



## A topological route choice model for metro

Sebastián Raveau<sup>a,\*</sup>, Juan Carlos Muñoz<sup>a</sup>, Louis de Grange<sup>b,1</sup>

<sup>a</sup> Department of Transport Engineering and Logistic, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile

<sup>b</sup> Department of Industrial Engineering, Diego Portales University, Vergara 432, Santiago, Chile

### ARTICLE INFO

#### Article history:

Received 6 August 2010

Received in revised form 6 December 2010

Accepted 10 December 2010

#### Keywords:

Route choice

Network topology

Angular cost

Perceptions

Model specification

### ABSTRACT

This article presents a route choice model for public transit networks that incorporates variables related to network topology, complementing those found in traditional models based on service levels (travel time, cost, transfers, etc.) and users' socioeconomic and demographic characteristics (income level, trip purpose, etc.). The topological variables represent concepts such as the directness of the chosen route and user knowledge of the network. For both of these factors, the necessary data is endogenous to the modelling process and can be quantified without the need for information-gathering beyond what is normally required for building route choice models. Other novel variables in the proposed formulation capture notions of user comfort such as vehicle occupancy rates and certain physical characteristics of network stations. We conclude that these new variables significantly improve the explanatory and predictive ability of existing route choice specifications.

© 2010 Elsevier Ltd. All rights reserved.

### 1. Introduction

The purpose of this article is threefold: to advance our understanding of the behaviour of public transit users when choosing a route in a transit network, to quantify the impacts of the underlying factors that influence their decisions, and to improve the statistical inference and predictive ability of route choice models.

The route choice factors normally included in these models relate to the service levels of the route alternatives (e.g., in-vehicle travel time, waiting time, access time, number of transfers, train passenger density, etc.) and the socioeconomic and demographic characteristics of users (income level, purpose of trip, etc.) (Ortúzar and Willumsen, 2001).

However, there are good reasons to suspect the existence of other influences on the user's route choice process that existing models generally ignore. With this possibility in mind, we propose a new model based on three main hypotheses:

1. Transit users tend to penalize routes that deviate from a direct path to the final destination along certain segments. In other words, between two routes with exactly the same trip time, users prefer the most direct one, whether in geographical or topological (display map) terms.
2. Transit users tend to prefer routes that are better known or more heavily travelled.
3. Transit users consider different factors when choosing travel routes, besides traditional variables such as travel time or fare. Some of these factors may include comfort, reliability or physical characteristics from the vehicles and stations.

To test the first hypothesis our model defines and incorporates a novel explanatory variable called angular cost to represent the underlying effects of route segment deviation on user route choice in a transit network. The function expressing this

\* Corresponding author. Tel.: +56 2 354 4270; fax: +56 2 553 0281.

E-mail addresses: [sraveau@puc.cl](mailto:sraveau@puc.cl) (S. Raveau), [jcm@ing.puc.cl](mailto:jcm@ing.puc.cl) (J.C. Muñoz), [louis.degrange@udp.cl](mailto:louis.degrange@udp.cl) (L. de Grange).

<sup>1</sup> Tel.: +56 2 676 8118; fax: +56 2 676 8130.

variable is such that the more indirect is a given route, the higher will be the value of its angular cost variable. If the parameter of this variable is found to be statistically significant, the first hypothesis is confirmed. Various definitions and specifications of route directness and angular cost penalties may be found in Montello (1991), Tversky (1992), Turner (2001), Conroy-Dalton (2003).

The angular cost of a route is obtained directly from a map of the transit network. Here, however, an important issue arises, for the network maps published and displayed by transit systems typically incorporate design considerations that may distort the true geographical coordinates or spatial relationships of the network's stations. This implies that different map designs will generate different angular cost values and as a result may vary in their effects on user behaviour. By including the angular cost variable we are therefore also attempting to explain how travel decisions are affected by the way alternative routes are visually presented to the user.

To test the second hypothesis we need a variable that indicates how knowledgeable users are of a given route. Since this factor cannot be determined exactly, we will employ a variable measuring the use levels of the different routes as a proxy. Thus, our model is an endogenous formulation in which the probability of a route choice depends in part on how heavily used are its constituent segments.

To test the third hypothesis our model includes different variables related with level of service that are not usually included in route choice modelling. These variables are related to comfort on the vehicles, facilities on the stations such as escalators and level of usage, among others. The inclusion of this kind of variables allows the analysis of how the environment's characteristics impact the way alternative routes are perceived and, ultimately, the choices.

In the literature there are multiple route choice models based on users' socioeconomic and demographic characteristics and their perceptions of route attributes (Daganzo & Sheffi, 1977; Ramming, 2001; Prashker and Bekhor, 2004) such that user behaviour is determined by perceived costs. However, the attributes included in these models are all tangible and easily justified as significant factors in a rational individual decision-making process. Furthermore, the data on user characteristics are normally gathered through user surveys or through network measurements so that the models are completely exogenous. The result is that the choices they generate display considerable variability, a reflection of the inadequacy of the data incorporated by the modeller on the factors affecting users' route decisions. Needless to say, any additional data that could help explain this variability would be very welcome.

A possible source of such data is suggested by the well-established fact that transit user decision-making is affected by psychological considerations such as aesthetics, comfort and travel-time reliability (see Papinski et al., 2009). However, there are major difficulties inherent in integrating these factors into route choice modelling that stem from: (i) their subjectivity, given that each user perceives them differently; and (ii) their intangibility, since there is no scale for measuring them. Although some progress has been made in accommodating these phenomena in mode choice models (Ben-Akiva et al., 2002; Raveau et al., 2010) their applications in route choice contexts remain limited (Prato et al., 2009b).

In the present study we aim to improve the explanatory and predictive abilities of route choice models by incorporating variables that capture angular cost and route knowledge level. Our proposed model is a logit multinomial formulation that explicitly adds these novel explanatory variables to the ones traditionally found in existing designs. It is calibrated with data from the Santiago Metro system and will be compared to a base model that does not include angular cost, network knowledge or variables related to intangible factors of the level of service. Our conclusions will be that the added variables significantly improve the ability to explain and predict route choice processes in transit networks.

## 2. Specification of the model

The Santiago Metro system has a number of origin–destination pairs that can be travelled by more than one feasible route, as is evident from the map of the network (see Fig. 1a). To test our route choice model, therefore, we empirically analyze how system users travelling these pairs choose the transfer point or points between their origin and destination stations. The analysis is based on trips made during the morning (7am–9am) and evening (6pm–8pm) peak hours. During these periods users take approximately 700,000 trips, 44% of which include transfers.

Trip data were obtained from an origin–destination survey conducted at Metro stations in October 2008. Information was gathered on the peak-hour trips of 92,800 individuals, or about 12% of all users. Only those whose origin–destination pairs could have been taken by more than one route are included in our analysis. There were 16,029 such cases, amounting to about 40% of the individuals surveyed who transferred at some point on their trip.

Two different depictions of the Metro network are given in Fig. 1. The design on the left shows the true topology of the various stations and lines (georeferenced data) while the one on the right, which is the map actually displayed in the stations and consulted by users, contains serious distortions both in the actual location of the stations and the distances between them. This means that users' angular and geometric perceptions also diverge from the real ones, which may induce them to choose routes that are not the optimal or lowest cost ones. Though the extent of this distortion has increased with the considerable growth of the Santiago system over the last decade, it is still a long way from the complexity and high density of the networks in the major cities of Europe and the United States where the angular and geometric distortions of the display maps are much more significant (see, for example, the London system maps in Appendix A). The scale of the potential impact in these systems lends particular significance to the study of how graphical presentations of network information effect user behaviour.

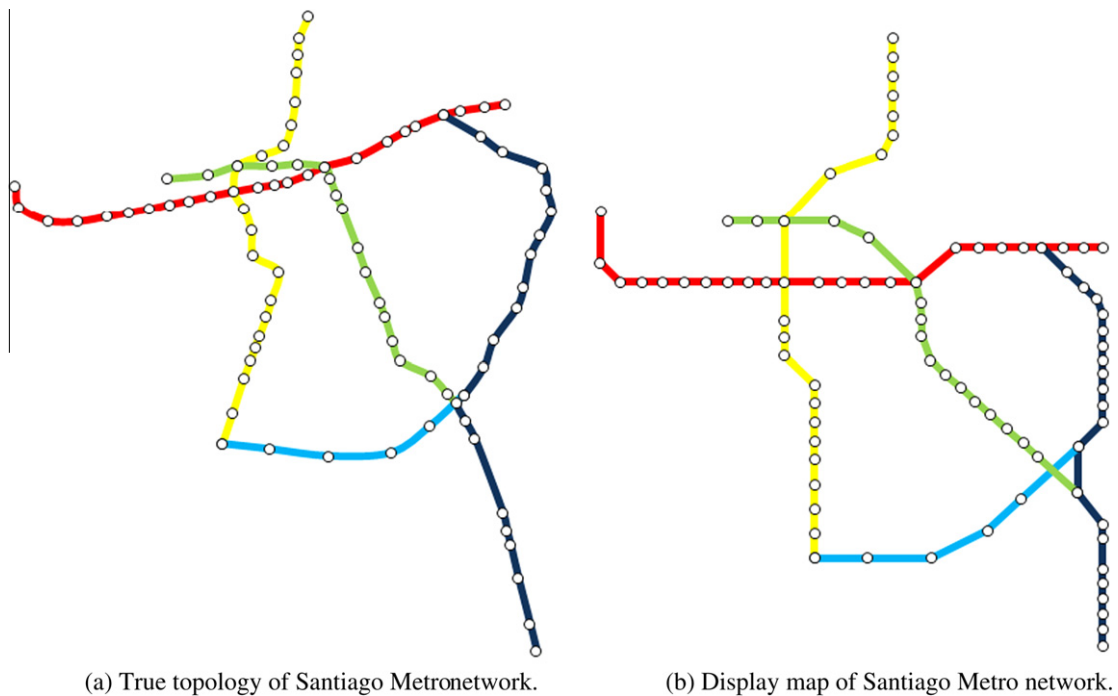


Fig. 1. Santiago Metro network maps.

The Santiago Metro network has five lines and 85 stations, 7 of which are transfer points. Of the 7140 possible origin–destination station pairs, 4985 (70%) require transferring between lines. For 1365 of these there is more than one route that is “reasonable,” by which we mean a route that was taken by at least one user in our survey database. The origin–destination pairs with more than one reasonable route are broken down in Table 1 by the number of such routes. Although in the majority of cases the route choice is between just two reasonable alternatives, there are also some pairs for which three and even four different options were chosen by some of the individuals surveyed. In denser transport networks, both the quantity and percentage of origin–destination pairs with alternative route possibilities would no doubt rise.

### 2.1. General description of the model

As just noted, our route choice model is a logit multinomial design in which the user utility function contains the standard explanatory variables representing network service levels plus additional ones for angular cost and route knowledge levels. Since the socioeconomic and demographic variables usually included are not relevant for the purposes of this study, they have been left out of our specification. Nor have we incorporated ticket prices since the Santiago Metro uses a flat-fare system. The variables are discussed in detail below.

The base model used for comparison purposes is a restricted or nested formulation containing only traditional service level variables (that is, without angular cost, route knowledge and variables related to intangible factors of the level of service). The estimation of the parameters is performed using the maximum likelihood method.

### 2.2. Explanatory variables

The standard explanatory variables in our model are *in-vehicle travel time* from origin to destination, total *waiting time* at origin and transfer stations, and the *number of transfers* for each route alternative as an indicator of additional (walking) time

**Table 1**  
Origin–destination pairs in Santiago Metro with more than one reasonable route alternative.

Observed route alternatives	% of O–D pairs	% of trips
2	97	93
3	3	7
4	<1	<1

between platforms and the disutility of changing train lines. No fare variable is included since under the Metro's flat-fare policy, user monetary cost is the same regardless of route length or transfers and thus has no effect on route choice.

To strengthen the explanation of line changes provided by the number of transfers reported, three explanatory variables were added to explicitly represent line-change characteristics: *walking time* between different train lines, whether or not platforms are accessed by *escalators*, and whether or not a given transfer involves an *ascending level change*. Data for these variables were collected via a field survey of transfer stations.

Also, since detailed information was available on the load profiles of each Metro line for different times of day, we were able to include an explanatory variable indicating the *average occupancy rate* for the different routes. This variable represents the degree of passenger crowding on train carriages, and indirectly also captures train capacity restrictions and the probability a user is unable to board the first-arriving train due to overcrowding and has to await the next one (a frequent occurrence on the Santiago Metro at peak hours). The variable itself is the average occupancy level (i.e., the ratio of load to capacity) of the arcs constituting the alternative routes, and is weighted by the route distances. By definition the rate can vary between 0 (for a train travelling empty along all arcs of a route) and 1 (for a train travelling fully loaded along all arcs).

Two more variables reflecting factors related to extreme occupancy rates at origin and transfer stations are also integrated into the model. If the rate is greater than or equal to 85%, not only is comfort reduced due to the high passenger density but there is a chance users will be *unable to board* the first train. If, on the other hand, the rate is less than or equal to 15%, chances are that users can *ride seated*. Thus, for the stations where passengers board, the two percentages are the occupancy rate thresholds at which user satisfaction changes. These particular values were chosen based on two criteria: (i) they reflect reasonable ranges given the various factors involved; (ii) they provide the best fit of the model to the data. Note also that the “unable to board” and “ride seated” variables are both dummies.

As regards network or route knowledge levels, since the Santiago Metro was developed in various stages over the last few decades (the oldest line being inaugurated in 1975 while the newest one opened in 2006), familiarity with the different route possibilities varies considerably and users tend to prefer the ones they know best. In addition, the more a given line is used the more likely it is to be better known than the others. To incorporate this effect in the proposed model we defined a *network knowledge* proxy variable that is simply the average passenger volume on each route during peak hours.

As regards the network topology factor, we assume users prefer the most direct routes from origin to destination. To reflect this preference, and to facilitate our evaluation of the effects of network map distortions of true route geography on route choice, we define the *angular cost* variable for each route as a penalty indicator that must satisfy the following three characteristics:

1. The penalty is at a minimum for route segments that head directly toward the final trip destination (i.e., the angle with respect to the destination is  $0^\circ$ ), and at a maximum for segments heading in the opposite direction (angle =  $180^\circ$ ).
2. Marginal changes at the extremes ( $0^\circ$  and  $180^\circ$ ) generate small penalty variations whereas a large variation is produced at the point where a route segment turns toward or away from the destination (this occurs when the angle is  $90^\circ$ ).
3. The penalty is angularly symmetric in the sense that if, for example, the angle of a segment is  $15^\circ$ , the penalty is the same as for an angle of  $345^\circ$ .

A penalty function that has all three of these characteristics is  $\sin(\theta/2)$ , where  $\theta$  is the angle formed by a straight line from the initial point of a segment to the final trip destination and another straight line from the initial point of the segment itself. An example of the estimation of angular cost is given in Fig. 2. The penalty for a segment is weighted by its length to reflect how far the deviation distances users from their destination. In the case of the two Santiago Metro network maps, note that the angular costs for the routes as depicted in the true topology version (Fig. 1a) will be different than those for the Metro display map (Fig. 1b). This factor can therefore be tested separately for the two designs.

One of the characteristics of our angular cost specification is that it is not topologically symmetric. In other words, the angular cost of travelling from O to D is not the same as for travelling from D to O. This is consistent with the structure of route choices observed in the Santiago Metro. A dispersion graph of the proportions of route choices in both directions

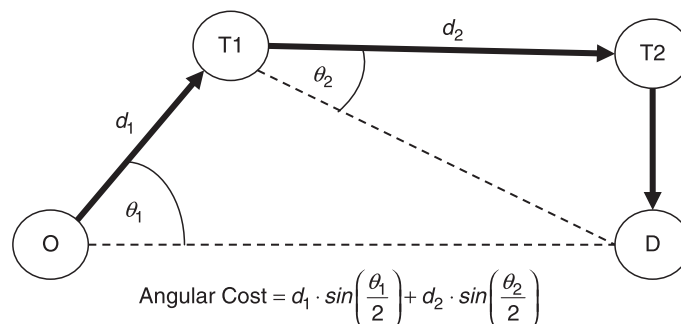


Fig. 2. Angular cost estimation.

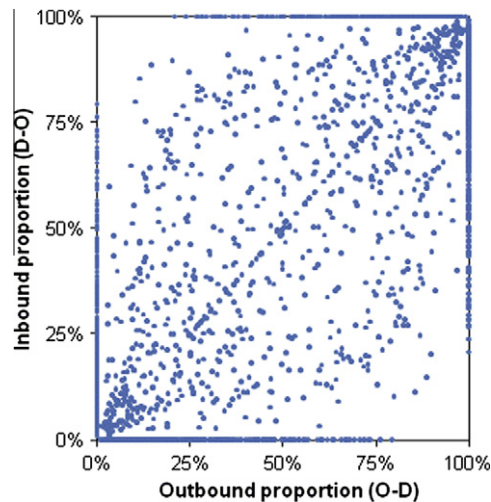


Fig. 3. Asymmetry of user route choices.

for all O–D pairs with at least 50 observed trips in our survey sample is set out in Fig. 3. For each route between two stations that has more than one possible route, the graph shows the proportion of outbound (O–D) trips versus that of inbound (D–O) trips. If user route choices were symmetrical, the points on the graph would tend to be concentrated around the 45° diagonal. But as can be seen, this is not the case. The absence of topological symmetry is thus consistent with actual user route choices and is therefore not a shortcoming of the proposed specification.

Finally, and also in relation to network topology and the angular cost concept, we incorporated two dummy variables first defined by Dial (1971) that capture another relevant aspect of route geometry. The first variable identifies routes with a transfer station (if any, otherwise the destination) that is closer to the origin than the station immediately before it, creating the impression of *turning back to the origin*, while the second one identifies routes that have a transfer station further from the destination than the station immediately before it, giving the impression of *turning away from the destination*.

### 3. Results

Based on the data gathered on route choice in the Santiago Metro network we estimated both our proposed logit multinomial model, built around the complete set of explanatory variables just described, and a base model (which only considers the standard explanatory variables) for comparison purposes. The sample consisted in route choices of 28,961 users on peak hour, over 1365 different OD pairs. The number of route alternatives varied from 2 to 4, as presented in Table 1. The results of the estimation are summarized in Table 2. Note that two versions of the proposed model were solved, one based on angular cost for a true topological map of the Metro and the other for the distorted display map.

As may be observed, all of the parameter estimates for the proposed models have the right sign and are statistically significant at the 95% level. Of particular interest is the result for the network knowledge variable, measured by a route's aver-

Table 2

Estimates of route choice models, showing parameters and critical values.

Variable	Base model	Proposed model (true distances)	Proposed model (display map distances)
In-vehicle travel time	−0.144 (−46.61)	−0.119 (−27.94)	−0.117 (−27.29)
Waiting time	−0.203 (−3.68)	−0.111 (−3.17)	−0.121 (−3.17)
Walking time	–	−0.240 (−7.14)	−0.229 (−6.87)
Number of transfers	−1.220 (−17.06)	−0.449 (−3.95)	−0.420 (−3.69)
Ascending level change	–	−0.426 (−7.89)	−0.432 (−8.02)
No escalators	–	−0.438 (−11.95)	−0.457 (−12.34)
Occupancy rate	–	−1.430 (−3.53)	−1.250 (−3.05)
Unable to board	–	−0.426 (−6.00)	−0.413 (−6.06)
Ride seated	–	0.126 (2.16)	0.106 (1.81)
Network knowledge	–	0.031 (7.14)	0.030 (6.81)
Angular cost	–	−0.024 (−7.14)	−0.038 (−6.75)
Turn back to origin	–	−0.516 (−10.96)	−0.503 (−10.65)
Turn away from destination	–	−0.505 (−11.43)	−0.512 (−11.59)
Sample size	28,961	28,961	28,961
Log-likelihood	−7416	−7142	−7136

age passenger volume, whose effect is positive. This being so we could easily specify an endogenous or fixed-point route choice model in which the probability of choosing a particular route depends on the number of users choosing it. Such a formulation would be needed for use in making predictions.

Regarding angular cost and the variables on turning back to the origin or away from the destination, the results are more statistically significant when the proposed model is constructed using the distorted display map distances (Fig. 1b) as opposed to the true topological distances (Fig. 1a). This strongly indicates the significance of the way in which route information is presented to users. To the extent the display map distortions induce users to choose certain routes over others, this phenomenon could be exploited to modify the use of certain network lines or route segments.

Another point demonstrated by these results is that the traditional explanatory variables travel time, waiting time and number of transfers, although certainly important (travel time is the most statistically significant), do not fully explain the decision-making process. Indeed, the absence of the information represented by the proposed new variables may bias the parameter estimates of the traditional ones (note the large difference in the waiting time and number of transfers parameters).

Also confirmed by the results in Table 2 are the two hypotheses stated in Section 1: users perceive an additional cost for routes that have segments which turn away from the destination, and tend to choose the routes they are most familiar with or are heavily used by other users.

In light of the foregoing, the inclusion of non-traditional variables is clearly needed. A particular advantage of adding factors relating to network topology such as those suggested here is that the necessary information is readily obtained at low cost as part of the usual data-gathering process for building route choice models.

As regards goodness of fit, the proposed model, with 10 more explanatory variables than the base model, delivers a significant improvement over the latter of more than 200 points in log-likelihood. This leads us to reject the null hypothesis that the two specifications are equivalent ( $2 \times 280 = 560 > \chi^2 = 23.2$  at 99% confidence with 10 degrees of freedom). The mean square error, given in Table 3, indicates that the proposed model is also superior to the base model in terms of predictive ability. The prediction error for both models increases as the number of alternative routes grows, but so do the differences between them.

To quantify the improvement in predictive ability brought about by the new variables in the proposed model, we analyzed the route choices between a group of 11 stations on a single line of the Santiago Metro and another group of nine stations on a different line (this is, a set of 198 O–D pairs). According to our data, trips taken between the two groups during peak hours numbered some 7300 and followed four different route patterns (see Fig. 4). This example is particularly suitable for our purposes because the four routes use all seven transfer stations in the Metro network and can be clearly described in topological terms, which are given below:

1. Route 2 is the most direct route and has no segment that turns back to the origin or away from the destination.
2. Route 1 has the highest angular cost on certain segments, especially between transfer stations 1 and 2. However, the segment between transfer station 2 and the destination is relatively direct.
3. Routes 3 and 4 are less direct and require transferring at stations far from the destination. However, they have lower occupancy rates, particularly Route 3, and Route 4 is the fastest one to arrive at the destination line (but not necessarily the fastest one to arrive at the destination station).
4. Routes 1 and 2 both pass through transfer stations 1 and 3. On Route 2 the section between those two stations is direct but the occupancy rate is the highest of any part of the network. On Route 1 that leg is made through transfer station 2, but the occupancy rate is lower and there are fewer intermediate stations (i.e., lower travel time).

The trip assignment results for our application example are summarized in Table 4, showing: (i) the observed route choices, (ii) the route choices predicted by the base model, (iii) the route choices predicted by the proposed model, and (iv) the minimum-time assignments (an all-or-nothing loading in terms of travel, waiting and walking times). Note that the minimum-time assignments do not assign the entire flow to a single route. This occurs because train frequencies during the peak periods for the different routes vary and no single route is always the fastest for all of the O–D pairs.

As can be seen in Table 4, Route 2, the most direct route, is generally the fastest and therefore the most frequently used. Nevertheless, the trips assigned to Route 2 in the proposed model are fewer than those assigned considering minimum time due to a high occupancy rate, which is perceived as a negative attribute. Route 1, though faster than Route 4, is used considerably less frequently, because the segment between transfer stations 1 and 2 has a high angular penalty and turns away

**Table 3**  
Mean square error, route choice models.

Origin–destination pairs	Base model	Proposed model
All	13.3	10.9
With 2 routes	7.9	6.7
With 3 routes	20.0	10.7
With 4 routes	27.8	15.9

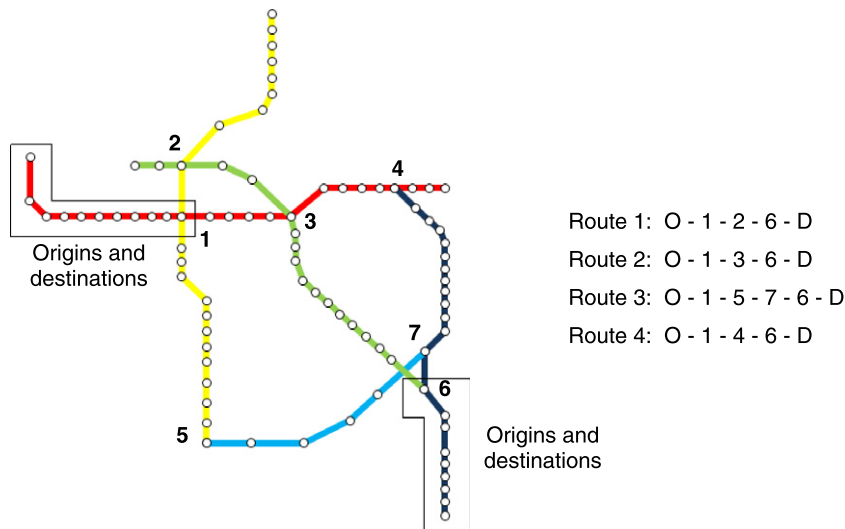


Fig. 4. Model application example.

Table 4

Trip assignment for application example.

Assignment	Route 1	Route 2	Route 3	Route 4
Observed trips	99	5362	30	1854
Minimum time	668	6676	0	0
Base model	361	3884	33	3067
Proposed model	270	4446	25	2603

Table 5

Marginal rates of time substitution.

Variable	Base model	Proposed model
In-vehicle travel time	1.00 (base)	1.00 (base)
Waiting time	1.41	1.07
Walking time	–	1.79

from the destination. The proposed model generates an assignment that is closer to the observed trips than that of the base model, thus demonstrating the proposed model's superior predictive ability.

Although our model does not produce monetary valuations for the various attributes (e.g., the subjective value of time) given that no cost variable was included due to the Metro's flat-fare system, marginal rates of substitution between the various time factors can be derived using the in-vehicle travel time as the baseline. The values so obtained are presented in Table 5.

Waiting time in the base model is 41% more highly valued than in-vehicle travel time whereas in the proposed model, the two variables are similarly valued statistically. In the case of walking time, not included in the base model, the proposed model values it 79% more than in-vehicle travel time. Also notable is that the waiting time value from the base model is close to the average of the walking time and waiting time values from the proposed model. The explanation for this last result may be that in the base model the waiting time value captures part of the walking time effect, reflecting the fact that the two variables are correlated since they are both directly related to, and simultaneously increased by, transferring. As a result, the base model parameters are biased due to the absence of significant variables that are incorporated in the proposed model.

As for transfers, the base model can only generate a single value of 8.5 travel minutes per transfer. In the proposed model, however, the components associated with this factor are disaggregated and values vary depending on the transfer station. The valuation for walking time between line platforms was given in Table 5 while the valuations for the specific station characteristics are indicated in Table 6. The valuation the different transfer cases is obtained by considering all the parameters related to each scenario (i.e. the base parameter associated to transfer and the marginal parameters associated to the possibilities of boarding or getting a seat, the possibility of using escalators and the possibility of descend/ascend when trans-

**Table 6**

Transfer valuations by transfer station characteristic.

Characteristic		Without escalators	With escalators
Chance of riding seated (occupancy $\leq$ 15%)	Descending	2.7 min of travel	6.6 min of travel
	Ascending	6.4 min of travel	10.3 min of travel
Intermediate (15% < occupancy < 85%)	Descending	3.6 min of travel	7.5 min of travel
	Ascending	7.3 min of travel	11.2 min of travel
Chance of being unable to board (occupancy $\geq$ 85%)	Descending	7.9 min of travel	11.2 min of travel
	Ascending	11.0 min of travel	14.9 min of travel

ferring). These figures can be interpreted as the additional travel time users will accept to avoid a transfer. Due to its omission of significant information the base model delivers an average transfer value that does not resemble any of the 12 possible cases valuated by the proposed model.

Finally, many public transit systems experience serious problems of congestion during peak periods and the Santiago Metro is no exception, exhibiting high load factors during the hours of heaviest use over much of the network. As has been demonstrated, the way in which map-based network information is presented to users can induce them to change their route choice behaviour. Such changes can be positive, tending to decongest overloaded lines, or negative, overloading them even further. The design of graphic information displays is thus an important issue that must be handled carefully in order to ensure the effects on user behaviour are the desirable ones.

#### 4. Conclusions

Route choice modelling is traditionally based only on tangible explanatory variables. However, the perceptions of transport users regarding available route alternatives are such that they do not always choose what the modeller would consider as the “lowest cost” option. In this study we have demonstrated and quantified the influence of non-traditional factors that also impact individual user decision-making. Specifying certain aspects of the trip environment (e.g., passenger densities and the physical characteristics of Metro stations) is also shown to improve the explanation of route choices.

Among the non-traditional variables incorporated in our proposed model are a number that relate to the topology of the transit network, representing the angular cost (i.e., directness) of the various route alternatives and their geometry (whether a segment turns back to the origin or away from the destination). One of the main advantages of adding these factors is the low cost of collecting the corresponding data. Other novel variables included in our specification express user knowledge of the network’s route alternatives.

The results generated by our model lead us to conclude that when the non-traditional variables are absent the estimated parameters are biased. This is manifested in inaccurate marginal rates of substitution and route choice models with insufficient explanatory and predictive abilities. It is therefore highly recommendable that field surveys be undertaken of networks’ physical characteristics (e.g., certain aspects of Metro stations) and operating features (e.g., user information) in order to obtain the data necessary for specifying these proposed new variables in route choice formulations that will produce more accurate estimators and better predictions.

Finally, we have shown that the way in which network information aimed at users of a transit system is presented can influence their decision-making. Distortions in the representation of a network’s geographical relationships can induce trip assignments that reduce service levels for all users. At the same time, these distortions can be exploited to induce optimal trip assignments and thereby making the best use of transport system capacity.

In judging our findings it should also be recalled that the Santiago Metro is a small network with relatively few route alternatives for any given origin–destination pair compared to the systems of major cities such as London, New York or Madrid. The results in Table 3 show that as the number of alternative routes grows, the proposed model improves significantly compared to a standard route choice model without the additional variables introduced in this study. The superior predictive ability of our formulation should therefore be even more apparent with denser networks. The distortion of the true spatial relationships by the Santiago network’s display map, though by no means negligible, is not overly significant. For large and highly complex networks with many route alternatives, however, the distortions in these user information maps would take on considerable importance. Our findings could assist in orienting the design considerations for the maps these systems publish in order to induce socially optimal behaviour in their users.

A natural extension of the work presented here would be to use the proposed model for making predictions. This would require solving the fixed-point problem given that the variables relating to network knowledge and occupancy rates are endogenous to the modelling process.

#### Acknowledgments

This research was supported by the Across Latitudes and Cultures – Bus Rapid Transit Centre of Excellence funded by the Volvo Research and Educational Foundations (VREF), Project ACT-32 Anillos Tecnológicos of Conicyt and Fondecyt Iniciación 11100024.



## Appendix A. London underground network maps

See Figs. A.1 and A.2.

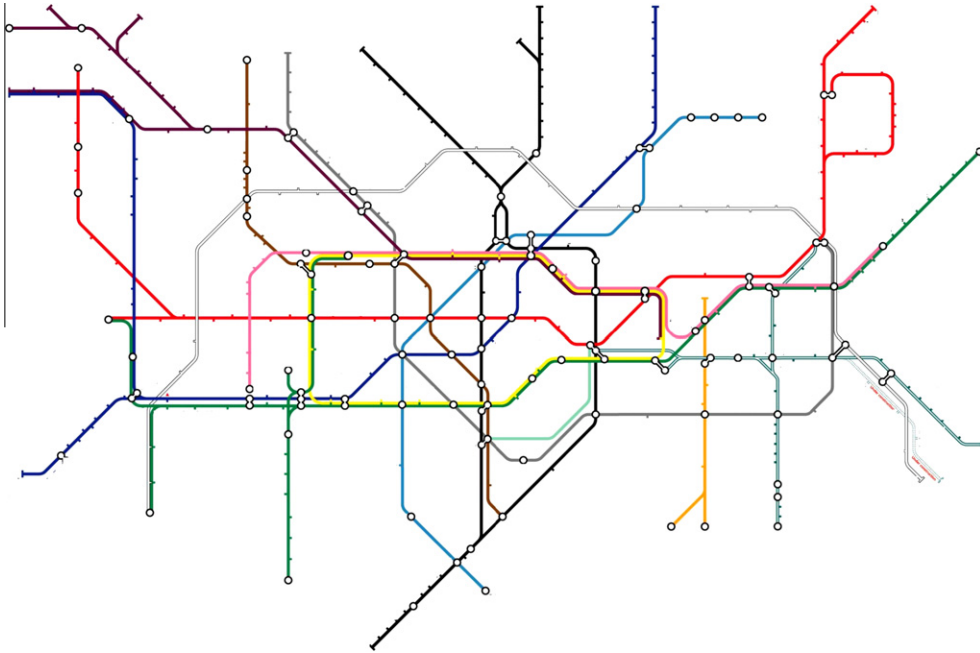


Fig. A.1. Display map of the London underground.



Fig. A.2. True topological map of the London underground.

## References

- Ben-Akiva, M.E., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 2002. Integration of choice and latent variable models. In: Mahmassani, H.S. (Ed.), *In Perpetual Motion: Travel Behaviour Research Opportunities and Challenges*. Pergamon, Amsterdam, pp. 431–470.
- Conroy-Dalton, R., 2003. The secret is to follow your nose: route path selection and angularity. *Environment and Behavior* 35, 107–131.
- Daganzo, C.F., Sheffi, Y., 1977. On stochastic models of traffic assignment. *Transportation Science* 11, 253–274.
- Dial, R.B., 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research* 5, 83–113.
- Montello, D.R., 1991. Spatial orientation and the angularity of urban routes: a field study. *Environment and Behavior* 23, 47–69.
- Ortúzar, J.de.D., Willumsen, L., 2001. *Modelling Transport*. John Wiley and Sons, Chichester.
- Papinski, D., Scott, D.M., Goherty, S.T., 2009. Exploring the route choice decision-making process: a comparison of planned and observed routes obtained using person-based GPS. *Transportation Research Part F* 12, 347–358.
- Prato, C.G., Bekhor, S., Pronello, C., 2009. Latent variables and route choice behavior. In: *12th International Conference on Travel Behaviour Research*, Jaipur, India.
- Prashker, J.N., Bekhor, S., 2004. Route choice models used in the stochastic user equilibrium problem: a review. *Transport Reviews* 24, 437–463.
- Ramming, M.S., 2001. *Network knowledge and route choice*. Unpublished Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Raveau, S., Álvarez-Daziano, R., Yáñez, M.F., Bolduc, D., Ortúzar, J.de.D., 2010. Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. *Transportation Research Record* 2156, 131–139.
- Turner, A., 2001. Angular analysis. In: *Proceedings of the 3rd International Symposium on Space Syntax*, Georgia Institute of Technology, Atlanta, pp. 30.1–30.11.
- Tversky, B., 1992. Distortions in cognitive maps. *Geoforum* 23, 131–138.