ESTIMATION OF A DISAGGREGATE PUBLIC TRANSPORT OD MATRIX FROM MASSIVE SMART CARD DATA

Submission date: May 13, 2011

Word count: 5.767 words

5 figures, 2 tables

Marcela Munizaga
Universidad de Chile, Casilla 228-3, Santiago, Chile
mamuniza@ing.uchile.cl

Carolina Palma (corresponding author)
Transantiago, Moneda 975, Santiago, Chile
carolina.palma@transantiago.cl

Daniel Fischer, Universidad de Chile
Casilla 228-3, Santiago, Chile
dfisher@ing.uchile.cl

ABSTRACT

A good quality Origin-Destination matrix is a fundamental prerequisite for any serious strategic transport system analysis. However, is not always easy to obtain it, as OD matrices are expensive and difficult to obtain. This is particularly relevant in large cities, with congested networks, where detailed zonification and time disaggregation require large sample sizes and complicated survey methods. Therefore, the incorporation of information technology in some public transport systems around the world is an excellent opportunity for passive data collection. In this paper we present a methodology for estimating an OD matrix from smartcard and GPS data for Santiago, Chile. We applied the proposed method to a one-week database, obtaining detailed information for time and position of boarding, time and position of alighting for 80% of the 36 million boarding transactions. The results are available at any desired time-space disaggregation. After some post processing, and incorporating expansion factors to account for unobserved trips we build OD matrices disaggregated at bus stop level.

Keywords: public transport, origin-destination matrix, passive data.

1. INTRODUCTION

In the late 90s smartcard payment systems were incorporated in some cities such as Washington (Smartrip) and Tokyo (Suica). This new technology rapidly spread to other cities, and nowadays it has become an important part of the public transport fare collection system. For example the Oyster card was implemented in London in 2003, with discount fares (compared with buying single tickets) and it is currently the most popular payment method. In Chicago, (Zhao et al, 2007) the Chicago card has a very high penetration rate. Other examples, with different ways of implementation and different levels of penetration are: San Francisco (Buneman, 1984), Portland (Furth et al., 2006), New York (Barry et al., 2002), Netherlands (Muller and Furth, 2001; Furth et al., 2006), Changchun (China) (Lianfu et al., 2007) and Quebec (Gatineau, Quebec, Canadá) (Trépanier et al., 2007; Chapleau and Chu, 2007; Chapleau and Chu, 2007; Chapleau et al., 2007; Chapleau et al., 2008). In all these cities, the smartcard is used as one of the payment possibilities. In Santiago (Chile) it is the only available payment system in buses, and by far the most important in the Metro (99%); therefore, the penetration rate is very near 100%.

The research challenge of obtaining valuable information from the data generated by smartcard transactions has been taken by several researchers, who recognize its potential to improve public transport planning and operation. The MADITUC research group has developed several methods to obtain the information and to improve its quality. Chapleau and Chu (2007) propose a method to identify and replace incorrect or suspicious observations from the automatic fare collection system. Trepanier et al (2007) propose a method to estimate the alighting point of a trip, in a system where users only validate when boarding. Lianfu et al (2007) propose a method to build an Origin-Destination (OD) matrix at bus-stop level, using the data generated in Changchun, China. Zhao et al (2007) develop a method for inferring rail passenger trip Origin-Destination (OD) matrices from an origin only automatic fare collection system, where the position of buses in known through an Automatic Vehicle location system.

The research efforts have focused in the integration and enrichment of the information available from different passive sources (such as automatic fare collection systems, automatic vehicle location systems, passenger counts), detection and correction of information errors, estimation of alighting or destination point, identification of transfers, generation of origin destination matrices from the information available.

This paper presents a methodology that goes a step further in terms of the dimension and complexity of the public transport system, as is applicable to large scale multimode public transport systems. The rest of the paper is organized as follows: in the next section the Transantiago public transport system and the data available are described, in section 3 the methodology proposed is presented, section 4 contains the results of the application. Section 5 concludes.

2. BACKGROUND

The data available comes from Transantiago, the public transport system available in Santiago, Chile since February 2007. The system is based on a trunk-feeder structure, where the Metro (underground) is an important component. It has nine feeder operation areas, each serving one part of the city. There are also six trunk operators, which operate larger routes across the city. Trunk operator 1 is Metro, the rest are bus lines. The payment system is such that each passenger pays a fare when entering the system that allows him/her to make up to three combinations within two hours. The payment structure is slightly different in buses and metro. In the buses, the only payment method is the contactless smartcard bip!, while in the Metro it is possible to buy a single ticket or to use the bip!; however, the percentage of users who buy single ticket in Metro is very small (around 1%). The fare is also different, slightly higher for Metro at peak hours; buses have flat fare. If a passenger uses a bus first and metro afterwards, the difference between both fares is charged when entering the metro. All metro lines are connected and internal changes between lines are made without showing the bip! card again.

As a general description of Santiago, it is the capital city of Chile, with nearly 6 million inhabitants. The distribution of the population is not homogeneous; with clearly identifiable wealthier and poorer neighborhoods. The city has a circular shape, with a large proportion of trips going from the suburbs to the centre in the morning, and from the centre to the suburbs in the evening. According to the 2001 Origin Destination survey (Dictuc, 2003) there were 16 million trips in a working day, and from them 10 million were motorized trips (38.6% of trips were walking or bicycle trips). The average household size was 3.81, and the trip rates were 2.82 trips per person, 10.76 trips per household. The market share of public transport was by then 53%.

There were severe problems at the beginning of the operation of new the system, but it is now operating normally, with some persisting problems in certain areas. An evasion problem has been detected in the buses, geographically biased towards poor neighborhoods located far from the city centre. Evasion is almost inexistent in Metro. The system contains over 300 bus routes, and nearly 6,000 buses. It has more than 10,000 bus stops and 85 Km of Metro rails. More than 11 million bip! cards have been issued. There are 150 bus-stations with very light infrastructure (basically a fence) equipped with extra vehicle payment system where passengers pay when entering the station, to increase the boarding efficiency. These bus-stations, called “Zonas paga”, operate during peak hours at congested points.

All Bip! Transactions are recorded in a database that contains information about the operator and the instant when the transaction was made. Each passenger has to make a transaction (put his/her card close to a payment device) when entering a bus, bus station or metro station. Each payment device has an id and is associated to a bus, metro station or bus station. The information recorded for each transaction includes the card id and type, bus or site where the transaction was made, time, date, and amount of money paid. Every week there are around 35 million bip! transactions, made by over 3 million bip! cards.

Another database contains geo-coded information of all buses, such as latitude, longitude, time, date and instant speed. Each bus is identified with plate number and operator code. In most observations, the position is available every 30 seconds. In some cases this period is longer, and shorter in others. Every week of data contains around 80 million GPS observations. The geocoded bus routes, and position of Metro stations, bus-stops and bus stations is also known and valuable information. There are timetables associated to bus and metro services, and also for bus stations, but they mainly indicate the operative hours and the frequency of each service (no scheduled services are provided).

On the other hand, bus assignment information is stored in a database that contains information about the service each bus is giving in a certain period. This information is generated by the Transantiago authority using a triangulation procedure. They define three points of each service route, if a bus passes through the three points then it is assigned to the service. This process has been evaluated by the Transantiago Authority and by the operators and proved to be reasonably reliable.

3. STATISTICAL DESCRIPTION OF THE DATA

A descriptive statistical analysis is conducted for one particular week (20-26 March 2009) from the database. Looking at the transactions information, it can be observed that 43.6% of the bip! transactions (boarding) are made in buses from trunk operators, 31.6% in metro stations, and 24.9% in buses from feeder operators. From the bus transactions, 8.6% are made in bus-stations, the rest are made directly in buses. The number of transactions along the week show similar numbers for working days of around six million transactions per day, during the weekend this number falls to less than four million on Saturday, and less than 2.5 million on Sunday. The relation between total and first stage transactions is 1.65, which gives an initial idea of the number of stages per trip.

Figure 1 shows the time distribution of boarding transactions along a working day by mode. Morning peak and evening peak can be clearly observed at 8 AM and 7 PM respectively. A much smaller but still noticeable midday peak is observed between 1 and 2 PM. The metro and bus peaks occur at different times (bus peak earlier). Saturday and Sunday (not shown here) have much less transactions. Saturday has a pattern that starts early in the morning (as early as in a working day), but rises only up to 60,000 transactions per hour during the afternoon. Sunday shows very little activity, with less than 40,000 transactions per hour and no clear peak. On the right hand side of Figure 2 we show the boarding profile by type of user. It can be seen that time profiles are different for students and adult passengers. The morning peak and evening peak are slightly earlier for students. This information is valuable for policy measure evaluation; however, more processed information such as load profiles is required for planning and design purposes.

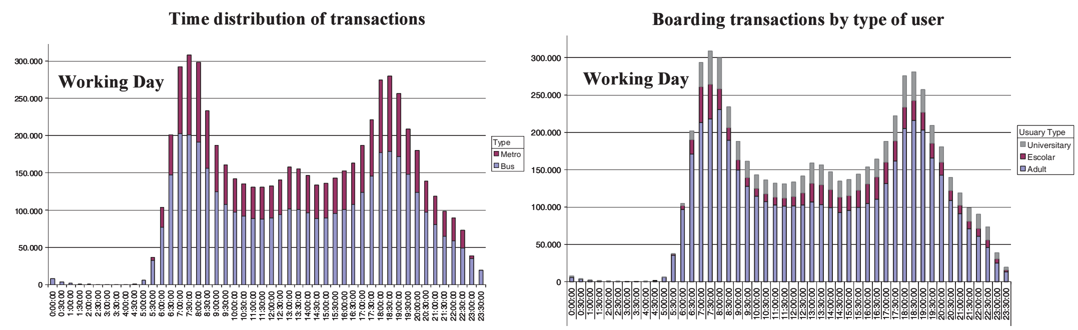

FIGURE 1 - Time distribution of transactions along a working day

Matching the Transactions and Positions databases through bus plate or Metro/Bus-station code and time, it is possible to identify the position where the transaction was made in 98.5% of the cases. This information can be used to make a spatial analysis of transactions, as shown in Figure 2 for boarding transactions at bus stops. The aggregate analysis of all routes over time shows that morning peak is different in different locations, remarkably earlier in poorer neighborhoods. The amount of information gathered permits as much time and space disaggregation as required to perform this kind of analysis.

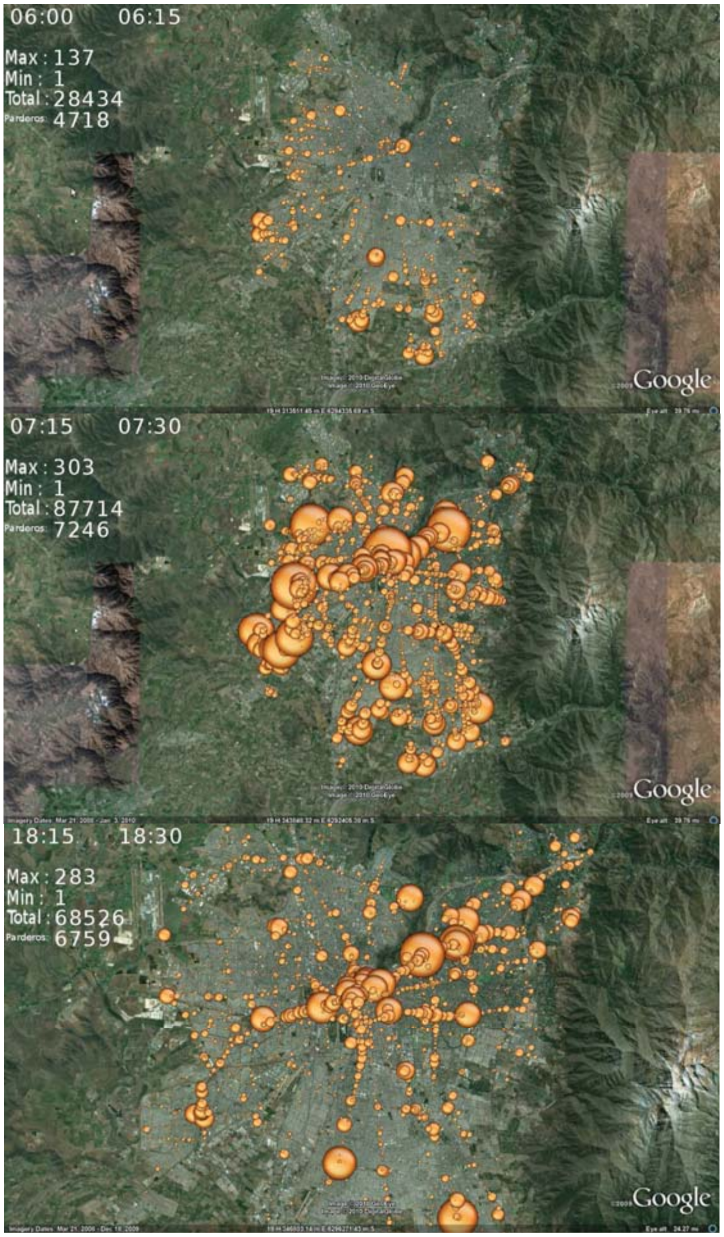

FIGURE 2 - Boarding transactions at different times

4. METHODOLOGY

Based on the work by Trepanier et al (2007) and a preliminary analysis conducted over a small subsample of cards (users), a methodology suitable for large public transport systems, such as Transantiago is proposed. Some definitions are required to explain the proposed method. Let us define a trip as a movement from a point of origin to a point of destination (Ortúzar and Willumsen, 1994). Each trip can have one or more stages, which are movements in a particular service (bus or metro). Origin and destination are the positions where the trip begins and ends, respectively. Boarding and alighting points are the positions where the stage begins and ends, respectively.

The main objective of the proposed method is to reconstruct the trip chain of users behind bip! cards, by estimating the destination points from the information available. Once this is available, it is possible to analyze behavior, build origin-destination matrices, estimate vehicle load profiles, and many other tasks in a simple and direct way. The proposed model has several components, as shown in Figure 3. The inputs of the model are three main databases: transactions (boarding) from automatic fare collection system, vehicles position from the automatic vehicle location system and geocoded definition of the public transport network. After matching these three databases, it is possible to obtain the position of the transactions, and then estimate alighting point. The estimation procedure is described below. It is different for transactions in buses, bus stations and Metro stations, but in all cases the result is an estimate of the position-time coordinates of the alighting point. Then, using this information, our proposed method includes a module to distinguish transfer from destination, identifying trip stages. As a result of this procedure trips and trip stages are obtained for a proportion of the sample. Furthermore, in some of those cases the method will be able to estimate the alighting point of all boarding transactions of a particular card in a particular day. Those cases are very valuable, because they allow building the public transport trip chain of the person behind that card. On the other hand, there are some cases where the estimation of the alighting point is not possible.

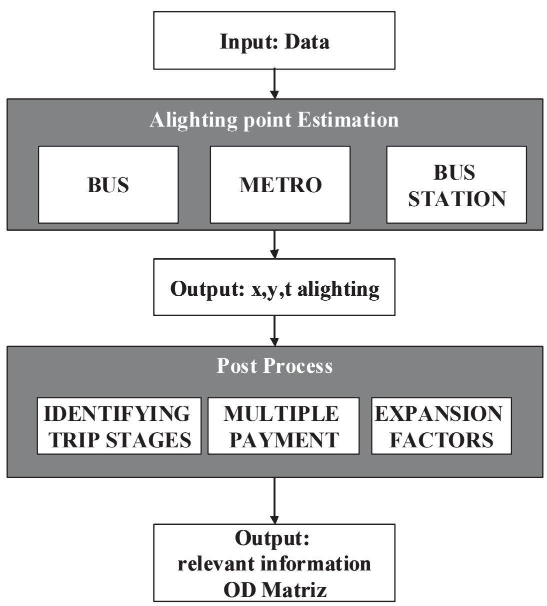

FIGURE 3 - Proposed Method Structure

a) Alighting point estimation

To estimate the alighting point it is assumed that the next bip! transaction is posterior to the alighting. Following Trepanier et al (2007) we also assume that the alighting stop is close to that of the next boarding. This is only possible to apply when both the current transaction and the next one have position information (from the AVL database). In the case of the last transaction of the day, we assume that its destination is close to the point where the first trip of the day began, finishing the daily trip cycle for that particular user (card). If there is only one trip per card, no imputation is possible with single day information. The model is shown in Figure 4, where the three possible cases are illustrated: next transaction in bus, metro station or bus station.

In a complex network such as the one in Santiago the Trepanier et al (2007) methodology of identifying the point of the previous trip route closest (distance) to the position of the next boarding cannot be applied directly, because in many cases erroneous points will be identified. An example of this is when a bus route uses the same street in both directions, a point from the return direction might be the closest, but the bus already passed very near the next boarding in the initial direction. To overcome this difficulty, it is proposed to use generalized time instead of distance, as the function to minimize.

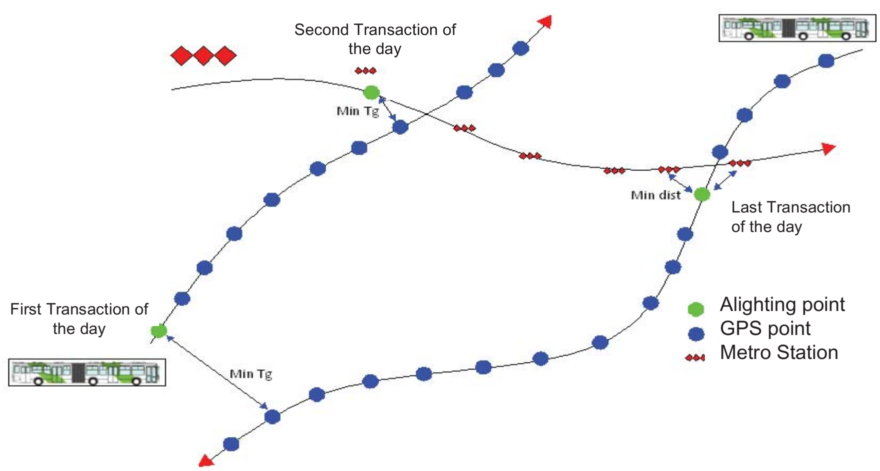

FIGURE 4 - Alighting Estimation Model.

If the previous trip or trip stage is in a bus, the alighting point is searched along the trajectory of that bus, known from the GPS database. The position-time alighting estimate (x_a,y_a,t_a) is that of the bus trajectory that minimizes the generalized time distance with the next boarding time-position. Generalized time (Tg_i) is defined en equation (1) as the time associated to position i t_i, plus the distance between position i and the next position identified by sub index post: $d_{i\text{-}post}$ divided by the average walking speed s_w and multiplied by a weight factor f_w representing the disutility of walking time as a proportion of in vehicle travel time.

$$Tg_i = t_i + f_w \cdot \frac{d_{i-post}}{s_w} \qquad (1)$$

The search is conducted over all positions of the bus trajectory that are within walking distance (d) from the next transaction position. Therefore, the optimization problem can be written as:

$$\text{Min } Tg_i \qquad (2)$$

$$s.t.$$

$$d_{i\text{-}post} < d$$

This will identify a case where the bus is sufficiently close to the destination to alight and walk, avoiding the aforementioned problem of two way routes, where the minimum distance point can be very inconvenient in time. This situation is illustrated in Figure 5, where a passenger boards a line that goes from left to right. The route of that bus goes up to a certain point to the right, and returns to the left. If the route goes in both ways along the same street, or even if they are not the same but close streets, a passenger whose destination is the point designated with X in Figure 5 will not remain in the bus along the whole route to alight exactly at the closest point of his/her next boarding, s/he will rather alight at the more convenient *i* point considering travel and walking time.

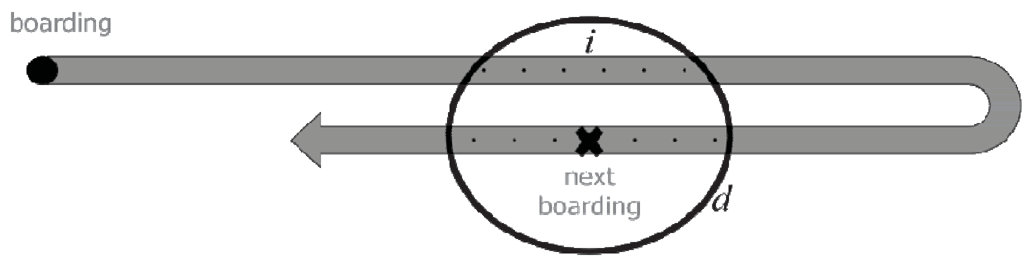

FIGURE 5 - Search procedure

To implement the method in an efficient and feasible way, a time window is defined for the search in the bus trajectory from the instant when the user boards the bus. This is a parameter of the model, which can be set at different levels for trunk and feeder routes, depending on the characteristics of both types of services. If this constrain becomes active, the limit is doubled, because then the closest point is likely to be further away along the bus trajectory.

Another parameter required by the model is the distance assumed as walking distance d. It probably depends on the type of person, type of city, weather and other factors. This parameter was initially set to 1,000m.

If no solution is found for equation (2) within the maximum distance threshold, then it is assumed that there is a missing trip or stage, probably in another transport mode or using another bip! card. In that case that trip is labeled as “no alighting point estimation”.

a. Metro

In the case of Metro stages, the boarding and alighting points are metro stations. The boarding station is known directly from the data, and the alighting station is estimated as that closer (in distance) to the next boarding, within a circumference defined by the walking distance d. If there is no station within that distance, it is assumed that there is a missing part of information and the alighting point cannot be estimated.

For those cases where an alighting metro station is found within the *d* ratio, the instant when that alighting occurred must be estimated. As only the boarding station is known, a Dijkstra (1959) shortest path procedure is implemented to estimate the route followed by the user to go from the boarding to the alighting station. The travel time between stations, detention time at stations, and walking time inside the station are parameters of this procedure. The total travel time in Metro is the sum of the corresponding components.

b. Bus station

Probably the most difficult case is that of individuals who board at a bus station, where the bip! transaction is made when entering the station, and the user can then board into any of the buses from routes that use that bus station as bus-stop. Therefore, in this case there is an additional problem to be solved: to assign a bus to each transaction made at the bus station. Once a bus has been assigned, the aforementioned buses procedure can be applied.

As a first stage all those routes that use that bus station and have at least one bus stop within walking distance from the position of the next bip! transaction are identified. If only one route in that situation is found, then it is assumed that the user will probably board the first bus of that route that passes through the bus station after the bip! transaction is made. If there is no route that has at least one bus stop within walking distance of the next boarding point, it is not possible to estimate the alighting point. Finally, if there are two or more such routes, an assumption have to be made on which bus the user boards. To do this, the common bus lines concept proposed by Chriqui and Robillard (1975) is applied. The user is assumed to choose a set of routes that minimize his/her expected travel time, and board the first bus of that optimum set. Observed frequencies are used to implement this procedure.

Once the set of common lines is found, the user is assumed to take the first bus observed after his/her arrival at the bus stop, from any of the common lines. The previously described procedure used to estimate the bus alighting point is used.

b) Post Processing

Once the alighting point is estimated, we can calculate travel times, time between alighting and next poarding, load profiles and other relevant variables. However, to build an OD matrix, it is necessary to differentiate trips from trip stages, identifying the trip destination as the place where a person goes to conduct an activity. At this stage of the research, we propose to use a simple rule, and assume that if a person (card) stays for longer than 30 min in a particular point, then it is a destination, and otherwise, it is a transfer point. In the cases where the alighting point of the previous transaction could not be estimated, we assume that if the time elapsed to the next transaction is over two hours then this next transaction is the first stage of a new trip.

a. Multiple Payment

Even though it is unusual, there are some cases where two or more persons use the same bip! card. This behavior is unusual because the transfer discount is not applicable in this case, but, as some people do it anyway, we developed a procedure to identify when two or more persons use the same card to travel. If two or more bips of the same card are registered in the same bus in a very short period of time, then those trip stages are regrouped as part of the same trip. If two or more bips of the same card are registered in very near locations, but not the same, like for example a bus stop and metro station, in a very short time period, then we assume that we can only try to estimate the alighting point of the last bip! and the other(s) fall in the category of boarding without known destination.

b. Expansion Factors

As not all trips are identified with this procedure it is necessary to build expansion factors to account for unobserved trips. There are three cases that require different treatment: i) trips associated to a bip! transaction where the origin is known but the destination could not be estimated, ii) trips detected through a bip! transaction where neither the origin nor the destination are known and iii) trips not detected though bip! transactions (fare evasion).

For the first case, we shall assume that the distribution of trips with unknown destination is the same as that of other trips with the same origin. Therefore, we build an expansion factor associated to trips of a particular origin/time period that accounts for all those trips with unknown destination, as f_{ijt} defined in equation (3).

$$f_{it} = \sum_{j} \frac{Trips_{ijt}}{\sum_{j \neq null} Trips_{ijt}}, \tag{3}$$

where i is the trip origin, j is the destination according to the zonification of the studied zone (in our case 600 ESTRAUS zones) and t is the time period. A similar factor could be applied to trip stages to build a trip stages OD matriz.

To account for trips detected through a bip! transaction where neither the origin nor the destination are known, we build a general expansion factor, similar to the previous one, but including only the time disaggregation, as shown in equation (4).

$$f_{t} = \sum_{ij} \frac{Trips_{ijt}}{\sum_{\substack{i \neq null \\ j \neq null}} Trips_{ijt} \cdot f_{ijt}}, \tag{4}$$

A similar treatment should be applied to account for evasion, however, good quality additional information is required to be able to do that. Unfortunately, the effect of fare evasion and or avoidance is not homogeneous, and apart from underestimating the OD matrix could induce some biases. For example in the case of Transantiago there are two types of fare evaders: the casual evader, that usually does not validate in the first stage of the trip at the feeder service, because of the lack of charge points, but in the next stage of the trip, usually in the Metro, charges his/her card and validates; and the hard evader, that has the strong intention to evade, and will not validate in any of the trip stages. Both types of evasion described imply different biases. The first type of evasion described will contribute to underestimate trips in feeder services, especially those with origin in zones with low commercial activity, and to overestimate trips originated at Metro stations. The latter would probably imply an underestimation of trips in a group of services.

c) Computational implementation

To process these big databases all the information is stored in PostgresSQL 8.3 files. Several indexes are used to improve the consultation speed. The code to process the data is build in C++, ssing libpqxx as interface with the database. The results are visualized on Google-Earth using API for KML format. All the stages of the above-described procedure are implemented as described, with parameters that can be varied by the user, such as walking distance, time thresholds, walking speed, among others.

The implementation of the bus-station alighting estimation procedure presented some difficulties. As in that case the exact bus that the user take is not known, the procedure has to find the most likely, and sometimes the information is not enough to make a good estimation. To solve the common routes problem, the frequency of the available routes is a key element, and when a bus-station is located at the beginning of a route path, the frequency information is not reliable. Furthermore, due to the triangulation procedure used to determine which route is serving a particular bus, some buses are not assigned to the route from its geometric beginning.

The overall implementation of the alighting estimation procedure was very time consuming, however, the problem is separable, so it can be distributed in various computers running simultaneously.

5. APPLICATION AND RESULTS

The method described was applied to 36 million observations corresponding to one week of operation of the system (March 20-26, 2009). The method was able to estimate the alighting point for a global 79.9% of trip stages (bip! transactions) and 72.1% trips (destinations). In Table 1 we present the success percentages by day of the week and by type of transaction. As expected, the lower success rate is for transactions made in bus stations, and the highest is for transactions made in Metro stations. Also, weekend days including Friday have lower success rates than working days, probably due to more variations in the ways of travelling during these days. The last column shows the success rates for entire trips.

Table 1: Percentage success in alighting estimation

Day	Trip stage				Trip %success
	Bus %success	**Metro %success**	**Bus station %success**	**Global %success**	
Monday	78.8	87.6	65.7	**80.1**	**73.3**
Tuesday	79.5	88.0	66.8	**81.3**	**74.4**
Wednesday	79.6	87.9	66.9	**81.4**	**74.3**
Thursday	71.2	87.5	66.2	**81.0**	**73.7**
Friday	78.1	85.4	64.8	**79.6**	**71.8**
Saturday	74.4	80.6	58.6*	**76.1**	**66.2**
Sunday	72.3	77.3	38.7*	**73.4**	**62.4**
Global	**78.0**	**86.1**	**65.8**	**79.9**	**72.1**

* Bus station transactions are less than 1% of total transactions on weekend days

Given the alighting point estimation, we built the expansion factors f_{it} and f_t to distribute trips without known origin and/or destination. The average f_{it} over all periods in the entire week, and over all origins i was 1.6, with a variation coefficient of 0.66; the average overall correction factor f_t was 1.008. The multiple payment situation described above was observed in 1.91% of trip stages, and processed accordingly. In Table 2 we present the estimated Origin-Destination matrix at an aggregate level for one working day, using the same aggregate zonification used to present aggregate results of the matrix built with information from the last OD survey in Santiago (Dictuc, 2003). The small number bellow is the percentage difference between the normalized values of the matrix obtained with the proposed methodology and those of the OD survey matrix. It can be seen that the structure is very similar, with small percentage differences in all cells. Nevertheless, we observe that the larger percentage differences are in the intrazonal cells, something that could be due to a change in travel behavior in response to the new fare structure. The totals values are not compared, because some further analysis us required beforehand.

Table 2: Working day aggregate OD matrix. Thousand trips.

$O_i \backslash D_j$	North	West	East	Centre	South	South-East	Total
North	154 2.5	32 0.1	51 0.2	92 0.1	23 0.1	18 0.0	371
West	32 0.2	267 3.6	113 0.7	179 0.7	36 0.2	27 0.0	654
East	48 0.1	101 0.5	288 1.8	191 2.1	72 0.5	143 0.9	843
Centre	83 0.2	172 0.4	172 1.8	195 2.6	115 0.2	109 0.5	846
South	24 0.1	33 0.3	82 0.6	126 0.5	184 4.9	59 0.2	508
South-East	17 0.1	25 0.0	149 1.1	115 0.8	57 0.3	239 3.3	602
Total	358	631	856	898	487	595	3,824

6. CONCLUSIONS

A method to obtain information from automatically generated data in a large and complex public transport system such as Transantiago has been presented. The alighting estimation method is quite robust. It was applied to a one-week database, obtaining 72% success in trip destination estimation and 80% in trip stage alighting point estimation. After the post processing a matrix at any desired level of aggregation can be obtained. We believe this will change the way public transport planning is conducted. Santiago is a privileged case study, as smartcard penetration is almost 100%, and all buses are equipped with GPS device; however, this will probably be the standard in many cities in the near future. The size of the databases is a challenge, as some processing can take several hours to run in powerful computers; therefore, only simplified processes can be done in real time. There are some limitations of the method that are unlikely to be overcome without additional information. One of them is the case where only a single transaction is observed for a particular card in a particular day. Another limitation is that there is no information available about non-integrated modes such as shared taxi, taxi and car. If users take one of these to reach the metro or bus network, then there is a missing piece of information that will induce an error in the estimation of the trip chain of those users. Also, there is the fare evasion problem that apart from the financial implications, can bias the trip structure obtained in the OD matrix.

In terms of further research, first of all the method and its results have to be validated, which is something we are already working on. Secondly, some components of the model can be further developed. For example, the module that distinguishes transfers from destinations must be refined. There is information available that can help to this process, such as frequency of bus services and land use at the position. A 20 min stop in a commercial zone with very frequent bus services is probably a short activity, while a 20 min stop in a zone without commercial activities and infrequent bus services might be a very bad connection. The Metro route choice procedure can also be improved, using models developed by other researchers such as Raveau et al (2010) who incorporate perceptual aspects to passenger route choice in the Metro network. In terms of the global validation of the procedure, we to use a control sample; it is very important to contrast the results of our method with the users behind those cards. The other important issue for further research is how to build expansion factors to obtain Origin-Destination matrix correcting for all potential biases. We offer first stage simple solution, but certainly this issue can be further sophisticated, although it would probably require additional information. The whole evasion problem, for example, cannot be analyzed without additional information. Finally, there is a great potential of the use of data mining techniques, to analyze variability of behavior, to complete missing information, or to identify behavioral patterns.

ACKNOWLEDGEMENTS

This research was partially funded by Fondecyt grant 1090204, PBCT Redes Urbanas, the Institute Complex Engineering Systems (ICM P-05-004-F, CONICYT FBO16) and Transantiago. We specially thank the collaboration of Mauricio Zúñiga.

REFERENCES

Barry, J.J., Newhouser, R., Rahbee, A. and Sayeda, S. (2002), Origin and destination estimation in New York City with automated fare system data. Transportation Research Record 1817, 183-187.

Buneman, K. (1984), Automatic and passenger-based transit performance measures, Transportation Research Record 992, 23-28.

Chapleau, R. and Chu, K.K. (2007). Modeling transit travel patterns from location-stamped smart card data using a disaggregate approach. Presented at the 11th World Conference on Transportation Research, June 24-28 2007, Berkeley, CA.

Chapleau, R., Trépanier, M. and Chu, K.K. (2008).The ultimate survey for transit planning: Complete information with smart card data and GIS. Presented at the 8th International Conference on International Steering Committee for Travel Survey Conferences, Lac d'Annecy, France.

Chriqui, C. and Robillard, P. (1975) Common bus line. Transportation Science 9, 115-121.

DICTUC (2003) Actualización de encuestas Origen Destino de viajes, V Etapa**.** Informe Final a Sectra, Santiago, 2003.

Dijkstra, E.W. (1959) Note on two problems in connection with graphs (spanning tre, shortest path). Numerical Mathematics 1 (3), 269-271.

Furth, P.G., Hemily, B.J., Muller, T.H.J. and Strathman, J.G. (2006) Uses of archived AVL-APC data to improve transit performance and management: Transportation Research Board, TCRP Report No. 113.

Lianfu, Z., Shuzhi, Z., Yonggang, Z. and Ziyin, Z. (2007). Study on the method of constructing bus stops OD matrix based on IC card data. Wireless Communications, Networking and Mobile Computing WiCom 2007, 3147-3150.

Muller, T.H.J. and Furth, P.G. (2001) Trip time analyzes: Key to transit service quality. Transportation Research Record 1760, 10-19.

Ortúzar, J. de D. and Willumsen, L.G. (1994) Modelling Transport, Second edition. Wiley, Chichester.

Raveau, S., Muñoz, J.C. y De Grange, L. (2010) El Efecto de la Topología de la Red y las Percepciones en la Elección de Ruta RAVEAU. XVI Congreso Panamericano de Ingeniería de Tránsito y Transporte. Lisboa, 11-15 Julio.

Trépanier, M., Tranchant, N. and Chapleau, R. (2007) Individual trip destination estimation in a transit smart card automated fare collection system. Journal of Intelligent Transportation Systems 11, 1-14.

Zhao, J., Rahbee, A. and Wilson, N. (2007) Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. Computer-Aided Civil and Infrastructure Engineering 22, 376-387.