International Business Management 10 (19): 4659-4664, 2016
ISSN: 1993-5250

Creation of a Business Origin-Destination Matrix for Building of the Regional Transport Model Based on Tax Reporting Data

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

Abstract: A necessary and important step in the development of a transport model is to determine the demand for the transport in a modeled area which is usually described by the origin-destination matrix. The aim of the work which results are presented in this study is to develop methods to create business origin-destination matrices based on open data including official government data. The raw data used were publicly available tax records with a breakdown to municipal entities allowing to determine the total number of economic entities. The proposed method using the principles of "origin-destination gravity models" allows relevant business origin-destination matrices to build for subsequent use in creation of transport models of Russian regions with use of different complexes of transport modeling. The technique for creating origin-destination matrices described in the study is also applicable in the construction of predictive traffic models. The obtained results can be used both by domestic developers and by professionals working in countries with transitional type of economy (India, Pakistan, Argentina, Mexico, Brazil, some former Soviet republics and so forth) having difficulties in obtaining raw data according to standard procedures. The authors believe that the method to create the origin-destination matrices represented in this study may also be useful to foreign suppliers of transport modeling systems to account for the specifics of the Russian Federation in their products.

Key words: Origin-destination matrix, transport modeling, informational transport model, business correspondences origin-destination correspondences, work correspondences, traffic flow

INTRODUCTION

Transport systems of cities and regions are essential to create the most favorable conditions for the functioning of all branches of the economy to ensure the smooth, safe and efficient movement of people and goods. Predictive traffic models of cities and regions are built to receive information about the state of a transport system as well as the possible allocation of traffic flows along the road network depending on the input parameters (Safronov, 2005; Shvetsov *et al*., 2009). A transport model is a mathematical description of the transport system allowing on the basis of information on transport demand, transport supply and way of their interaction to make the calculation of the transport network load. It consists of two main components: the transport demand model and transport supply model, as well as method of their interaction (distribution model).

This study discusses the possibility to create a transport demand model based on open data including official government data. To create a model of transport demand in the first place, it is necessary to structure the demand: to form functionally specific groups or "layers" of transport demand, from the general nomenclature of possible transport movements. The quality of the final transport model depends on the detailed elaboration of the demand structure (the number of layers of demand); upon that all the diversity of the objectives of origin-destination correspondences can be conditionally divided into several major segments of demand:

- Work correspondence (trips "home-work" and vice versa)
- Cargo transportation
- Business travels (travels on official matters when the generator of origin-destination correspondence is "work")
- Social travels (generator of origin-destination correspondence,as a rule is "home"; education, health care, social security, service facilities, etc may act in the capacity of the purpose of the travel)
- Trips with recreational and tourist purposes

Construction of the transportation demand model uses Origin-Destination Matrices (ODM) which describe

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

the communication between the corresponding points in terms of “origin-destination” on the basis of the raw data characterizing the movement using motor vehicles (Shvetsov and Ushchev, 2008; Batishchev *et al.*, 2016; Ortuzar and Willamsen, 2001; Vuchic, 2007). Own origin-destination matrix is created for each layer of a transport demand.

One of the primary operations for construction of the transport model is allocation of corresponding points or zoning: division of the considered territory in the transportation districts. Determining of the Transportation Districts (TD) is based on the following basic principles: low dispersion of district areas; the ability to accurately describe the functionality of each district in the socio-economic structure of the examined object; availability of data on the traffic demand of each district (Abrahamsson, 1998). Ways to build transport models depend strongly on the modeling scale.

The question of forming an origin-destination matrix as part of a city (metropolitan area) transport model has been well studied and is methodologically clear. The basis of intracity origin-destination correspondences is made by movements of citizens, primarily in the “rush hour” which form the work, social, educational and similar layers of demand. Transit cargo flows affect them only slightly because movements are performed in bypass routes. A city is an all-sufficient object in relation to transport of cargoes and the work origin-destination correspondences.

Regional (in relation to the administrative and territorial structure of the Russian Federation) transportation systems are characterized by a number of factors which distinguish them from urban systems. The main of them are the density of population, the nature of production (industry and agriculture), the level of production, the value of the service area. All this leads to a difference of methodological and technological approaches in building of transport model of the region and the city transport model. The majority of works devoted to the problems of transport in the region pay main attention to the question of synchronization of cargo transportation by different modes of transport (Shvetsov, 2003; Bekmagambetov, 2012). Significantly less attention is paid to the issues of creating their own regional transport model.

Regional transport models are characterized by a limited number of layers of demand on movements. It is usually assumed that the main transportations in a region are business origin-destination correspondences and cargo transportations. This statement is based on the fact that a work origin-destination correspondence is limited to the territory which borders are located at a distance from the city limits of a half-hour accessibility by individual or public transport, i.e., not >30 km. In fact, this is area of an urban agglomeration. Since the nominal duration of the working day in most urban economic entities is 8 h, the cost of >2 h a day for travel from home to work and back may be considered unreasonable. Given the traffic conditions on urban road network where the average speed in the “rush hour” is practically not >10-15 kmh^{-1}, the statement on the limitation of areas of work origin-destination correspondence by boundaries of the urban agglomeration can be considered quite reasonable.

Business travels (trips) as a segment of the demand for transportation are neglected in the practice of building the transport models compared with other types of origin-destination correspondences. At the same time, a large number of representatives of economic entities makes on their own daily business travels within the region as well as in the neighboring regions in conditions of weak logistics of transport companies as well as for the purpose to perform a number of administrative, financial, executive and other procedures. The accuracy of the resulting business origin-destination matrix as one of the main layers of the demand for mobility has a significant impact on the adequacy of the regional transport model as a whole. As noted earlier origin-destination matrices describe the communication between the corresponding points (transportation districts) on the basis of the raw data characterizing the movement with use of motor vehicles. Therefore, the question of comprehensive and qualitative raw data available by the developers of transport models is extremely relevant, regardless of the hardware and software tools and techniques used in the modeling process. In the conditions of absence in Russia an established mechanism for obtaining from the public authorities the necessary statistical and accounting information so far, the availability of raw data for transport modeling is a problem throughout the country. The solution for this problem by standard methods requires the use of significant manpower and is time consuming (Yakimov, 2013).

The experience gained by the authors in the process of practical activities in the field of transport modeling confirms the existence of this problem and shows that the use of raw data as an appropriate source of publicly available information from various authorities is unreasonable because of its insufficiency and often irrelevancy. However, one of the most appropriate sources of information can be considered the reports of the Federal Tax Service of the Russian Federation published as freely available in accordance with the requirements of the legislation. This fact led to the main direction of research the results of which are presented in this study. These studies have been devoted to developing methods of creating business origin-destination matrices on the basis of tax reporting data. To do this, the methodological approaches to the choice of sources and methods for the preparation of statistical raw data as well as to their treatment have been developed.

MATERIALS AND METHODS

Description of the problem on collection of raw data for transport modeling: Development of mathematical transport models requires a large amount of raw data which collection is related with serious difficulties both in terms of quality and in terms of the preparation (Yakimov, 2013). Over the past 100 years in the developed countries there were systematically conducted numerous sociological studies in a wide range of directions, including in the transport sector, the results of which are freely available to experts which obtain various qualitative, systematic and chronologically grouped data. In countries with a transitional type of economy (India, Pakistan, Argentina, Mexico, Brazil, some former Soviet republics and so forth) and also in the Russian Federation both due to absence of these studies in the necessary volume and dynamic changes in economic conditions, raw data collection for transport modeling presents a critical problem that requires scientific solution. To solve the problem on the development of methodological approaches to the choice of sources and methods for the preparation of statistical raw data, let's consider the situation with information support foru models of development in existing conditions, having identified possible sources of raw data for transportation modeling and methods for their collection in the Russian Federation with the example of a typical region.

An analysis of publicly available information points that in the capacity of raw data there can be used the information resources of the state statistics, state and regional authorities as well as statistical reports contained in the official sites of local governments. The practice of collecting raw data tested by the authors allows us to formulate a number of common problems:

- The necessary information is not presented in a systematic way in the reliable public sources
- As a rule, regional administrations do not have the necessary data in full and in some cases, these data are not available at all
- Interaction with the authorities to provide the necessary data is usually associated with a considerable amount of time and with the need to overcome various bureaucratic hurdles
- A significant portion of the obtained raw data requires preprocessing for its use in order to create a transport model

Preparation of raw data obtained from the sources mentioned above may be carried out in two ways:

- Updating the information received from the reliable public sources through formal requests to various specialized agencies
- Creation of special techniques to obtain the necessary raw data from available aggregated statistics and data of other types

Option of official clarification of raw data obtained from public information resources of the region is attractive from the standpoint of improving the quality of information, but contains disadvantages relating mainly to the request lead times, including the need to direct new requests in the case of necessity an additional data clarification. The second direction of data preprocessing involves availability of special techniques for data aquisition from available aggregated statistics as well as other types of information. These techniques should have unlimited applicability and usability. The development of special techniques within the frameworks of the research defined the scientific nature of the work and innovative approach to solving the problem. Reports of the Federal Tax Service of the Russian Federation were selected in the capacity of a source of basic statistical data due to previously mentioned reasons.

Methods of preparation of the statistics raw data on the basis of the federal tax service reports: Considering the particularities of business origin-destination correspondences layer in the regional transport model, it is logical to assume that the demand for mobility is proportional on the average to the number of economic entities in the transportation district. It should be noted that the use of government statistics data in this case is complicated. As part of their methodology, bodies of state statistics annually track the number of large and medium-sized businesses and the number of small businesses and individual entrepreneurs (IP) only on a sample survey every 5 years. This statistic is conducted mainly for the entire region as well as for major cities and districts.

Within the framework of the chosen approach, the authors 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 approach the essence of which is as follows. In the Russian Federation, Federal tax service represents on its internet site reporting data generated by the subjects of the federation broken down by municipalities and allowing to estimate the transport demand for business travels in the region. As mentioned above, at the stage of zoning the basic principles of allocation of traffic districts are the ability clearly to describe the functional purpose of each area in the socio economic structure of the object under consideration (in this case in the region of the Russian Federation) and the availability of state social statistics data in all areas (Abrahamsson, 1998). Therefore, when creating a regional transportation model it is reasonable to use the division of the region by Municipal Formations (MF). In some cases, it is possible to combine several municipal formations. This allows the use for data collection the following statistical tax reporting adopted in the Russian Federation:

- A report on the tax base and structure of charges on Corporate Property Tax (CPT) with a breakdown of municipal formations
- A report on the tax base for the Unified Tax on Imputed Income for individual activities (UTII) with a breakdown of municipal formations
- A report on the tax base for the tax paid by taxpayers in connection with the use of a Simplified Tax System (STS) with a breakdown of the municipalities

The choice of these tax reports is based on the following reasons. Report on CPT shows the number of organizations which are payers of corporate property tax. It should be noted that such a tax is paid only by organizations that fall under the general taxation system. Also, this report shows the number of institutions of government, education, health, etc. having privilege on this tax. Reports on UTII and STS contain data about the number of enterprises and individual entrepreneurs using special tax regimes. Thus, the total number of active business entities in the municipal formation and the dynamics of their change in a few years can be defined.

Let's consider formation of the raw data on the following example. Fragment of a report on the tax base and structure of charges on corporate property tax for a specific municipal entity is given in Table 1. This report makes it possible to determine the number of organizations working in the municipal formation, owning property and applying the common form of financial statements. In this case, these are 2186 companies of which 424 public institutions having privileges on this tax. The fragment of UTII report is shown in Table 2, the fragment of a report on STS is shown in Table 3.

Table 1: Fragment of a report for corporate property tax

Indicators	Values
Number of taxpayers (units)	2 186
Charged the tax payable, without applying tax privileges	1 762
Applying tax privileges	424

Report on the tax base and structure of charges on corporate property tax

It follows from the three reports that in the municipal formation there operate 23147 business entities of which 10,693 organizations and 12,454 individual entrepreneurs. The information obtained can be used later at the stage of creation of origin-destination matrices in the course of building a regional transport model.

The technique of creating an origin-destination matrix: The proposed method allows calculating the first approximation of an origin-destination matrix in the conditions of using the minimum raw data volume. The calculation uses data on the number of business entities as part of a municipal formation and the assessment of its administrative status. In the absence of information about the types of business entities we assume that each of them generates one origin-destination correspondence that provides on average sufficient accuracy of the calculation of the relative movement coefficients between transportation districts.

We propose the following sequence of actions to realize the methodical approaches for creation of a business origin-destination matrix between N transportation districts which number is equal to n. In the first stage, on the basis of the raw data on the number of economic entities with a breakdown by transportation districts, the n-dimension vector Ob is formed and each its element is the number of organizations in the ith transportation district and actually determines the transport traffic being a part of the transportation district. The next step is to build a $n \times n$ dimensional matrix K which elements are the coefficients of association k_{ij} between the ith and jth corresponding transportation districts. The values of coefficients of association are determined on the basis of the provisions of the "guidelines for predicting traffic volume on the roads" depending on the administrative status and mutual subordination of central settlements in the transportation districts. Further, the

Table 2: Fragment of UTII report

Indicator name	Values of indicators: Individual entrepreneurs	Organizations	Total
The number of taxpayers submitting tax statements for the unified tax of imputed income for certain types of activity (unit/pers.)	7757	1744	9501

Report on the tax base and the structure of charges for the unified tax on imputed income for certain types of activities

Table 3: Fragment of STS report

Indicator name	Values of indicators: Individual entrepreneurs	Organizations	Total
The number of taxpayers submitting tax statements for the tax paid in connection with the use of the simplified tax system (units/pers.)	2	2	2

Report on the tax base and the structure of charges for the tax paid in connection with the use of the simplified tax system

n×n dimensional matrix Db is formed which elements characterize outgoing traffic flows from the traffic district and are calculated using the vector Ob and the matrix K elements according to Eq. 1:

$$Db_{ij} = Ob_i \times k_{ij}, i = \overline{1,n}, j = \overline{1,n} \quad (1)$$

Then, on the basis of the scheme of the road map in the region a distance matrix L is created which elements are the values of the distances L_{ij} between the respective transportation districts along the roads connecting them with the shortest distances. The next step is to create a preliminary n×n matrix G (a priori probability of generation of an origin-destination correspondence) (Shvetsov, 2003) with calculation of its elements according to the Eq. 2:

$$G_{ij} = \frac{Db_{ij} \times Ob_j}{L_{ij}^2}, i, j = \overline{1,n} \quad (2)$$

The use of this formula is caused by application of gravity approach to creation of preliminary origin-destination matrix. “Gravity model” is historically one of the first mathematical models proposed to estimate the inter-district origin-destination correspondences (Wilson, 1967, 1971). For the purposes of this study it is preferred because of its versatility and simplicity (Shvetsov, 2003). Upon creation of the final origin-destination matrix N elements of the matrix G should be normalized according to Eq. 3:

$$N_{ij} = \frac{G_{ij}}{\eta_j}, i = \overline{1,n}, j = \overline{1,n} \quad (3)$$

where, η_j is a normalization factor which is calculated according to Eq. 4:

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

This normalization ensures that the total number of transport means leaving the transportation district corresponds to the existing demand. It should be taken into account the dynamics of the business daily movements for the final formation of the business origin-destination matrix. Business origin-destination correspondences within the region are almost always performed within one day, i.e., a travel ends at the point of exit. Thus the trip can be considered as a travel “out and return”. Then the total number of travels between transportation districts are:

$$SN_{ij} = N_{ij} + N_{ji}, i = \overline{1,n}, j = \overline{1,n} \quad (5)$$

Using the methodology described above allows the calculation of the business origin-destination matrices to perform on the basis of publicly available data of tax reporting.

RESULTS AND DISCUSSION

Method of creating an origin-destination matrix represented in this study allows efficient obtaining the matrices reflecting the average daily movement of transportation means serving business origin-destination correspondences. During development of the technique for creating a business origin-destination matrix using open public data of tax reporting, methodological approaches to the choice of sources and methods for the preparation of statistical raw data and also methodical approaches to preparation of statistical raw data based on reports of tax authorities have been developed. The technique described 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 the building of an 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 business origin-destination matrices. Thus, the proposed technique allows in very effective and versatile way to create business origin-destination matrices on the basis of public data of the tax authorities. The technique provides a minimal dependence of a developer on human resources and enterprises databases the availability of which in different regions and countries can be very different.

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) RFMEFI58814X0001".

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.

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.

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.

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.

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.