2nd Internatıonal Conference on Problems of Logıstıcs, Management and Operatıon in The East-West Transport Corrıdor (PLMO 2023) Baku, Azerbaijan
DOI:10.54381/plmo2023.34

Analysis of the features of the practical application of some methods for traffic prediction

Ramin RZAYEV*, Emil AHMADOV
Institute of Control Systems, Baku, Azerbaijan
Intelligent Transport Management Centre of the Ministry of Internal Affairs of the Republic of Azerbaijan, Heydar Aliyev Ave. 137, Baku, Azerbaijan
**Corresponding author.* E-mail: raminrza@yahoo.com

Abstract. The article analyses the problem of practical modelling of traffic flows within the existing road infrastructure of an urban agglomeration. As a methodological basis for predictive analytics, well-studied extrapolation methods for constructing matrices of interborough movements and making forecasts based on them are considered. The features of the practical application of extrapolation methods by a single growth coefficient, average growth coefficients, Detroit and Fratara approaches are analysed. As a result of the analysis of methods and models for constructing correspondence matrices in relation to the transport system of the urban agglomeration, the disadvantages and advantages of the above extrapolation methods are revealed.

Keywords: *transport system; traffic flow; correspondence matrix; extrapolation methods*

1. INTRODUCTION

Absolutely all megacities of the world permanently suffer from traffic queues, which negatively affects the business environment and the economy as a whole. One of the reasons for the formation of congestion is the lack of adequate transport infrastructure or when it works at the limit of its capabilities. However, in most cases, ineffective traffic management remains the main cause. Based on this premise, the importance and relevance of the tasks of strategic planning for the development of transport infrastructure and traffic management with the widespread use of adaptive control elements become obvious. To solve these problems, it is advisable to use traffic patterns that are able to provide management structures with useful information in the form of long-term forecasts for strategic planning of transport infrastructure development, as well as in the form of short-term forecasts to support operational decision-making on traffic management in real time. These forecasts are built on the basis of quantitative characteristics of the state of traffic flows (speed and intensity of traffic in each section of the transport network, travel time, volumes of interborough movements, etc.), which are generated by appropriate mathematical models at the preliminary stage of the study.

The current level of development of information and computer technologies allows developers to quickly build adequate transport models for cities, regions and countries on the whole. At the moment, a huge number of approaches to building models of traffic flows are described in the specialized scientific literature. This thematic review considers predictive models that answer questions about future volumes of interborough movements (correspondence between individual borough) and their distribution (routing) along the transport network graph. In most cases, the results of calculations obtained using predictive models regarding future volumes of demand for movement are input characteristics of dynamic models of traffic flows.

2. TRANSPORT DEMAND AND CORRESPONDENCE MATRIX

By analogy with the consumer market, in the theory of traffic modelling, the main concepts are also supply and demand. A transport offer is understood as a set of means of delivery that are available in a particular territory. Transport demand, which is of particular interest in the context of this review, is considered as an aggregate assessment of the need for the movement of people or goods along a specific transport network using various modes of transport – public, personal (including various types of individual mobility) and freight. In particular, the satisfaction of the transport demand for passenger transportation in cities is quantified by the so-called indicators of the transport mobility of the urban population.

Transport demand modeling is a paramount task when analyzing the current transport situation and forecasting the load on the transport network and/or its individual sections. So, to determine the volumes of transport flows and loading of elements of the transport network, various mathematical models are applied, the inputs of which are data on the number of people and goods moving around the city. This initial information, as a rule, is systematized in the form of the so-called matrix of interborough movements or, simply, correspondence matrix, and the dimension of the matrix is determined by the number of boroughs in the transport network. For the urban agglomeration consisting of n transport boroughs, the square correspondence matrix is generally presented in the form of Table 1, in which r_{ij} symbolizes the volume of movements (correspondences) from the i-th transport area to the j-th area, that is, the volume of the flow of people and goods from the i-th source-area to the j-th sink-area.

Tab. 1

Transport correspondence matrix

	1	2	…	j	…	n	$\sum_{j=1}^{n} r_{ij}$
1	r_{11}	r_{12}	…	r_{1j}	…	r_{1n}	
2	r_{21}	r_{22}	…	r_{2j}	…	r_{2n}	
…	…	…	…	…	…	…	
i	r_{i1}	r_{i2}	…	r_{ij}	…	r_{in}	
…	…	…	…	…	…	…	
n	r_{n1}	r_{n2}	…	r_{nj}	…	r_{nn}	
$\sum_{i=1}^{n} r_{ij}$							

Based on the elements of the correspondence matrix, the load on individual sections of the transport network is determined. This allows, firstly, to identify the most intensive nodes and, as a result, redistribute traffic flows for their unloading, and, secondly, to evaluate the intensity of transport and passenger flows to optimize the mode of public transport. The final calculation of the load – the excess of freight and passenger traffic in a separate section of the transport network is carried out step by step using a model to calculate the volume of transport demand. The hierarchical structure of such model for the common case is shown in Fig. 1 (see [1]) and is described in detail in [2, 3]. In particular, in [1] it is noted that this model is widely used in the preparation of transport projects covering both medium and long-term planning horizons.

At the top, tentatively speaking, at the first level of the model, based on statistical data and mobility coefficients that reflect the correlation between the socio-economic indicators of the given urban agglomeration and mobility, identified from surveys of respondents, common information is generated on the volumes of arrivals and departures for each transport area. Based on this information, at the next second level of the model, the correspondence matrix is formed in the form of Table 1. Further, at the third level, the correspondence matrix is “splitting” into separate matrices for various types of transport. As a result, at the last fourth level of the model, the desired solution is determined relative to the distribution of freight and passenger flows over the network, which makes it possible to calculate the load on each section of the transport network.

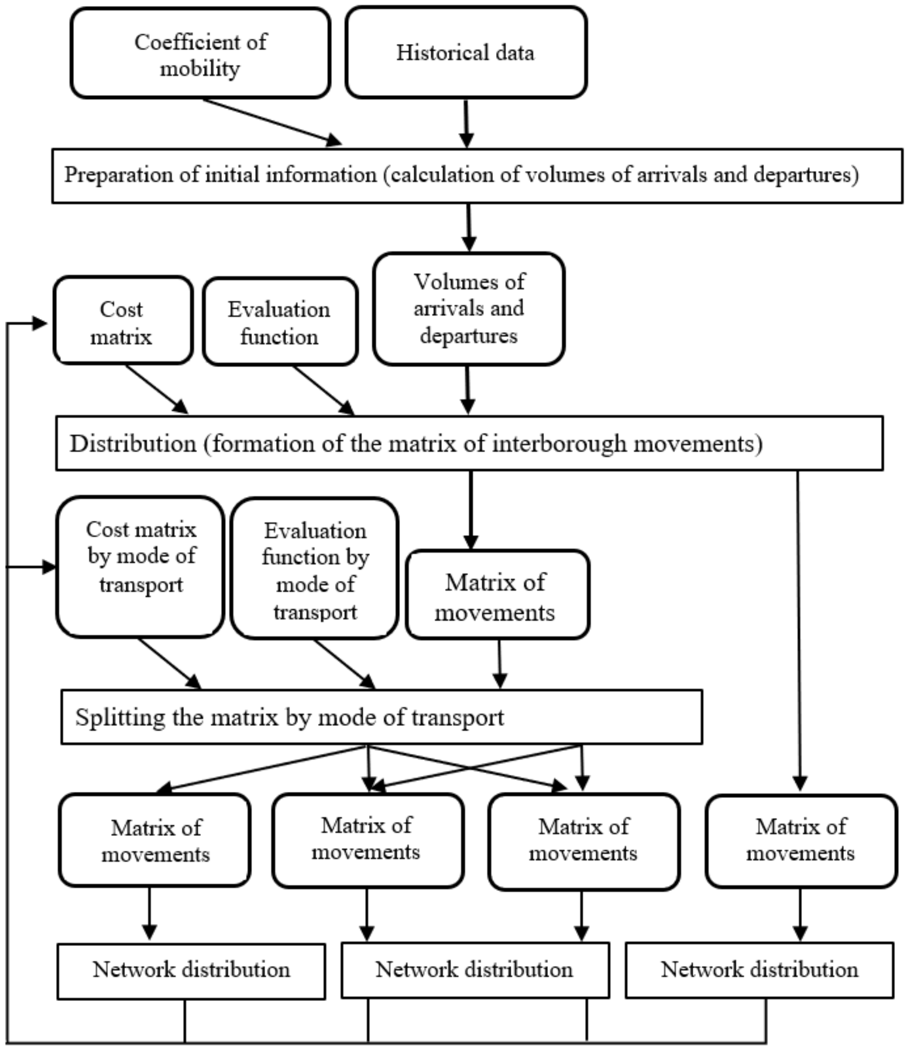

Fig. 1. Hierarchical structure of the model for calculating transport demand

2. PREDICTIVE APPROACH TO THE FORMATION OF THE CORRESPONDENCE MATRIX

Predictive analytics is primarily a set of methods of statistics, data mining and game theory that are used to study regularities between the historical data of a particular indicator in order to predict its values in the future: short-term, medium- and long-term periods. For successful predictive analysis it is recommended to strictly adhere to the following stages: goal-setting, obtaining data from various sources, preparing data, creating a predictive model, evaluating the model, implementing the model, monitoring the effectiveness of the model. Without taking into account the last three points, the step-by-step procedure for modeling transport demand described above is fully integrated into the scheme for implementing predictive analytics.

The key link in the transport demand model is the correspondence matrix, the elements of which are the quantities of movements from source-areas to sink-areas. To determine the quantitative characteristics of movements, various predictive methods can be applied, using actual data on interborough movements and forecasts of their growth as initial information. The following is an overview of some of the extrapolation methods that are described in so many specialized scientific publications that it is impossible to identify the original sources.

2.1. The method of the single coefficient of transport demand growth

In a series of predicative approaches, this method is the most trivial. It is implemented through the establishment of the so-called growth coefficient, which is determined by the following ratio [1, 4]

$$C = \frac{F}{A}, \tag{1}$$

where F is the total predicted volume of interborough movements; A is its actual value.

In general, the growth coefficient is an economic multifactorial category, that is, it depends on changes in many factors, for example, the population in the urban agglomeration (A_p), the average per capita income (A_{inc}), the level of motorization (A_{ml}), etc. Then, taking into account the above factors, by analogy with (1), the growth coefficient can be established from the following ratio [4]:

$$C = \frac{f(F_p, F_{inc}, F_{ml})}{f(A_p, A_{inc}, A_{ml})}, \tag{2}$$

where $f(\cdot)$ is a certain function that converts the values of socio-economic indicators into the value of the total volume of traffic flows within the given urban agglomeration. Obviously, in order to calculate the growth coefficient for the agglomeration under consideration using formula (2), it is necessary, firstly, to identify the function f: $R^3 \to R^1$ and, secondly, to use the apparatus of predictive analytics to analyze and predict the dynamics of changes in the indicated socio-economic indicators.

Thus, assuming the value of the growth coefficient to be known, it is possible to determine all elements of the predicted correspondence matrix by following equalities

$$r_{ij}^* = C \cdot r_{ij}^0, \tag{3}$$

where hereinafter in the text, ‘$*$’ is the index of the predicted volume of movements from the i-th transport region to the j-th region of the agglomeration, that is, the predicted volume of the flow of people and goods from the i-th source-area to the j-th sink-area; ‘0’ is the index of the current value of the corresponding amount of movement.

Building a predictable correspondence matrix using the method of a single growth coefficient according to formulas (1) and (3) is very rough operation, since this approach does not take into account the dynamics of development between individual socio-economic indicators of the urban agglomeration and has low reliability. However, in practice, this method may be required for approximate estimates of the volume of traffic flows in the short term when designing a separate part of the city.

2.2. Method of average growth coefficients

As in the previous case, the method of average growth coefficients is based on the predictive analysis of the volumes of interborough movements. Here, as the basic parameters of the predictive model, the growth coefficients C_i and C_j are used for each i-th and j-th transport area, respectively, which, similar to (1), are calculated by the following ratios

$$\begin{cases} C_i = \dfrac{r_i^*}{r_i^0}, \\ C_j = \dfrac{r_j^*}{r_j^0}, \end{cases} \tag{4}$$

where $i, j = 1 \div n$; n is the number of transport areas in the urban agglomeration; r_i^0 and r_j^0 are actual (current) volumes of traffic flows in the i-th and j-th areas, respectively; r_i^* and r_j^* are their predicted volumes. Then the predicted distribution of traffic flows in the cells of the correspondence matrix according to the method of average growth coefficients is carried out on the basis of the formula

$$r_{ij} = C(0) \cdot r_{ij}^0 , \tag{5}$$

where $C(0) = (C_i + C_j)/2$.

Since the values of traffic flows between n transport areas, determined from ratios (5), as a rule, do not satisfy the equalities

$$r_i^* = C(0) \cdot \sum_{i=1}^{n} r_{ij}(1) , \ r_j^* = C(0) \cdot \sum_{j=1}^{n} r_{ij}(1) , \tag{6}$$

the method of average growth coefficients for predicting the volume of interborough movements is implemented iteratively in several steps according to the formula

$$r_{ij}(t+1) = C(t) \cdot r_{ij}(t) , \tag{7}$$

where $C(t) = [C_i(t) + C_j(t)]/2$ are average values of the growth coefficients for the calculated moments $t = 0, 1, 2$ ...

The calculation results at each iteration step form the initial data for the computational procedure of the next step. The process will continue until equality is reached between the predetermined value of the transport turnover of the area and the amount of movements obtained as a result of the calculation for this area using formulas (6).

It is not difficult to see that the method of average growth coefficients takes into account the different dynamics of the development of certain areas of the urban agglomeration. However, with a significant increase in the mobility of the population due to the emergence of new residential areas, large shopping centers and/or large industrial enterprises, the application of this method leads to unacceptable errors, which does not allow it to be used for practical modeling of traffic flows.

2.3. Detroit method

Unlike the previous methods, the Detroit method, in addition to the growth coefficients of individual areas of the urban agglomeration, takes into account the growth coefficient of the entire urban agglomeration. In order to maintain the correspondence between the predicted and calculated volumes of interborough movements, the predicted volumes of interborough movements are also determined iteratively in several steps. At the first step, these values are established as

$$r_{ij}(1) = \frac{C_i(0) \cdot C_j(0)}{C} \cdot r_{ij}(0) ,$$

where $i, j = 1 \div n$; n is the number of areas; $C_i(0)$ and $C_j(0)$ are the growth coefficients for each i-th and j-th transport area, respectively, which are calculated in the form of (4); C is the growth coefficient of the entire urban agglomeration, determined in the form of (1).

At the next steps, the calculation of the volumes of predicted interborough movements is is carry out by the following equalities

$$r_{ij}(t) = \frac{C_i(t-1) \cdot C_j(t-1)}{C} \cdot r_{ij}(t-1) , \ t = 1, 2 \ \ldots \tag{8}$$

The Detroit method of predicting the volume of interborough movements is also trivial in terms of calculations. In practice, this method provides greater reliability of forecasts compared to previous methods of predictive analytics. However, the accuracy of forecasts deteriorates markedly in cases where the growth rate of the individual transport area differs significantly from the growth rate of the urban agglomeration as a whole.

2.4. Fratar method

According to the Fratar method, the predicted volume of movements from the i-th transport area to the j-th one for some future point in time is proportional to the actual volume of movements from the i-th transport area, multiplied by the growth coefficient of transport demand in the j-th area:

$$r_{ij}(1) = C_i(0) \cdot C_j(0) \frac{K_i(0) + K_j(0)}{2} \cdot r_{ij}(0),$$

where

$$K_i(0) = \frac{\sum_{k=1}^{n} r_{ik}}{\sum_{k=1}^{n} C_k r_{ik}} \text{ and } K_j(0) = \frac{\sum_{k=1}^{n} r_{jk}}{\sum_{k=1}^{n} C_k r_{jk}}$$

are the coefficients of growth in movements in the k-th arrea, due to the development of demand in the i-th and j-th transport areas, respectively; C_k is the demand growth factor for transport area k.

The approximate desired solution by the Fratar method is also determined iteratively based on the following expression

$$r_{ij}(t) = C_i(t-1) \cdot C_j(t-1) \frac{K_i(t-1) + K_j(t-1)}{2} \cdot r_{ij}(t-1),\ t = 1, 2\ \ldots \tag{9}$$

In practical modeling of traffic flows, the Fratar method has become most widespread as the most effective extrapolation method for forecasting correspondence matrix.

3. DISCUSSION

In practice, traffic flow management occurs under uncertainty, one of the factors of which is the incompleteness and/or inaccuracy of information relative to the actual volumes of interborough movements of human and freight resources, which is a “stumbling block” in the formation of correspondence matrices. Within the framework of predictive analytics, the possibility of overcoming this type of uncertainty is increasingly being considered in the application of new approaches. One such approach bears on a more adequate representation of the available information.

Regardless of the level of formalization, all observations of the volume of interborough movements within the transport network, in fact, are carried out at the level of so-called "soft measurements", the results of which are reflected in the form of weakly structured estimates [5], for example, in the form of intervals. At the moment, the most adequate formalisms of such assessments are fuzzy sets (fuzzy numbers) [6], which also allow overcoming semantic uncertainty in the value evaluative judgments of specialized experts. Thus, the fuzzy representation of relevant data makes it possible to apply the methodology of contextual fuzzy relations for the formation of actual and forecast correspondence.

When working with fuzzy numbers (sets), various membership functions can be chosen. For example, to describe the actual and predicted values of the elements of the correspondence matrices, as well as the growth and development coefficients, triangular fuzzy numbers are used in [7]. However, from the point of view of the realism and adequacy of the fuzzy description in the study of transport systems, the most convenient formalisms are still trapezoidal functions, or even better, membership functions of the Gaussian or "bell-shaped" type. In particular, if a weakly structured estimate of the volume of the traffic flow between the i-th area-source and the j-th area-sink is expressed as the interval $[a_{ij}, d_{ij}]$, then it can be represented as the fuzzy set $\widetilde{A}_{ij} = \widetilde{A}(r_{ij})$ with the trapezoidal membership function of the following form (see Fig. 2)

$$\mu_{\tilde{A}}(r_{ij}) = \begin{cases} \dfrac{r_{ij}-a_{ij}}{b_{ij}-a_{ij}}, \text{ if } a_{ij} \le r_{ij} \le b_{ij}, \\ 1, \qquad \text{if } b_{ij} \le r_{ij} \le c_{ij}, \\ \dfrac{d_{ij}-r_{ij}}{d_{ij}-c_{ij}}, \text{ if } c_{ij} \le r_{ij} \le d_{ij}, \\ 0, \qquad \text{in other cases} \end{cases}$$

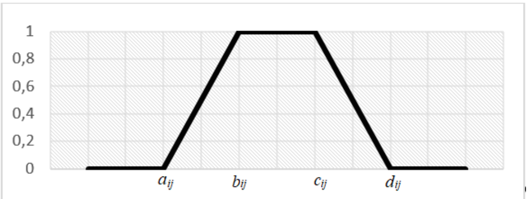

Fig. 2. Trapezoidal membership function

Then, by analogy with [7], in each cell of the correspondence matrix the actual volumes of interborough movements can be reflected by the corresponding fuzzy sets $\tilde{X}_{ij} =< a_{ij}, b_{ij}, c_{ij}, d_{ij} >$ and the volumes of predicted correspondence can be found in the form of the fuzzy set $\tilde{X}^*_{ij} =< a^*_{ij}, b^*_{ij}, c^*_{ij}, d^*_{ij} >$, respectively. Of course, based on the rules of elementary operations on fuzzy sets [6], in accordance with (1) it is possible to represent the growth coefficients in the form of corresponding fuzzy sets, for example, as it was proposed in [7]

$$\tilde{C} = \frac{\tilde{F}}{\tilde{A}} =< c_1, c_2, c_3, c_4 >,$$

where c_k are the parameters of the fuzzy growth coefficient obtained from the results of operations on the corresponding fuzzy sets, and, further, carry out calculations according to the schemes of extrapolation methods in the usual manner with crisp α-cuts of contextual fuzzy formalisms.

It seems to us that the formation of the predictable correspondence matrix within the fuzzy information environment can be implemented in a more interesting manner, i.e. not by simply "immersing" the formulas of extrapolation methods (1) – (8) into the fuzzy environment, but using, for example, a Fuzzy Iinference System, both for case (1) and for a more objective analysis of the growth in transport demand according to formula (2). To begin with, it is necessary to correctly select discrete universes, identify appropriate membership functions, and fuzzify the available weakly structured data relative to interborough movements of the transport network.

Another approach to the formation of the predictable correspondence matrix in a fuzzy paradigm can be the analysis of the dynamics of historical data of interborough movements by modeling the corresponding fuzzy time series with their subsequent forecasting in nominal numbers after defuzzification [5, 8, 9]. In particular, assuming historical data on transport demand in previous periods of observation to be weakly structured, it is possible through modeling the corresponding fuzzy time series to obtain forecasts in nominal units for short-term and even medium-term periods, and then calculate the value of the growth coefficient, and with it all the rest parameters of extrapolation methods for forming the correspondence matrix. But this is the subject of our next research.

4. CONCLUSION

In practical modeling of traffic flows, extrapolation methods are not widely used, primarily due to the irregular dynamics of the structural development of the urban agglomeration. Therefore, the extrapolation methods discussed above are usually used to form correspondence matrix with the forecasting horizon of no more than 5 years. However, the main reason for the inefficiency of extrapolation methods is still the lack of verified (adequate) information relative to the actual volumes of interborough movements, the values of which, as noted above, depend on numerous factors of very different nature. This problem requires a separate consideration, and one of the ways to overcome it lies in the plane of using fuzzy logic elements.

References

1. *Тематический обзор Ассоциации: Математические методы моделирования транспортных потоков.* СПб: ООО «Издательско-полиграфическая компания КОСТА». Выпуск 5. 2019. 72 p. [In Russian: *Thematic review of the Association: Mathematical methods for modeling traffic flows. St. Petersburg: LLC Publishing and Printing Company KOSTA*]
2. Ortizar, I.D., Williumsen, L.G. *Modelling Transport*. *Forth Edition*, Wiley: Hoboken. 2011. 586 p.
3. Горев, А.Э., Бёттгер, К., Прохоров, А.В., Гизатуллин Р.Р. *Основы транспортного моделирования*. СПб: ООО «Издательско-полиграфическая компания КОСТА». 2015. 168 p. [In Russian: Gorev, A.E., Boettger, K., Prokhorov, A.V., Gizatullin R.R. *Fundamentals of Transport Modeling*. *St. Petersburg: LLC Publishing and Printing Company KOSTA*]
4. Селиверстов, Я.А., Селиверстов, С.А. Методы и модели построения матриц транспортных корреспонденций. *Научно-технические Ведомости СПб ГПУ: Информатика. Телекоммуникации. Управление*. 2015. № 2-3 (217-222). P. 49-70. [In Russian: Seliverstov, Ya.A., Seliverstov, S.A. Methods and models of the construction of transport correspondence matrix. *Scientific and Technical Bulletin of St. Petersburg State Polytechnical University: Informatics. Telecommunications. Control*]
5. Рзаев, Р.Р. *Интеллектуальный анализ данных в системах поддержки принятия решений*. Saar-bruchen (Germany): Palmerium Academic Publishing. 2016. 127 p. [In Russian: Rzayev, R.R. *Data Mining in Decision Support Systems*. Saarbruchen (Germany): Palmerium Academic Publishing]
6. Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences.* 1965. Vol. 8. No. 3. P. 199–249.
7. Кошуняева, Н.В. Применение нечетких характеристик при построении матриц корреспонденций экстраполяционными методами. *Современные Наукоемкие Технологии*. 2022. № 5. P. 209-213. [In Russian: Koshunyaeva, N.V. Application of fuzzy characteristics in the construction of correspondence matrices by extrapolation methods. *Modern High Technologies*].
8. Rzayev, R.R., Abdullayev Kh. Kh., Alizada P.E., Salmanov F.M. Fuzzy time series forecasting by the example of the Dow Jones Index dynamics. *Lecture Notes in Networks and Systems.* 2022. Vol. 438. P. 216-228.
9. Rzayev, R.R., Alizada P.E., Mehtiyev T.Z. Volatile time series forecasting on the example of the dynamics of the Dow Jones index. *Problems of Information Society*. 2023. Vol. 14. No. 1. P. 14-25.