MERGING AFC, APC, GPS AND GIS-T DATA TO GENERATE PRODUCTIVITY INDICATORS AND TRAVEL DEMAND MODELS IN PUBLIC TRANSIT

Robert CHAPLEAU (corresponding author)
Professor, Department of Civil, Geological and Mining Engineering
École Polytechnique de Montréal, Université de Montréal
P.O. Box 6079, Station Centre-ville, Montréal, Québec, Canada. H3C 3A7
Email: rchapleau@polymtl.ca
Telephone: +1-514-340-4711 ext 4809

Bruno ALLARD
Research Assistant, Department of Civil, Geological and Mining Engineering
École Polytechnique de Montréal, Université de Montréal
P.O. Box 6079, Station Centre-ville, Montréal, Québec, Canada. H3C 3A7
Email: ballard@polymtl.ca
Telephone: +1-514-340-4711 ext 4221

ABSTRACT

Public transport service planning has, for a long time, been determined by a limited amount of relevant data and travel demand knowledge. The cost of maintaining a large and consistent amount of onboard surveys has been prohibitive. Emerging technologies implementations such as APC (Automatic Passenger Counting), AFC (Automatic Fare Collection by Contactless Smart Cards), GPS (Global Positioning System) location at a regular 4-second intervals, and GIS-based operational databases are frequently observed. Very often, taken separately, each of the data sources lies. Authors regularly report data recovery rates as poor as 25 to 75%. This fact is also confirmed by our own experience. In that context, the actual research is an attempt to realise the convergence of the four tools in order to provide coherent, consistent, and complete ridership information merging for several corridors. The archived databases for the case of Montreal, where more than 200 vehicles are equipped with independent APC, GPS and Smart Cards systems, are scrutinized for the development of intelligent algorithms to deduct correct and complete information: imputation of GPS signal loss and spatial identification of the weak GPS signature, correction of double counting from the passenger sensors, imputation of the most probable alighting stop of the riders, etc. Combined visualization techniques (GIS maps, pseudo-geo-time-space diagram) and dataset processing contribute to an enriching procedure where all relevant attributes are determined to measure productivity factors and level of service. The demonstration of and experimentation with the processing of huge datasets are part of the exercise where some kind of data synthesis is sought for the future.

Keywords: AFC, APC, GPS, smart card, public transit, Montreal, archived data

INTRODUCTION

Seemingly, truth does not come from measure, but from redundancy. Or convergence. Or coincidence. In a large system, when instruments are used without paranoiac maintenance, strict measurement often comes with questionable ambiguity. Taken separately, actual technologies like AFC (Automatic Fare Collection), APC (Automated Passenger Counting), GPS (Global Positioning System) and GIS (Geographic Information System) are producing meaningless and huge amounts of data. Then, without desperation, in the transit industry, data are archived at an exponential rate while cheaper equipment becomes continuously available. In this context, this project examines the proper characteristics of each instrument and explores some empirical strategies for conducting practical data analyses which may benefit from the merging of multiple information technologies. While the main objective of this paper is to illustrate some of the difficulties encountered with the different data sources, the article is organized as follows: first, a short literature review on archived data potential analysis, followed by a practical and summary demonstration of data inconsistencies and consequent challenges, and then, presentation of concepts and according visualization of fundamental data statistics capable of characterizing transit service.

POTENTIAL OF AUTOMATED DATA COLLECTION SYSTEMS

Literature review

Recent literature has put a large emphasis on the potential of archived data processing. In particular, the book on *Schedule-based Modelling of Transportation Networks*, edited by Wilson and Nuzzolo, devotes core chapters on the subject. Wilson and al. (2009) presents the large spectrum of functions within a transit agency that could benefit from an effective use of automated data: operations planning, performance management, customer information and service planning. Few details are provided on the difficulties and issues emerging from the actual technologies and embedded in the associated data processing. Slavin and al. (2009) address, in the case of the AFC, the advent of copious quantities of data that should be processed individually to improve the understanding of route choice in large transit systems; the paper shows clearly the importance of peripheral data such as bus schedules, GIS transit networks, bus trip log information connected with boarding transactions.

Smart cards, in particular, have been considered in several studies for its potential in deriving travel demand origin-destination matrices. Utsonomiya and al. (2006) had the privilege of working on good-quality data and showed some interesting aspects of the potential for conducting a detailed analysis of fare card data. Working on a set of data covering a month of smart card transactions –necessary to derive a complete one-weekday-, Chu and al. (2007, 2008, 2009, 2010) have insisted, on a continuous basis, in the development of procedures to alleviate the problems related to incomplete, erroneous and missing data frequently found in the archived transit data; data correction and enrichment procedures were necessary to obtain a sufficient consistency when delivering disaggregate trip information (age category, travel purpose, origin, destination, boarding and alighting stops and times, and links with trip generators). Most of the time, as mentioned in Furth and al. (2006), automated data collection systems contain a very large portion of incorrect data, and it becomes essential to develop strategies to overcome these difficulties and to realize the full potential of the technology.

Availability of operational archived data as a modelling framework

The following research is based on some databases typical of a North American transit operator. In the actual case, a random sample gathered during one day (Wednesday, March18th, 2009) served as the kernel of the investigation of the scope for the determination of data processing duties of the various automated data collection systems. The data come from the STM (*Société de Transport de Montréal*), and are taken unfiltered at the beginning. For a specific weekday, the following operational data sources are:

- Automatic Passenger Counting (APC) system installed on more than 200 bus vehicles (about 15% of the bus fleet) randomly distributed on 187 day bus routes and 20 night routes; it consists of a combination of GPS-Odometer-DoorSensors to essentially measure 3 types of events (Stop, 4-second Cyclic, Passenger);
- GIS description of the bus route network: sequences of bus stops, with Lon-Lat coordinates and linked to a line-based transit network definition used for transit trip assignment;
- Planned Schedule of bus trips (route, direction, departure time) containing a subset of stops, called time-checkpoints associated to planned scheduled times; every trip (outbound or/and inbound) may have a specific itinerary;
- Matched APC data synthesis report, resulting from the data processing of the raw events in reference with the planned bus route definition.

The scope of the operational transit network comprehends, for an average weekday:

- 300,000 checkpoints with scheduled times for 1100 coded locations,
- 84,000 bus route departures (trips) on a line-based reference,
- about 16,000 line bus stops

For our research, the characterisation of the system activities is demonstrated with a sample bus, called here vehicle **BusV**. According to the diagram shown (figure 1), archived data from the APC system comprise the following variables – for a specific vehicle and a specific day:

- Cycle: about every 4-second interval, an event is created and provides these variables: **X**-Lon, **Y**-Lat, **T**-Time, **Od**-Odometer; for a typical bus operation with a spread of 12 hours, it means 10,000 records per bus….
- Stop: an event is triggered for any halt (recognized through a zero speed signal) and is providing: **X**-Lon, **Y**-Lat, **AT**-ArrivalTime, **DT**-DepartureTime, **Od**-Odometer, **DO**-DoorOpening; for one day of a bus operation, one observes around 600 stops including around 450 door openings…
- Passenger: for this event, variables are: **T**-Time, DoorNumber, **B1** or **B2** -Boarding, **A1** or **A2** -Alighting; regular 40-feet bus has a front and a rear doors, where boardings are supposed to be concentrated at the front to ensure proper fare collection; the logic imposes that this event comes at a stop and with the door open

event (on an average day, 600 riders will board a specific vehicle); some of the gathered data, less than 1%, seem to demonstrate something else.

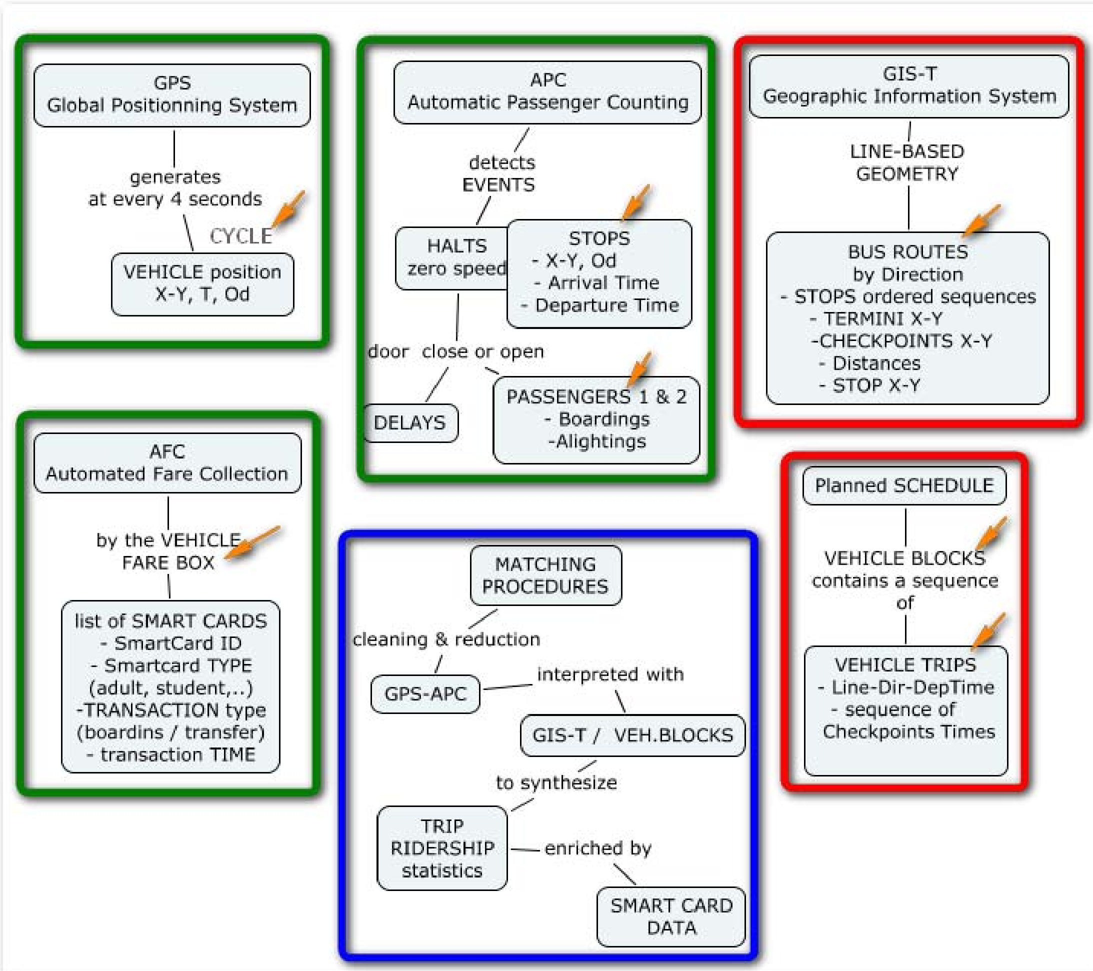


Figure 1Procedural diagram composed of archived raw data, reference planning data and data processing steps

The deployment of technology responds to usual statistical criteria. In Montreal, in complement to a 68-station subway system, the transit system is organized around 7 bus garages where, for a typical actual weekday, the following figures are encountered:

	ANJOU	ST-DENIS	LASALLE	MT-ROYAL	LEGENDRE	FRONTENAC	ST-LAURENT	total
Veh blocks	385	263	410	221	312	205	391	2187
Vehicles	**230**	**159**	**270**	**138**	**187**	**132**	**244**	**1355**
BL/VEH	1.67	1.65	1.52	1.60	1.67	1.55	1.60	1.61
GPS-BLOCKS	65	36	46	37	49	30	61	324
GPS-veh	**34**	**24**	**34**	**24**	**33**	**18**	**41**	**208**
% blocks	17%	14%	11%	17%	16%	15%	16%	15%

The temporal representativity of GPS-APC equipped vehicles is confirmed (around 14-17%) all the day long, as shown in figure 2.

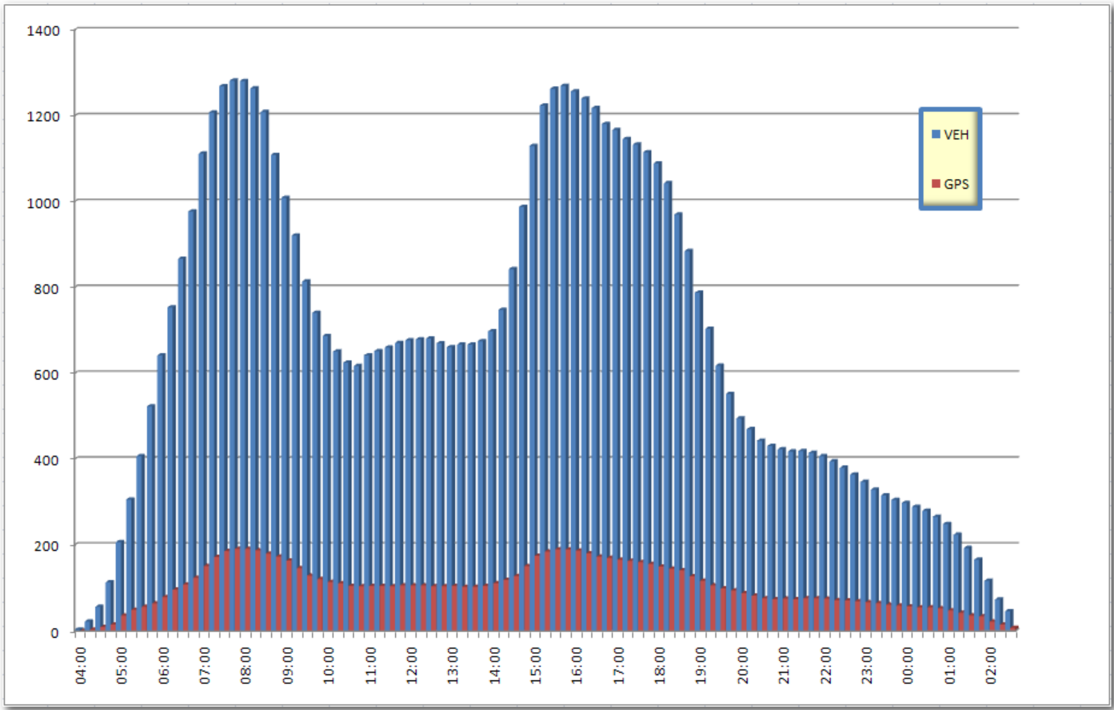


Figure 2Number of vehicles in service per 15-minute period and GPS-APC equipped vehicles

Data problems encountered

As mentioned before, archived GPS data represent a serious data processing challenge. For the Montreal transit operator, the order of magnitude of the files to be processed everyday is the following (figures are given for the chosen weekday):

- **CYCLE file**: 2,123,409 records, from which 15% do not have X-Y GPS signal, and (may include X-Y=0 records) 45% are related to a non-moving vehicle (no change for the odometer, at 1-meter level of resolution).This latter fact leads to a primary elimination of intermediate non-moving records. From then, we keep zero Lon-Lat (X-Y=0) records and make a first analysis to characterize what may be called a “GPS signal loss map” signature, as illustrated in figure 2. GPS problems are twofold: signal loss and doubtful X-Y. Signal losses are shown (in green) from start–to-the-end with a width proportional to the number of X-Y=0 records in between, while circles (in red) represents distance from consecutive Lon-Lat coordinates without correspondence with odometer differentials. The latter occurs primarily in the central business district where high-rise buildings generate diverse signal reflections. Incidentally, the garages are well identified by the green traces when the vehicles are making their pull-out.

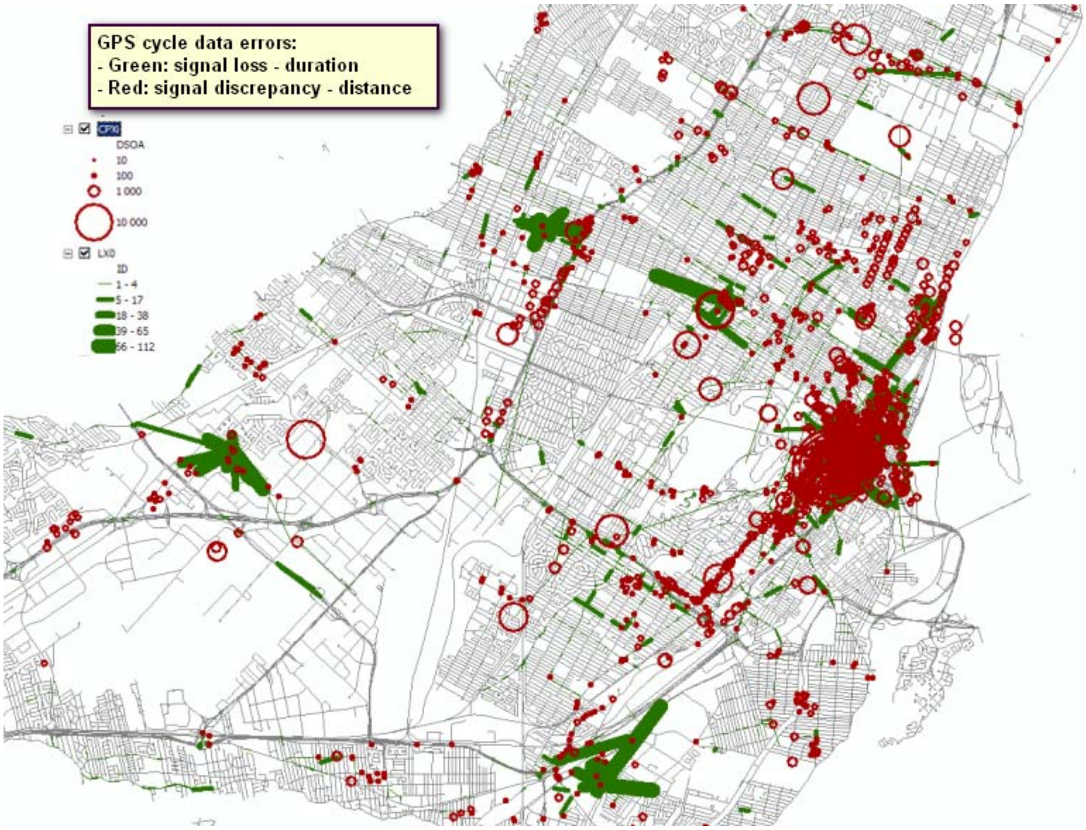


Figure 3 GPS cycle data errors: a kind of terrain signature

Figure 2 -GPS cycle data errors

- **STOP file:**111,594 records, where 60% of the halts are measured when the door is open for passengers' transactions. According to these figures, because of the impossibility of determining when the vehicle is or is not in service, the estimation of 84 seconds per stop with passengers against 32 seconds per halt due to congestion or layover and time recovery is irrelevant. There is a need to decompose the vehicle activities before stating significant estimates. Figure 4 shows the spatial distribution of the vehicle halts, as issued from the stop file for the examined day.
- **PASSENGER file**: 132,153 records, where, at the exception of ineffective sensors, passengers are detected at the two bus doors:
    - **Front door**: 141,701 boardings and 70,577 alightings
    - **Rear door:** 1,311 boardings and 71,454 alightings
    - **About 1,000 riders discrepancy** (less than 1%)… more boardings than alightings…

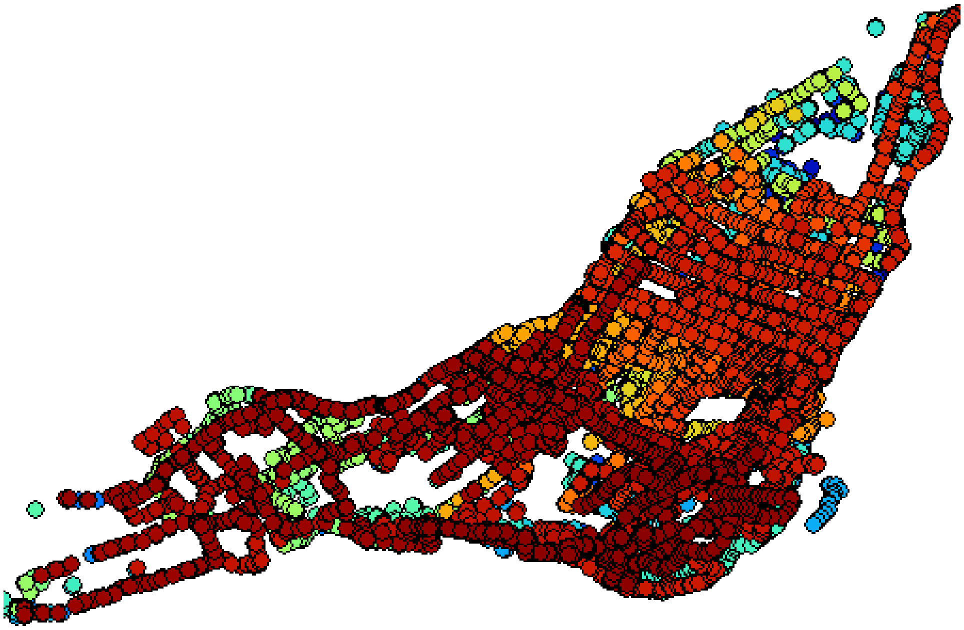


Figure 4 Representation of stops sampled by the GPS-APC equipped vehicles on March 18, 2009

In summary, raw data processing is cumbersome and is not providing relevant statistics. The GPS APC archived data files need to be filtered and interpreted by an intelligent algorithm, much as must be done for efficient map-matching. See Marchal and al. (2004).

Results from matched data with a GIS-based transit network

The best GPS-APC data processing strategy requires a spatial and temporal service planning schedule to make the extraction of relevant data suitable to the calculation of performance indicators. Passenger counting should also be corrected, because of several shifts in the boarding and alighting logics. At a minimum, it is required to adjust the data for each vehicle trip, when no passenger is admitted to go through the terminal.

When processed with adequate software, the resulting data follow the stated geometry of the bus route (ordered sequence of stops in a line-based definition). The overall vehicle trajectory (corresponding to a BLOCK) is then cut into several separate vehicle trips (line-direction with a scheduled departure time) in-between pull-out, interline and lay-up sequences.

For a typical weekday for the Montreal case, the matched service with the GPS-APC data provides the following numbers:

- a fleet of 227 vehicles, 10% in maintenance, and another about 10% with unmatchable raw data, to end with plausible statistics from 175 vehicles;
- the surveyed vehicles, after the application of a systematic matching of bus stops encountered and cycle-stop-passenger events, have produced these statistics (Table 1);

Table 1 – Indicators of operational service for 175 matched vehicle from GPS-APC data

	TRIPS	DISTANCE (km)	Duration (hr)	Length &Speed @(km/h)
PRODUCTIVE TRIPS	2069 73%	22678 84%	1414 89%	11.0 km @ 18.3
INTERLINING TRIPS	775	4414	175	5.7 km @ 25.2

- for the data coming from the sole APC (Table 2), after necessary adjustments for balancing the numbers of boardings and alightings on each vehicle trip:

Table 2 – Indicators of passenger counting for the 175 matched GPS-APC equipped vehicles

	STOPS	Boardings front& rear	Alightings front	Alightings rear
Scheduled Stops encountered by 175 buses	82693 Excluding Pull-in/out	112700 Average 1.36 / stop	55630 0.67/stop	57170 0.69/stop
Stats / vehicle for 175 buses	46020 with pass.	2.45 /stop	1.21/stop	1.24/stop

Transit objects as a modelling framework

GPS, APC and AFC data are intimately linked to the tasks and activities of a vehicle. When considering the paradigm of the travel survey, Chu and al. (2009) use the term “DABI survey” for Driver-Assisted Bus Interview travel survey as a very efficient method of developing substantial Origin and Destination person trip matrices. With this experience in mind, it appears that the GPS-APC data processing requires a conceptual framework tactically adapted to the measurement of relevant performance indicators. For the specific needs of operations planning, the level of resolution that should be defined in a spatial and temporal environment consists of the following objects associated to a bus regular route:

NODES:

- **TERMINUS**: for each terminus, two virtual points are to be defined to delineate between two events: the riders’ unloading at the terminus that will be followed by a layover time, and the repositioning of the vehicle at the first boarding terminus of the next TRIP, event that will be followed by a first boarding before the planned departure time;
- **Time checkpoints**: subset of stops for which time check-points are determined by the service planning department; bus drivers are supposed to be timely checked at these locations for some punctuality requirements (say, 1-minute early/ 3-minute delay interval window)
- **Regular stops**: bus stops, where riders are boarding and/or alighting; scheduled times are interpolated from time points;
- (**NETWORK nodes**): for simulation purposes and rider trip assignment, nodes are defined to represent boarding/alighting and transfer locations; it normally integrates a series of events? near bus stops or stations;

LINKS:

- **Inter-stops**: links defined to connect consecutive bus stop; length and travel time are attributes
- **SEGMENTS** (inter-checkpoints): level of aggregation of consecutive inter-stop links for which a specific travel time is attributed according to traffic conditions and time of day, reflecting the service planning target;

BLOCKS:

- **TRIPS** :a **vehicle block** is formed of a combination of sequential **trips** (commercial fixed-route trajectories between termini and going though tagged checkpoints) and interlines (unproductive travel from and to depot, and between trips)

TIME:

- Arrival Time
- Departure Time
- Transaction Time (boarding/alighting)
- Trip Travel time (productive: commercial)
- Block Travel Time (productive, effective: sum of productive and unproductive – layover, recovery and deadhead)

DISTANCE:

- Trip length
- Interline length
- Block length

With these definitions in context, one may look at the integration of GPS and APC available data for a specific case, without loss of generality for the potential of application to large and systematic databases. Figure5 illustrates the way diverse calculations are applied to the available data. Figure 6 shows the corresponding space-time diagram for a sequence of outbound and inbound trips. The focus is put on the virtual termini, where final alighting activity, vehicle repositioning and initial boarding before departure activity can be differentiated and implicitly measured. In the same spirit, the time checkpoints constitute events where a driver may adjust his pace to the planned schedule, while the segments are the control determinants of the total travel time.

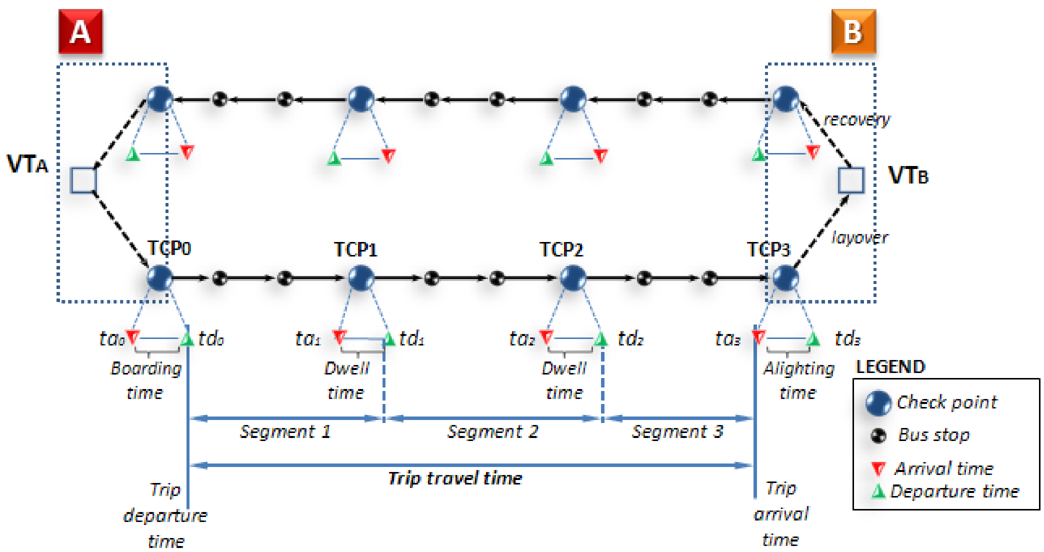


Figure 5Bus route objects: termini, stops, checkpoints, travel time

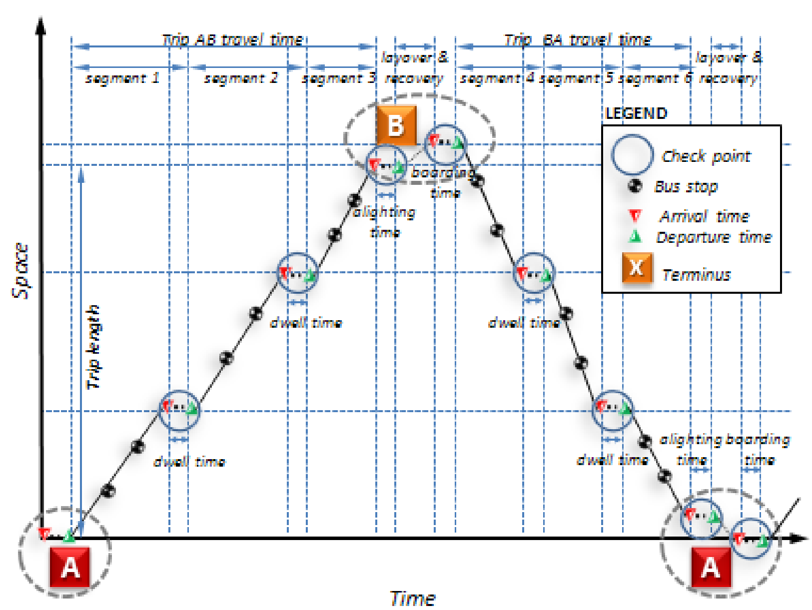


Figure 6 Space-Time diagram suitable for performance measurement

To make these concepts more explicit, the next paragraphs tell the detailed story of a vehicle equipped with GPS, APC and AFC.

A DAY IN THE LIFE OF BUS 22-369

The demonstration of analytical concepts is easier when applied to a specific example. Almost chosen at random (too complex cases have been discarded), the bus 22-369 (a 7-year old bus of a subset of 227 GPS/APC-equipped vehicles, part of a total bus fleet of 1500 vehicles) on

Wednesday March 18, 2009, has been selected for a detailed analysis. As shown in the following figure, it operates East-West in the centre of the Montreal Island and the main Route 161 is connected to 3 subway stations (figure 7).

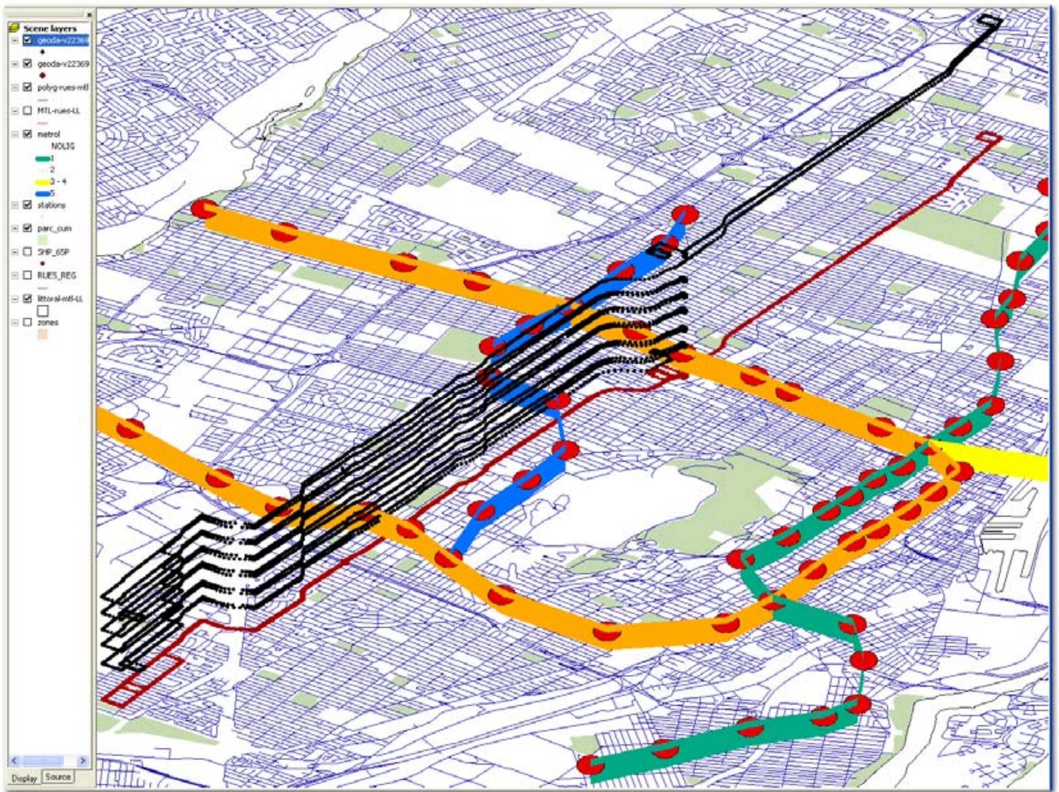


Figure 7 Aquarium of the vehicle block driven by the bus 22-369, a certain Wednesday: the Red trajectory is the projection of the regular Line 161, and the school extra Line 197.

As defined by the transit operator, the actual vehicle block is composed of 15 legs. Vehicle 22-369 is a member of the fleet of the St-Denis depot whose operations are illustrated in the figure 8. The next figure 9 shows the entire planned schedule of vehicle 22-369, and actual observed departure and arrival times for each leg, as measured by GPS and matched by the related software. It appears that on-time performance was not perfect, and several calculations have been made to see the extent of the non-schedule adherence.

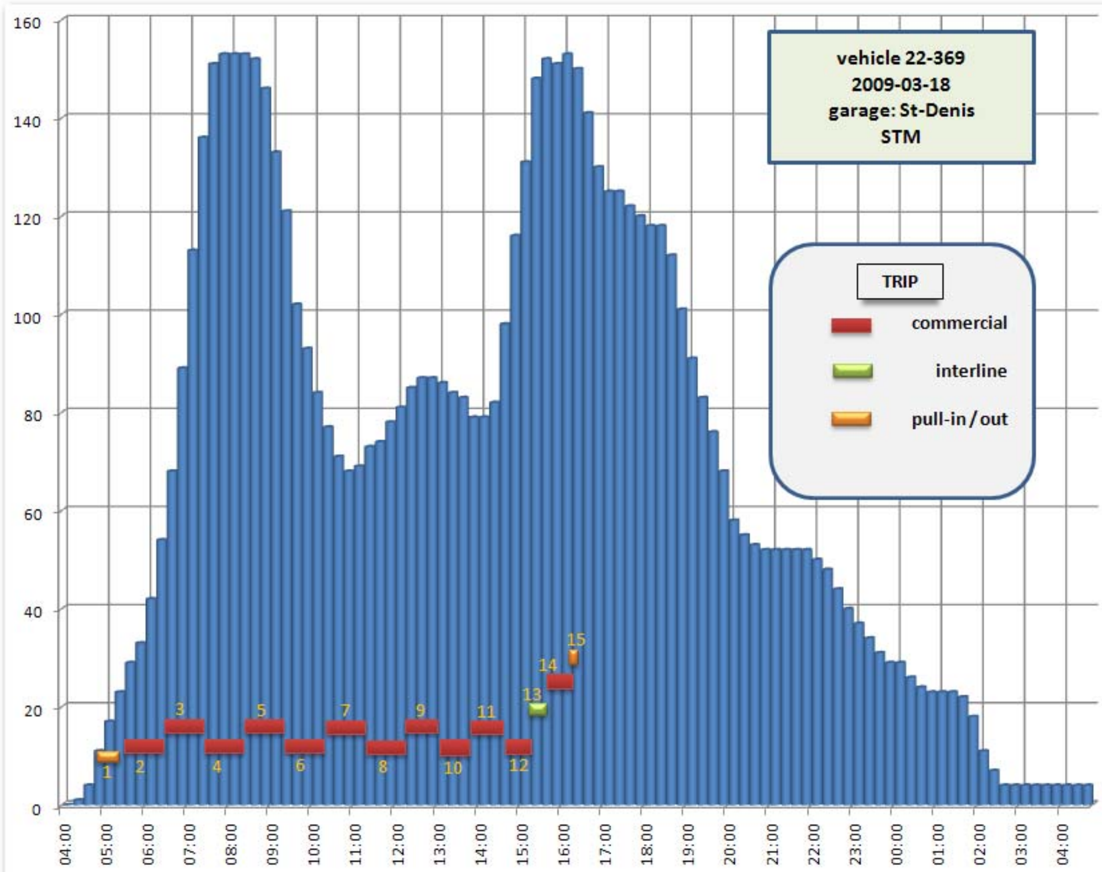


Figure 8 Vehicles in service from depot St-Denis, and vehicle block (15 legs) of the bus 22-369

TRIP	DI	G140	G050	M232	G241	M540	M26801	StartTime	distance(m)	ArrivalTime	CommSpeed
PULL-OUT 1	**from the DEPOT to the East terminus:**							4:57	10435	5:21	26.0
2	E	**5:26**	5:30	5:40	5:46	5:54	6:02	5:26	10472	5:59	18.8
3	W	6:59	6:47	6:36	6:27	6:22	**6:15**	6:15	12286	6:55	18.4
4	E	**7:09**	7:14	7:28	7:35	7:43	7:55	7:09	10465	7:51	14.7
5	W	8:52	8:40	8:28	8:19	8:12	**8:03**	8:03	12268	8:58	13.2
6	E	**9:11**	9:17	9:30	9:37	9:45	9:54	9:11	10452	9:52	15.1
7	W	10:49	10:37	10:26	10:18	10:10	**10:02**	10:03	12277	10:41	19.2
8	E	**10:54**	10:59	11:12	11:19	11:27	11:36	10:54	10479	11:32	16.4
9	W	12:31	12:19	12:08	12:00	11:52	**11:44**	11:43	12270	12:30	15.8
10	E	**12:39**	12:44	12:57	13:04	13:12	13:21	12:38	10481	13:22	14.6
11	W	14:18	14:06	13:54	13:44	13:36	**13:28**	13:28	12279	14:13	16.1
12	E	**14:25**	14:30	14:44	14:52	15:00	15:13	14:24	10414	15:06	15.1
Interline 13	**from the line 161->197**							15:06	6811	15:30	17.0
14	W	**school EXTRA service**						15:43	7417	16:13	15.1
pull-in 15	**to the DEPOT**							16:13	5	16:13	10.4

Figure 9 Vehicle block schedule at the time checkpoints and, and at right, observed departure and arrival times at termini for the vehicle 22-369.

Raw data exploration

When the first data are extracted for the chosen vehicle, the passenger counting file provides an inconsistent summary of riders, as shown in figure 10. Constraining data to stay non negative, and balancing boardings and alightings on each trip, are imperative tasks to be performed prior to any analysis.

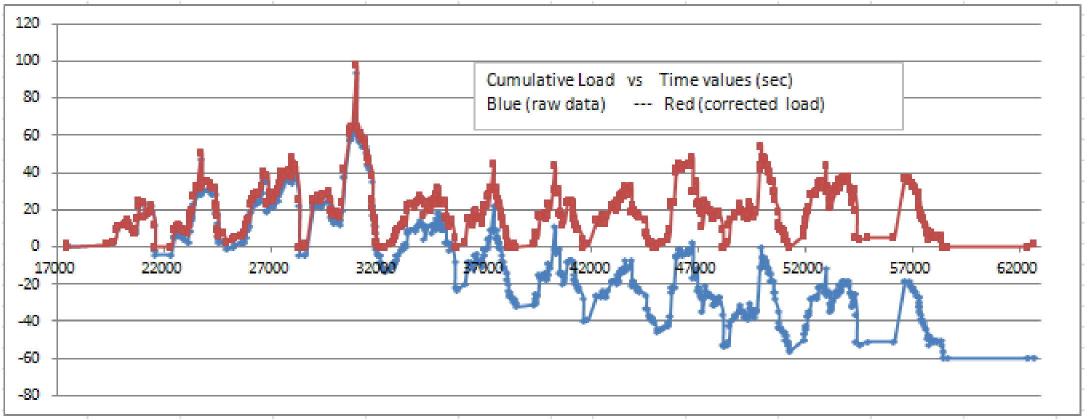


Figure 10 Cumulative load as reaveled by the raw data for the vehicle 22-369

By the same token, the GPS data exhibit a significant number of missing data which could be frequently interpolated with the help of odometer values and known geo-referenced locations. The figure 11 illustrates, with red points in a time-space diagram, the moments where the GPS signal has been lost for the vehicle 22-369, even if the trajectory is almost free of known obstacles.

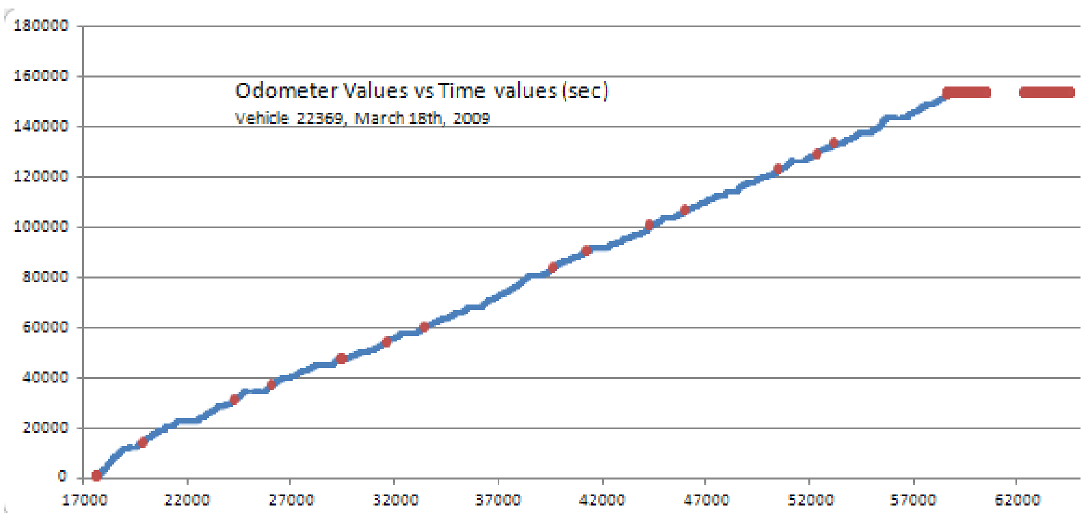


Figure 11 Time-space diagram (odometer vs. time (sec) with GPS signal losses (in red)

Productivity indicators with filtered and matched data

The planned service and the GIS-T database provide the referential framework to filter the raw data and to interpret some coherent spatial and temporal events such as departure and arrival times at passenger stops. Planned time checkpoints are targets for the bus driver and become the reference points for deriving punctuality indicators and variability estimates for travel time. The following paragraphs demonstrate some calculations conducted with the vehicle 22-369 GPS-APC data in combination with the programmed schedule and GIS network information.

		Checkpoints							
Direction	**Scheduled Departure**	**G140**	**G050**	**M232**	**G241**	**M540**	**M26801**	**Next depart**	**Total pass.**
		A					B		
Eastbound	526	0	-1	-1	0	1	3	615	52
	709	0	0	0	-2	-3	3	803	119
	911	0	1	0	-3	-2	1	1002	106
	1054	0	0	1	1	0	3	1144	106
	1239	0	-2	-2	-2	-3	-1	1328	120
	1425	0	-1	0	0	0	7		114
	Mean	0	-1	0	-1	-1	3		617
		A					B		
Westbound	615	4	0	0	-1	0	0	709	76
	803	-7	-9	-9	-5	-4	0	911	156
	1002	8	3	2	3	1	-1	1054	94
	1144	1	-3	-3	-2	-1	0	1239	91
	1328	4	0	0	0	1	0	1425	105
	Mean	2	-2	-2	-1	-1	0		522

Figure 12 Schedule adherence: time difference (delays) with planned departure times at checkpoints

From the aforementioned GPS-APC derived data, delays are calculated at each checkpoint (figure 12) and punctuality indicators are consequently derived, as illustrated in figure 13. In the case chosen – on-time interval of -3 and +1 minute-, the results are weighted by the number of passengers according to three cases: average load of the segments (inter-checkpoints), riders alighting along the segments, riders boarding within the segments. Interestingly, these measures take into account the situations where riders arrive earlier (alightings in advance).

Delay (minutes)	Mean load		Aligthings		Boardings	
real - schedule	East	West	East	West	East	West
<-3	0%	27%	0%	32%	0%	24%
[-3;1]	100%	61%	82%	52%	100%	66%
>1	0%	12%	18%	15%	0%	10%
Mean delay	-0.8	-1.8	0.1	-2.1	-0.6	-1.5

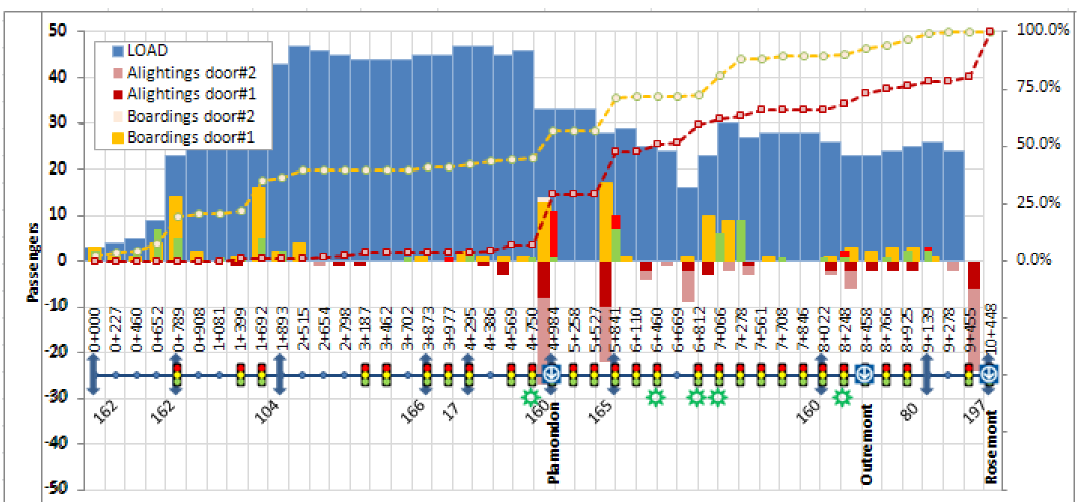
Figure 13 Weighted on-time performance of a subset of trips on the bus route 161

DERIVED OPERATIONAL STATISTICS

Trip load profile from the matched data

The ultimate goal of gathering and processing such large databases is to concretize summary statistics on the service productivity. The following figure 14 (load profile of a vehicle trip) is built around all the available attributes coming from the combination of GPS, APC, GIS-T and programmed service. The X axis contains the sequence of fixed bus stops for which many details are referenced: stops, traffic signals, transfer routes, subway stations, and main trip generators with distance from the starting terminus. From the APC, boardings and alightings at each stop allow the calculation of link volumes and passenger-kilometers and passenger-minutes.

Subsequently, statistics are estimated for the specific bus departure (Line 161-eastbound, 12:39): commercial speed is equal to 14.6 Km/h, maximum load of 47 passengers, 120 revenue passengers, for an average distance 2.64 km and in-vehicle travel time of 11 minutes per passenger, passengers travelling at an average of 14.4 km/h. Average occupancy for this trip is 30.3 passengers.



Line-direction	161-E	Departure	1239	Bus #	22-369
# Stops	44	Departure time	12:38:58	Load_max	47
Length (km)	10.45	Arrival time	13:22:03	Load/Capacity	0.723
Stops spacing	0.237	Speed (km/h)	14.55	Pass.-km	317.1
Total time	00:43:05	Passengers	120	Pass.-min.	1319.1

LEGEND
0+000 line chaining
Metro station
bus connection (principal line) (XX)
traffic ligth
trips generator

Figure 14 Load profile of the eastbound trip departing at 12:39, with the respective boardings and alightings

In order to augment the monitoring of the vehicle productivity on a bus route, summary statistics are elaborated for each trip departure, thus serving as a diagnostic tool of the quality of service. As an example (figures 15 and 16), one can follow the 11 legs of the vehicle 22-369 on the bus route 161, and an additional leg to serve a school extra. Incidentally, the location of the maximum load points varies slightly over the day while the utilization factor stays relatively high.

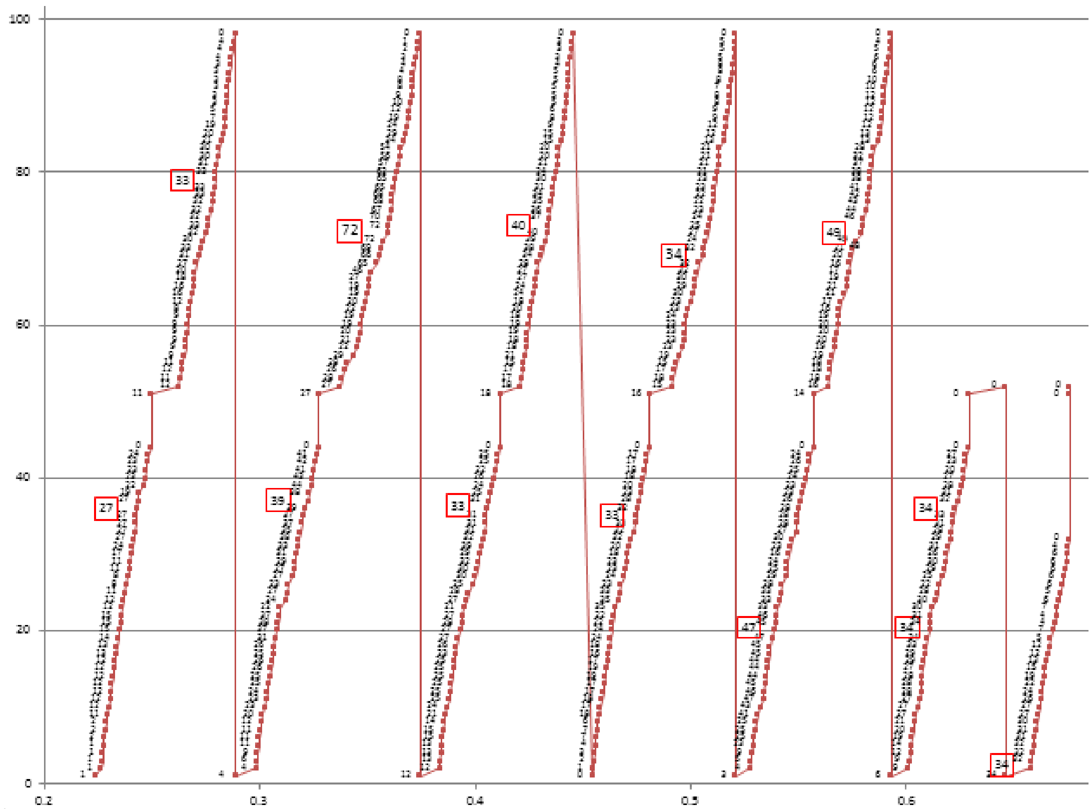


Figure 15 Schematic space-time diagram (based on stops) of the vehicle block (12 commercial legs) with maximum load points;

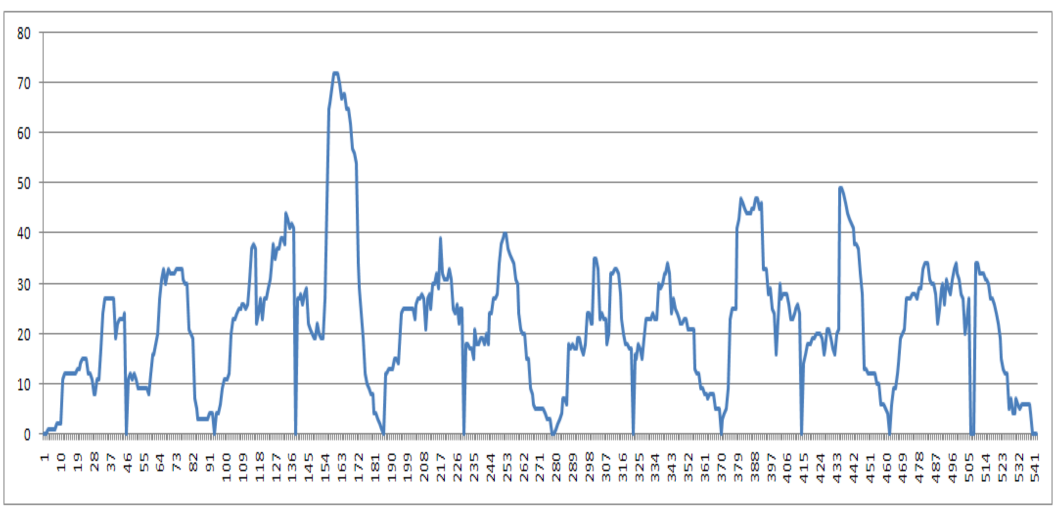


Figure 16 Evolution of the load of the bus 22-369 over the day (sequence of stops)

GLOBAL STATISTICS FROM PROGRAMMED SERVICE

From a more global standpoint, the planning decisions follow a sequence of requirements necessitated by a series of factors conditioned by:

- The geometry and the direction of a bus route for a time period
- The locations of termini
- The travel time needed to deliver the service under given conditions
- The crew scheduling constraints (spread, straight runs, etc…)
- The anticipated occupancy (load factor) for a time period
- Many other factors (school extras, special events)

The resulting situation creates driver runs and consequent vehicle blocks that are structured around the shown distribution (figure 17). Frequency peaks occur at 4-hour, 8-hour, 12-hour and 20-hour blocks. Consequently, a large proportion of the vehicle-hours – the costliest factor- is consumed on more than 11-hour vehicle blocks; this strategy could minimize deadheads.

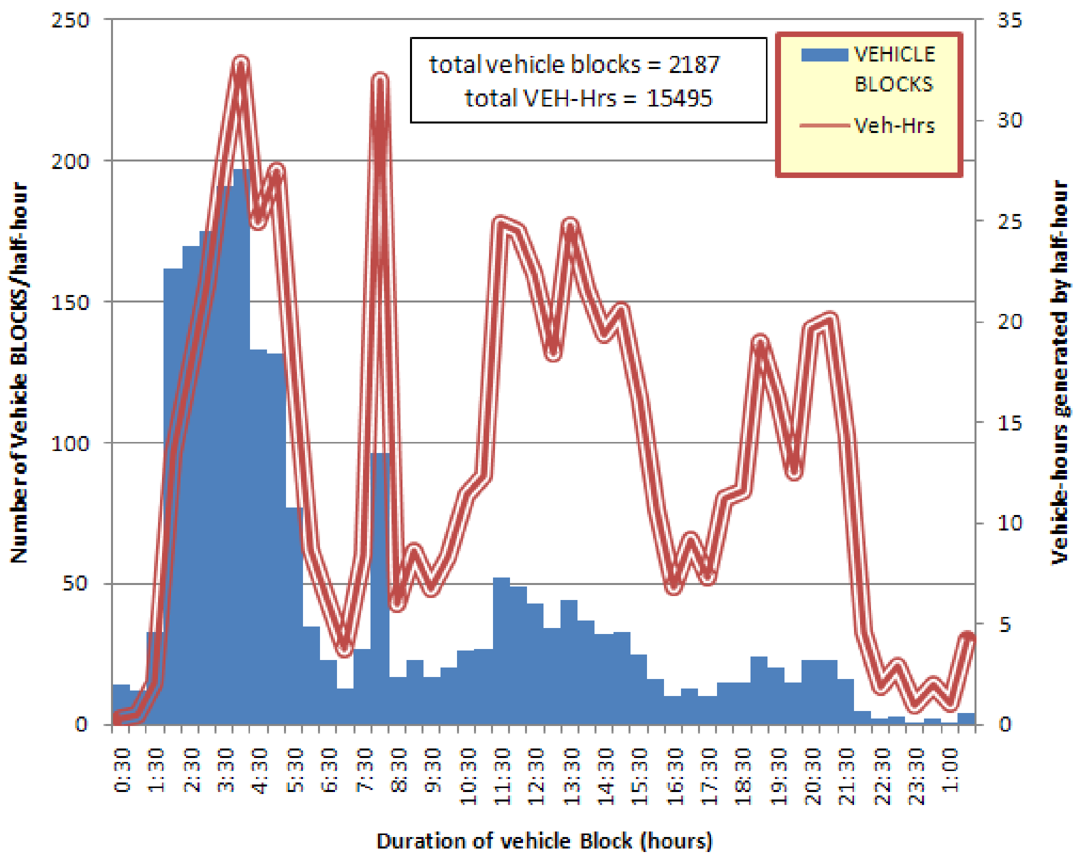


Figure 17 Planned service in terms of vehicle blocks: frequency and vehicle-hours (per half-hour)

In parallel, the coverage at a 15% sampling level seems spatially well-distributed, as demonstrated by the next figure 18. This 3D representation illustrates both the actual level of service and the location where the APC-equipped vehicles are deployed.


Figure 18 Spatial distribution of transit service (grey) and 15% of GPS-APC equipped vehicles (blue)

COMPARISON BETWEEN AFC AND APC

The last experiment, for the actual state of research, considers the comparison and/or the merging of AFC data with the GPS-APC data. In the case of the Montreal bus network, every bus (fleet of 1500 buses) is equipped with an electronic registering farebox which can process contactless smart cards and cash. These boxes are not linked to any GPS, but the time of the transaction (boarding only) can be recorded (clock precision unknown).
In our experiment a match between APC and AFC is done for the 175 GPS-APC-equipped buses that were active the chosen day.

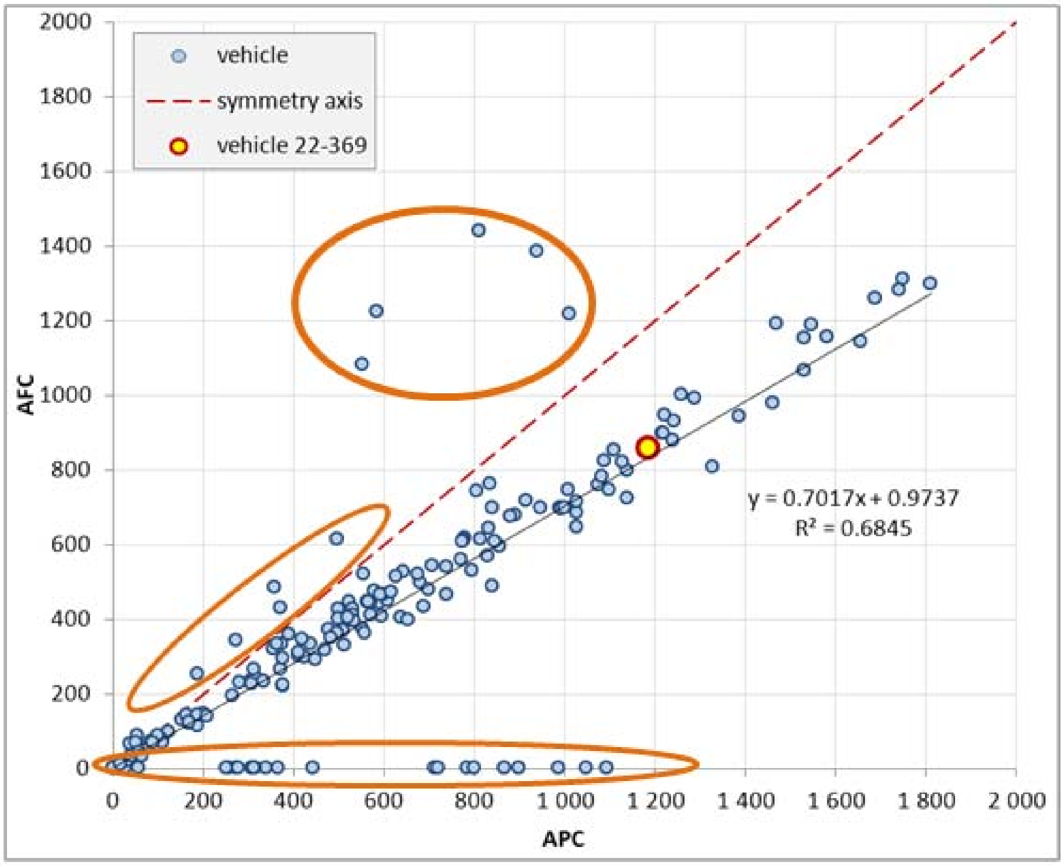


Figure 19 AFC and APC comparison of boardings for 175 vehicles: error detection and model

The figure 19 illustrates certain types of problems encountered when in practice a matching is made based upon the vehicle identification. Because we do know that the number of riders detected by the APC sensors must be greater than the number of AFC smart card transactions (cash fares are accepted), situations A, B and C mean that there is a mix-up in the assignment of fareboxes to vehicles. A series of procedures have to be implemented to tackle this problem in the management of data. For the moment, there is no easy way to untangle each trip data, but errors can be estimated to a level of about 10-15% of the transaction. Overall, as demonstrated by the simple regression model (D), the smart cards represent around 70% of total transactions, and a little more when reduced fares (students) are involved. Notably in the figure, our studied vehicle 22-369 is marked with a red-and-yellow dot.
Beyond these difficulties which partially affect the precise matching between AFC and APC systems, one may observe that in a temporal perspective the tracking of the demand is maintained with some uniformity. The figure 20 shows a stable rate of 15% from the presence in the network of the APC vehicles, and curiously, of a similar rate from one vehicle depot as part of the entire fleet.

To monitor the adequacy between the supplied service in terms of vehicle in service, for 15-minute periods, one may combine the evolution of the number of boardings for the total network, these being data coming from the AFC system. With some extrapolation, and making the hypothesis of travel time of around 15 minutes per boarding (in fact, average travel time per boarding is 13.0 minutes for a distance of 3.36kms), one may establish an average load factor varying over the day, from 17.3 boardings/vehicle in peak period to around 8.0 in off-peak periods for a mean of about 10.5. The best estimate of the average occupancy all over the day is approximately 9.1, a true global indicator for the entire day.

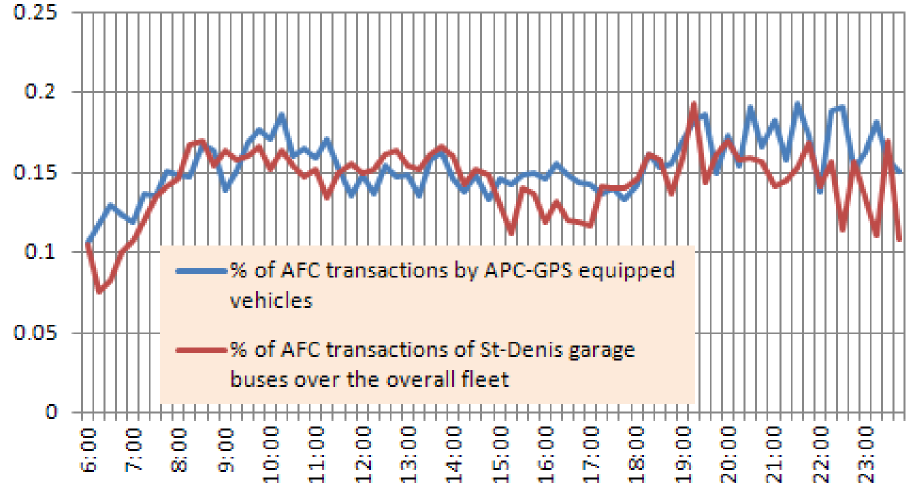


Figure 20 Temporal monitoring of smart card transactions by the GPS-APC equipped vehicles

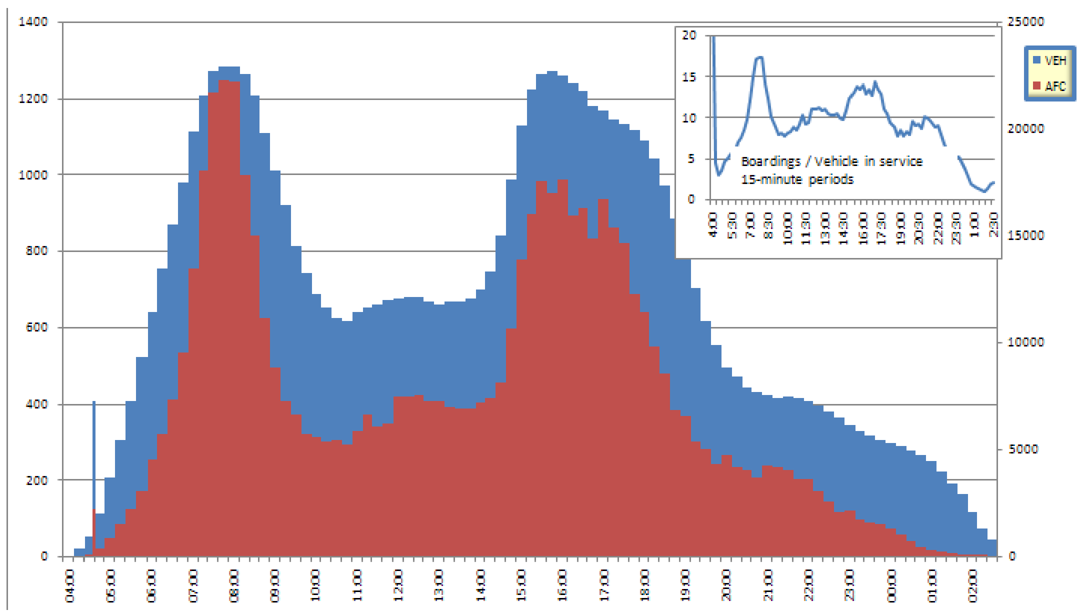


Figure 21 Comparison of the evolution of vehicles in service and number of boardings (by AFC) per 15-minute periods; calculation of the indicator "boardings/vehicle in service"

CONCLUSIONS AND FUTURE DIRECTIONS

From a specific experiment conducted with the newly available data of the Montreal Transit Authority, this paper has demonstrated some of the potentialities embedded in the GPS-APC and AFC data when contextualized with GIS-T and programmed service. Potential for calculation of productivity indicators and demand monitoring is confirmed.
At the same time, the analysis revealed that many more efforts have to be invested to alleviate procedural misconducts in the data management of huge databases and to produce more intelligent, or at least adequate, software capable of correcting, imputing, enriching the available APC and AFC data. Once completely coherent and consistent data are achieved, it

will be possible to develop ridership modelling methods which are sufficient to plan a more effective and more efficient public transport service.

ACKNOWLEDGEMENTS

We thank the Montreal Transit Authority (STM- Société de Transport de Montréal) for providing the data and partially supporting this research. The authors want to acknowledge the contribution of Daniel Piché and Guillaume Bisaillon at different steps of the huge data processing involved.

REFERENCES

Chu, K.K.A. and R. Chapleau (2007). *Imputation Techniques for Missing Fields and Implausible Values in Public Transit Smart Card Data.* Presented at the 11th World Conference on Transportation Research, Berkeley, CA.

Chu, K.K.A. and R. Chapleau (2008). *Enriching Archived Smart Card Transaction Data for Transit Demand Modeling*. In Transportation Research Record: Journal of the Transportation Research Board, No. 2063, TRB, National Research Council, Washington, D.C., pp. 63–72.

Chu, K.K.A., R. Chapleau and M. Trépanier (2009). *Driver-Assisted Bus Interview: Passive Transit Travel Survey with Smart Card Automatic Fare Collection System and Applications*. Transportation Research Record: Journal of the Transportation Research Board, 2105, pp. 1-10. TRB, National Research Council, Washington, D.C.

Chu, K. K. A., & Chapleau, R. (2010). *Augmenting Transit Trip Characterization and Travel Behavior Comprehension with Multi-day Location-Stamped Smart Card Transactions.* Paper presented at the Transportation Record Board - 89th Annual Meeting Compendium of Papers CD-ROM Washington, D.C.

Furth, P.G., B. Hemily, T.H.J. Muller and J.G. Strathman (2006). *TCRP Report 113: Using Archived AVL-APC Data to Improve Transit Performance and Management*, TRB, Washington D.C.

Marchal, F., J. Hackney, and K.W. Axhausen (2004). *Efficient map-matching of large GPS data sets - test on a speed monitoring experiment in Zurich. Arbeitsbericht Verkehrs- und Raumplanung,* Institut f¨ur Verkehrsplanung und Transportsysteme, ETH Zurich, Zurich

Utsunomiya, M., J. Attanucci and N. Wilson (2006). *Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning*. Transportation Research Record: Journal of the Transportation Research Board, 1971, TRB, National Research Council, Washington, D.C., pp. 119-126.

Wilson, N. H. M., J. Zhao, and A. Rahbee. (20090 The Potential Impact of Automated Data Collection Systems on Urban Public Transport Planning. In *Schedule-Based Modeling of Transportation Networks.* Springer, New York, pp. 75–99.

Slavin, H., A. Rabinowicz, J. Brandon, G. Flammia, and R. Freimer. (2009)

Using Automatic Fare Collection Data, GIS, and Dynamic Schedule Queries to Improve Transit Data and Transit Assignment Models. In *Schedule-Based Modeling of Transportation Networks* (N. H. M. Wilson and A. Nuzzolo, eds.). Springer, New York, pp. 101–118.