ESTIMATION OF A ZONAL ORIGIN-DESTINATION MATRIX FROM OBSERVED PUBLIC TRANSPORT TRIPS FOR SANTIAGO DE CHILE

Sebastián Tamblay, Pontificia Universidad Católica de Chile, sjtambla@uc.cl
Patricia Galilea, Pontificia Universidad Católica de Chile, pga@ing,puc.cl
Juan Carlos Muñoz, Pontificia Universidad Católica de Chile, jcm@ing.puc.cl

ABSTRACT

This article presents a generic methodology that allows to infer the origin and destination zones for an observed trip between two public transport stops, using land use information. Thus, it enables the estimation of a zonal origin-destination matrix for Santiago de Chile's morning peak. Additionally, an original access arcs creation methodology is presented, along with a new zonal system for the city. The resulting matrix will allow to evaluate (through the Project D10I1049 FONDEF computational tool) major changes to the transport system that modify the stop where users start or end their trips; such as new bus or metro lines.
Keywords: public transport; zonal origin-destination matrix; smartcard data.

RESUMEN

Este artículo presenta una metodología genérica que permite inferir las zonas de origen y destino para un viaje observado entre dos paraderos de transporte público, utilizando información de uso de suelo. Así, permite estimar una matriz origen-destino zonal para la punta mañana de Santiago. Adicionalmente, se presenta una metodología original de creación de arcos de acceso y una nueva zonificación para la ciudad. La matriz resultante permitirá evaluar (a través de la herramienta computacional del proyecto FONDEF D10I1049), cambios mayores al sistema de transporte que modifiquen los paraderos escogidos por los usuarios, como nuevos recorridos o líneas de metro.
Palabras claves: transporte público, matriz origen-destino zonal; tarjetas inteligentes.

1 INTRODUCTION

One of the key elements for correct urban transport planning is the knowledge of a reliable origin-destination (O-D) matrix of the city of interest (Ortúzar and Willumsen, 2011). However, in fast paced changing cities with detailed zoning systems this information might be hard, costly, and time consuming to acquire and keep up to date. The latter evidences the need to develop more efficient and inexpensive methods of collecting information about citizens' travel patterns.

On the other hand, the last decade has seen an increasing trend of implementing information technologies such as Automated Fare Collection and Automatic Vehicle Location Systems, which provide very valuable information about trip patterns.

In systems where passengers are forced to tap-in and out in every vehicle (bus or train), a full stop-to-stop (bus stops or metro stations) trip matrix would be readily available. In systems where passengers only tap-in, then the exiting stop of every trip leg could be estimated as has been done for Santiago (Munizaga and Palma, 2012) and London (Gordon et al., 2013). However, this information is incomplete to estimate the impact of changes in the transit system that could change the stop in which users start or end their trips; e.g. extending routes, relocating their bus stops, designing new routes or changing the fare structure. To study this kind of changes, a zonal O-D matrix that allows modelled users to choose their initial and final stops is needed.

In this line, this study presents a methodology that allows us to estimate a zonal origin-destination matrix for a city, given a stop-to-stop trip matrix. Our proposed model infers a probability for the zone of origin and destination for each trip between two stations, using a Logit model. Model inputs are land use and public transport information, for any city's period of interest.

The methodology is then applied to estimate the model for Santiago de Chile's morning peak period. It is important to note that the resulting zonal inference model for Santiago, along with its respective estimated zonal O-D matrix, will be used as input for the Project D10I1049 of the Fondo de Fomento al Desarrollo Científico y Tecnológico (FONDEF) public transport planning computational tool. The proposed software seeks to predict the reactions of public transport users towards network changes, allowing better design and operation of Santiago's transport system.

The remainder of this paper is structured as follows. Section 2 provides the theoretical basis of our model, presenting its structure along with the proposed estimation and zonal O-D matrix reconstruction procedure. Section 3 introduces and describes the available data for the case of Santiago de Chile's morning peak, relating it to the zonal inference model when appropriate. Additionally, it describes the proposed methodology to determine zones, centroids, and access arcs. Then, Section 4 presents the final zonal inference model and applies it to estimate the zonal origin-destination matrix, along with exploring alternative formulations. Finally, in Section 5 we draw our conclusions, present future research guidelines, and provide policy applications of this study.

2 MODEL FORMULATION

This section presents the zonal inference model, which assigns a probability for the zone of origin and destination for each entry in a stop-to-stop trip matrix. In other words, the model infers the access (and egress) walks for each trip between two bus stops (or metro stations), considering land use and network information of zones within walking distance. Once deduced, an estimation procedure using survey data and the Maximum Likelihood method is proposed.

2.1 Zonal inference model

The proposed zonal inference model allows us to obtain the probability that an observed trip using access stations k and l (as their boarding and alighting points, respectively), was originated in zone i and had zone j as its destination. This probability is defined as $\text{Prob}\left({}^{\text{ij}}/_{\text{kl}}\right)$.

In order to obtain these probabilities for each bus stop, metro station, and zone, we follow Daly (1982) and formulate this particular gravitational model:

$$T_{ij}^{kl} = A_i B_j f_{ij}^{kl} \;\; \forall (i,j) \wedge (k,l) \tag{1}$$

Subject to the constraints:

$$\sum_m \sum_n T_{mn}^{kl} = T_{kl} \;\; \forall (k,l) \tag{2}$$

where A_i and B_j are measures of the generating and attracting powers of zone i and j, linked to the population and land use information for these zones; T_{ij}^{kl} is the (unknown) number of trips made from zone i to zone j that use access stations k and l as their boarding and alighting point, respectively; f_{ij}^{kl} is an inverse measure of the "cost" of choosing those detentions to make the trip; T_{kl} is defined as the observed trips between public transport detentions k and l, which is supposed known for every k and l; lastly, sets m and n are the zones of influence for detentions k and l, respectively (i.e., zones within walking distance from each bus stop or metro station).

This way, for a given trip using bus stops (or metro stations) k and l (as their boarding and alighting points, respectively), the probability that its origin was zone i and its destination zone j can be expressed as:

$$Prob\left({}^{ij}/_{kl}\right) = \frac{T_{ij}^{kl}}{T_{kl}} \tag{3}$$

Furthermore, by replacing Equations (1) and (2) into (3), we obtain:

$$Prob\left({}^{ij}/_{kl}\right) = \frac{A_i B_j f_{ij}^{kl}}{\sum_m \sum_n A_m B_n f_{mn}^{kl}} \tag{4}$$

If we now define:

$$V_{mn/kl} = ln(A_m B_n f_{mn}^{kl}) \tag{5}$$

The latter can be represented by a disaggregate Logit model (refer to Daly, 1979; or McFadden, 1979; for more details on the Logit formulation), as follows:

$$Prob\left({}^{ij}/_{kl}\right) = \frac{e^{V_{ij/kl}}}{\sum_m \sum_n e^{V_{mn/kl}}} \qquad (6)$$

It is worth noting that this Logit model including size variables is obtained without any loss of generality and imposing no restrictions to the specifications of the cost function implied (Daly, 1982).

Finally, defining T_{ij} as the trips from zone i to j, we can reconstruct each element of the desired zonal O-D matrix from the observed travel behaviour between access stations as shown by Equation (7):

$$T_{ij} = \sum_k \sum_l Prob\left({}^{ij}/_{kl}\right) * T_{kl} \quad \forall (i,j) \qquad (7)$$

Therefore, the remaining task is to find appropriate forms for the parameters A_m, B_n, and f_{mn}^{kl} to be included in the "utility" functions expressed by Equation (5).

2.2 Maximum likelihood estimation

Different model specifications can be estimated using survey data and the Maximum Likelihood method (Ortúzar and Willumsen, 2011). Therefore, we must first clearly define each observation q taken from survey entries. In this line, each one of the registered trips must have zones and access stations for their origin and destination assigned. Also, we must find sets m and n as defined by Equations (6), i.e. those zones within walking distance and connected by access arcs to the initial k and final l bus stops (or metro stations). Once this is done, the likelihood equation can be formulated as follows:

$$L(\theta) = \prod_q Prob\left({}^{i_q j_q}/_{k_q l_q}\right) = \prod_q \frac{e^{V_{i_q j_q/k_q l_q}}}{\sum_{m_q} \sum_{n_q} e^{V_{m_q n_q/k_q l_q}}} \qquad (8)$$

Where subscript q indicates that origin and destination choice sets, zones, and stations differ for each observation q. Later, Equation (8) is maximised on the parameters θ of the utility function V_{ij} (i.e., the parameters related to A_m, B_n and f_{mn}^{kl}) and different specifications for the zonal inference model can be deduced and compared.

3 DATA AND MODEL INPUTS FOR SANTIAGO, CHILE

This section presents the different sources of information that were gathered and unified in order to estimate and apply the zonal inference model for Santiago's morning peak. A brief description of data origins and processing, along with the procedure followed to determine the zonal system and its access arcs, is explained next.

3.1 Smart-card origin-destination data

The zonal inference model works by assigning observed trips between two public transport stops to nearby zones, through a Logit model. In this study, we considered observed trips as those belonging to the stop-to-stop trip matrix estimated through the methodology presented in Munizaga and Palma (2012) (and enhanced in Devillaine et al., 2012; and Munizaga et al., 2014), for April 2013 Santiago's network. These trips were used in T_{kl} variables defined in the previous section.

It is important to note that the way these matrices were built translates into some limitations for this study. Particularly, the assigned stop-to-stop matrix does not explicitly includes non-integrated modes (such as share taxis and private vehicles) nor fare evasion, which distorts some of its trips; this limitation is extended to the zonal O-D matrix when assigned by the zonal inference model.

3.2 Land use information

Generating and attracting powers of each zone (A_i and B_j, respectively) are linked to the land use information of each one of them. The latter allows the model to capture that residential zones generate a larger amount of trips or business centres attract more trips, for instance.

Land use information was included for Santiago de Chile for 2014, which contains the number of units and total squared meters dedicated to each land use classification in each zone. Some of the most relevant classifications available are: residential, commercial, educational, industrial, health and offices buildings. This information is gathered and updated periodically by the Chilean Internal Revenue Service (SII, 2014) with taxation and regulatory purposes, meaning that it is available to transportation planners at practically no cost.

3.3 Access station survey

In order to estimate the zonal inference model through the maximum likelihood procedure, survey entries were included. These observations were taken from the revealed preferences survey conducted in bus stops and metro stations in November 2013 in Santiago during the morning. Each of these passengers was asked, amongst other questions, her origin, destination, and intended route; allowing us to obtain her zone of origin and destination, along with her initial and final bus stop (or metro station). After digitation and processing, a total of 628 valid entries remained (i.e., which had clearly identified routes, origins, and destinations).

3.4 Zonal system

As previously specified, the proposed model allocates observed trips into nearby zones, following the probabilities resulting from the Logit calculations. This way, one of the key elements to determine is the city's zonal system that will be used.

Given that the model and zonal O-D matrix resulting from this study will be used in the aforementioned computational planning tool, it is essential that the zonal system is detailed enough so that each zone is walkable and not too large, allowing the application of a route choice model from its centroid without major aggregation errors.

As a result, the chosen zoning system is based in the one proposed in SECTRA (2008) origin destination study, which divides the city into 779 zones. However, there have been many urban developments in Santiago in recent years, and the zonal boundaries had to be updated to reflect them properly. Additionally, some zones were considered too large and were divided, resulting in an original zonal system with 1,171 zones.

3.5 Access arcs

The proposed zonal system did not possess centroids nor their corresponding access distances to bus stops and metro stations, so they had to be added in order to allow the representation of the alternatives considered by users and further application of the zonal inference model. This way, two sequential tasks had to be accomplished: (i) decide which stops were connected by access arcs to each zone; and (ii) assign each one of them a distance value.

The first was solved by connecting centroids to bus stops and metro stations that were inside the zonal boundaries, but also inside an imaginary extension of its borders defined as the influence radius r (Figure 1). Influence radii were obtained from survey data, allowing different walking tolerance towards each mode, resulting in a measure of 250 metres for bus stops and 750 metres for metro stations. These values allowed that 95% of surveyed travellers use bus stops (and stations) connected to their origin and destination zones, without creating an excessive amount of access arcs that leads to extremely long execution times in the computation planning tool.

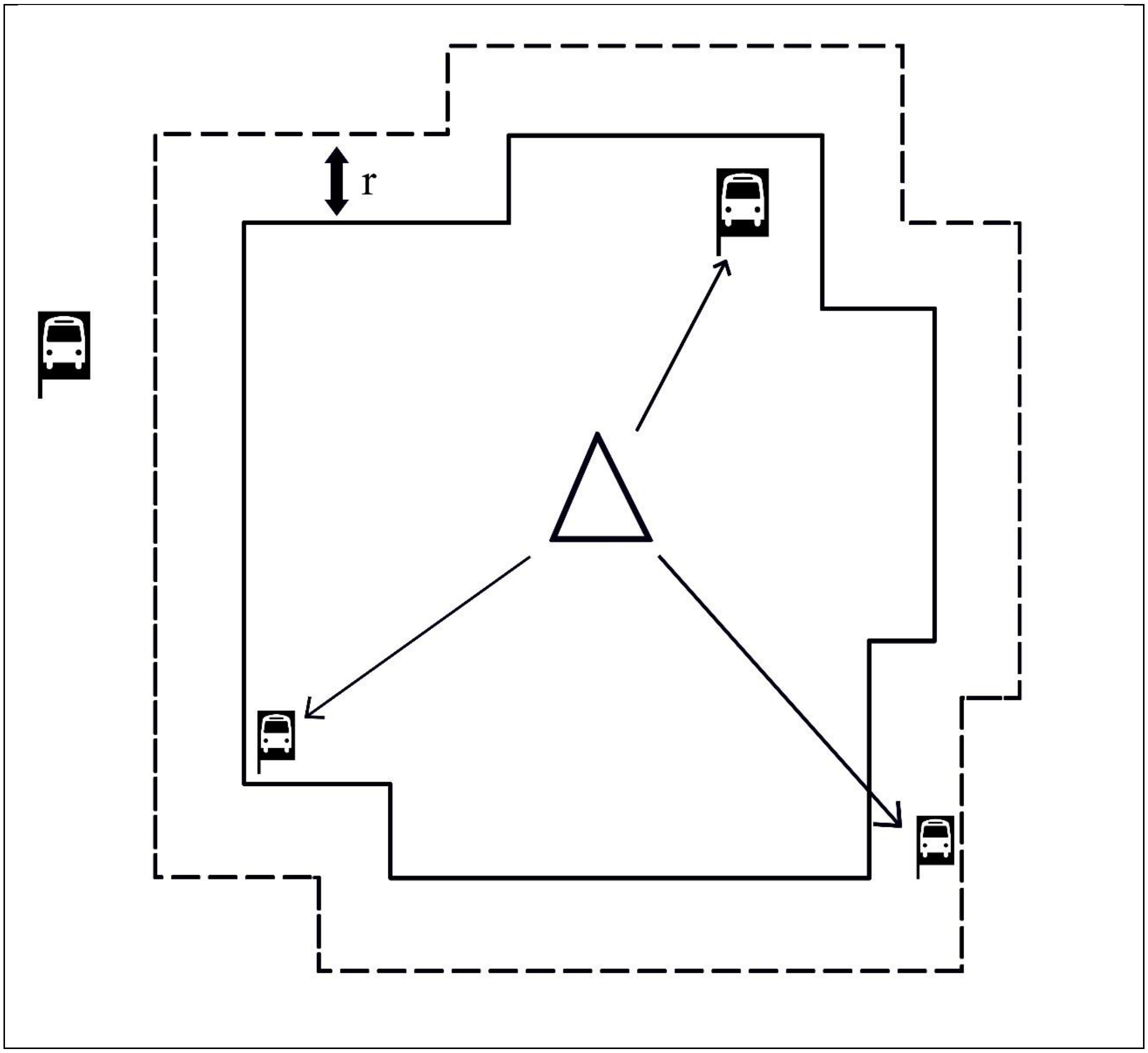

Figure 1: Access arcs construction criteria

Once defined which zones were connected to which stops (both for the software and zonal inference model), we proceeded to assign a distance value to each access arc. In this study, access distances were calculated considering land use variables from the city blocks of each zone. The purpose was to obtain accurate values for access times as each zone is not necessarily homogenous and, for instance, population may be concentrated in particular areas of the zone.

This way, each city block M was assigned a point in its geometrical centre with two different weights linked to land use characteristics; one for its potential as a trip generator (w_M^g , linked to residential land use, i.e., housing squared metres), and the other as a trip attractor (w_M^a , linked to commercial, industrial, residential, health, educational, and offices total squared metres). It is worth highlighting that the relative attraction importance of each land use type was obtained from preliminary zonal inference models, through an iterative procedure (rapidly convergent) in which the model attraction parameters were used to build new access arcs, which in turn lead to new models with updated parameters.

Then, distances are measured from each one of these points to each public transport detention inside the zone; finally, weighted averages of these distances are calculated using the generating and attracting weights of each city block. Thus, two distances are calculated for each detention: one corresponding to the generator link and other to the attractor link of the zone. Note that this methodology allows access time from zone i to station k to be different from the access time from station k to zone i, giving a better representation of heterogeneous zones (where population and trip attractors are located in different sectors).

4 FINAL MODEL AND ZONAL MATRIX ESTIMATION

Finally, only the survey entries in which the user chose stops connected by access arcs to her origin and destination zones were deemed valid. Consequently, the model was estimated with 569 valid entries (remember that 95% of the access walks are represented by the access arcs of the network). The best estimated model is presented next, which was obtained using BIOGEME (Bierlaire, 2003).

4.1 Final zonal inference model

Different specifications for the zonal inference model were estimated and compared, following the maximum likelihood procedure and a Multinomial Logit formulation, using a utility function of the general form:

$$V_{mn/kl} = ln(A_m) + ln(B_n) + \ln(f_{mn}^{kl}) \tag{9}$$

More specifically, the proposed functions were expressed in the following manner:

$$\begin{aligned} V_{mn/kl} = {} & \theta_{\mathrm{O}} * \ln\left(\theta_{u_{h_O}} * U_{h_{Om}}\right) + \theta_{\mathrm{D}} * \ln\left(\sum_P \theta_{m2_{P_D}} * M2_{P_{Dn}}\right) \\ & + \ln\left(\frac{e^{(\beta_O * d_{m-k})}}{(d_{m-k})^{\gamma_O}}\right) + \ln\left(\frac{e^{(\beta_D * d_{l-n})}}{(d_{l-n})^{\gamma_D}}\right) \end{aligned} \tag{10}$$

Where $\mathrm{M2}_{\mathrm{H}_{\mathrm{O_m}}}$ are the total housing squared metres in the zone of origin m; $\mathrm{M2}_{\mathrm{P_{Dn}}}$ are the total squared metres of land use classification P in destination zone n; $\mathrm{d}_{\mathrm{m-k}}$ and $\mathrm{d}_{\mathrm{l-n}}$ are access distances in kilometres from origin and destination zones to the chosen initial and final detentions, respectively; finally, θ, β, and γ are parameters to be estimated. The first and second sum terms are related to the generating and attracting powers of zones m and n, respectively; while the last two correspond to the inverse cost measure $\mathrm{f}_{\mathrm{mn}}^{\mathrm{kl}}$, where we used a gamma function and explicitly allowed different impacts of access and egress distances on utility.

It is important to note that, for identifiability reasons, we cannot estimate a parameter for each size variable in origin and destination, so two of them are fixed to a value of one. Furthermore, both θ_{O} and θ_{D} are fixed to unitary values, in order to ensure that the model is independent from the zonal system used for estimation (Daly, 1982). Table 1 shows the estimation results for this formulation.

Table 4.1: Zonal inference model

Associated to	**Parameter**	**Value**	**t test**
Generation	θ_O	1	Fixed
	$\theta_{m2_{H_O}}$ - Habitational m^2	1	Fixed
Attraction	θ_D	1	Fixed
	$\theta_{m2_{C_D}}$ - Commercial m^2	0.228	2.53
	$\theta_{m2_{E_D}}$ - Educational m^2	0.689	2.59
	$\theta_{m2_{H_D}}$ - Habitational m^2	0.0623	2.39
	$\theta_{m2_{I_D}}$ - Industrial m^2	0.0315	1.04*
	$\theta_{m2_{O_D}}$ - Offices m^2	0.198	1.93
	$\theta_{m2_{S_D}}$ - Health m^2	1	Fixed
Cost	γ_O	2.83	14.75
	β_D	1.71	1.96
	γ_D	2.48	5.24
Fit	**Null log-likelihood**	**-1,612.9**	
	Final log-likelihood	**-1,327.3**	
	Likelihood ratio test	**571.1**	

From results above we observe that all included parameters have the expected signs, and most of them are statistically significant; however, industrial land use attraction resulted non-significantly different from zero (at 90% confidence). In this respect, even though our survey entries were not sufficient to identify significant effects of this low density attractor, we decided to show it in the model, as it has the correct sign and is a commonly relevant variable in distribution models. Moreover, with respect to the attracting power part of the utility function, the model shows that the most attractive land use are health facilities and educational buildings, followed by commercial and offices buildings. Residential and industrial uses attract a relatively low number of trips. Regarding the inverse cost function, parameter β_O dropped out due to its low significance and the fact that it is not fundamental (cost function maintains the expected trend without it).

4.2 Zonal origin-destination matrix

Lastly, the zonal inference model was used to estimate a public transport zonal O-D matrix for Santiago's morning peak. Remembering Equation (7), we can obtain the trips between each zone i and j, using probabilities $\mathrm{Prob}\left({}^{ij}/_{kl}\right)$ (from the zonal inference model) to assign T_{kl} entries from the stop-to-stop trip matrix (presented in Section 3.1).

Figure 2 shows the total number of public transport trips generated and attracted (left and right, respectively) by each zone, for Santiago's morning peak period. In both maps, zones with higher number of trips are coloured in darker tones.

Trip origins map shows some of the generating centres of the city and a larger amount of trips starting from zones adjacent to the metro lines (where there is higher residential densities). However, this concentration is not as clear as the one produced by the attraction centres of the city. The destination map shows a distinct pattern of people travelling to the city's centre and north-east area.

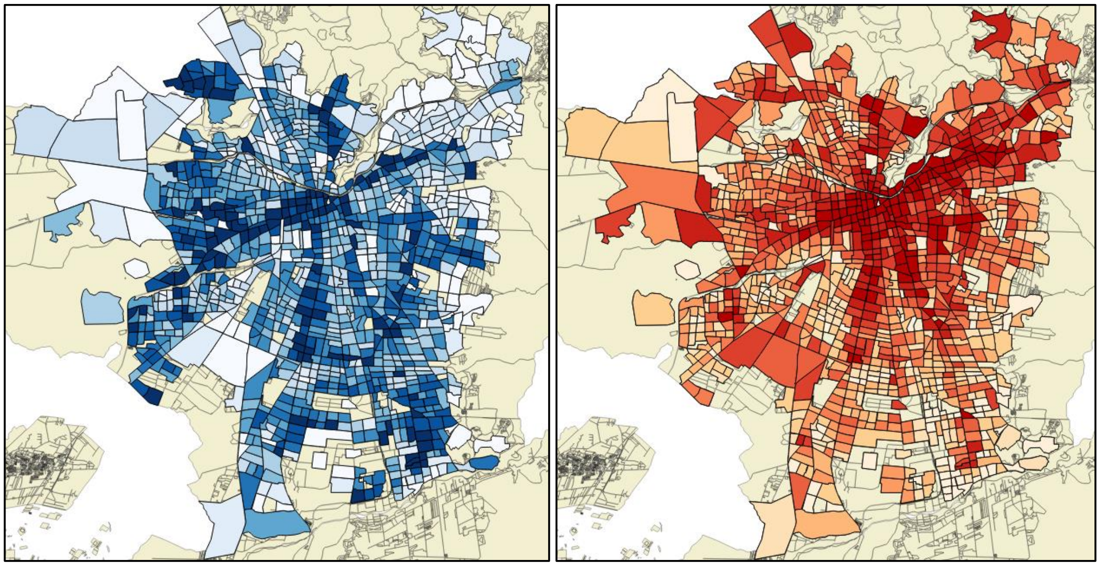

Figure 2: Total trips generated and attracted by zone, according to the model
Source: Prepared using QGIS

4.3 Alternative formulations

First, as advised in specialized literature, we analysed distinguishing travellers attracted to different destination alternatives. In this line, considering that students in Santiago are given special smart-cards (which allow them to pay lesser fares), we studied models in which student cards received a different treatment. This way, we included an interaction between the total educational squared metres and smart-card type, using a dummy variable with unitary value for student cards. The results of this model are shown in Table 2.

Results show that this differentiation allows us to separate the attraction effect of educational buildings to each traveller type. The attraction parameter associated to the interaction between student cards and total educational squared metres has a much higher value than the one obtained for regular smart-cards. Thus, we confirm that educational land use attracts students (or, more precisely, travellers using student smart-cards) more intensely.

It is worth noting that this model is presented in an exploratory manner and cannot be applied directly in the computational planning tool, because trip information included in it does not differentiate between smart-card types. However, Transantiago's payment system is capable of such distinction, allowing the possibility of separating this trips and apply a model of this kind in the future.

Table 2: Model with student cards differentiation

Associated to	Parameter	Value	t test
Generation	θ_O	1	Fixed
	$\theta_{m2_{H_O}}$ - Habitational m^2	1	Fixed
Attraction	θ_D	1	Fixed
	$\theta_{m2_{C_D}}$ - Commercial m^2	0.228	2.56
	$\theta_{m2_{E_D}}$ - Educational m^2	0.240	1.97
	$\theta_{m2_{Est-E_D}}$ – Educational m^2 students	2.93	2.03
	$\theta_{m2_{H_D}}$ - Habitational m^2	0.0515	2.35
	$\theta_{m2_{I_D}}$ - Industrial m^2	0.0332	1.07*
	$\theta_{m2_{O_D}}$ - Offices m^2	0.196	1.92
	$\theta_{m2_{S_D}}$ - Health m^2	1	Fixed
Cost	γ_O	2.83	14.75
	β_D	1.75	2.00
	γ_D	2.49	5.26
Fit	**Null log-likelihood**	**-1,612.9**	
	Final log-likelihood	**-1,313.6**	
	Likelihood ratio test	**598.5**	

On the other hand, one of the limitations of the model estimated for Santiago is that our surveys were taken in a period of time longer that the morning peak, for which we aim to estimate zonal origin-destination matrices. This consideration is worrying, because travel patterns in different periods are not necessarily equal and we could obtain imprecise models.

Consequently, we estimated models that allowed different attraction powers for each land use type and time period. Specifically, we included an interaction between the total squared metres built of each land use classification and survey hour, using a dummy variable with unitary value for observations taken before 9:00 am. The results of the best model estimation with this consideration are presented in Table 3.

From Table 3 we note that this model presents some parameters with very different values, compared to the ones presented in Section 4.1. Particularly, interactions between commercial and offices land use resulted significant, with lesser and higher attraction powers for the morning peak period, respectively. Furthermore, it is worth highlighting that our sample did not allowed us to capture statistically significant differences for educational, habitational, industrial, and health land uses. On the other hand, the attraction parameter for offices in morning off-peak period (after 9:00 am) presented a t-test of 0.70 and was discarded.

Table 3: Model with morning peak surveys differentiation

Associated to	**Parameter**	**Value**	**t test**
Generation	θ_O	1	Fixed
	$\theta_{m2_{H_O}}$- Habitational m^2	1	Fixed
Attraction	θ_D	1	Fixed
	$\theta_{m2_{C_D}}$ - Commercial m^2	0.393	2.49
	$\theta_{m2_{9am-C_D}}$ – Commercial m^2 morning peak	-0.341	-2.25
	$\theta_{m2_{E_D}}$ - Educational m^2	0.697	2.66
	$\theta_{m2_{H_D}}$ - Habitational m^2	0.0573	2.40
	$\theta_{m2_{I_D}}$ - Industrial m^2	0.0354	1.09*
	$\theta_{m2_{9am-O_D}}$ – Offices m^2 morning peak	0.359	1.87
	$\theta_{m2_{S_D}}$ - Health m^2	1	Fixed
Cost	γ_O	2.83	14.75
	β_D	1.76	1.99
	γ_D	2.51	5.24
Fit	**Null log-likelihood**	**-1,612.9**	
	Final log-likelihood	**-1,321.5**	
	Likelihood ratio test	**582.7**	

5 CONCLUSIONS

In conclusion, we have presented a methodology that allows us to estimate a zonal origin-destination matrix from an observed stop-to-stop O-D matrix. This methodology is applied to Santiago's morning peak, but it is completely general and could be used in any public transport system with sufficient information.

Another relevant contribution of this work is the development of an updated zonal system for Santiago, along with its centroids and access arcs, required in the computational planning tool. Access arc definition was accomplished using novel methodologies that benefits from survey data and attracting powers of each land use type (obtained from preliminary zonal inference models). Regarding potential improvements for the model, we highlight the opportunity of including differentiations between different smart-cards, survey time, or both. Furthermore, including new cost measures is expected to benefit the model. In order to maintain consistency, access arcs will be recalibrated considering these three effects, once the route and access station choice model is available in its definitive formulation.

Concerning the limitations of this study, it is important to note that the resulting zonal matrix is an assignment of the stop-to-stop trip matrix. In this sense, it does not explicitly include non-integrated modes (such as share taxis and private vehicles) nor fare evasion. There are also some limitations with the survey data used for model estimation, both in sample size and representativeness.

In this respect, a survey specifically designed for the purposes and period of this study would have been beneficial for the zonal inference model estimation. Alternatively, we suggest researching how to use the latest large scale origin-destination survey for Santiago (SECTRA, 2015) to re-estimate the model for morning peak and other periods.

Also, it is interesting to study the possibility of including new measures of the cost of choosing a stop, more complex and precise than the distance of its associated access arc. In this line, we propose for future research an alternative procedure in which f_{ij}^{kl} variables will be related to the probability that a trip originated in zone i with its destination in zone j uses access stations k and l, as their boarding and alighting points (defined as $\mathrm{Prob}\left({}^{kl}/_{ij}\right)$), since when a route alternative is perceived as costly the probability of choosing its related detentions must be low for a rational user (and vice versa for low cost routes).

This way, a route and access station hierarchical Logit choice model, currently under development (with preliminary results presented in Abud, 2015), can also be included in the calibration procedure. In this choice model, users decide which route and bus stops (or metro stations) they will use in order to travel between two particular zones.

Finally, we highlight the numerous policy implications of our research. The inclusion of the resulting matrix in the planning software will allow the evaluation of major changes to the transport system; such as route extensions, relocation of bus stops, or designing new bus or metro lines. Previously, transportation projects that modified the stop where users start or end their trips could not be properly evaluated, because trip assignment was made from a stop-to-stop matrix and, for example, it was not possible to know how many trips would start or end in a new metro station. So, the proposed methodology and model becomes a key ingredient of the public transport planning software for Santiago, allowing the analysis of trips in a new zonal level, resulting in better and more flexible predictions.

ACKNOWLEDGEMENTS

This research was supported by FONDEF D10I1049 “Una herramienta táctico-estratégica de gestión y planificación de sistemas de transporte público urbano”, and CONICYT-PCHA/MagísterNacional/2014-22141442 through the funding of Master Studies. The authors gratefully acknowledge the research support provided by the Centre for Sustainable Urban Development (CEDEUS, CONICYT/FONDAP/15110020) and the Bus Rapid Transit (ALC-BRT) Centre of Excellence funded by the Volvo Research and Educational Foundations (VREF).

REFERENCES

Abud, I., 2015. *Modelos de Elección de Paradero, Modo y Ruta para Herramientas de Planificación de Transporte Público*. Tesis de Magíster en Ciencias de la Ingeniería, Pontificia Universidad Católica de Chile.

Bierlaire, M., 2003. *BIOGEME*: A free package for the estimation of discrete choice models. *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

Daly, A. J., 1979. Some developments in transport demand modelling. In *Behavioural Travel Modelling* (Edited by D.A. Hensher and P.R. Stopher). Croom Helm, London.

Daly, A. J., 1982. Estimating choice models containing attraction variables. *Transportation Research Part B: Methodological*, *16*(1), 5-15.

Devillaine, F., 2012. *Estimación de Viajes y Actividades en Base a Sistemas Tecnológicos de Transporte Público*. Tesis de Magíster en Ciencias de la Ingeniería, mención Transporte, Universidad de Chile.

Gordon, J. B., Koutsopoulos, H. N., Wilson, N. H. M. and Attanucci, J. P., 2013. Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle Location Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2343 (1), 17-24.

McFadden, D., 1979. Quantative methods for analysing travel behaviour of individuals. In *Behavioural Travel Modelling* (Edited by D.A. Hensher and P.R. Stopher). Croom Helm, London.

Munizaga, M., Palma, C., 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.

Munizaga, M., Devillaine, F., Navarrete, C., Silva, D., 2014. Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies*, *44*, 70-79.

Ortúzar, J. de D., Willumsen, L. G., 2011. *Modelling Transport* (Fourth Edi.). Chichester: John Wiley & Sons, Ltd.

SECTRA, 2008. *Informe Final del estudio "Actualización y Recolección de Información del Sistema de Transporte Urbano, III Etapa".* Secretaría Interministerial de Planificación de Transporte, Santiago.

SECTRA, 2015. *Informe Final del estudio "Encuesta Origen Destino de Viajes 2012"*. Secretaría Interministerial de Planificación de Transporte, Santiago.

SII, 2014. *Roles de la Región Metropolitana de la Segunda Serie No Agrícola y Superficie Construida.* Datos obtenidos en virtud de la Ley Nro. 20.285. Servicio de Impuestos Internos, Santiago.