	42	**905**	132	56	116	114	31	56	148	73	42	63	83	60	**974**
2381	69	45	49	41	30	48	22	36	30	15	19	13	**417**	27	88	16	50	50	51	8	33	27	35	7	24	**416**
2361	54	78	60	72	70	77	62	69	142	46	67	44	**841**	71	16	111	79	51	122	81	13	121	74	102	11	**852**
2341	15	32	34	34	17	7	7	13	18	13	8	10	**208**	7	23	30	44	28	3	7	8	21	15	12	8	**206**
2321	26	91	58	47	48	19	37	17	43	18	25	11	**440**	56	68	58	76	17	50	31	22	29	36	17	18	**478**
2301	7	3	2	2	5	0	0	6	3	2	0	2	**32**	15	2	3	3	3	2	0	4	4	1	2	2	**41**
2281	0	2	2	2	3	1	0	1	1	1	1	2	**16**	0	1	1	3	3	1	1	1	1	1	0	3	**16**
2241	11	10	11	2	7	3	9	10	1	3	5	4	**76**	8	18	11	4	7	2	6	11	4	4	3	5	**83**
2221	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	0	**0**
TOTAL	**3061**	**3068**	**3434**	**3359**	**2783**	**3063**	**3041**	**2810**	**2797**	**2452**	**2408**	**2209**	**34485**	**3521**	**3276**	**3515**	**3573**	**2648**	**3242**	**2565**	**3062**	**2702**	**2579**	**2509**	**1786**	**34978**

Figure 6 - Spatiotemporal distributions of demand on subway platforms

In, Figure 7 the rows of the table represent the platforms and the columns represent the transit service. The platforms are listed in scheduled order in the leftmost column. Rows with black boarders indicate platforms in stations where a transfer is possible with other subway lines. Each column represents a specific train running on Line 2 in the direction of the Côte-Vertu terminus (platform 2221). Every second train starts its journey at Montmorency (platform 2881) while the remaining trains are short-turns originating at Henri-Bourassa (platform 2801). A clearly visible effect of this operational regime is that the long-run trains are considerably more crowded than the short run trains. In addition, the capacity problems on the first half of the line are readily apparent. We have used 1,024 passengers as a plausible estimate of the number passengers that can fit inside a nine-car subway train. Several trains depart multiple consecutive stations at capacity indicating a high probability that some riders are left waiting on station platforms for the next train. Finally, the variability of riders' transferring behaviour is discernible through an examination of the platforms at transfer stations. Transfers at platform 2721 are primarily toward Line 2 since the load of all trains increases at this platform. Platform 2601 exhibits a different pattern: some trains experience a net gain in passengers while others experience a net loss. The reason is that 2601 (Berri-UQAM) is a transfer point with Line 1 and Line 4. The two other locations where a transfer is possible, 2461 and 2361, are downstream from downtown and consequently experience much lower passenger volumes during the a.m. peak period. It should be noted that the simulation does not consider passenger decisions regarding seating. At stations near the beginning of short runs (2801, 2781, 2681 for example), passengers may be disinclined to board the first train that arrives since the probability of finding a seat is much higher on short-run trains.

	RUN NUMBER													
PLATFORM	**21006**	**21007**	**21008**	**21009**	**21010**	**21011**	**21012**	**21013**	**21014**	**21015**	**21016**	**21017**	**21018**	**AVG**
2881	270		218		358		264		357		264		232	**280**
2861	351		310		418		335		716		341		300	**396**
2821	700		510		756		647		812		498		626	**650**
2801	805	141	672	167	797	171	762	186	941	137	562	131	810	**483**
2781	854	223	715	278	874	242	750	289	926	271	667	215	806	**547**
2761	847	309	734	332	878	303	716	334	907	332	719	285	818	**578**
2741	897	349	752	404	922	356	748	369	955	399	764	317	899	**625**
2721	963	365	825	562	905	579	761	475	1024	480	937	516	881	**713**
2701	977	433	879	624	978	632	790	548	1024	519	1010	553	952	**763**
2681	979	431	924	633	989	653	866	550	1024	590	1024	590	1015	**790**
2661	1020	502	984	656	995	704	865	549	1019	639	986	674	941	**810**
2641	1024	547	1024	688	1024	773	939	573	1024	656	1024	717	1024	**849**
2621	997	570	994	661	1009	752	911	572	1000	649	998	706	950	**828**
2601	960	797	1002	497	683	1024	865	775	894	535	1024	960	767	**829**
2581	913	769	970	462	660	962	806	711	818	518	986	920	733	**787**
2561	768	632	736	358	441	797	678	589	647	434	827	777	593	**637**
2541	559	444	551	247	309	584	472	399	419	316	589	566	382	**449**
2521	426	403	407	223	309	457	346	341	274	199	388	420	235	**341**
2501	388	375	367	211	284	413	322	313	253	187	363	382	212	**313**
2481	400	394	377	238	286	426	341	320	268	198	360	377	218	**323**
2461	477	399	360	387	446	409	387	467	323	247	312	357	328	**377**
2441	437	373	335	360	428	372	362	412	287	246	283	292	326	**347**
2421	533	418	342	379	432	395	380	406	269	253	266	369	321	**366**
2381	508	399	334	406	418	382	367	405	249	261	251	337	318	**357**
2361	295	198	185	248	207	274	156	244	242	136	109	284	148	**210**
2341	263	189	183	263	213	273	161	237	226	131	102	268	140	**204**
2321	235	187	214	290	229	274	158	241	201	128	107	251	134	**204**
2301	207	163	188	267	207	245	143	203	171	104	92	175	118	**176**
2281	189	150	166	237	185	220	124	179	154	85	77	148	99	**155**
2241	100	66	83	114	72	103	60	99	91	53	48	95	51	**80**
2221	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
AVG	**611**	**379**	**545**	**377**	**557**	**473**	**516**	**399**	**584**	**322**	**533**	**433**	**513**	**480**

Figure 7 - Subway train load profiles for thirteen consecutive departures on line 2 between 07:22 and 07:58 in the direction of peak travel

Figure 8 represents the same information as Figure 7 but displayed as a continuous three-dimensional surface. Figure 8 is an example of the type of results synthesis that can be obtained only when the data and simulation methods have a high spatiotemporal resolution. The figure highlights the dramatic differences in simulated passenger loads between consecutive trains and effectively illustrates two problematic elements of the present methodology: first, the unobserved choice made by riders to either board the first train that arrives or to wait for next one; and second, the behaviour of the simulation algorithm itself in which the choice of transit line depends on the planned (scheduled) service rather than the service that is actually offered. It is important to consider this feature of the modelling process when, as in the case of Line 2 of the Montreal subway, a single platform is served by regular and short-run trains which are coded as separate lines. While the schedule-based trip plan indicates that the traveller is to board a long-run train, congestion effects generated during microsimulation may cause the traveller to be unable to do so. The direct assignment algorithm avoids these problems, but does not account for capacity constraints.

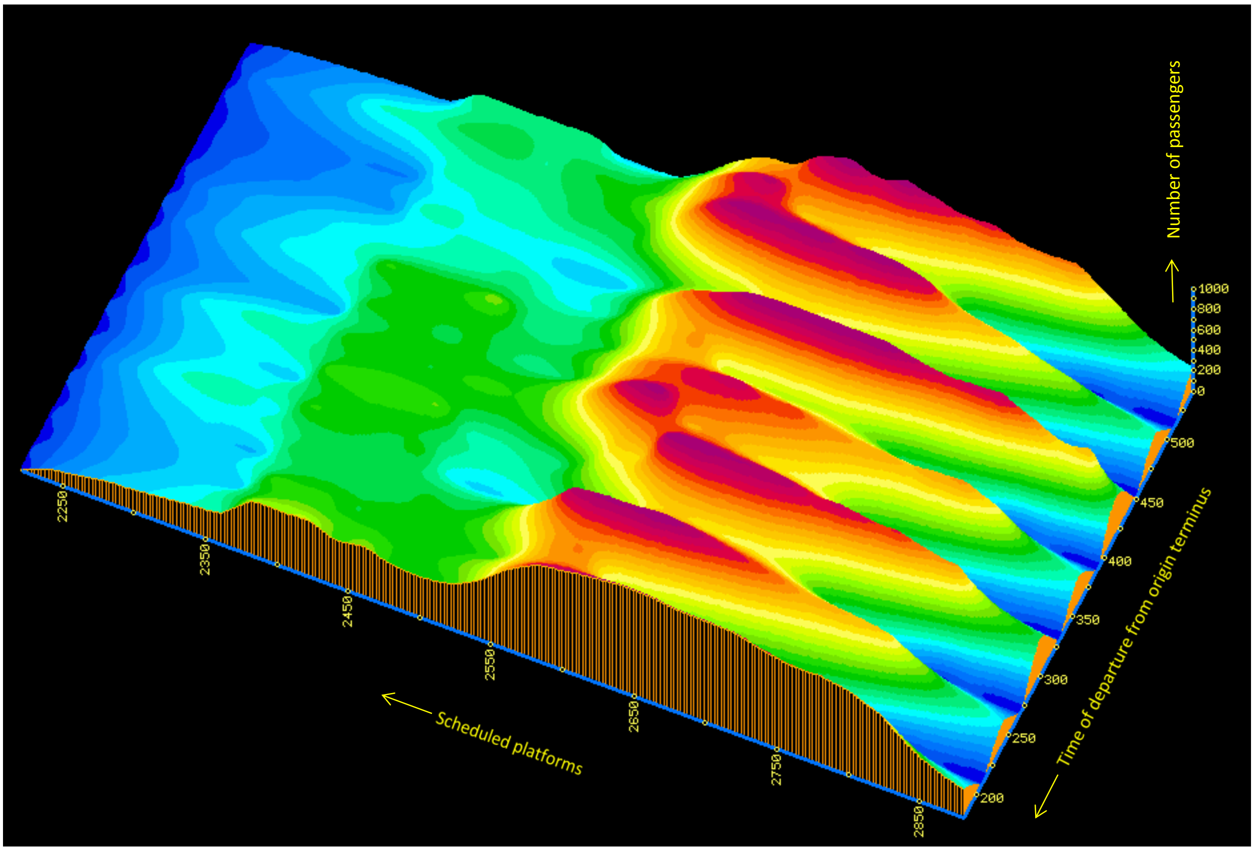

Figure 8 – Three-dimensional load profile of subway trains on line 2 in the Côte-Vertu direction

DISCUSSION AND CONCLUSION

The experiment described in this paper has demonstrated a methodology for incorporating large data sets into a totally disaggregate public transit modelling framework with a high spatiotemporal precision. The dynamic characteristics of public transportation services are represented using large quantities of schedule information. Meanwhile, a non-synthetic population of traveling agents is derived from smart card transaction data collected using an AFC system that is entry-only. These data afford an opportunity to improve our understanding of travel behaviour within a transit system, conditional on the ability to convert transactions into trips with a high degree of certainty. Finally, a totally disaggregate dynamic approach to network modelling, using either direct assignment or agent-based microsimulation methods, provides a high-resolution portrayal of a major transit infrastructure. For example, while the results of a conventional approach to the modelling of the Montreal subway network would be limited to load profiles on eight directional lines and 68 stations, a dynamic disaggregate assignment can estimate the specific demand for each of 1500 vehicles runs at 30,000 distinct stop times. This high-resolution approach to network modeling will facilitate future investigations into the evaluation of important elements of passenger experience such as crowding and transfer optimisation. Because it permits a precise estimation of vehicle, station and platform occupancy at any given moment in time, the method could also be applied to the development of evacuation plans and the development of emergency response strategies. Finally, with the application of a capacity constraint to each subway train, it will be possible to estimate the delay caused by congestion effects.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the collaboration of the Agence métropolitaine de transport and the Société de transport de Montréal as well as the technical assistance provided by Bruno Allard of Groupe MADITUC at the École Polytechnique de Montréal. The comments of several anonymous reviewers are greatly appreciated.

NOTE

The opinions, facts and comments in this paper are the sole responsibility of the authors and do not necessarily represent the views of the collaborating institutions. Numerical results are presented solely for the purpose of demonstrating the methodology.

REFERENCES

Bhat, C. R. and J. Y. Guo (2007). Population synthesis for microsimulating travel behavior. Transportation Research Record 2014: 92-101.

Blais, S. (2007). Modélisation espace-temps du métro de Montréal. M.Sc.A., École Polytechnique de Montréal.

Bradley, M., J. L. Bowman and B. Griesenbeck (2010). SACSIM: An applied activity-based model system with fine-level spatial and temporal resolution Journal of Choice Modelling 3(1): 5-31.

Chapleau, R. (1992). La modélisation de la demande de transport urbain avec une approche totalement désagrégée, *World Conference on Transportation Research Proceedings*, Lyon, volume II : 937-948.

Chapleau, R. (2002). A Method for Measuring Impacts on Customers of a Subway Breakdown. 4th Transportation Specialty Conference of the Candian Society for Civil Engineering. Montreal, Canada, Canadian Society for Civil Engineering.

Chapleau, R., M. Trépanier and K. K. A. Chu (2008). The ultimate survey for transit planning: Complete information with smart card data and GIS. 8th International Conference on Survey Methods in Transport: Harmonisation and Data Comparability. Annecy, France.

Chu, K. K. A. and R. Chapleau (2008). Enriching Archived Smart Card Transaction Data for Transit Demand Modeling. Transportation Research Record: Journal of the Transportation Research Board 2063: 63-72.

Eom, J. K., M. H. Choi and J. Lee (2012). Evaluation of Metro Service Quality using Transit Smart Card Data. 91st Annual Meeting of the Transportation Research Board. Washington, DC.

Erath, A., P. Fourie and M. van Eggermond (2012). Large-scale agent-based transport travel demand model for Singapore. 13th International Conference on Travel Behaviour Research. Toronto.

Farzin, J. M. (2008). Constructing an Automated Bus Origin-Destination Matrix Using Farecard and Global Positioning System Data in São Paulo, Brazil. Transportation Research Record: Journal of the Transportation Research Board 2072: 30-37.

Google Developers. (2012). "General transit feed specification reference." Retrieved April 1, 2012, from https://developers.google.com/transit/gtfs/reference.

Moeckel, R., B. Schwarze, K. Spiekermann and M. Wegener (2007). Simulating interactions between land use, transport and environment. World Conference on Transport Research. Berkeley, CA.

Müller, K. and K. W. Axhausen (2011). Population synthesis for microsimulation: State of the art. 88th Annual Meeting of the Transportation Research Board. Washington, DC.

Munizaga, M. and C. Palma (2012). Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. Transportation Research Part C: Emerging Technologies 24: 9-18.

Pendyala, R. M., K. C. Konduri, Y.-C. Chiu, M. Hickman, N. Hyunsoo, P. A. Waddell, L. Wang, D. You and B. Gardner (2012). An integrated land use - transport model system with dynamic time-dependent activity-travl microsimulation. 91st Annual Meeting of the Transportation Research Board. Washington, DC.

Pritchard, D. R. (2008). Synthesizing agents and relationships for land use / transportation modelling. MaSC, University of Toronto.

Salvini, P. and E. Miller (2003). ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems. 10th International Conference on Travel Behaviour Research. Lucerne, Switzerland.

Scherr, W., G. Burton and C. Puchalsky (2011). A Paradigm Shift in Travel Forecasting: Let Web 2.0 Feed the Network Model 90th Annual Meeting of the Transportation Research Board. Washington DC.

Sokolov, V. (2010). The Schedule-Based transit model of the Chicago Metropolitan Area. TRANSIMS: Applications and Development Workshop, TRACC-Argonne, Argonne National Laboratory.

Spurr, T., Chapleau, R., & Piché, D. (2013). Animation tools for the microsimulation of a public transport network. Presented at the 13th World Conference on Transport Research, Rio de Janeiro, Brazil.

Trépanier, M., N. Tranchant and R. Chapleau (2007). Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 11(7): 1-14.

Volosin, S. E., S. Paul, R. M. Pendyala, B. Grady and B. Gardner (2012). The application of microsimulation model system to the analysis of a light rail corridor: insights from a TRANSIMS deployment. 91st Annual Meeting of the Transportation Research Board. Washington DC.

Wolfson, O. and B. Xu (2010). Spatio-temporal databases in urban transportation. IEEE Data Engineering Bulletin 33(2): 18-25.

Zhao, J., A. Rahbee and N. H. Wilson (2007). Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems. Computer-Aided Civil and Infrastructure Engineering 22(5): 376-387.

Zorn, L., E. Sall and D. Wu (2012). Incorporating crowding into San Francisco activity-based travel model. 91st Annual Meeting of the Transportation Research Board. Washington DC.