Research Journal of Applied Sciences 11 (8): 722-728, 2016
ISSN: 1815-932X

Use of Tax Reporting Data in a Freight Flows Correspondence Matrix When Building a Regional Transport Model

P.V. Loginov, S.N. Kuznetsov, A.N. Zatsepin and V.A. Pavlov
"TFG Kvanteks" LLC, Kudryavtseva Str., 34, Ryazan, Russian Federation

Abstract: Each transport model is based on a correspondence matrix. To achieve an adequacy of a transport model to the real conditions, all of its layers should be described in the structure of the transport demand each of that layers has its own correspondence matrix. When developing a macro model of a regional transport system the most urgent layers of the transport demand are business and commercial correspondences. The purpose of the work the results of which are presented in this study is the development of methods for creating freight flows correspondence matrices. The source of the original data are publicly available tax records with a breakdown to sub-federal entities and municipal entities which make it possible to determine the total number of trucks by the engine power characteristics groups as well as information about economic entities in municipal entities. The analysis of a raw data composition to build matrices of freight flows correspondence was carried out and a methodical approach to collection of certain types of raw statistical data based on the Federal Tax Service reports was developed. The proposed technique makes it possible to build current matrixes of freight flows correspondence that could be used in the future to create models of transportation districts at the regional level with the use of different sets of traffic simulation. The described technique is also applicable in the construction of predictive traffic models. The results can be used by both domestic developers and specialists working in countries with transitional type of economy (India, Pakistan, Argentina, Mexico, Brazil, some former Soviet republics and so forth) in the conditions of limited capacity to collect statistical and accounting source data by standard techniques. The researchers believe that the technique presented in the study on creating correspondence matrices can also be useful for foreign suppliers of traffic simulation systems to account for the specifics of the Russian Federation in their products.

Key words: Correspondence matrix, traffic simulation, cargo transportation, information transport model, freight flows correspondences, business correspondences, transport correspondences, labor correspondences, traffic stream

INTRODUCTION

One of the primary conditions for creation of a modern socio-economic basis as well as a comprehensive dynamic development of cities and regions are their transport systems (Safronov, 2005). Predictive traffic models of cities and regions are built to obtain information about the state of a transport system as well as the possible distribution of traffic streams over the road network according to the different input parameters (Safronov, 2005; Shvetsov *et al*., 2009)

Given the trends of continuous development of automotive engineering and logistics systems as well as structural changes in the Russian economy and in order to ensure compliance of the transport infrastructure of cities and regions with the growing demands on transport movements, activities on monitoring and prediction of transport systems development become important. In this regard, a need to create predictive models of transport of cities and regions grows also (Shvetsov *et al*., 2009).

A transport model includes transport demand and supply models as well as the distribution model which describes the way they interact. Using modern methods of mathematical modeling, transport model allows the calculation of the transport network loading to carry out. This study is devoted to the creation of a transport demand model.

When modeling a transport demand we must first make its structuring which consists in the allocation of the specific groups, "layers" of transport demand in possible transport movement options. The demand structure detalization (number of demand layers) directly determines the level of compliance of the final transport model to the real situation in a simulated territory. In order to build models transport demand, the possible types of transport movements can be roughly grouped into several main

Corresponding Author: P.V. Loginov, "TFG Kvanteks" LLC, Kudryavtseva Str., 34, Ryazan, Russian Federation

layers of demand: labor correspondence (daily trips from home to work and back); cargo transportation; business trips (trips to the workplace on business matters); trips from home with social (household) purposes.

The transport demand models basically include Correspondence Matrixes (CM) describing the movement of road transport between the corresponding points in terms of "Origin-Destination" and are the result of raw data processing that characterize the movements (Batishchev *et al*., 2016; Schnabel and Lohse, 1997; Ortuzar and Willamsen, 2001; Vuchic, 2007). An appropriate correspondence matrix is built for each layer of the transport demand.

At the initial stages of building a transport model, selection of corresponding points or zoning is performed, i.e., division of the territory in transportation districts (Kotikov, 2012). The bases of determining the boundaries ofTransportation Districts (TD) in the course of zoning are the following basic principles: compactness of a district territory; possibility to accurately describe the functionality of each district in the socio-economic structure of the simulated territory; availability of data on transport demand with a breakdown to transportation districts (Abrahamsson, 1998). The choice of the method of building transport models significantly depends on a simulation scale.

The issue of building of the city transport model in almost all its aspects, including the methodology and features of correspondence matrix creation was studied in detail and described in the literature (Shvetsov *et al*., 2009). The basic transport correspondences within the city (agglomeration) borders are movements of people, mostly in the morning and evening "rush hours" that are grouped in labor, education, social, household and similar layers of demand. Impact of transit traffic on the urban transport system is negligible because those movements occur mainly on roundabout routes. Thus, concerning freight traffic and labor correspondences a city is a self-sufficient object.

Transportation systems of the Russian Federation regions are characterized by a number of factors which determine the difference between them and the transport systems of cities and urban agglomerations. The main factors include: population density; structure of the economy, including industry and agriculture; levels of production development by industries; transport network placement area. All this leads to a difference of methodological and technological approaches in building a region transport model and a city transport model (Safronov, 2005; Kotikov, 2012).

Currently, experts in the field of regional transport issues give the basic attention to problems of synchronization of cargo transportation by different modes of transport (Shvetsov, 2003; Bekmagambetov, 2012). Thus, the main areas of contemporary research activities in the field of traffic simulation at the regional level are utilitarian applied researches. Much less attention is paid to research of creation of regional transport system models.

When creating a regional transportation model the low level of detalization of transport demand should be taken into account; it is due to the fact that the main transport movements across the region are business correspondences and cargo transportations. Labor correspondences are virtually confined by the area within the boundaries of the urban agglomeration, including the very territory of the city and 25-30 km long zone around city limits. Given the traffic conditions on urban road network where the average speed in "rush hours" is practically not >10-15 km^{-1}, the time of 2 h spent on daily labor correspondence (at rated working day 8 h) may be considered as a top of rationality to a worker.

In practice of models building, insufficient attention is given to a freight transport as a transport demand layer in comparison with other types of correspondence. At the same time, in the conditions of structural changes in the economic system of the country, economic diversification, development of small and medium-sized business segments, increased mobility of the population, the demand for transportation continuously increases. This applies both to the volume of goods transported, geography of correspondences and to increase in the number of types and improve the performance of freight vehicles. In this connection, there are also increased accuracy requirements posed to freight flows correspondence matrix as one of the main layers of the demand for mobility in the region what largely determines the adequacy of regional transport model as a whole.

As noted earlier, the correspondence matrixes describe communications between the point pairs (transportation districts) on the basis of the raw data characterizing the movements using motor vehicles. Accordingly, high demands of model developers to the sufficiency and quality of the raw data are reasonable since this information is the basis of the simulation regardless of the use of software and hardware tools and techniques. The lack in the conditions of modern Russia of an established mechanism for access to the necessary statistical information makes the problem from the collection of raw data for the traffic simulation in almost the entire territory of the country. The solution to this problem using standard techniques requires involvement of significant manpower and is time consuming (Yakimov, 2013).

Practical activities of the researchers in the field of traffic simulation confirms the existence of the above problem and shows that the use of publicly available information of various authorities in the capacity of a valuable and reliable source of raw data is unreasonable because of its insufficiency and quite often irrelevancy. At the same time we can use in the capacity of the most appropriate source of information the reports of the Federal Tax Service of the Russian Federation published in accordance with the requirements of the legislation (Anonymous, 1995). This assumption has determined the direction of research the results of which are described in this study.

These researches focus on the development of a technique for creating a freight flows correspondence matrix based on data from tax reporting data. For these purposes there were designed methodical approaches to select sources and methods for obtaining raw statistical data and for processing the information received.

Description of a raw data collection problem for traffic simulation: Development of mathematical transport models deals with a large amount of raw data collection of which is accompanied by serious difficulties both in respect of the information quality and in terms of its production (Yakimov, 2013).

Developed European countries, the USA and Canada over the past 100 years systematically conduct numerous sociological studies involving almost all spheres of life and activity of people, society and the state including in the transport sector. Results of those researches are freely available to specialists that yields a variety of quality, systematic and chronologically grouped data (Shvetsov and Ushchev, 2008). In countries with a transitional type of economy (India, Pakistan, Argentina, Mexico, Brazil, some former Soviet republics and so forth) including the Russian Federation, taking into account absence of such studies there as well as the dynamics of the economic situation, the collection of raw data for traffic simulation is a significant challenge and solution of the problem requires a scientific approach.

In view of solving the problem of development of methodological approaches to the choice of sources and methods for preparation of statistical raw data, let's consider the current situation in this sphere as an example of a typical region of the Russian Federation.

An analysis of publicly available information leads to the conclusion that the source of the original data can be information resources of state statistics authorities as well as statistical reports in the official websites of local authorities. The practice of collecting raw data that was obtained by the researchers allows describing the current situation as follows:

- The necessary information presented is in open reliable sources in unstructured form
- Generally, regional administrations do not have necessary data in full and in some cases the data are completely absent
- Interacting with the authorities on provision of the necessary information is usually associated with significant delays and various bureaucratic obstacles
- A significant portion of the resulting raw data does not satisfy the requirements for them to be used for creation of transport models and requires additional preparation
- Preparation of raw data obtained from the sources indicated above for their use in the traffic simulation can be carried out in two directions
- Clarification of the information received from public reliable sources and through formal requests to various specialized agencies
- Creation of special techniques to obtain the necessary raw data from available accumulated statistics and other data types

The option of official clarification of raw data has positive aspects reflected in high quality of the information but contains flaws relating mainly to the timing of requests execution.

The second direction of data preparation provides for availability of special techniques which allow receiving the required raw data from the available accumulated statistical data as well as other types of information and have unlimited applicability and usability. The development of special techniques within the framework of the research has determined the scientific nature of the work and innovative approach to solving the problem. Reports of the Federal Tax Service of the Russian Federation have been selected in the capacity of a source of raw statistics for the foregoing reasons.

MATERIALS AND METHODS

Methods of preparation of the statistics raw data on the basis of the federal tax service reports: The demand layer reflecting a cargo transportation is one of the main components of any regional transport system demand model. However, it should be noted that cargo transportation in the said aspect are poorly understood and described in the Russian literature (Shvetsov, 2003). This is due to: firstly with an exponentially growing volume of cargo transportation by motor transport on a national scale as well as the number and nomenclature of the vehicle fleet; secondly with the development of freight motor transport both in Russia and abroad and with the qualitative improvement in the technical performance of vehicles. At the same time, the number of trucks in the region, concerning payload is one of the main data source for building of the regional transport model. This information is not available in the reports of state statistics bodies. Accounting and registration of vehicles is carried out by relevant units of road-traffic safety inspection of the Ministry of Internal Affairs (State Road Traffic Safety Inspectorate). Thus, the data relating to vehicles registered in the region can be obtained only at the official request to the Inspectorate division of the region.

At the same time, the inspections as part of inter-agency cooperation are required to send the following data to the authorities of the Federal Tax Service:

- Information about all vehicles and persons for which they are registered as of December 31 of the preceding calendar year
- Information about the sale and the change of the car owner within 10 days after registration or de-registration of vehicles

Within the framework of the chosen approach, the researchers propose to use in the capacity of raw data the open reporting information of the tax authorities, previously prepared with the use of the developed methodological approaches, the essence of which is as follows.

The Federal Tax Service of the Russian Federation reports on its web-site the data formed per the sub-federation entities with a breakdown to municipal entities and allowing assessment of the demand for freight transport in the region.

It is a question of the report on the tax base and structure of charges on Transport Tax (TT) which is presented with a breakdown to the sub-federal entities and to municipal entities.

This report includes data on commercial vehicles registered for the citizens and organizations in the corresponding territory. Thus, using the records of the tax authorities in open access, it is possible to get actual raw data for creating freight flows correspondence matrix. Reported data on commercial vehicles are formed into five groups depending on the engine power:

- Up to 100 hp (up to 73.55 kW) inclusively
- Over 100 hp up to 150 hp (>73.55 and 110.33 kW) inclusively
- Over 150 hp up to 200 hp (>110.33-147.1 kW) inclusively
- Over 200 hp up to 250 hp (over 147.1-183.9 kW) inclusively
- Over 250 hp (>183.9 kW)

Table 1: Payload parameters of commercial vehicles depending on their engine power

Engine power (kW)	Vehicle payload (t)
≤73.55	≤3.06
73.55-110.33	3.06-4.6
110.33-147.1	4.6-6.13
147.1-183.9	6.13-7.66
>183.9	>7.66

The specified raw data on commercial vehicles are sufficient for building both a general freight flows correspondence matrix and individual correspondence matrices for commercial vehicle systems of trucks formed by payload parameters shown in Table 1.

The technique of creating a correspondence matrix: The proposed method allows upon using a minimum number of raw data to perform CM calculation at a first approximation. The calculation used data on the number of commercial vehicles registered in the region, the number of economic entities within the municipal entity, as well as an assessment of its administrative status. Considering that cargo transportation is the prerogative of the economic entities and in the absence of information about the types of these entities, it is advisable to use the average figure for the region of the number of commercial vehicles owned by a single economic entity. Based on these raw data, let's suggest that each economic entity generates a number of freight flows correspondences equal to the number of existing serviceable vehicles owned by it that on average provides sufficient accuracy of the calculation of the relative coefficients of movements between transportation districts.

To create a matrix of freight flows correspondences N between the transportation districts which accounts for n, we propose the following sequence of actions.

At the first stage the vector O_t is formed with dimension of n each element of which O_{tj} is equal to the number of economic entities in the transportation district and is proportional to the transport stream incoming in the transportation district that is to transport demand.

Number of economic entities with a breakdown to transportation districts B_i and the total number of economic entities in the region B_0 is determined on the basis of the reports by the Federal Tax Service according to the technique developed by the researchers of this study and described by Vuchic (2007).

The next step is to build a connection matrix K with dimension of n×n which elements are connectivity coefficients k_{ij} between ith and jth corresponding transportation districts. The values of connectivity coefficients are determined on the basis of the provisions from “guidelines for predicting traffic volume on the roads” (Mot, 2003), depending on the administrative status and mutual subordination of central settlements in the transportation districts.

Then, a traffic stream matrix D_t with dimension of n×n is formed, the elements of which are equal to the transport streams coming from the TD and are calculated using the numbers of commercial vehicles in the transportation districts and the matrix K elements according to the equation:

$$Dt_{ij} = A_i \times K_{ij},\ i = \overline{1, n},\ j = \overline{1, n} \quad (1)$$

where, A_i the number of commercial vehicles in the ith transportation district that are suitable for making freight flows correspondences; the number is determined under the equation:

$$A_i = A_o \times K_{ispr} \frac{B_i}{B_o},\ i = \overline{1, n} \quad (2)$$

Where:

- A_o = The number of commercial vehicles registered in the region that is determined from the report of the Federal Tax Service on the tax base and structure of transport tax charges with a breakdown to the sub-federal entities
- K_{ispr} = The coefficient of serviceability of vehicles showing their readiness to enter the route to perform a correspondence. In accordance with (Loginov *et al*., 2015a) the value of Kispr is taken equal to 0.85 (i.e., 15% of the vehicle fleet is constantly under repair, maintenance, etc)

The next step is to create a provisional matrix G of the so-called “a priori probability of correspondence origination” (Shvetsov, 2003) with dimension of n×n and calculation of the matrix elements under equation:

$$G_{ij} = \beta \frac{Dt_{ij} \times Ot_j}{L_{ij}^2},\ i, j = \overline{1, n} \quad (3)$$

Where:

- β = The balancing factor being the ratio of the number of commercial vehicles registered in the region A_o to a number of economic entities in the region B_o and necessary to ensure the balance of incoming and outgoing flows of the transportation district
- L_{ij} = Elements of the matrix L that represent the values of the distances between the respective TDs along the roads connecting them with the shortest length (based on the regional road map)

Equation 3 actually describes the gravitational approach to creating a provisional correspondence matrix. “Gravity model” is historically one of the first models proposed for an estimation of inter district correspondence (Carrothers, 1956; Voorhees, 1955; Wilson, 1967, 1971). For the purposes of this study, it is preferred because of its versatility and simplicity (Shvetsov, 2003). In addition, the use of this approach in the technique of building correspondence matrixes for the transport demand layer “labor correspondence” as described by Loginov *et al.* (2015b) has shown its relevance in building of adequate correspondence matrix as the first approximation.

When creating the final correspondence matrix N as a result of which the value of freight flows correspondence N_{ij} would be obtained, it is necessary to make the normalization of the matrix elements G:

$$N_{ij} = \frac{G_{ij}}{\eta_j}\ i = \overline{1, n}, j = \overline{1, n} \quad (4)$$

using normalization factor η_j which is calculated by the Eq. 5:

$$\eta_j = \frac{\sum_{i=1}^{n} G_{ij}}{Ot_j},\ j = \overline{1, n} \quad (5)$$

This normalization ensures that the total number of departing commercial vehicles from this TD complies with the existing demand.

For the final formation of the freight flows correspondence matrix it should be borne in mind that the loaded mileage proportion (the ratio of a loaded mileage of a vehicle to its general “back and forth” mileage) as a rule has a value of about 0.5 t, i.e., each travel with a load falls to a single unladen travel. It is important when planning a transport network load. Typically, in the simplest case, the unladen travels are accounted for as a certain proportion of laden travels but in the opposite direction. Then the total number of travels between the ith and jth TDs in one direction can be determined by the following equation:

$$SN_{ij} = N_{ij} + \lambda N_{ji}\ i = \overline{1, n},\ j = \overline{1, n} \quad (6)$$

Where X loaded mileage proportion (“unladen”) determined empirically but mostly this value is in the range of 0.7-0.9 (Loginov *et al*., 2015).

Using the technique presented above allows making calculations of freight flows correspondence matrices based on open public data including official government data.

RESULTS AND DISCUSSION

The technique for creating correspondence matrixes described in this study allows effectively to obtain the matrixes which reflect the average daily movements of vehicles which provide cargo transportation services in the territory of a region.

During development of the technique for creating freight flows correspondence matrixes based on open and public tax reporting data, methodological approaches to the choice of sources and ways for obtaining statistical raw data as well as for preparation of statistical raw data based on reports of tax authorities have been developed.

The technique presented in this study was tested within the framework of the applied research on the topic "Development of technology for building information model of transport systems for a typical Russian region including the development of the algorithm for calculating the efficiency of operation of toll and other roads of federal, regional and local level based on world-class software system PTV Vision® VISUM/VISSIM (Germany)" for building the informational transport model for a typical Russian region. The results obtained have confirmed the relevance of the developed techniques for building regional transport models.

CONCLUSION

It can be concluded on the basis of these results that the use of methodological approaches described in the study for the preparation of statistical raw data based on the tax authorities' reports can successfully solve the above indicated problem of raw data collection in order to build freight flows correspondence matrixes.

Thus, the proposed technique allows in very effective and versatile way to form a freight flows correspondences matrix based on open data of the tax authorities. The technique provides a minimal dependence on a developer on a database for accounting of vehicles the availability of which in different regions and countries can differ greatly.

ACKNOWLEDGEMENTS

The studies were conducted in accordance with the Agreement No. 14.588.21.0001 dd. September 26, 2014 with the Ministry of Education and Science of the Russian Federation to grant subsidies for the performance of work on the theme "Development of technology for building of information model on transport systems of a typical Russian region including the development of the algorithm for calculating the efficiency of operation of toll and other roads of federal, regional and local level on the basis of a world-class PTV Vision® VISUM/VISSIM software package (Germany)" together with the company "A+S Consult GmbH Forschungund Entwicklung" (Dresden, Germany) as part of the applied research in priority areas with the participation of research organizations of EU member-states within the framework of the Provision 2.2 of the Federal Target Program "Research and development in priority directions of scientific-technological complex of Russia for 2014 - 2020. Unique identifier for Applied Scientific Research (project) RFMEFI58814×0001.

REFERENCES

Abrahamsson, T., 1998. Estimation of origin-destination matrices using traffic counts-a literature survey. Interim Report IR-98-021, International Institute for Applied Systems Analysis, May, 1998, pp: 1-27.

Anonymous, 1995. Federal Law on information, informatization and the protection of information. Federal Law No. 24-FZ. http://fas.org/irp/world/russia/docs/law_info.htm.

Batishchev, S.V., K.V. Ivkushkin, T.V. Iskvarina, V.A. Nikiforov and P.O. Skobelev, 2016. Analysis of the possibilities for application of multi-agent technology in transport logistics problems. Institute for Control of Complex Systems RAN, February 2016.

Bekmagambetov, M.M., 2012. Analysis of modern transport modeling software. J. Automotive Eng., 5: 25-29.

Carrothers, G.A.P., 1956. An historical bedew of the gravity and potential concepts of human interaction. J. Am. Inst. Planners, 22: 94-102.

Kotikov, Y.G., 2012. Aspects of the use of ArcGIS technology in transport model of St. Petersburg. Mod. Probl. Sci. Educ., Vol. 3.

Loginov, P.V., S.E. Shirobakin and S.N. Kuznetsov, 2015a. Creation of a business correspondence matrix for building the regional transport model based on tax reporting data. Fundam. Applied Res. Mod. World, 11: 33-39.

Loginov, P.V., S.E. Shirobakin and S.N. Kuznetsov, 2015b. Creation of labor correspondence matrix on the basis of public statements for the development of computer transport model of the region. Int. J. Applied Eng. Res., 10: 45449-45453.

MOT., 2003. Guidelines for predicting traffic volume on the roads. The Order of the Russian Federation Ministry of Transport, No. OS-555-p, June 19, 2003. http://aquagroup.ru/normdocs/6562.

Ortuzar, J. and G. Willamsen, 2001. Modelling Transport. Wiley, New York, USA., ISBN: 0-471-86110-3.

Safronov, E.A., 2005. Transport Systems of Cities and Regions: Manual. ASV Publisher, USA., ISBN: 5-93093-345-6, pp: 12-25.

Schnabel, W. and D. Lohse, 1997. Grundlagen der Straâenverkehrstechnik und der, Band 2: Verkehrsplanung. Verlag Bauwesen, Berlin, ISBN-13: 9783345005671, Pages: 432.

Shvetsov, V.I. and F.A. Ushchev, 2008. Modeling tools PTV VISION® VISUM as a basis of transport systems control technology Organization and traffic safety in the major cities. Proceedings of the 8th International Scientific and Practical Conference, November 5-6, 2008, Baku.

Shvetsov, V.I., 2003. Mathematical modeling of traffic flows. Automation Remote Control, 64: 1651-1689.

Shvetsov, V.L., A.V. Prokhorov and I.V. Ilyin, 2009. Transport models in the public administration system. Sci. Technical Bull. Saint Petersburg State Polytechnical Univ., 85: 20-28.

Voorhees, A.M., 1955. A General Theory of Traffic Movement. Automotive Safety Foundation, Washington, DC., USA.

Vuchic, V.R., 2007. Urban Transit Systems and Technology. John Wiley and Sons, USA., ISBN: 9780471758235, Pages: 602.

Wilson, A.G., 1967. A statistical theory of spatial distribution models. Transp. Res., 1: 253-269.

Wilson, A.G., 1971. A family of spatial interaction models and associated developments. Environ. Plann. A, 3: 1-32.

Yakimov, M.R., 2013. Transport Planning: Creation of Transport Models of Cities. Monograph, Moscow, ISBN: 978-5-98704-729-3.