Proceedings Paper:

Nadudvari, T, Liu, R orcid.org/0000-0003-0627-3184 and Balijepalli, NC orcid.org/0000-0002-8159-1513 (2018) The reasonable route choice set in large and complex metro networks; an implementation of the K-shortest path algorithm for the London Underground. In: Proceedings of the 21st International Conference of Hong Kong Society for Transportation Studies, 10th-12th December 2016. HKSTS 2016 - Smart Transportation, 10-12 Dec 2016, Hong Kong. , pp. 247-254. ISBN 9789881581457

THE REASONABLE ROUTE CHOICE SET IN LARGE AND COMPLEX METRO NETWORKS; AN IMPLEMENTATION OF THE K-SHORTEST PATH ALGORITHM FOR THE LONDON UNDERGROUND

Tamas NADUDVARI [a], Ronghui LIU [b] Chandra BALIJEPALLI [c]
[a] Institute for Transport Studies, University of Leeds, United Kingdom
Email: tstn@leeds.ac.uk
[b] Institute for Transport Studies, University of Leeds, United Kingdom
Email: r.liu@its.leeds.ac.uk
[c] Institute for Transport Studies, University of Leeds, United Kingdom
Email: n.c.balijepalli@leeds.ac.uk

ABSTRACT

Numerous models have been developed to estimate route choice from smart card data. However they require route choice set as an input. In this paper we propose a method to determine that automatically. Firstly, we find the K shortest routes. Secondly, we set cut-off criteria to find the reasonable routes, by comparing the in-vehicle, wait and access egress interchange (AEI) cost of each route with the shortest route's. We work with 7 origin destination (OD) pairs of the London Underground for which more routes were surveyed. Among them 5 of them had similar properties and from that it resulted that a route is considered reasonable, if its cost components are up to 1.71, 1.45 and 1.26 times the shortest route's respectively. From the 2 outliers we understood that we need to include OD specific attributes when setting cut-offs: route with minimum interchange and lines available at origin and destination.

Keywords: metro network, route choice set, K shortest path, cut-off criteria

1. INTRODUCTION

Transport operators of large and complex metro networks need to know time period specifically the link flows and bottlenecks in the system; so they could give the appropriate response, both on the supply and on the demand side (Halvorsen, 2015). Recent research is towards methods and technologies to determine link flows in real time (Nuzzolo and Comi, 2016).

Estimating link flows from smart card data is advantageous as it has low marginal cost, can cover a large sample of passengers and it can be stored and processed automatically. The challenge is that it tells us only the entry/exit station, but not the chosen route. To know link flows, it is necessary to apply a transit assignment model (TAM) which estimates route choice of passengers and assign them to transit links.

The classical way to estimate route choice, is to use stated preference, based on the network attributes and on parameters coming from a calibration. (Dial, 1971, Ben-Akiva and Lerman, 1974, McFadden, 1974). Using revealed preference, such as smart card data, is more advantageous in many aspects, as we can rely on observations on what the passengers have actually chosen. Fu (2014) worked with the distribution of journey times between an OD pair, and decomposed it as a mixture of I_r probability density functions according to the I_r routes. Sun et al. (2015) observed the mean and variance of journey times and applied this as an additional input for a logit model. Paul (2010) and Zhu (2014) used disaggregate smart card data and train operator logs to assign every individual to a train. Holleczek et al. (2015) deployed a traffic measurement system and Poonawala et al. (2016) joined mobile phone and smart card data.

In our research we apply Fu (2014) in a TAM. In order that it could give correct results for the route choice probabilities, it requires the number of routes I_r as an input, which is not easy to know, as in a complex metro network, theoretically there are many possible routes, but only few (4-5) of them are considered in the choice set of passengers, we call this set here: the reasonable route choice set.

In this paper we present a method which determines automatically the set of reasonable routes for a given OD pair. Firstly we apply the K shortest path (Yen, 1971) to find the first K shortest routes. Secondly we set a cut-off criteria that can automatically find the I_r number of reasonable routes among them. We set the criteria based on route choice observations for 7 OD pairs coming from a passenger survey. We apply our proposed method on one of the most complex metro systems in the world, the London Underground (LU) network.

2. METHODOLOGY

2.1 The K shortest path algorithm

The K shortest path algorithm (Yen, 1971) first calculates the shortest path (Dijkstra, 1959) and then the second, third, … K-th shortest paths by removing one link from the shortest path and calculating the shortest path for that modified network. It considers only paths without loops. The purpose here is to apply an existing pathfinding algorithm to present the concept of finding automatically the reasonable route choice set, not to find the most one. Therefore we have chosen to work with a simpler algorithm, for which we had the program code available[1].

2.2 Case study network

We apply the K shortest path algorithm on the London Underground (LU), however we work on a simpler problem instead of the realistic network. We include only the Central London part of the network, without considering the common line problem. This simpler case study network has 9 lines and 57 stations. (Figure 1). To set the cut-off criteria for the reasonable route choice set, we select OD pairs for which more routes were surveyed with relatively larger sample of data. We work with the OD pairs in Fu (2014) (Figure 1, Table 1).

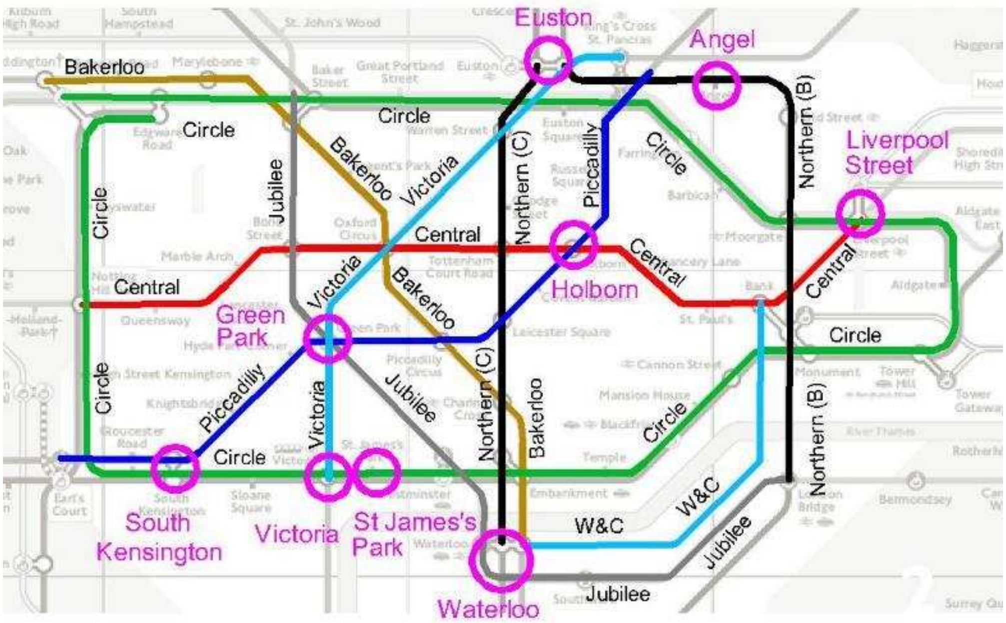


Figure 1 Case study network and the OD pairs of our case study

[1] http://uk.mathworks.com/matlabcentral/fileexchange/32513-K_shortest-path-yen-s-algorithm

Table 1 OD pairs of the case study

Case	Origin	Destination	RODS sample size
1	Victoria	Holborn	562
2	Euston	St. James's Park	437
3	Victoria	Liverpool Street	557
4	Angel	Waterloo	77
5	Liverpool Street	Green Park	196
6	Euston	South Kensington	209
7	Victoria	Waterloo	386

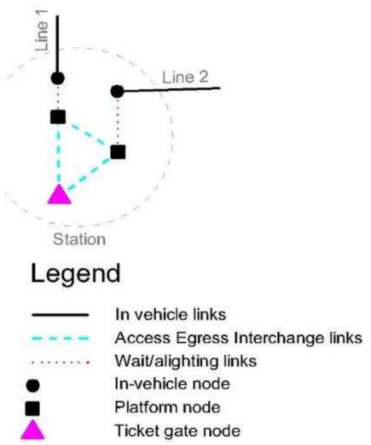
2.3 Data sources

To build our network model (2.4), for the in-vehicle times and headways we use timetable[2] data; and for the Access Egress Interchange (AEI) times we use a 4 week survey data conducted by TfL in 2011, called the AEI survey. For AEI movements that were not surveyed we estimated their time from the distances reported in The Nationwide Access Register (Direct Enquires)[3].

To set the criteria for reasonable routes, the results of the K shortest path algorithm were compared with a survey on passenger route choices (2.8), called the Rolling Origin and Destination Survey (RODS)[4], having observations from the period of 1998-2010. The sample size for each OD pair is reported in Table 1.

2.4 Network representation

We represent the LU network with in-vehicle, platform and ticket gate nodes in order to distinguish the different type of link costs (in-vehicle, wait and access egress interchange (AEI) time) in the model. At every station we define 1 in-vehicle node for every line, 1 platform node for every pair of platforms and 1 ticket gate node (Figure 2 a). The matrix of link costs is explained in Figure 2 b.



O\D	In vehicle nodes	Platform nodes	Ticket gate nodes
In vehicle nodes	In vehicle links	Alighting links (zero value)	INF
Platform nodes	Wait links	Interchange links	Egress links
Ticket gate nodes	INF	Access links	INF

Figure 2. a) Representation of a station with nodes and links
b) Allocation of link types in the matrix of link costs

[2] https://www.tfl.gov.uk/travel-information/timetables/
[3] http://www.directenquiries.com/londonunderground.aspx
[4] https://tfl.gov.uk/info-for/open-data-users/our-feeds

2.5 Notation

We use the following notation throughout the paper:
i: index for the i-th shortest route, n: index for OD pair
j: index for time/cost component: in-vehicle, wait and access egress interchange (AEI).
t_j: time of the j-th component [min]
C_j: generalised cost of the j-th component [min], C: total generalised cost of the route [min]
θ_j: parameter of the j-th cost component, IC: parameter for interchange experience [min]
R: set of routes, containing I number of routes ($r_1,\ldots,r_I$) with their cost components ($C_{j,1},\ldots,C_{j,I}$)
K, r, s, ns: indices for R, I and q meaning: K shortest, reasonable, surveyed and non-surveyed
$q_{j,i}$: the ratio between j-th cost component of the i-th route and of the shortest route: $q_{j,i}=\frac{C_{j,i}}{C_{j,1}}$
N: Number of OD pairs. N^*:number of OD pairs after taking out the outliers.

2.6 Generalised costs of links

In this paper we consider for the total generalised cost of a route:

$$C = C_{in\text{-}veh} + C_{wait} + C_{AEI} = t_{in\text{-}veh} + t_{wait}\theta_{wait} + t_{AEI}\theta_{AEI} + IC \quad (1)$$

From our data sources (2.3) we know t_j. For θ_j and IC we apply the results of a recent calibration done on the LU network (Raveau et al., 2014). Supposing that trips are done in weekdays, morning peak and with restrictive purpose, they are: $\theta_{wait} = 1.93$, $\theta_{AEI} = 1.30$. IC depends on the level and assistance of the interchange movement (Table 2).

Table 2 Applied parameters of interchange experience (IC) depending on the level and assistance of the interchange movement

Characteristics		IC
Level	Assistance	[min]
Ascending	Assisted	5.71
Ascending	Semi-Assisted	6.84
Ascending	Non-Assisted	7.32
Even	N/A	2.39
Descending	Assisted	4.87
Descending	Semi-Assisted	5.97
Descending	Non-Assisted	6.49

2.7 Implementation of the K shortest path algorithm on the case study network

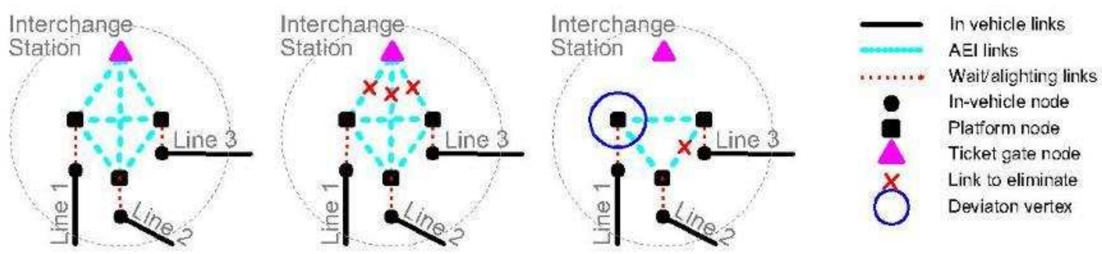


Figure 3 Modification of the K shortest path algorithm. a) Original network containing all AEI links, b) eliminating I links at O and D and AE links at all other stations c) if dev vertex at platform node, eliminate interchange links from other platforms

If we run the K shortest path algorithm on the network described in (2.2), the results will contain many routes which differ only in their AEI movements (Figure 3a). In order to avoid this, it is necessary that the algorithm could eliminate some of the AEI links automatically depending on the OD pair and the route. Firstly, depending on the OD pair, the modified algorithm eliminates interchange links at origin and destination stations as well as access and egress links at all other stations (Figure 3b). Secondly it checks the results of the Dijkstra (1959) algorithm, whether it gives the shortest possible interchange movement at every station. Finally, at every iteration, if the deviation vertex is set at the platform node of a station, it eliminates interchange links from other platforms in order to avoid that the next shortest path would be the same route with different type of interchange movement (Figure 3c). The algorithm will give $R_{K,n}$ as an output for each n OD pair.

2.8 Cut-off criteria for the reasonable route choice set

In this paper the purpose is to find the cut-off criteria, $q_{j,lim}$, which tells for the i-th route of the n-th OD pair, $r_{i,n}$, whether it is reasonable (considered in the choice set of passengers) or not:

$$r_{i,n} \in R_{r,n}, \text{ if for all j } q_{j,i,n} \leq q_{j,lim} \quad (2)$$

We can estimate $q_{j,lim}$ from the $q_{j,i,n}$ ratios in $R_{s,n}$ and $R_{ns,n}$ for each n OD pair.

$$q_{j,lim} = \frac{\max_{n=1}^{N^*}\left(q_{j,s,max,n}\right) + \min_{n=1}^{N^*}\left(q_{j,ns,min,n}\right)}{2} \quad (3)$$

We find $q_{j,s,max,n}$ for every j cost component within $R_{s,n}$:

$$q_{j,s,max,n} = \max_{i=1}^{K_{s,n}}\left(q_{j,i,n}\right) \quad (4)$$

We find $q_{j,ns,min,n}$ in two steps. First, for all j cost component we create a subset of $R_{ns,n}$, called $R_{j\_high,ns,n}$ where we include those $q_{j,i,n}$ values which are considered high:

$$R_{j\_high,ns,n} = \left\{q_{j,i,n} | q_{j,i,n} \text{ is high}\right\}_{i=K_s+1,n}^{K,n} \quad (5)$$

The practical meaning of (5) is to find for each $r_{i,n}$ route which component was high, so that it was not chosen by the passengers Now we do it heuristically: observing results and map, however our further aim is to automatize it. The second step is to find the minimum within $R_{j\_high,ns,n}$:

$$q_{j,ns,min,n} = \min_{i=K_s+1,n}^{K,n}\left(R_{j\_high,ns,n}\right) \quad (6)$$

Having obtained $q_{j,s,max,n}$ and $q_{j,ns,min,n}$ for the N OD pairs, we check whether there are outliers among them. To set $q_{j,lim}$ we work only with those N^* OD pairs which are not considered outliers. Among the outliers we observe, what attribute this OD pair has, so that it gives different results. And we discuss how we should consider those attributes in a later version of the model.

3. RESULTS

The case study network, which contains 9 lines and 57 stations, is represented by 243 nodes. Among them there are 93 in-vehicle nodes, 93 platform nodes and 57 ticket gate nodes. The K shortest path algorithm was implemented for this network to find the 10 shortest routes for 7 OD pairs.

We present the methodology described in chapter 2 through a numerical example for OD pair 1 (Victoria – Holborn). The 10 shortest routes are presented in Figure 4 with their cost components in Figure 5. For this OD pair, the set of first 2 routes obtained by the K shortest path algorithm (r_1, r_2)

coincide with the set of 2 surveyed routes from RODS data ($r_{s,1}$,$r_{s,2}$).

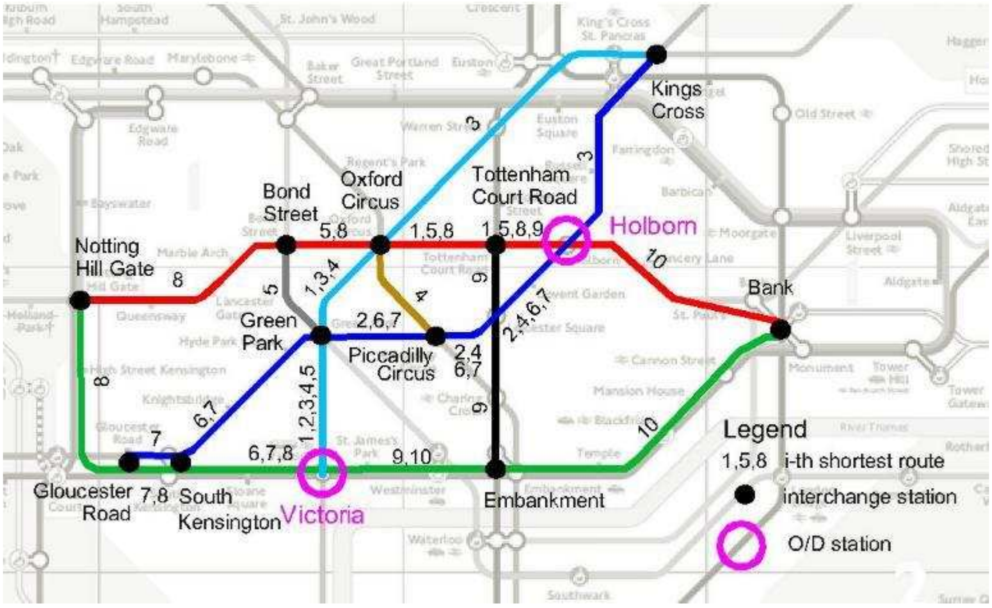


Figure 4 The 10 shortest routes (Victoria and Holborn)

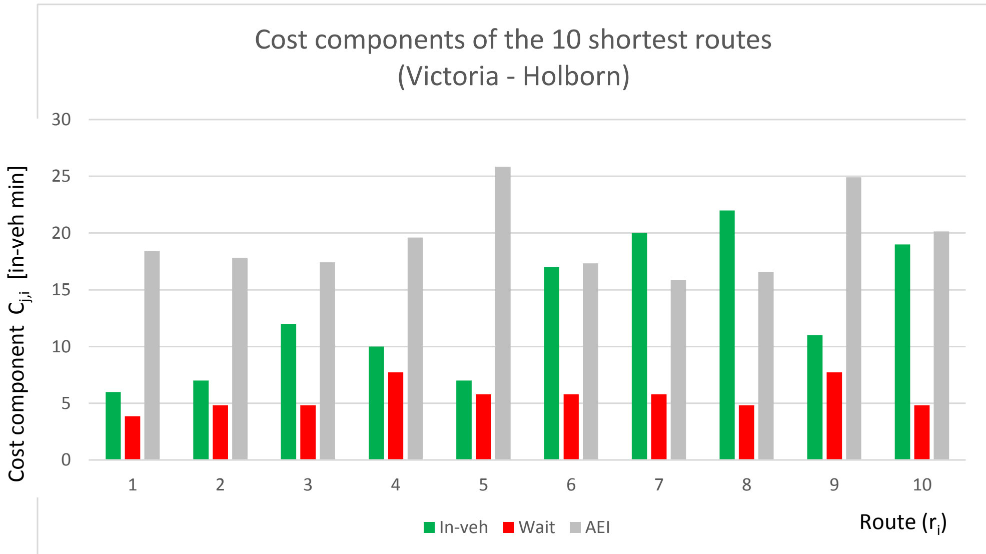


Figure 5 Cost components of the 10 shortest routes (Victoria - Holborn)

As $R_{s,1}$ has only 2 surveyed routes, for all j cost components

$$q_{j,s,max,1}=q_{j,2,1}=[1.17,\ 1.25,\ 0.97] \tag{7}$$

For $R_{ns,n}$ for each cost component we find the $R_{j\_high,ns,n}$ set:

$$R_{in\text{-}veh\_high,ns,n} = \{r_3, r_6, r_7, r_8, r_{10}\},\ R_{wait\_high,ns,n} = \{ r_4, r_5, r_6, r_7, r_9\},\ R_{AEI\_high,ns,n} = \{r_5, r_9 \} \tag{8}$$

Then for all j cost component we find $q_{j,ns,min,1}$ within each $R_{j\_high,ns,n}$ set according to (6)

$$q_{j,ns,min,1}=[2.00,\ 1.50,\ 1.35] \tag{9}$$

Repeating the same steps for each n OD pair, we resume the results in Figure 6.

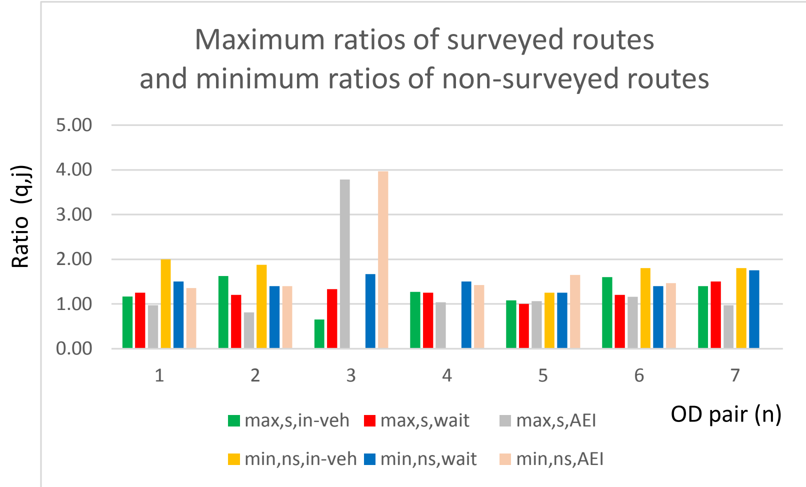


Figure 6 Max/min ratio of surveyed/non-surveyed routes

Among the results for the 7 OD pairs, we find that OD pair 3 and 5 is an outlier. OD pair 3 has much higher $q_{AEI,s,max,3}$ and $q_{AEI,ns,min,3}$ values. This is due to the fact that the shortest route is a direct route, therefore $C_{AEI,1,3}$ is very low as it has only access and egress, but no interchange cost. Being all the other routes indirect, it makes $q_{AEI,s,max,3}$ and $q_{AEI,ns,min,3}$ high. OD pair 5 has lower $q_{j,ns,min,5}$ values. This is due to the fact that many lines are available both at origin and at destination station, therefore there are many possible routes with 1 interchange. Apart from the 3 surveyed routes there are 5 other routes which have low $C_{j,i,5}$ values, which makes $q_{j,ns,min,5}$ low, however these routes were not surveyed. These observations on outliers imply, that apart from cost components for each route ($C_{in\text{-}veh}$, C_{wait}, C_{AEI}) we need to consider OD specific attributes, such as minimum number of interchanges and number of lines available at origin and destination station. Excluding these outliers, we obtain $q_{j,lim}$ for each j cost component according to (3) (Table 3)

Table 3 Cut-off criteria for each cost component

Cost component	Ratio		
j	$q_{j,s,max}$	$q_{j,ns,min}$	$q_{j,lim}$
In-vehicle	1.63	1.80	1.71
Wait	1.50	1.40	1.45
AEI	1.16	1.35	1.26

Results show that $q_{in\text{-}veh,lim}$ is quite high with respect to $q_{wait,lim}$ and $q_{AEI,lim}$, which means that passengers are likely to consider in their choice set, routes with longer in-vehicle time, where they don't have to change twice.

4. CONCLUSIONS, FURTHER WORK

In this paper, we presented a method to find the reasonable route choice set in metro networks. Firstly, how the K shortest path (Yen, 1971) can be implemented. Secondly, what criteria we can set for cut-offs to find the set of reasonable routes among the results.

The cut-offs for the in-vehicle, wait and AEI cost component, were 1.71, 1.45 and 1.26 respectively. This means that passengers consider a route reasonable with longer in-vehicle time, where they don't have to change twice. These criteria were set only based on OD pairs within Central London, where distances are shorter, services are more frequent and stations are more complex. For OD pairs from/to Outer London, we might have had different results.

Some of the OD pairs had special attributes and because of this they resulted outliers. From this, we understood that apart from the route-specific cost attributes, it is necessary to include also OD specific attributes, such as minimum number of transfer and lines available at O and D station.

In this paper we considered only travel time components and interchange experience for the generalised costs, however we acknowledge, that also the perception of the map is also an important attribute to consider, especially for the London Underground where the map is quite distorted (Guo, 2011).

When we set the cut-off criteria, we did it deterministically, based on the surveyed routes. However the fact that a route was not surveyed, does not necessarily imply that it was not considered in the route choice set of passengers. For this, is necessary to work towards probabilistic choice modelling methods to determine the cut-offs (Swait, 2001, Martínez et al., 2009).

5. REFERENCES

Ben-Akiva, M. and Lerman, S. R. (1974) Some estimation results of a simultaneous model of auto ownership and mode choice to work, *Transportation,* 3(4), pp. 357-376.

Dial, R. B. (1971) A probabilistic multipath traffic assignment model which obviates path enumeration, *Transportation Research,* 5(2), pp. 83-111.

Dijkstra, E. W. (1959) A note on two problems in connexion with graphs, *Numerische mathematik,* 1(1), pp. 269-271.

Fu, Q. (2014) Modelling route choice behaviour with incomplete data: an application to the London Underground. Ph.D. Thesis, University of Leeds, United Kingdom.

Guo, Z. (2011) Mind the map! The impact of transit maps on path choice in public transit, *Transportation Research Part A: Policy and Practice,* 45(7), pp. 625-639.

Halvorsen, A. A. F. (2015) Improving transit demand management with Smart Card data: general framework and applications, M.Sc. Thesis, Massachusetts Institute of Technology, U.S.A.

Holleczek, T., Anh, D. T., Yin, S., Jin, Y., Antonatos, S., Goh, H. L., Low, S. and Shi-Nash, A. Traffic measurement and route recommendation system for mass rapid transit (MRT). *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1859-1868.

Martínez, F., Aguila, F. and Hurtubia, R. (2009) The constrained multinomial logit: A semi-compensatory choice model, *Transportation Research Part B: Methodological,* 43(3), pp. 365-377.

McFadden, D. (1974) The measurement of urban travel demand, *Journal of Public Economics,* 3(4), pp. 303-328.

Nuzzolo, A. and Comi, A. (2016) Advanced public transport and intelligent transport systems: new modelling challenges, *Transportmetrica A: Transport Science,* 12(8), pp. 674-699.

Paul, E. C. (2010) Estimating train passenger load from automated data systems: application to London Underground. M.Sc. Thesis, Massachusetts Institute of Technology, U.S.A..

Poonawala, H., Kolar, V., Blandin, S., Wynter, L. and Sahu, S. Singapore in Motion: Insights on Public Transport Service Level Through Farecard and Mobile Data Analytics, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, USA. 2939723: ACM, 589-598.

Raveau, S., Guo, Z., Muñoz, J. C. and Wilson, N. H. M. (2014) A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics, *Transportation Research Part A: Policy and Practice,* 66, pp. 185-195.

Sun, L., Lu, Y., Jin, J. G., Lee, D.-H. and Axhausen, K. W. (2015) An integrated Bayesian approach for passenger flow assignment in metro networks, *Transportation Research Part C: Emerging Technologies,* 52, pp. 116-131.

Swait, J. (2001) A non-compensatory choice model incorporating attribute cutoffs, *Transportation Research Part B: Methodological,* 35(10), pp. 903-928.

Yen, J. Y. (1971) Finding the K Shortest Loopless Paths in a Network, *Management Science,* 17(11), pp. 712-716.

Zhu, Y. (2014) Passenger-to-Train Assignment Model Based on Automated Data. M.Sc. Thesis, Massachusetts Institute of Technology, U.S.A.