THE NETWORK OF CITY PUBLIC TRANSPORT AS THE BASE FOR TRIP LENGTH DISTRIBUTION DETERMINING

P. Horbachov, Prof., D. Sc. (Eng.), S. Svichynskyi, T. Asst.,
Kharkov National Automobile and Highway University

Abstract. The up-to-date methods of modelling the demand for public transport services require an objective estimation and improvement. Such an improvement can be achieved by taking into account the trip length distribution during trip matrix calculation that requires determining the reasons of regularities occurrence in city population trip lengths.

Key words: trip length distribution, distribution law, stop co-ordinate, distance between stops, trip matrix.

Introduction

The modern theory of the travel demand prognosisication uses in general only one characteristic of route as the criterion of choosing the Origin-Destination (O-D) pair by a person. At the same time, ignoring the other factors, which do not refer to travel means or impedance, makes travel to be not the means to hit the target, but the purpose in itself.

However, in spite of seemingly such the self-evident truth that the purpose of majority trips cannot be the movement process per se, all well-known methods of travel demand modelling are based on transport factors only. The main arguments for this are the obvious desire of a person to minimize travel cost. Also the results of various investigations demonstrate a clear correlation between travel distance and its probability [1–10]. In transportation science the correlation is called a trip length distribution (TLD). The reason of its occurrence is the desire of people to live near the main places to satisfy their needs [7–10]. Determination of existing regularities in allocation of Origins and Destinations of city
dwellers’ travels will clarify the possibilities of transport demand modelling on the basis of travel independence between Origin and Destination points and provide additional information to improve the transportation planning accuracy and efficiency.

**Publication review**

Transport demand models are of primary importance in transport planning because they determine the accuracy of a transport model and efficiency of planning results. The transport demand models are usually presented as trip matrices [5, 6, 11–13]. To make transport demand modelling most accurate the passenger travels in city transport systems are divided into groups by day period, purposes, socio-economic status of people, etc. The main group of city population travels is labour travels, which create the greatest load on transport network and cause peak periods.

That’s why their modelling is required during transport system models development [1, 3–5, 7–10, 12]. In addition, labour travels are fundamental in majority of spatial interaction models as they have strong influence on the choice of residences or workplaces [1, 3, 4, 14]. Meanwhile, all the authors [1, 3, 4, 7–10, 12, 15, 16] consider that the basis for trip matrix formation and determination of city population allocation character accordingly is the choice of the «Home – Workplace» pair by population. It makes labor travels to be the universally recognized base for determining the TLD of urban population.

Among existing approaches to modelling the distribution of population demand for travelling over the city territory the gravity models of interaction between passenger Origins and Destinations are wide spread [1–5, 12]. The use of the gravity type models is based on the assumption about dependence of spatial interaction between transport attractors on their size and travel impedance between them. For example, a large city or an enterprise generates more traffic and attracts more labour than small one. Concerning travel impedance, any spatial interactions between objects such as labour travels, exchange of product or information is realized on some distance. It is logical to consider that the greater the distance, the lower the interaction intensity and hence the probability of travel on this distance [2].

During the determination of Origin and Destination transport districts, transportation costs should be paralleled with the positive part of travel, that is, the purpose which a person realizes through travel. On the base of such paralleling in labour travels, the paper [15] shows that when a person chooses the workplace as continual transport attractor, transport factors are not decisive and the main role is played by other factors which are called social. This is true for individual behaviour, but doesn’t explain underlying conditions of collective, overall behaviour of people that is illustrated by dependence of travel probability on impedance of connection between transport districts, i.e. the existence of TLD.

The main role in the study of the TLDs in cities belongs to scientists from the former socialist countries. Theoretical evidence of Soviet scientists is mainly based on assumption of people’s desire to choose the residence or workplace that is reasonably close to each other to minimize the time of regular travels between them. The result of such collective behaviour is the TLD. It’s general form in most empirical cases can be illustrated by gamma distribution with shape parameter which is more than 1, Fig. 1 [7–10].

The form of this curve is typical for most cities investigated and differs only by the slope of curve. As the modal value of distribution is shifted to the left, it leads to conclusion about the substantial dependence of the probability of travel on its impedance [7–10].

The results of the formal testing of the hypothesis of coincidence between empirical data and a theoretical gamma-distribution curve were not found in literature but similarity of empirical and theoretical data may be considered as suffi-
cient. The more important task is finding the reasons of existence of such regularities.

All the known TLD curves for the cities of the former USSR are the same as in Fig. 1. They were built by approximation of a great number of empirical data [7–10]. Such investigations require considerable resources and a strong administrative impact on enterprises in all economic sectors. It explains the greater effectiveness of socialist scientists in investigation of TLD functions in cities.

It should be noted that the usage of such popular theoretical models of the trip matrix calculation as gravity and entropy ones results in matrices which give identical population allocation curves [15]. It would be an advantage of the methods, but matrices which are obtained in another way lead to the similar results: for example, randomly filled matrices with known transport districts capacities; matrices of extreme travel impedances or matrices that minimize and maximize passenger-kilometres [15]. The last two types of matrices are considered to be theoretical and they are practically impossible. On the basis of this information a preliminary conclusion about small dependence of population allocation functions on the method of receiving the trip matrix can be made. At the same time a large number of possible trip matrix variants with constant transport districts capacities indicate incomplete knowledge and a need for additional investigation of TLDs.

In general the TLD curve is the result of proper processing of data that are contained in a trip matrix when correspondences are grouped into segments depending on the proper intervals of travel time or distance. Lack of real trip matrices for modern cities makes the receiving of empirical allocation functions very difficult in practice, and the use of theoretical matrices that are obtained by any of the known methods, for constructing the TLD functions is senseless because it will not characterize the factual population allocation completely.

Investigations of H. Sheleikhovskyi can be treated as the first theoretical studies of people allocation in cities [10]. They are completely based on transport factors and their perception by person, which the author described as the psychophysical Weber-Fechner law that is generally used to describe a human response to a physical stimulus. The effect of this law is valid at the average stimulus impact on senses and significantly degrades at the large or boundary ones [17]. Regarding to the distance or time of travel between residence and workplace it is not quite acceptable. H. Sheleikhovskyi accepted the population allocation curve that had been existing (Fig. 1) but on the basis of his own theory he argued that any TLD would follow up to the «normal» according to the city evolution that is presented in Fig. 2 [10].

Such a theoretical TLD looks like the exponential distribution, which is a special case of gamma distribution. This fact provides the visual conformity between Sheleikhovskyi’s theory and empirical data in an extreme case.

This theory and some other allocation theories imply that, if the residential real estate and job vacancies are in excess, any person would seek to change the residence in order to live nearby their workplace and vice versa – change the workplace for the same reason [7, 8, 10]. Such assumptions are necessary to explain the results of city population collective behaviour and to justify the using of travel impedance between transport areas as the only factor of formation the travel demand model. However, these assumptions are not theoretically justified and contradict the real state of both labour and real-estate markets as well as behaviour of people in situations of collective decision making, e.g. family decision making. Therefore, the factors of origin of allocation function, which is shown in Fig. 1 and describes empirical data quite well [7–10], cannot be considered theoretically grounded what causes a need to search and test other hypotheses.

One of these hypotheses is contained in work [11]. The authors of [11] consider transport and route networks as self-organized hierarchical
structures which formation is influenced by interests of government, large business owners, network users and other subjects and, therefore, can be considered in many aspects as random despite of some plans [11, 18]. Attempts to explain the results of such interaction were made on the basis of different analogies but fundamental progress in explaining the existing TLDs was not achieved.

Drawing a line under publication review, it can be considered that there are two main views on the origin of the TLD function: 1) it is the result of people deliberate behaviour to locate as close as possible to their workplaces or 2) it is a natural process of transport demand distribution over the city that is caused by the city growth under influence of many factors. The first view is already developed in the existing allocation theory; the second one needs the theoretical and experimental testing.

**Problem statement**

At present it is possible to consider that general TLD plot, which is shown in Fig. 1, is known. At the same time, there are no considerable differences when grouping on a criterion of travel distance or time, as they cover more than 95% of a total travel time variation [15]. It should be pointed out that these data concern only public transport (PT) and they are more than relevant for Ukrainian cities as 70-80% of city dwellers use PT services during labour travels.

The publication review has shown that the route network (RN) of PT (RNPT) and its characteristics are not directly considered in the known theory of TLD obtaining. Upon that, PT is seen only as a factor which defines travel impedance on its basic directions according to own development level, which can be described by velocities, vehicles comfort level, network branching degree etc. [7, 10].

The forming of the RNPT is performed under the influence of a large number of factors. Their interaction causes some randomness of RN development [11, 18]. The most influential factor is passenger demand for transportation because its satisfaction is the main purpose of a RN. Therefore, the location of the RN infrastructure objects, that’s the transport supply, is determined by transport demand. At a considerable part of route passenger transportation, which is typical for Ukrainian cities, the RNPT is a very good tool to study the TLD. It is stipulated by such RN elements as stops because they are places of urban trips concentration. All other characteristics of the RNPT including the Distances Between Adjacent Stops (DBAS) and the Distance Between A Pair of Stops (DBPS) are derived from the stops spatial location. The first parameter which is determined by the stops location is the DBAS regardless of their belonging to different PT types – bus, trolleybus or tram.

In terms of walking distance to and from a stop it can be assumed that they can be negligible for TLD. Therefore, distribution of labour travels by PT can be taken as the basis to determine the reasons of TLD as an example of Ukrainian cities of the 21st century.

TLD function formation can be presented as a process of transforming the square matrix of DBPSes in a linear travel distance set that can be described by proper distribution law. The gist of transformation process is that each distance $l_{ij}$ between stops $i$ and $j$ is repeated in a linear travel distance set as many as $h_{ij}$ times if $h_{ij}$ is a quantity of trips between stops $i$ and $j$. Unfortunately, absence of a real trip matrix for any Ukrainian city makes this transformation impossible.

So, to achieve this goal it is necessary to investigate and explain the reasons of DBASes distribution and then compare last one with the known TLD curves. In case of the similarity of these distributions it can be considered that the trip matrix has a negligible influence on the TLD function and the main reason of the origin of TLDs is the natural processes of enterprises and population allocation in the vicinity of the city centre. Otherwise, it is necessary to determine conditions of transforming DBASes into TLD function, i.e., those trip matrix variants that will provide such a transformation. Analysis of these matrices will provide opportunities to put forward the hypotheses of the distribution of transport area capacities and check the likelihood of the existing models of trip matrix calculation.

To investigate regularities in DBASes distribution it is necessary to connect DBASes and stops location on the city territory that will describe existing dependence between spatial characteristics of a RN infrastructure.
This investigation is based on the hypothesis of randomness of stop plane co-ordinates in the city which are influenced by many factors including surroundings, individual and collective activity of population and enterprises, history, minerals etc. Thus, the applicable coordinates can be neglected.

**Theoretical background**

To achieve an aim it is necessary to theoretically determine the type of distribution of distances between stops \(i\) and \(j\) for every possible pair of stops, number of which is \(N (= 1, N)\). Such distances are formed as a set of DBASes that are on the route from stop \(i\) to stop \(j\):

\[
l_{ij} = \sum_{k=1}^{n_{ij}} l_k,
\]

where \(l_{ij}\) is a distance between stops \(i\) and \(j\) (e.g. DBPS), km; \(n_{ij}\) is a quantity of DBASes between stops \(i\) and \(j\); \(l_k\) is a factual DBAS \(k\), km.

DBAS \(l_k\) is a distance on a RN. That’s why determining the \(l_k\) properties should be based on the investigation of regularities of stops distribution over the city. It is this distribution that is the starting point for the formation of required regularities in DBPSes. According to the hypothesis in Problem Statement, distribution of stops over the city can be characterized by the bivariate distribution of random coordinates \((X; Y)\). Taking into account a great number of factors that determine stops location co-ordinates, it is reasonable to assume that a hypothetical law of their distribution is an asymptotically bivariate normal law.

Since the stops coordinates are a random variable then the distance between public transport stops will also be a random variable and in majority of cases can be presented as the shortest distance between adjacent stops. Individual cases of unstraight DBASes cannot significantly affect the random variable \(l_k\) distribution so they can be neglected in the theoretical part of investigation.

If each stop is the beginning of co-ordinates and measure existing distances to adjacent stops on RN avoiding repeated measurements of the same distances, the overlaying of all possible initial stops at the beginning of co-ordinates will result in the distribution of final stops in neighbourhood of a certain radius which equals the maximum DBAS in the city.

If the distribution of \((X; Y)\) is bivariate normal circular with mean values of both co-ordinates \(m_x = m_y = 0\) and standard deviations of co-ordinates \(\sigma_x = \sigma_y\), DBAS can be presented as \(l_k = \sqrt{X^2 + Y^2}\) and described by Rayleigh distribution [19]

\[
F(l_k) = 1 - e^{-\frac{l_k^2}{2\sigma^2}} \quad (l_k > 0),
\]

where \(\sigma\) is Rayleigh distribution parameter – standard deviation of \(l_k\) [19], km².

The first condition for the bivariate normal circular distribution of DBASes is satisfied automatically for distances between stops due to sequential investigation of stops of the DBASes for PT as the beginning of co-ordinates and their imaginary overlaying at one point. As for the second condition which defines the scale and shape of distribution, we could not theoretically determine it.

Despite of lack of the proof for the second condition, the hypothesis is that DBASes correspond to Rayleigh distribution. However, due to the specific formation of this random variable, when the initial and final stops of DBASes are close to each other, the requirement for bivariate normality of the initial co-ordinates distribution is not so rigid. It is important that this condition should be met after imaginary overlaying of initial stops of DBASes at the beginning of co-ordinates. In this case, the proximity of transformed coordinates to bivariate normal distribution can be confirmed by coincidence between empirical DBASes distribution and theoretical Rayleigh distribution.

However, in our case the Rayleigh distribution is not further suitable as analytical expression for the DBPS \(l_{ij}\) cannot be obtained by using it because the known analytical transformation of this distribution has not been found. The use of Rayleigh distribution as the basis for further calculations is sophisticated by the fact that the DBAS \(l_k\) enters the expression of \(l_{ij}\) in a linear form and in the relative Rayleigh distribution function they are squared. That’s why it is reasonable to converse to another distribution with linear DBAS as an argument. The conversion of
squared DBAS into linear ones can be performed using the Taylor expansion \[19\]

\[
l_k^2 = f(l_k) = f(a) + f'(a) \cdot (l_k - a), \tag{3}
\]

where \(a\) is a point of expansion; \(f'(a)\) is the first-order derivative value in point \(a\), i.e. \(f'(l_k) = (l_k^2)' = 2l_k\).

To make the expansion, the point-value of \(0.5\) km should be chosen. It will result in the shift of factual DBAS by \(0.25\) km. and the value of \(f'(a)\) is \(1\) after Taylor expansion

\[
l_k' = l_k - 0.25 = l_k - 0.25 > 0. \tag{4}
\]

If \(\lambda = 1/(2\sigma^2)\) Rayleigh distribution will converge into exponential:

\[
F(l_k') = 1 - e^{-\lambda(l_k - 0.25)} = 1 - e^{-\lambda l_k'}, \tag{5}
\]

where

\[
l_k' = l_k - 0.25 > 0. \tag{6}
\]

It is important for the linearization of squared DBAS to have the constant in equation (6). This constant determines the shift of the random DBAS to the right relating to the beginning of coordinates. Taking into account the constant mentioned, equation (1) looks like

\[
l_{ij} = \sum_{k=1}^{n_{ij}} l_k' + n_{ij} \cdot q, \tag{7}
\]

where \(q = \text{const}\) is the shift parameter of the exponential distribution of the continuous random variable \(l_k'\), km.

The first part of the equation (7) is convolution of the exponential distribution of the random part of DBAS, which is Erlang-\(n_{ij}\) distribution

\[
G(n_{ij})(l_{ij}) = \frac{\lambda^{n_{ij}} (l_{ij})^{n_{ij}-1} e^{-\lambda l_{ij}}}{(n_{ij} - 1)!}, \tag{8}
\]

\((l_{ij} > 0, n_{ij} = 1, 2, \ldots)\),

where \(\lambda\) is parameter of exponential distribution of the random part of DBAS [19].

The Erlang distribution is a special case of gamma distribution when shape parameter, in our case \(n_{ij}\), is an integer. But the quantity of DBASes is constant only when considering one distance between stops \(l_{ij}\). When considering all possible \(l_{ij}\) for the city, the quantity of DBASes that came on the route from stop \(i\) to stop \(j\) will be a random variable. Random variable \(l_{ij}\) is the probabilistic mixture of exponential distributions [20]. In general, summation of variables that have Erlang distribution and \(n_{ij} \cdot q\), according to formula (7), will result in the loss of shape parameter integrity and conversion of DBPSes distribution into gamma distribution with scale parameter \(b\) and shape parameter \(c\) [21]

\[
f(l_{ij}) = \frac{b^c l_{ij}^{c-1} e^{-b l_{ij}}}{\Gamma(c)} \quad (l_{ij} > 0, b > 0, c > 0). \tag{9}
\]

It indicates the known TLD at the stage of stops location in the city. It is necessary that the discrete part in expression (7) \(n_{ij} \cdot q\) has not influenced the shape of distribution \(l_{ij}\). The influence of this part of (7) upon the form of the DBPSes \(l_{ij}\) distribution curve can be different and even may cause several peaks in the resulting distribution. Such an effect is possible if DBASes exponential distribution shift parameter \(q\) is considerable. If this parameter is small, i.e. several times less than the mean value of \(l_{ij}'\), the discrete part will not cause the change of the distribution \(l_{ij}\).

To test this statement it is necessary to select the appropriate value of the \(l_{ij}'\) shift parameter and check the possibility of the description of the DBASes variation using exponential distribution. In this case it is logical to select the minimum DBAS in the city as a shift parameter

\[
q = \min(\forall l_{ij}, k \in \{1, M\}) = l_{min}, \tag{10}
\]

where \(M\) is the total quantity of DBAS in the city; \(l_{min}\) is the minimum DBAS in the city.

If the empirical data of DBASes having such a shift parameter are described by exponential distribution, the curve of gamma distribution of DBPSes \(l_{ij}\) will be the same as shown in Fig. 1. Thus, if all above theoretical background is practically proved, it will be proved that the common type of the TLD function in the cities is determined by distribution of PT stops over the
city and trip matrix doesn’t influence significantly on the TLD.

Experimental test of hypotheses

To receive empirical data the RNPT models of the cities Sumy, Kharkiv, Kyiv and Kryvyi Rih that were designed using PTV Vision® VISUM software are used. These models allow getting the stop co-ordinates, DBASes and matrices of distances between pairs of stops [13]. The first stage of experimental investigations is the testing of the correspondence between DBASes empirical distribution and theoretical Rayleigh distribution, which illustrated in Fig. 3 [22]. Results of such correspondence testing for other cities mentioned are analogous.

The hypothesis of correspondence between Rayleigh and DBASes distributions was correct for all the cities that confirms both the randomness of stops distribution over the city area and proper theoretical background. The parameters of Rayleigh distribution of DBASes in cities investigated are shown in Table 1.

Table 1 Characteristics of Rayleigh distribution of DBAS in Ukrainian cities

<table>
<thead>
<tr>
<th>Distribution parameter</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation, km</td>
<td>Sumy</td>
</tr>
<tr>
<td></td>
<td>Kryvyi Rih</td>
</tr>
<tr>
<td></td>
<td>Kharkiv</td>
</tr>
<tr>
<td></td>
<td>Kyiv</td>
</tr>
<tr>
<td>0.42</td>
<td>0.57</td>
</tr>
<tr>
<td>0.54</td>
<td>0.41</td>
</tr>
<tr>
<td>0.554</td>
<td>0.746</td>
</tr>
<tr>
<td>0.682</td>
<td>0.557</td>
</tr>
<tr>
<td>0.37</td>
<td>0.11</td>
</tr>
<tr>
<td>0.18</td>
<td>0.12</td>
</tr>
</tbody>
</table>

The next stage of experimental investigations is the testing of the usefulness of exponential distribution with a shift parameter which equals the minimum DBAS in each city \( q = l_{\text{min}} \) for the description of the DBASes set.

The example of such a distribution chart, which is analogous for all the cities investigated, is presented in Fig. 4; the distribution parameters are given in Table 2.

Theoretically such a result is evidence of gamma distribution of DBPSes but it requires experimental testing on the basis of a proper matrix. The example of the density of DBPSes distribution for Sumy is shown in Fig. 5.

Table 2 Characteristics of exponential distribution of DBAS in Ukrainian cities when a shift parameter is equal to \( l_{\text{min}} \)

<table>
<thead>
<tr>
<th>Index</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum DBAS ( l_{\text{min}} ), km</td>
<td>Sumy</td>
</tr>
<tr>
<td>0.043</td>
<td>0.079</td>
</tr>
<tr>
<td>Parameter ( \lambda )</td>
<td>2.16</td>
</tr>
<tr>
<td>Expectation, km</td>
<td>0.511</td>
</tr>
<tr>
<td>Dispersion, km²</td>
<td>0.540</td>
</tr>
<tr>
<td>Probability of ( \chi^2 ) criterion</td>
<td>0.77</td>
</tr>
</tbody>
</table>
The test of a possibility to describe DBPSes when using gamma distribution was made in other cities where the density charts of distribution are analogous to Fig. 5 [21]. The parameters of the distribution received are given in Table 3.

Thus, the experimental investigations do not contradict the theoretical background of population allocation regularities. In this case DBPSes in cities investigated are gamma-distributed that meets generally accepted city population allocation function.

| Table 3 Parameters of gamma distribution of DBPSes |
|-----------|-----------|-----------|-----------|
| Index     | Sumy      | Kryvyi    | Rih       | Kharkiv   | Kyiv     |
| Scale parameter λ | 2.20 | 7.43 | 3.59 | 4.10 |
| Shape parameter α | 3.40 | 2.53 | 3.25 | 3.40 |
| Expectation, km | 7,331 | 18,415 | 11,317 | 13,586 |
| Dispersion, km² | 14,466 | 122,860 | 32,611 | 44,816 |
| Probability of χ² criterion | 0.47 | 0.53 | 0.28 | 0.14 |

It means that trip matrix does not significantly influence on population allocation function.

**Conclusion**

The investigations have theoretically and experimentally confirmed that the characteristics of PT supply, i.e. distribution law of DBPSes, are the basis to determine the TLDs for Ukrainian cities. It indicates that these regularities are the result of spatial allocation of city RN elements rather than the influence of transportation factors on the formation of Origin-Destination pair for population. When modelling the demand for PT services, it is necessary to determine such states of trip matrix that do not converse gamma distribution of DBPSes into any other distribution and use the interval conception of transport demand determining.

**Література**


15. Горбачов П.Ф. Концепция формирования систем маршрутного пассажирского транспорта в миостах: дис. ... д-ра техн. наук: 05.22.01 / Горбачов Петро Федорович. – Х., 2009 – 370 с.


References


13. VISM 10.0 User Manual. CD-ROM 80 min / 700 MB.

Рецензент: Е.В. Нагорный, профессор, д.т.н., ХНАДУ.
Статья поступила в редакцию 17 апреля 2015 г.