

A Simple Method to Forecast Travel Demand in Urban Public Transport

Balázs Horváth

Széchenyi István University
Department of Transport
Egyetem tér 1, H-9026 Győr, Hungary
balazs.horvath@sze.hu

Abstract: The key to the planning of public transport systems is the accurate prediction of the traffic load, or the correct execution of the planning stage assignment. This requires not only a well-functioning assignment method, but also reliable passenger data. Reliable passenger data means a time-dependent origin-destination matrix.

To solve the problem of the lack of time-dependent passenger data, we have developed a forecasting method. It consists of three stages.

In the first stage, we collect full scope cross-section data. This can be done either with personnel or with an automatic counting system. If personnel are used it costs a lot, and there is the chance for many possible errors. However, the results in most cases are good enough. Automatic counting system can be either a counter machine or even a simple "Check in" E-ticketing system.

In the second stage, we link boarding and alighting. As result, we get the origin-destination matrix for each run. This method is based on the likelihood of alighting at a given stop.

In the third stage, we combine origin-destination matrices of the runs through transfers. At this stage we assume that the probability of a transfer between two runs in a given stop is proportional to the travel possibilities in this relation.

To view the entire method in practice we proved it in a Hungarian city (Dunaujváros). The results were reliable, so they could be used in the planning process.

Keywords: public transport; transport planning; demand; OD matrix

1 Introduction

One of the possible solutions to handling the anomalies in city transport is the preference of public transport. Conversely, it is important to plan and operate a high level public transport service. The bottleneck of the planning of such systems is the knowledge of user demand. Without this knowledge, even the smallest change in the system is only a guess work, and the effect is unpredictable.

The cognition's methods of travel demand have been known for a long time [1], but their use has its limitations. Through our research we have built up a model which is able to generate a time-dependent O-D matrix for public transport with the use of the present transport system's characteristics.

2 Difficulties of Travel Demand's Determination

There are several methods to discover travel demand. Such can be the use of questionnaires or the application of a "check in – check out" e-ticketing system.

The application of an e-ticketing system can collect very detailed and accurate time dependent data day by day. But the establishment of such a system is very expensive and for a small bus operator unrealistic.

Another possible way is to organise questioners, but this costs a lot of financial resources and needs a large number of employees. Furthermore, the accuracy and reliability of the data are not always perfect, because reliable data need a large sample several times a day.

As an example, we can see a Hungarian city with 30,000 trips /day. This means about 7,000 trips in the morning peak period. The city can be divided into 25 zones, which means 625 possible trip relations. Some of these possibilities are not realistic or are used by only a few passengers; therefore, the number of real transport connections are about 100. If it is needed to take a sample from the morning peak, we have a basis of 7,000 persons. We can discover only the relations with at least 70 passengers, which means the incidence ratio is $P=0,01$. With the use of the common reliability of 95% and 10% relative error, the sample size is as follows (Eq. (1)).

$$n = \frac{t^2 \cdot (1 - P)}{h^2 \cdot P} = \frac{1,96^2 \cdot (1 - 0,01)}{0,1^2 \cdot 0,01} = 38031,8 \quad (1)$$

where t reliability level of 95% ($t=1,96$)

P incidence ratio

h relative error

This value should be corrected because of the finite number of the basis (Eq. (2)).

$$n_0 = \frac{n}{1 + \frac{n}{N}} = \frac{38031,8}{1 + \frac{38031,8}{7000}} = 5911,8 \quad (2)$$

where N number of the elements in the basis

This means an 85% sample to have accurate and reliable O-D data for the morning peak. Under the same conditions for the whole day (the basis is 30,000 persons) it is necessary to ask 16,771 persons, which means 56% of the users should be questioned.

Literature recommends for such a network with home interviews a sample of 25% [2], [3]. This means 50% of the users under a modal-split of 50% should be asked. This means that every 4th household needs to be questioned.

It is easy to understand that this task (a sample of 50%) is hard or even impossible to perform. Therefore it is clear that there is a need for other methods to produce a time dependent O-D matrix for public transport.

3 Calculation of the O-D Matrix from a Cross-Section Counting

We have been dealing with the evaluation and planning of public transport systems for a long time. We are in almost cases we facing the problem: How to produce a reliable O-D matrix?. Considering that in all of the analyses there were cross-section countings, it was obvious to start our method with these countings. On the basis of this revelation we built up a method to calculate the O-D matrix for public transport.

The method has two major parts:

- Calculation of the O-D matrix for runs (services)
- Forecasting of the O-D matrix for the whole network considering transfers

3.1 Calculation of the O-D Matrix for Runs

After the performance of the full-scope cross-section counting we know the number of boarding and alighting passengers for each run in each stop (Fig. 1).

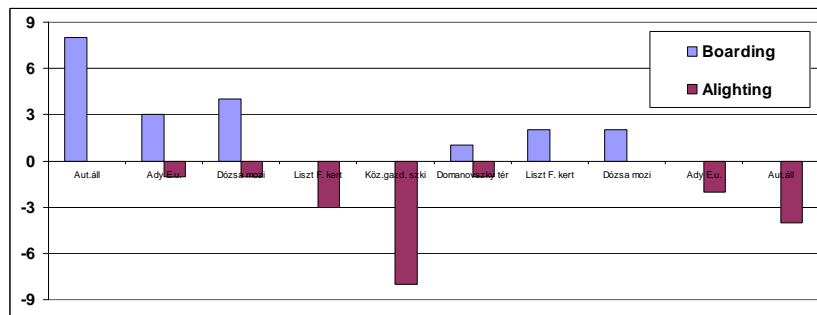


Figure 1

Number of boarding and alighting through the route of a line in one run

Assuming that the destinations (alighting) of the passengers boarding at a given stop are commensurate with the ratio of the alighting passengers of the remaining stops, we can forecast with likelihood the possible alighting stop of a passenger boarded at a given stop. The probability that a passenger boarded in stop i will travel to stop j can be calculated (Eq. (3)).

$$P_{i,j} = \frac{out_j}{\sum_{k=i+1 \rightarrow n} out_k} \quad (3)$$

where out_j number of passengers alighting at stop j

n number of stops on the line route

With the help of this probability, the estimated number of trip makers between stop i and j on a given run of a given line is as follows (Eq. (4)).

$$f_{i,j} = in_i \cdot P_{i,j} \quad (4)$$

where in_i number of passengers boarding in stop i

The method can be refined if we take into account the average trip length (l_i) on that given run. The average trip length can be calculated if we know the number of boarding and alighting and the stop distance (distance between stops - l_s) (Eq. (5)).

$$l_{t,avg} = \frac{\sum_{k=1 \rightarrow n-1} l_{s,k} \cdot p_k}{\sum_{r=1 \rightarrow n} in_r} \quad (5)$$

where $l_{s,k}$ length of the k^{th} link (distance between stop k and $k+1$)

n number of stops on the line route

p_k number of passengers on the k^{th} link

In this case the probability of someone's travelling from stop i to stop j is as follows (Eq. (6)):

$$P_{i,j} = \frac{1}{\left| l_{t,avg} - l_{i,j} \right| + 1} \cdot out_j}{\sum_{k=i+1 \rightarrow n} \frac{1}{\left| l_{t,avg} - l_{i,k} \right| + 1} \cdot out_k} \quad (6)$$

where $\left| l_{t,avg} - l_{i,j} \right|$ absolute value of the difference between average trip length and length of the analysed trip

Before we start the calculation of the probabilities to estimate the O-D matrix of the run, we can isolate some of the trips. These are the so-called definite trips. Either before or after these trips the vehicle is empty. This means that the vehicle is truly empty, or that a full passenger change has happened (after alighting the vehicle is empty, but there is also boarding). At this point, the line route of the run can be divided into sub-runs or sub-line routes.

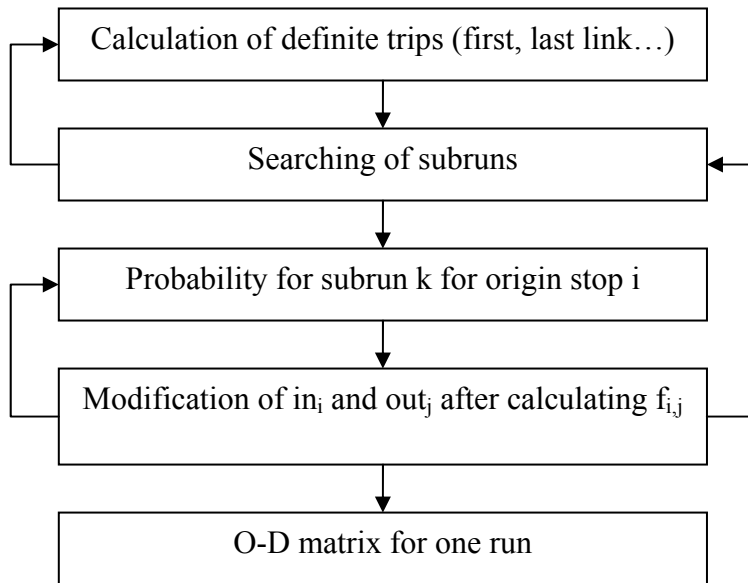


Figure 2

Process of the O-D matrix calculation for one run

Our task is to search the possible trips inside these sub-runs. This method gives much more accurate results than the calculation of the original one without any prudence. A special case is the first and the last link because before or after them the vehicle is empty, so the alighting after the first link, and the boarding before the last one are definite trips. The process of the model is shown in Figure 2.

3.2 Transfers between Runs

The O-D matrix of the runs should be corrected because a trip from i to j with a transfer in k will be two trips after the first step. These two trips are: i - k and k - j . The connection between these two trips is the transfer. In this second step of the method we must search this connection. To determine these trips with transfers, first we need to calculate the transfer ratio for stops and/or trips.

There are two different kinds of transfer ratio: the transfer ratio for alighting passengers and the transfer ratio for boarding passengers.

The transfer ratio for alighting passengers means the share between passengers who have reached their destination at a stop and the passengers who have alighted for transfer. Clearly it is the ratio of passengers who need to transfer and sum of alighted passengers.

The transfer ratio for boarding passengers is the other way around, and this means the share of passengers who are boarding because of a transfer among all of the boarding passengers.

If a trip from i to j needs to transfer at stop k , then the passengers of relation i,j has to be chosen from the passengers of the trips $i-k$ and $k-j$. If we know the share of passengers who are transferring and those reaching destination from i to k , then we can give an estimation of the number of passengers travelling from i to j . Similarly to this, there is a ratio for passengers travelling from k to j . With this ratio, we know the share of trip starters and transfer makers. On the basis of this, there is another estimation of passenger numbers of $i-j$. We have to choose the right one from these two estimations. In general if a trip needs n transfers there are 2^n possible passenger numbers.

The process of this estimation can be done as follows: first, all of the possible routes between i and j must be calculated. Afterwards, trips with transfers will be broken up into sub-trips (arms). For all of the arms the number of passengers will be calculated. For all of the transfer points the transfer ratio will be calculated. On the basis of these, all of the trips will have several possible passenger numbers. Finally, out of these possible numbers the right one will be chosen.

The calculation of this step starts with the calculation of the routes. Through this calculation it is important to write down all the stops which are reachable from a given origin point with a given run. If a stop can be reached with several runs the transfer possibilities will be noted, and new possible destinations will be recorded. In the practice it works as follows: first line's first run starts from first stop. From this stop the second stop of this run can be reached. If there is a new line's new run, the process starts again, if not, then the third stop can be reached etc.

This will result in the list of all transport connections. Some of them are wrong or unrealistic connections. They have to be deleted either through the generation or afterwards. A connection is unrealistic if:

- it starts earlier and finishes together or later than another connection
- in a T time period the number of transfers of this connection is $n+a$
- in a T time period it is k times or at least with t minutes longer than the shortest path
- the trip time is longer than T_{\max}

where n minimum number of transfers in a given relation

According to the literature [4] the parameters should be as follows:

- $T=10$ minutes
- $a=2$
- $k=1,5$
- $t=10$ minutes
- T_{\max} in Hungarian cities can be 50 minutes

In the previous step the number of passengers was calculated for the direct trips. In this step the task is to calculate the number of passengers for the trips with a transfer.

The goal is to calculate the number of passengers in the relation $i-j$ while the route leads through $i-k_1-k_2-j$ with a transfer in k_1 and k_2 . On the basis of the first step, the number of passengers for $i-k_1$, k_1-k_2 and k_2-j is known. The transfer ratios (calculation will be shown later) are known for both transfer points for alighting and boarding. It means for n transfers $2*n$ transfer ratios. It is impossible to use all of them separately because relation $i-j$ has only one number of passengers. There is a method to solve this problem:

- Trips with one transfer: the average of the two numbers of passengers must be used
- Trips with more than one transfer: the highest and lowest number of passengers will be deleted. The searched number is the average of the rest

The number of passengers calculated in this way is higher than one of the arms (e.g.: $i-k_1$, k_1-k_2 or k_2-j); then the smallest number of the arms must be used.

3.3 Calculation of Transfer Ratios

The key to the transfer correction (shown previously) is the accurate calculation of the transfer ratio. There are two ways to calculate these:

- Questionnaires
- Estimation

With the use of questionnaires the basic problem returns. In this case, the situation is much better because a smaller sample can give good results due to the limited number of possible answers at few transfer points.

It is enough to ask passengers at selected transfer points about their behaviour at that stop: Are they starting here or transferring? If transferring, which line have they arrived from, and which line will they travel with further? Theoretically a small sample can give accurate results.

The other possibility to estimate the transfer ratio is based on the principle that the operated public transport system describes more or less the travel demand. Accepting this idea, it is then possible to build up a basis matrix with reciprocates of the journey times. This matrix must be corrected with the known boarding and alighting numbers. On the basis of this matrix, it is possible to determine the transfer ratio [5].

Following this, the transfer ratio for relation i-j in transfer point k_1 can be calculated:

$$tr_{ijk_1out} = \frac{f_{ij}^n}{\sum_l f_{il}^n \cdot tp_{k_1ij} + f_{ik_1}^n} \quad (7)$$

where f_{ij}^n member of the basis matrix after correction (iteration in n step)
 tp_{k_1ij} coefficient for transfer: 1 if trip from i to j has to transfer in k_1 ; 0 if not

Afterwards, the first possible passenger number for relation i-j on run r can be:

$$f_{ij,r}^1 = f_{ik_1,r} \cdot tr_{ijk_1out} \quad (8)$$

It is possible to calculate all the potential numbers for relation i-j. With the application of the method written in the previous section, the number of passengers can be calculated for relation i-j.

3.4 Utility of the O-D Matrix

After the calculation of all the possible relations, in the result is a “travel diary”. This diary will contains all the passenger numbers of all connections. These connections are true only in the existing public transport system, and therefore the diary is useless for planning in this format.

To use it in the planning, it must be aggregated in time (time periods for e.g. peak) and space (zones).

If there is a detailed plan for the future it is possible to use it as 10-15 minute matrices. Another way it can be useful to have it in the format of hour matrices.

4 Matrix Estimation in the Practice

4.1 Raw Data for Estimation

In 2008 there was a full scope cross-section counting in the Hungarian city of Dunaujváros [6]. Through this counting, the whole public transport system was analysed. The resulting numbers of the counting were processed with the demonstrated estimation method.

To make the plans and improvements in the public transport system, we built up the model of the transport system in the software Visum. The model had two pieces of input data:

- Supply (transport service)
 - Transport network (routes, stops)
 - Public transport service (lines, timetable)
- Demand (travel demand)

While the supply side was clear, the demand side was only partly known.

4.2 Application of the Estimation Method

We estimated the O-D matrix for each run with the help of the above described method. To accelerate the process, we divided the area into 24 zones and aggregate the data not after but before the process. In this way we searched zone-zone connections instead of stop-stop ones.

In this city there are few transfers. The transfer correction was needed in only a limited number of cases. Therefore the errors caused by this correction (if) were not significant.

As a result of the process, there was a matrix with 8819 rows and 24 columns. All of the rows symbolised passenger movement at stops on the basis of runs. After aggregation of these 8819 rows, there were 19 O-D matrices each for one hour through the day.

It was possible to divide them into smaller parts but it was enough for the planning.

These 19 O-D matrices was implemented into the transport model created by the software Visum (Figure 3).

No.	Code	Name	Dec.	Random Round	Sum	Intrazone total	DSeg	DStratum
1		04-5:00	3	<input checked="" type="checkbox"/>	343.000	0.000	B1	
2		5:00-6:00	3	<input type="checkbox"/>	1535.000	0.000	B2	
3		6:00-7:00	3	<input type="checkbox"/>	2099.000	0.000	B3	
4		7:00-8:00	3	<input type="checkbox"/>	2407.000	0.000	B4	
5		8:00-9:00	3	<input type="checkbox"/>	1832.000	0.000	B5	
6		9:00-10:00	3	<input type="checkbox"/>	1813.000	0.000	B6	
7		10:00-11:00	3	<input type="checkbox"/>	1600.000	0.000	B7	
8		11:00-12:00	3	<input type="checkbox"/>	1605.000	0.000	B8	
9		12:00-13:00	3	<input type="checkbox"/>	1633.000	0.000	B9	
10		13:00-14:00	3	<input type="checkbox"/>	2490.000	0.000	B10	
11		14:00-15:00	3	<input type="checkbox"/>	2790.000	0.000	B11	
12		16:00-18:00	3	<input type="checkbox"/>	1994.000	0.000	B12	
13		18:00-17:00	3	<input type="checkbox"/>	1575.000	0.000	B13	
14		17:00-18:00	3	<input type="checkbox"/>	1462.000	0.000	B14	
15		18:00-19:00	3	<input type="checkbox"/>	972.000	0.000	B15	
16		19:00-20:00	3	<input type="checkbox"/>	539.000	0.000	B16	
17		20:00-21:00	3	<input type="checkbox"/>	300.000	0.000	B17	
18		21:00-22:00	3	<input type="checkbox"/>	320.000	0.000	B18	
19		22:00-04	3	<input type="checkbox"/>	427.000	0.000	B19	
20		night_nsp	3	<input type="checkbox"/>	27872.000	0.000	X	

Figure 3

Travel demand in the transport model

After implementation of this demand into the model, the next step was to calibrate the assignment method. To calibrate it, some assignment was done on the present network.

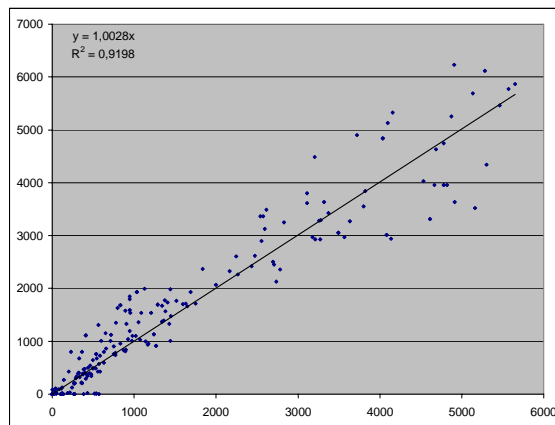


Figure 4

Calculated values in the function of measured values

In the calibration we used the timetable-based assignment method since the timetable was known for the present system.

The result of the assignment was checked on the basis of the links of the network. Theoretically the measured (counted) and calculated values should be equal. This means a graph of $y=x$, where y is the calculated and x is the measured (observed) value. In this study we get a graph of $y=1.0028x$ which is very good. The indicator of the accuracy, the correlation coefficient was $R^2=0.9198$ (Figure 4). After these the estimation can be declared as good enough for planning purposes.



Figure 5

Measured and calculated values on the link-bar graph

The values can be compared also on graphical way, with the help of a link-bar graph.

Further modification of the assignment model it was possible to reach a correlation coefficient of 0.9991 although the graph was worse in this case ($y=0.9245x$). It is important to note that both situations give good results for network planning purposes.

Conclusions

The bottleneck of the planning of public transport systems is the knowledge of user demand. This knowledge is usually missing or only partly known. To solve this problem we built up a method to estimate the travel demand in time and space with high reliability. It means the estimation of a time-dependent O-D matrix for public transport systems.

The method was checked in practice. It proved that the method is good enough to use it in the normal day-to-day work for planning public transport systems.

References

- [1] B. Horváth, G. Horváth, "Methodology of Public Transport Planning", Proceedings of European Society of Transport Institutes (ESTI) Young researcher's conference '99, Paris, France October 2-3, ESTI, 1999
- [2] G. Fülöp, B. Hirkó, J. Mátyus, I. Prileszky, L. Szabó, *Közlekedési üzemtan II.*, SZIF-Universitas, Győr, Hungary, 1999
- [3] E. Nagy, D. Szabó, *Városi közlekedési kézikönyv*, Műszaki Könyvkiadó, Budapest, Hungary, 1984
- [4] PTV AG, *Visum 10.0 User Manual*, PTV Planung Transport Verkehr AG, Karlsruhe, 2007

- [5] B. Horváth, I. Farkas, R. Horváth, Á. Winkler, *Városi közforgalmú közlekedési szolgáltatás javításának lehetőségei Zalaegerszegre adaptált modell segítségével*, Széchenyi István University, Győr, Hungary, 2008
- [6] I. Prileszky, G. Fülöp, B. Horváth, G. Horváth, I. Farkas, Á. Winkler, *Dunaujváros helyi tömegközlekedési szolgáltatásának fejlesztése komplex hatékonysági kritériumok alapján*, Universitas-Győr Nonprofit Kft, Győr, Hungary, 2009