



## AN EMPIRICAL MODELLING FRAMEWORK FOR FORECASTING FREIGHT TRANSPORTATION

Hakan Güler

*Dept of Civil Engineering, Faculty of Engineering, University of Sakarya, Turkey*

Submitted 26 March 2012; resubmitted 3 November 2012; accepted 10 January 2013

**Abstract.** This paper presents a framework which includes empirical modelling methods to estimate freight transportation between defined zones. In this method, observed origin and destination matrices for each type of freight are constituted based on the link counts and the roadside truck survey data. The gravity method is selected to estimate origin and destination matrices by using observed link flows, gross domestic product by provinces and interzonal distances. Advanced statistical techniques and regression analyses are used to estimate the coefficients of the gravity method. The final freight transportation matrix is calibrated with the link flows data by using iterative techniques. The developed method was applied to find the origin and destination matrix of the total freight transportation in Turkey and successful results were obtained.

**Keywords:** trip distribution; freight transportation; gravity method; empirical modelling; modelling freight distribution; statistical analysis.

**Reference to this paper should be made as follows:** Güler, H. 2014. An empirical modelling framework for forecasting freight transportation, *Transport* 29(2): 185–194. <http://dx.doi.org/10.3846/16484142.2014.930927>

### Notations

The following symbols are used in this paper:

$\tau$	– an overall scaling factor;	$V_a$	– observed flow at count site $a$ ;
$\theta$	– constant of proportionality;	$p_{ij}^a$	– proportion of trips from zone $i$ to zone $j$ travelling through link $a$ ;
$\alpha, \beta, \gamma$	– constants;	$at_{ij}^f$	– total highway freight transportations between $i$ and $j$ ;
$t_{ij}^o$	– initial estimate of trips made between zones $i$ and $j$ ;	$uct_{ij}^{Nf}$	– unit transportations between $i$ and $j$ ;
$p_{ij}^a$	– proportion of trips from zone $i$ to zone $j$ ;	$X'_a$	– vector of parameters after iteration;
$h_a$	– adjustment factor;	$X_a$	– vector of parameters to be estimated;
$at_{ij}^f$	– average general total transportations between $i$ and $j$ ;	$1/(d_{ij})^\beta$	– friction function;
$T_{jin}^f$	– final matrix for freight type $f$ ;	AADTT <sub><math>N</math></sub>	– Annual Average Daily Truck Traffic in total for both directions for section $N$ ;
$t_{ij}^{Nf}$	– freight transportations obtained from roadside surveys between $i$ and $j$ ;	$A_j$	– total trip attraction at $j$ ;
$t_{ij}^f$	– general total transportations between $i$ and $j$ ;	$A_j$	– total trip attractions;
$uct_{ij}^{Nf}$	– general total unit transportations between $i$ and $j$ ;	ANN	– Artificial Neural Networks;
$ct_{ij}^{Nf}$	– intermediary total freight transportations between $i$ and $j$ ;	AT	– total highway freight transportation matrix;
$t_{ij}^{Nf}$	– intermediary total freight transportations between $i$ and $j$ ;	AT <sup><math>f</math></sup>	– average general total matrices;
$at_{ij}^f$	– number of trucks carrying freight type $f$ between provinces of $i$ and $j$ ;	AVI	– Automatic Vehicle Identification;
		$C_N$	– coefficient of enlargement or diminishment for section $N$ ;
		$CT^{Nf}$	– modified intermediary total matrices;
		DFBETAs	– change in the regression coefficient that results from the deletion of the $i$ th case;
		DFFITs	– change in the predicted value when the $i$ th case is deleted;



DGTREN	– Directorate-General for Transport and Energy;
$d_{ij}$	– distance between province $i$ and province $j$ in kilometres;
EM	– Entropy Maximization;
ESA95	– The European System of national and regional Accounts;
$f$	– freight types;
GDP	– Gross Domestic Product in Turkish Liras;
$GDP_{fi}$	– sectored GDP by province for freight type $f$ and province (zone) $i$ in Turkish Liras;
$GDP_j$	– total GDP by province (zone) $j$ in Turkish Liras;
GLS	– Generalized Least Squares;
GVA	– Gross Value Added in Turkish Liras;
$i, j$	– provinces;
IM	– Information Minimization;
$k_p, A$	– coefficients;
KGM	– Republic of Turkey General Directorate of Highways;
$L$	– total number of traffic counts;
LSE	– Least Square Error;
$N$	– number of road section;
$n$	– roadside surveys;
$n_a$	– number of roadside surveys performed on a definite highway section;
NGLS	– Non Linear GLS;
O-D	– origin and destination;
$p$	– significance;
PFE	– Path Flow Estimator;
$P_i$	– total trip production at $i$ ;
$R^2$	– coefficient of determination;
$T^f$	– general total matrices;
$T_{fin}^f$	– final matrix;
$T_{fin}^f$	– final O-D matrix of total freight transportation;
$T_{ij}$	– trips produced at $i$ and attracted at $j$ ;
$T_{in}$	– initial matrix;
$T_N$	– total number of trucks and trailers in both directions for section $N$ ;
$T^{Nf}$	– intermediary total freight transportation matrices;
$T^{Nmf}$	– freight transportation matrices obtained from roadside surveys;
$UCT^f$	– general total unit matrices;
$UCT^{Nf}$	– unit matrices;
$V_a$	– traffic volume on link $a$ ;
$X_1, X_2, X_3$	– independent variables for regression analysis;
$Y$	– dependent variable for regression analysis.

## Introduction

Passenger and freight carriage is a dynamic process since its parameters frequently change in the course of time. The parameters of the carriage process are stochastic; however, their change is usually influenced by concrete conditions, the impact of which may be simulated and forecasted. There are many modelling concepts applied for estimation of freight transportation in the literature

although they are originally developed for passenger transport. De Jong *et al.* (2004b) presented a review of the literature related to freight transport models that have been developed since the 1990s for forecasting, policy simulation and project evaluation at the national and international levels. They state that there are 222 transport models in Europe. While sixty-five of those models are freight transport models, 29 of them are joint passenger and freight transport models.

Crainic and Laporte (1997) identified some of the main issues in freight transportation planning and operations. They presented appropriate Operations Research models and methods, as well as computer-based planning tools. The presentation was organized according to the three classical decision-making levels: strategic, tactic and operational. Garrido and Mahmassani (2000) proposed a multinomial probit model with spatially and temporally correlated error structure in order to carry out freight demand analysis for tactical/operational planning applications. The resulting model has a large number of alternatives, and the estimation was performed by using Monte-Carlo simulation to evaluate the multinomial probit likelihoods. De Jong *et al.* (2004a) presented a fast and approximate meta-model for passenger and freight transport in Europe on the basis of the outcomes of five disaggregate national models for passenger transport, four national models for freight transport and two European transport models. The model is EXPEDITE meta-model for passenger and freight transport. It was developed in a project for the European Commission, Directorate-General for Transport and Energy (DGTREN) (European Commission 2001). They stated that the model was a fast and relatively simple one that integrates results from a number of national and international models. Moschovou and Giannopoulos (2010) modelled freight modal choice behaviour in Greece investigated in research between 2004 and 2009. They involved a large-scale survey of various freight transport actors in Greece; a full statistical analysis of the results and a presentation to determine priorities, preferences, and detailed rankings of mode choice criteria; and a modelling exercise to produce models depicted the mode choice behaviour of Greek firms.

Transport researchers generally agree on the fact that the four-step transport modelling structure adapted from passenger transport can be successfully applied to freight transportation as well. Nevertheless, there are some important differences within each of the four steps of passenger transport. These differences include the diversity of decision-makers in freight, the diversity of the items being transported and the limited availability of data (De Jong *et al.* 2004b).

The four-step freight transport modelling system is briefly presented below, and a multi-step freight transportation planning model is demonstrated in Fig. 1 (De Jong *et al.* 2004b; Goulias 2002):

- *generation and attraction*: the amounts of goods generated by and attracted to the defined zones are determined in tonnes;

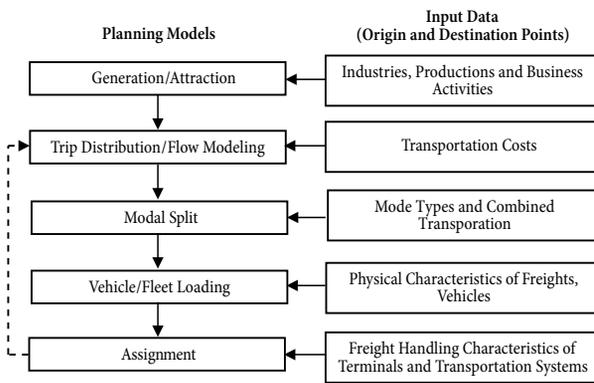


Fig. 1. Multi-step freight transportation planning model

- *distribution*: the flows of goods transported between the defined zones are designated in tonnes;
- *modal split*: the flows of goods are allocated to transportation modes which are motorways, railways, waterways and combined transportation etc.;
- *assignment*: freight flows are assigned to transportation network after converting the flows in tonnes to vehicle units.

Although there are many methods available for estimating freight transportation from traffic counts, they are all rather complicated. Here, it was intended to give practical guidance to practical engineers on how to estimate freight transportations between the provinces. For this purpose, a framework including empirical modelling methods was developed consisting a fourteen-stage progressive evaluation for a new procedure. The collected data, such as network data, road side surveys and type of freights, were formulated and the most crucial data, such as economical indicators (GDP and GVA) and friction factor (distance), were used to set up the gravity model based empirical model. Advanced statistical techniques were used to determine the coefficients of the model. Finally, the final freight transportation matrix is calibrated with the link flows data by using iterative techniques. The developed method was applied to find the origin and destination matrix of the total freight transportation in Turkey and successful results were obtained. This original method proved its performance by using less but the most important data analyzed in fourteen-stage progressive evaluation framework.

### 1. Estimation of Origin and Destination (O-D) Matrix

Transportation as a process can be described by a set of criteria, usually, speed, safety and costs of transportation are considered. All of the above criteria describe quality of transportation from a particular perspective, therefore, all of them should be used for selecting a particular traffic route. De Grange *et al.* (2010) states that trip distribution models are intended to produce the best possible predictions of travellers' destination choices on the basis of trip generation and attraction information for each travel zone and the level of interception or general-

ized cost of travelling between each pair of zones. There are a lot of O-D (Origin and Destination) estimation models in the literature. In this chapter, a brief literature survey is presented on estimation of O-D matrix.

Bell (1983) described a model which would, under certain circumstances, yield the most likely O-D matrix which was consistent with measurements of link traffic volumes. The GLS approach to the estimation of O-D matrices permits the combination of survey and traffic count data in a way that allows the relative accuracy of the two data sources. Bell (1991) also proposed an algorithm to solve the GLS problem subject to inequality constraints. Cascetta *et al.* (1993) suggested different dynamic estimators using time-varying traffic counts to obtain time-varying O-D flows or average O-D flows. The proposed two types of estimators included simultaneous estimators and sequential estimators. Cascetta and Russo (1997) examined Bayesian statistical inference techniques and they evaluated the statistical performances of the NGLS estimators on a test network and on a real urban network. They found the results in general satisfactory, showing the capability of the proposed estimator to reduce errors in initial estimates significantly. Ashok and Ben-Akiva (2000) examined two different approaches for real-time estimation/prediction of time dependent O-D flows which are a state-vector in terms of deviations in O-D flows instead of the O-D flows themselves and a state vector in terms of deviations of departure rates from each origin and the shares headed to each destination. Asakura *et al.* (2000) presented the formulation of origin and destination (O-D) matrices estimation model using the observed data with the AVI system. The results of license plate matching between a pair of AVI cameras were involved as the input variables. The formulated model was a least squares model and yielded to the linear transformation of the partly observed O-D matrices. The model was applied to the Kobe corridor line in the Han-Shin expressway network. Timms (2001) presented a philosophical structure for classifying methods that estimate O-D matrices using link counts. A classification structure is built up by using concepts of realism, subjectivity, empiricism and rationalism. Ashok and Ben-Akiva (2002) presented a new set of models which was the explicit modelling and estimation of the dynamic mapping between time-dependent O-D flows and link volumes. Celik (2004) modelled inter-regional commodity flows for 48 continental states of the US with three different ANN. Chen *et al.* (2005) examined the capability of PFE in capturing the total demand of the study network as well as individual O-D demands when proper observations, in terms of the number and their locations, were provided. Dixon and Rilett (2005) used the information from AVI systems to help estimate short-term trip O-D matrices in an urban environment. Castillo *et al.* (2008) proposed a combination (bi-level) of an O-D – pair matrix estimation model based on Bayesian networks, and a Wardrop-minimum variance model, which identifies origins and destinations of link flows. Sharma *et al.* (2011) identified three efficient solution-approximation approaches in stochastic network

design to obtain efficient and high-quality solutions by using small number of demand samples reducing the computational effort without much compromise on the solution quality. The application and the performance of these alternative approaches were reported. They concluded that the results from their study would help in deciding suitable approximation techniques for network design under demand uncertainty.

Among the commonly used calculation models, O-D matrix can be obtained by using the gravity model because it can ensure more practical results to transport researchers. Levine *et al.* (2009) developed an optimization model to estimate route flows and a corresponding multi-modal origin-destination table for containers by synthesizing data on international trade and railcar movements with a gravity model for the demand of container traffic. Veenstra *et al.* (2010) introduced a new trip distribution model for destinations that was not homogeneously distributed. The model is a gravity model, which incorporates the spatial configuration of destinations in the modelling process.

Travel time or travel distance can be used as the interceptive parameters in the gravity model. The gravity model is similar to Newton's law of gravity. The gravity model states that the trips between an origin and a destination  $T_{ij}$  depend directly on the total trip productions  $P_i$  and the total trip attractions  $A_j$  and inversely the friction factor  $d_{ij}$  called distance, travel time or cost (Rogers 2008). The formula of Gravity Model is given in Eq. 1:

$$T_{ij} = \theta \frac{P_i A_j}{(d_{ij})^\beta}, \quad (1)$$

where:  $\theta$  and  $\beta$  are proportionality and calibration constants respectively;  $i$  and  $j$  indices are origin and destination provinces.

## 2. The Empirical Modelling Method

In order to determine the observed highway freight matrices, the roadside surveys may be used. These roadside surveys contain the O-D information. Highways administrations perform the roadside surveys on different highway sections every year. The roadside surveys are carried out by stopping the trucks and filling in the specially designed questionnaire for every truck. In the questionnaire, some questions such as type of freight carried, O-D points of transportations, travelled distance and type of vehicle are asked, and the answers obtained are recorded. The freight transportation classification scheme of the Republic of Turkey General Directorate of Highways (KGM) consists of the recorded freight types which are agricultural products, ores, construction materials, animal products, manufactured materials, livestock and forest products (Vitoşoğlu 2006). In addition to the roadside surveys, the traffic count data giving the number of trucks on the highway sections are also useful information for O-D estimations. This paper presents a framework which includes empirical modelling methods to estimate freight transportation between

defined zones. The developed framework for estimation of freight transportation is illustrated in Fig. 2.

In this study, *Visual Basic* (VBA) macros in *Microsoft Excel* were written for implementation of the developed model. Several macros were written for each stage presented in Fig. 2. The produced trip matrices were transferred to the next stages step by step to obtain the final trip matrices.

There are several software packages available to help develop a four-step travel demand model. The common forecasting packages widely used on the world include *Visum*, *TransCAD*, *Emme*, *Cube*, QRS II (Quick Response System II), TMODEL (Travel Demand Model), FSUTMS (Florida Standard Urban Transportation Model Structure), TRANPLAN (Transportation Planning) and *Synchro*. These software packages are macroscopic simulation tools and they distribute the trips using either theoretically based methods, e.g. the standard gravity model, or growth factor methods, e.g. the Fratar method (Ullah *et al.* 2011). In this study, the standard gravity model was preferred but some important economical indicators were originally included in the model. The detailed explanations are given in Fig. 2.

The roadside surveys generally include the data in Table 1. The information contained in the roadside surveys performed every year on different highway sections are written into a matrix format in order to obtain the matrices for each freight type. Consequently, from each roadside interview, the matrices for each freight types [ $T^{Nnf}$ ] are obtained for the zones (Eq. 2).  $t_{ij}^{Nnf}$  corresponds to each freight types obtained from roadside surveys.  $N$ ,  $n$  and  $f$  indices are road section, road side survey and freight type respectively. The procedure is presented in Fig. 2, item 1.

$$[T^{Nnf}] = \begin{bmatrix} t_{11}^{Nnf} & \dots & t_{1j}^{Nnf} \\ \vdots & \ddots & \vdots \\ t_{i1}^{Nnf} & \dots & t_{ij}^{Nnf} \end{bmatrix}. \quad (2)$$

Normally the roadside surveys are carried out on the same highway sections in different years. Therefore, the roadside surveys performed on the same highway sections may be combined together. This operation defi-

**Table 1.** Roadside surveys' data

Type of data
1) No
2) Traffic plate
3) Type of vehicle
4) Empty weight
5) Maximum load
6) Model of vehicle
7) Age of vehicle
8) Type of freight
9) Origin point
10) Destination point
11) Travelled distance
12) Type of axle
13) Axle load

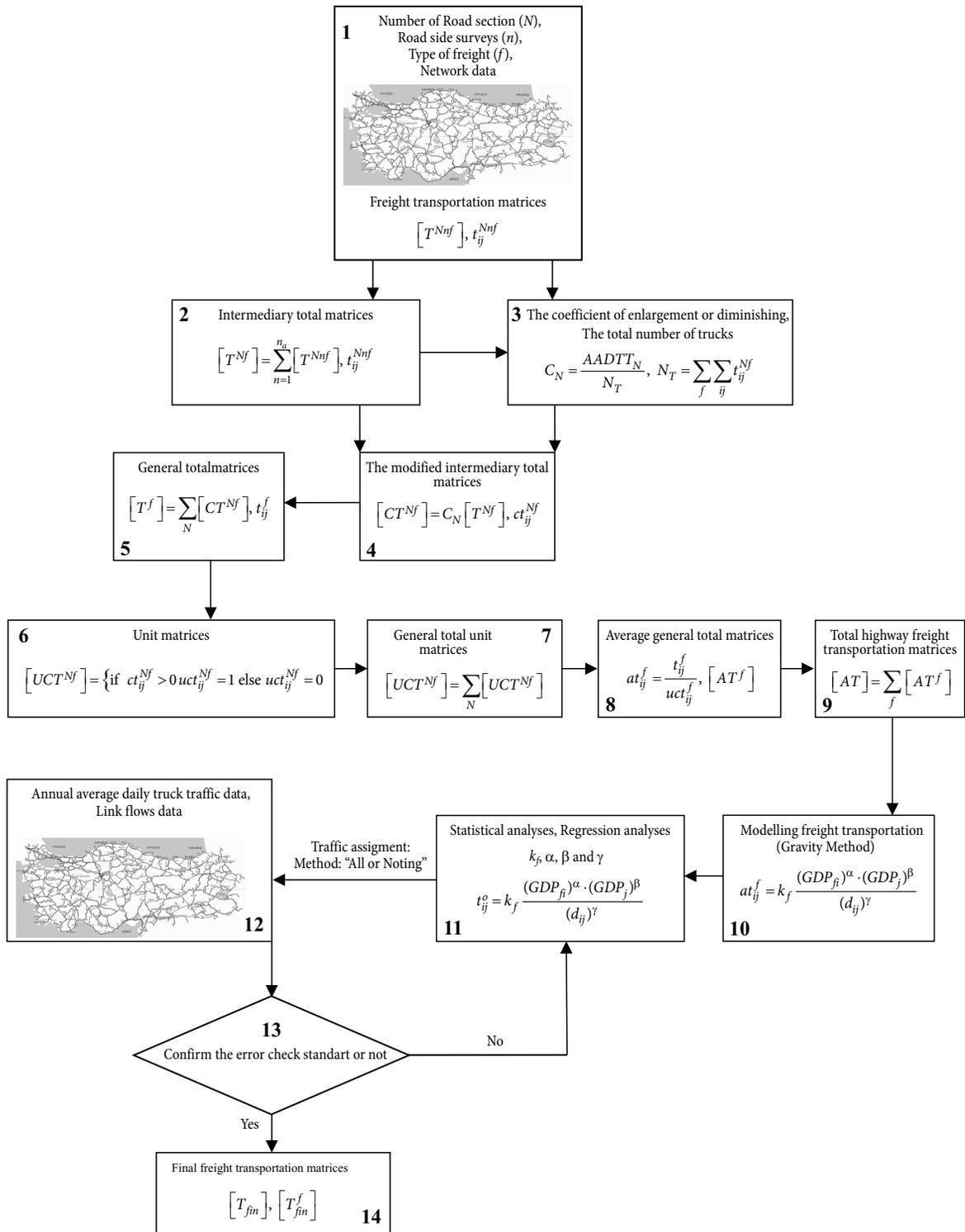


Fig. 2. A framework including empirical approach modelling methods

nately incurred the risk of significant changes in the flow during the period being obscured. Therefore, it is obviously better to use data from the same year. Whereas, it is apparent that this number is not enough to find meaningful results since the problem requires much more data. Therefore, the data from different years may be used in order to overcome this difficulty. As a result, the matrices obtained from the roadside surveys performed

on the same sections in different years are added up, and intermediary total matrices  $[T^{Nf}]$  are formed (Eq. 3).  $n_a$  corresponds number of roadside surveys performed on a definite highway section. This procedure is presented in Fig. 2, item 2.

$$[T^{Nf}] = \sum_{n=1}^{n_a} [T^{Nnf}] \tag{3}$$

These intermediary total matrices obtained for each highway section are then enlarged or diminished by multiplying a coefficient called  $C_N$ . Here, the coefficient of  $C_N$  is the ratio of the AADTT $_N$  on the highway section studied to the total number of trucks in both directions  $T_N$ . The statement mentioned above can be summarized by using Eq. 4. This procedure is presented in Fig. 2, item 3.

$$C_N = \frac{AADTT_N}{T_N}. \quad (4)$$

It is obvious that the total number of trucks and trailers surveyed in the combined roadside survey for a definite highway section  $T_N$  can be calculated by first the summation of the cells in the intermediary total freight transportation matrix  $[T^{Nf}]$  obtained for each freight type  $f$  and then the addition of these sums obtained for every freight type  $t_{ij}^{Nf}$ . Eq. 5 summarizes this statement. This procedure is presented in Fig. 2, item 3.

$$T_N = \sum_f \sum_{ij} t_{ij}^{Nf}. \quad (5)$$

The modified intermediary total matrices  $[CT^{Nf}]$  are obtained for every highway section and the freight type (Eq. 6). This procedure is presented in Fig. 2, item 4.

$$\begin{aligned} [CT^{Nf}] &= C_N [T^{Nf}]; \\ ct_{ij}^{Nf} &= C_N t_{ij}^{Nf}. \end{aligned} \quad (6)$$

After the modified intermediary total matrices belonging to the highway sections  $[CT^{Nf}]$  are obtained, the general total matrices for the different freight types  $[T^f]$  are determined by adding up the O-D matrices that are obtained for all road sections  $N$  (Eq. 7). This procedure is presented in Fig. 2, item 5.

$$\begin{aligned} [T^f] &= \sum_N [CT^{Nf}]; \\ t_{ij}^f &= \sum_N ct_{ij}^{Nf}. \end{aligned} \quad (7)$$

In the event that any O-D transport is in the list of two or more survey stations, the averages of these counts that gave the number of trucks carrying freight between this O-D pair and that are obtained for different survey stations are calculated. The purpose of this procedure is to prevent the risk that performing the general addition operation as a simple accumulation calculates the freight volumes transported between distant regions much larger than in reality. Consequently the unit matrices  $[UCT^{Nf}]$  are formed by writing '1' for filled cells and '0' for empty cells in the intermediary total matrices multiplied with the coefficients of  $C_N$  (Eq. 8). This procedure is presented in Fig. 2, item 6.

$$[UCT^{Nf}] = \begin{cases} uct_{ij}^{Nf} = 1, & \text{if } ct_{ij}^{Nf} > 0; \\ uct_{ij}^{Nf} = 0, & \text{otherwise.} \end{cases} \quad (8)$$

Then, the general total unit matrices  $[UCT^f]$  are calculated using the unit matrices as presented in Eq. 9. This procedure is presented in Fig. 2, item 7.

$$\begin{aligned} [UCT^f] &= \sum_N [UCT^{Nf}]; \\ uct_{ij}^f &= \sum_N uct_{ij}^{Nf}. \end{aligned} \quad (9)$$

Finally, the average general total matrices  $[AT^f]$  are obtained for the freight types by dividing the cells in the general total matrices  $[T^f]$  to the corresponding cells in the general total unit matrices  $[UCT^f]$  as presented in Eqs 10 and 11. In this way, in the event that any O-D transport is in the list of two or more study stations, the averages of these O-D carriages are calculated. The average general total matrices obtained for all freight types are, in conclusion, the observed matrices that give the number of trucks carrying freight between provinces. This procedure is presented in Fig. 2, item 8.

$$at_{ij}^f = \frac{t_{ij}^f}{uct_{ij}^f}; \quad (10)$$

$$[AT^f] = \begin{bmatrix} at_{11}^f & \dots & at_{1j}^f \\ \dots & \dots & \dots \\ at_{i1}^f & \dots & at_{ij}^f \end{bmatrix}. \quad (11)$$

Then, as the last step in this phase of the study, the total highway freight transportation matrix  $[AT]$  is found by adding up all the average general total matrices obtained for all freight types  $[AT^f]$  as presented in Eq. 12. This procedure is presented in Fig. 2, item 9.

$$[AT] = \sum_f [AT^f]. \quad (12)$$

The initial matrices for each type of freight are constituted based on the matrices obtained from the link counts and the roadside truck survey data. For this purpose, the principles of Gravity Method can be used by using observed link flows  $at_{ij}^f$ , GDP by provinces and interzonal distances  $d_{ij}$ . GDP by provinces named regional GVA estimates use national accounts definitions and concepts. GVA measures the contribution to the economy of each individual sector in an area. There is a link between GDP and GVA. In summary, GVA plus taxes on products less subsidies on products equals GDP.

The ESA95 collects comparable, up-to-date and reliable information on the structure and developments of the economy of the Member States of the European Union and their respective regions (source Eurostat – <http://epp.eurostat.ec.europa.eu>). The GDP series have been compiled according to ESA95 which is comprehensive and integrated set of accounts (Council Regulation (EC) No 2223/96). The procedure is summarised in Eqs 13–16. This procedure is presented in Fig. 2, item 10.

$$at_{ij}^f = k_f \frac{(\text{GDP}_{fi})^\alpha (\text{GDP}_j)^\beta}{(d_{ij})^\gamma}; \quad (13)$$

$$\log[at_{ij}^f] = \log \left[ k_f \frac{(\text{GDP}_{fi})^\alpha (\text{GDP}_j)^\beta}{(d_{ij})^\gamma} \right]; \quad (14)$$

$$\log[at_{ij}^f] = \log(k_f) + \alpha \log(\text{GDP}_{fi}) + \beta \log(\text{GDP}_j) - \gamma \log(d_{ij}); \quad (15)$$

$$Y = A + \alpha X_1 + \beta X_2 - \gamma X_3. \quad (16)$$

The beginning phase of the procedure is performing first multivariate statistic analyses. For this purpose, the first multivariate statistic analyses are carried out. The obtained regression model is tested by using some hypothesis tests. Variance analysis is performed to determine whether there is a linear relationship between dependent and independent variables. *F*-test is used for this purpose. Significances of the parameters in the analysis are tested with Student test (*t*-test). This procedure is presented in Fig. 2, item 11.

### 3. Calibration of O-D Matrix from Traffic Counts

Using conventional methods based on home surveys or roadside interviews disrupting traffic in order to estimate O-D matrices is generally expensive, time consuming, and labour intensive. The life of data is very short in developing countries, where rapid changes occur in land use and demographic structure. It is therefore necessary to revise frequently the data obtained by using relatively inexpensive methods. Various methods that are cheaper and do not require intensive labour have been developed in order to form and to revise present and future O-D matrices (Chen *et al.* 2005).

Vehicle counts on highways can be viewed as a function of a trip matrix and a route-choice pattern. Therefore, they provide information about all O-D pairs that use the counted links. In addition, traffic counts are very attractive data sources because they could be obtained in a relatively inexpensive and automatic way without disrupting traffic. As a result, since the beginning of 1980s, the idea of estimating trip matrices and developing demand models from traffic counts has attracted serious attention of researchers, and various methods have been suggested on this subject.

If it is assumed that *N* zones are interconnected by a road network, which consists of a series of nodes and links, it will be clear that the trip matrix is made up of *N*<sup>2</sup> cells. If intra-zonal trips can be disregarded, the number of cells in the trip matrix is *N*<sup>2</sup> - *N*. In order to find out these *N*<sup>2</sup> - *N* cells constituting O-D matrices from traffic counts, it is necessary first to identify paths followed by trips from each origin to each destination. If *p*<sub>ij</sub><sup>a</sup> is defined as the probability or proportion of trips from zone *i* to zone *j* travelling through link *a*, the flow in this link *V*<sub>a</sub> will be the summation of portions of all trips between zones using link *a*. Mathematically, this expression can be summarised in Eq. 17.

$$V_a = \sum_{ij} T_{ij} p_{ij}^a, \quad 0 \leq p_{ij}^a \leq 1. \quad (17)$$

The probability of *p*<sub>ij</sub><sup>a</sup> can be obtained by using various trip assignment techniques of which their degree of complexity increases from an all-or-nothing assignment to an equilibrium assignment. As a result, when all the *p*<sub>ij</sub><sup>a</sup> proportions and all the observed traffic counts *V*<sub>a</sub>

are given, there will be *N*<sup>2</sup> - *N* unknown *T*<sub>ij</sub>'s of the problem to be estimated from a set of *L* simultaneous linear equations, where *L* is the total number of traffic counts.

In principle, *N*<sup>2</sup> - *N* independent and consistent traffic counts are necessary for determining uniquely the trip matrix *T*. On the other hand, in practice, the number of traffic counts is much less than the number of unknown *T*<sub>ij</sub>'s. Therefore, it is impossible to find out a unique solution for the problem of estimating an O-D matrix. In general, more than one trip matrix that are consistent with the observed traffic counts will be found when they are assigned onto the network. Two basic approaches can be utilised in order to resolve this problem. In the first approach, the set of feasible solutions for the matrix to be estimated can be restricted by imposing a particular structure, which is provided by gravity or a direct demand model. In the second approach, general principles like maximum likelihood or entropy maximisation are utilised in order to provide the minimum additional information required for estimating an O-D matrix.

Assignment methods used for estimating trip matrix from traffic counts are classified under two main groups. In the assignment methods belonging to the first group, it is assumed that the proportion of drivers choosing each route does not depend on flow levels in links. The most common example of assignment methods in this group is all-or-nothing assignment, and the probabilities of *p*<sub>ij</sub><sup>a</sup> are defined in this case as follows:

$$p_{ij}^a = \begin{cases} 1, & \text{if trips from origin } i \text{ to destination } j \text{ use link } a; \\ 0, & \text{otherwise.} \end{cases}$$

Pure stochastic methods are also included in the first group. However, in these cases, the probabilities of *p*<sub>ij</sub><sup>a</sup> can take values between 0 and 1.

Assignment methods in the second group, on the other hand, take account of congestion effects. Therefore, the probability of trips made between each O-D pair using any link also depends on traffic flow in that link. Equilibrium and stochastic user equilibrium assignment methods are included in this group (Ortúzar, Willumsen 2011).

Generally, the IM model (Snickars, Weibull 1977; Van Zuylen, Willumsen 1980), GLS model (Cascetta 1984), the EM model (Wilson 1970), the Bayesian model (Mahmassani, Sinha 1981), the LSE model (Cochran 1963) and the PFE (Bell, Shield 1996) are used for O-D matrix estimation.

Al-Deek and Eman (2006) states that the reliability of travel time measure can be used as a tool to provide travellers with accurate information about the most reliable paths connecting origins and destinations. They developed and applied a new methodology to estimate the travel time reliability of a transportation network and its paths during the peak period in which links can degrade in a multimode, statistically dependent manner. Each model has its characteristics and application conditions based on the applied theory. The model developed by Bell is fundamentally similar to the princi-

ple of the modified IM model. The IM model based on the information minimization theory was developed by Van Zuylen and Willumsen (1980). The most likely O-D flows can be estimated through an iterative process with this model until all flow constraints are satisfied, i.e. estimated traffic flows are equal to the respective detected traffic flows. If the route choice proportions are not completely known and partially duplicated, the IM model is not suitable and the respective estimation results are not stable. For that reason, Van Zuylen and Willumsen modified the IM model to adjust the difference between the total number of historical trips and the actual trips. The form of the modified IM model used by Bell is shown in Eq. 18 (Van Zuylen, Willumsen 1980; Wang, Friedrich 2009).

$$V_a = \sum_{ij} t_{ij}^o \tau \left( \prod_a X_a^{p_{ij}^a} \right) p_{ij}^a. \quad (18)$$

These two parameters,  $\tau$  and  $X_a$ , and the most likely O-D matrix can then be solved with the satisfaction of the link flow constraints in an iterative process. The vector of parameters  $X_a$  is initially set to unity, and unless other values for  $\tau$  have been defined by the user, the value of  $\tau$  is calculated by using Eq. 19. The value of  $\tau$  remains set at the level defined or calculated by the formula above during later phases. As for the solution procedure, it involves improving upon the initial estimated values of  $X_a$  by carrying out many iterations.

$$\tau = \frac{\sum_a V_a}{\sum_a \sum_{ij} p_{ij}^a}. \quad (19)$$

For every count site, an adjustment factor  $h_a$  is calculated in each iteration. This adjustment factor is then added to the prior estimate of  $X_a$  in order to obtain the value of  $X'_a$  as shown in Eq. 20. Eq. 21 is used to calculate the value of  $h_a$ .

$$X'_a = X_a + h_a; \quad (20)$$

$$h_a = \frac{V_a - \sum_{ij} t_{ij}^o \tau \left( \prod_a X_a^{p_{ij}^a} \right) p_{ij}^a}{\sum_{ij} t_{ij}^o \tau p_{ij}^{a2} \left( \prod_{b \neq a} X_b^{p_{ij}^b} \right) X_a^{(p_{ij}^a - 1)}}. \quad (21)$$

The process of iteratively calculating the values of  $h_a$  for each count site continues until the difference between observed and estimated values is close to a minimum value defined by the user. After the final values of  $X_a$  are determined for every link, the cells of trip matrix are calculated by Eq. 22. Finally, the trip matrix has been formed by determining all its  $T_{ij}$  elements (Halcrow Fox & Associates 1986). This procedure is presented in Fig. 2, items 12 and 13.

$$T_{ij} = \tau t_{ij}^o \prod_a (X_a)^{p_{ij}^a}. \quad (22)$$

#### 4. Implementation of the Model

The suggested model was tested to find the O-D matrix, only for the total freight transportation, between 81 provinces in Turkey. The roadside surveys carried out on 63 different highway sections by KGM were used (Republic of Turkey General Directorate of Highways 2010). These roadside surveys carried out between 2008 and 2010 years contained the O-D information of 58935 trucks. The highway sections on which the roadside surveys have been performed are presented in Fig. 3. The total freight transportation matrix was obtained by the way as already explained in Fig. 2 and items 1 to 9. Gravity method was selected to model freight transportation between the defined provinces. For this purpose, first multivariate statistic analysis was carried out and the results are presented in Table 2. The coefficient of determination was found as 0.628.

After first multivariate statistic analysis, case statistics and case analyses were performed to determine outlier and influential points. Standardized values, Stu-

Table 2. Multivariable statistical analysis results

Determination of coefficient	Independent variables	Coefficients	p
First analysis $R^2 = 0.628$	$GDP_i$	0.539	0.00
	$GDP_j$	0.521	0.00
	$d_{ij}$	1.011	0.00
	$k$	5.654	0.00
Second analysis $R^2 = 0.712$	$GDP_i$	0.525	0.00
	$GDP_j$	0.502	0.00
	$d_{ij}$	1.002	0.00
	$k$	5.531	0.00

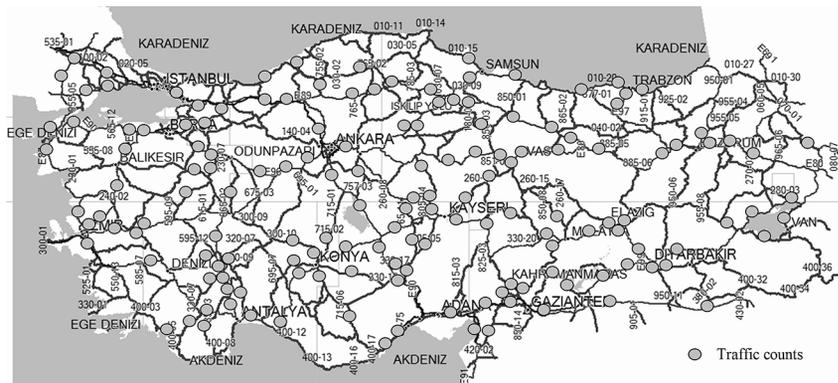


Fig. 3. The highway sections on which the roadside surveys have been performed

dentized values and Mahalanobis distance were used to determine outliers at first. Mostly, observations were accepted to scatter between  $-2$  and  $+2$  for Standardized and Studentized values. The observations above  $+2$  and below  $-2$  were accepted as high leverage points (Montgomery *et al.* 2012). Studentized values were plotted with estimation values. After high leverage points were determined, these points were tested whether they were influential on some variables. The DFBETAs (Change in the regression coefficient that results from the deletion of the  $i$ th case), Cook distance, DFFITs (Change in the predicted value when the  $i$ th case is deleted) and covariance ratios tests showed that there were no influential points on the parameters. The outlier points were excluded from the analyses as suggested and supported by Montgomery *et al.* (2012), and Draper and Smith (1998). Then, the second multivariate statistic analyses were performed for all freight types.  $F$ -test results were found lower than the  $p$ -value. Significances of the parameters in the analysis were tested with Student test ( $t$ -test) and the confidence level was found to be high because the  $p$ -value is lower than 0.05. The results are summarised in Table 2. The coefficient of determination was found as 0.712.

Finally, from second multivariate statistic analyses,  $k_f$ ,  $\alpha$ ,  $\beta$ , and  $\gamma$  calibration coefficients were found. These coefficients were placed into gravity type equations to determine the initial matrix  $[T_{in}]$ .

In order to determine the final intercity freight transportation matrix for the year of 2010, the gravity based freight transportation matrix considered as initial matrix was adjusted with the link flows data by using iterative techniques as already explained in Chapter 3. *Vmat Subprogram*, which is a part of TRANSPORT software, was used to determine the final matrix (Coombe 1989). The main purpose of this program is to identify the most likely trip matrix from input files of traffic counts at individual sites or groups of sites and from route probabilities, using Newton-Raphson Technique (Halcrow Fox & Associates 1986). Consequently, the final O-D matrix of the total freight transportation  $[T_{fin}]$  was obtained for 81 provinces.

## Conclusions

Although many methods are used for obtaining O-D matrices, the suggested method using basic traffic information, such as network data, road side surveys and type of freights, and economical indicators, such as GDP and GVA, is not expensive, time consuming, and labour intensive. This suggested method is mainly an O-D estimation method for road freight transport based on road counts. The O-D matrices may be easily updated in case of any planning phase by using relevant traffic information. In addition, the coefficients of the gravity model including economical indicators and distance may be easily calculated by using advanced statistical techniques and regression analyses. The suggested empirical modelling method proved its performance on how to estimate freight transportations between the provinces

in a fourteen-stage progressive evaluation framework. This framework enables practical engineers to understand the way they work, to evaluate the efficiency of the existing methods, and gives them particular data they can use to get better. In this study, the developed model using empirical modelling methods are suggested to determine the O-D matrices for a wide diversity of freight types. Here, the coefficient of determination for the total freight transportation was found in acceptable level ( $R^2 = 0.712$ ). The suggested method proved to be useful to determine the O-D matrices for a wide diversity of freight types in case of having available road survey data over time. This method has also a potential for expansion to estimating passenger O-D flows for road transport as well as freight and passenger O-D flows for other modes of transport such as railways and airways. Consequently, freight transportation models which are fast and easy to use on the base of relevant data can be developed. In addition, these types of models can include different kinds of freights for intermodal freight transportation and logistics activities.

## References

- Al-Deek, H. M. P. E.; Emam, E. B. 2006. Computing travel time reliability in transportation networks with multistates and dependent link failures, *Journal of Computing in Civil Engineering* 20(5): 317–327.  
[http://dx.doi.org/10.1061/\(ASCE\)0887-3801\(2006\)20:5\(317\)](http://dx.doi.org/10.1061/(ASCE)0887-3801(2006)20:5(317))
- Asakura, Y.; Hato, E.; Kashiwadani, M. 2000. Origin-destination matrices estimation model using automatic vehicle identification data and its application to the Han-Shin expressway network, *Transportation* 27(4): 419–438.  
<http://dx.doi.org/10.1023/A:1005239823771>
- Ashok, K.; Ben-Akiva, M. E. 2002. Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows, *Transportation Science* 36(2): 184–198.  
<http://dx.doi.org/10.1287/trsc.36.2.184.563>
- Ashok, K.; Ben-Akiva, M. E. 2000. Alternative approaches for real-time estimation and prediction of time-dependent origin-destination flows, *Transportation Science* 34(1): 21–36.  
<http://dx.doi.org/10.1287/trsc.34.1.21.12282>
- Bell, M. G. H. 1991. The estimation of origin-destination matrices by constrained generalised least squares, *Transportation Research Part B: Methodological* 25(1): 13–22.  
[http://dx.doi.org/10.1016/0191-2615\(91\)90010-G](http://dx.doi.org/10.1016/0191-2615(91)90010-G)
- Bell, M. G. H. 1983. The estimation of an origin-destination matrix from traffic counts, *Transportation Science* 17(2): 198–217. <http://dx.doi.org/10.1287/trsc.17.2.198>
- Bell, M. G. H.; Shield, C. M. 1996. A log-linear model for path flow estimation, in *Applications of Advanced Technologies in Transportation Engineering: Proceedings of the 4th International Conference*, June 27–30, 1995, Capri, Italy, 695–699.
- Cascetta, E. 1984. Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator, *Transportation Research Part B: Methodological* 18(4–5): 289–299. [http://dx.doi.org/10.1016/0191-2615\(84\)90012-2](http://dx.doi.org/10.1016/0191-2615(84)90012-2)
- Cascetta, E.; Inaudi, D.; Marquis, G. 1993. Dynamic estimators of origin-destination matrices using traffic counts, *Transportation Science* 27(4): 363–373.  
<http://dx.doi.org/10.1287/trsc.27.4.363>

- Cascetta, E.; Russo, F. 1997. Calibrating aggregate travel demand models with traffic counts: estimators and statistical performance, *Transportation* 24(3): 271–293. <http://dx.doi.org/10.1023/A:1004968411792>
- Castillo, E.; Menéndez, M. A.; Sánchez-Cambronero, S. 2008. Traffic estimation and optimal counting location without path enumeration using Bayesian networks, *Computer-Aided Civil and Infrastructure Engineering* 23(3): 189–207. <http://dx.doi.org/10.1111/j.1467-8667.2008.00526.x>
- Celik, H. M. 2004. Modeling freight distribution using artificial neural networks, *Journal of Transport Geography* 12(2): 141–148. <http://dx.doi.org/10.1016/j.jtrangeo.2003.12.003>
- Chen, A.; Chootinan, P.; Recker, W. W. 2005. Examining the quality of synthetic origin–destination trip table estimated by path flow estimator, *Journal of Transportation Engineering* 131(7): 506–513. [http://dx.doi.org/10.1061/\(ASCE\)0733-947X\(2005\)131:7\(506\)](http://dx.doi.org/10.1061/(ASCE)0733-947X(2005)131:7(506))
- Cochran, W. G. 1963. *Sampling Techniques*. 2nd edition. John Wiley & Sons Ltd. 413 p.
- Coombe, R. D. 1989. Review of computer software for traffic engineers, *Transport Reviews* 9(3): 217–234. <http://dx.doi.org/10.1080/01441648908716725>
- Council Regulation (EC) No 2223/96 of 25 June 1996 on the European system of national and regional accounts in the Community. 1996. 469 p. Available from Internet: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31996R2223&from=EN>
- Crainic, T. G.; Laporte, G. 1997. Planning models for freight transportation, *European Journal of Operational Research* 97(3): 409–438. [http://dx.doi.org/10.1016/S0377-2217\(96\)00298-6](http://dx.doi.org/10.1016/S0377-2217(96)00298-6)
- De Grange, L.; Fernández, E.; De Cea, J. 2010. A consolidated model of trip distribution, *Transportation Research Part E: Logistics and Transportation Review* 46(1): 61–75. <http://dx.doi.org/10.1016/j.tre.2009.06.001>
- De Jong, G.; Gunn, H.; Ben-Akiva, M. 2004a. A meta-model for passenger and freight transport in Europe, *Transport Policy* 11(4): 329–344. <http://dx.doi.org/10.1016/j.tranpol.2004.03.001>
- De Jong, G.; Gunn, H.; Walker, W. 2004b. National and international freight transport models: an overview and ideas for future development, *Transport Reviews* 24(1): 103–124. <http://dx.doi.org/10.1080/0144164032000080494>
- Dixon, M. P.; Rilett, L. R. 2005. Population origin–destination estimation using automatic vehicle identification and volume data, *Journal of Transportation Engineering* 131(2): 75–82. [http://dx.doi.org/10.1061/\(ASCE\)0733-947X\(2005\)131:2\(75\)](http://dx.doi.org/10.1061/(ASCE)0733-947X(2005)131:2(75))
- Draper, N. R.; Smith, H. 1998. *Applied Regression Analysis*. 3rd edition. Wiley-Interscience. 736 p.
- European Commission. 2001. *White Paper: European Transport Policy for 2010: Time to Decide*. 119 p. Available from Internet: [http://ec.europa.eu/transport/themes/strategies/doc/2001\\_white\\_paper/lb\\_texte\\_complet\\_en.pdf](http://ec.europa.eu/transport/themes/strategies/doc/2001_white_paper/lb_texte_complet_en.pdf)
- Garrido, R. A.; Mahmassani, H. S. 2000. Forecasting freight transportation demand with the space–time multinomial probit model, *Transportation Research Part B: Methodological* 34(5): 403–418. [http://dx.doi.org/10.1016/S0191-2615\(99\)00032-6](http://dx.doi.org/10.1016/S0191-2615(99)00032-6)
- Goulias, K. G. 2002. *Transportation Systems Planning: Methods and Applications*. CRC Press. 448 p.
- Halcrow Fox & Associates. 1986. *TRANSPORT User Guide: Micro Computer Transportation Planning Package*. London, England.
- Levine, B.; Nozick, L.; Jones, D. 2009. Estimating an origin–destination table for US imports of waterborne containerized freight, *Transportation Research Part E: Logistics and Transportation Review* 45(4): 611–626. <http://dx.doi.org/10.1016/j.tre.2008.11.001>
- Mahmassani, H. S.; Sinha, K. 1981. A Bayesian updating of trip generation parameters, *Journal of Transportation Engineering* 107(5): 581–589.
- Montgomery, D. C.; Peck, E. A.; Vining, G. G. 2012. *Introduction to Linear Regression Analysis*. 5th edition. Wiley. 672 p.
- Moschovou, T. P.; Giannopoulos, G. A. 2010. Investigation of inland freight transport modal choice in Greece: parameters and attributes influencing freight mode choice, *Transportation Research Record* 2168: 43–52. <http://dx.doi.org/10.3141/2168-06>
- Ortúzar, J. de D.; Willumsen, L. G. 2011. *Modelling Transport*. 4th edition. Wiley. 606 p.
- Republic of Turkey General Directorate of Highways. 2010. *Traffic and Transportation Statistics*. Available from Internet: <http://www.kgm.gov.tr>
- Rogers, M. 2008. *Highway Engineering*. 2nd edition. Wiley-Blackwell. 352 p.
- Sharma, S.; Mathew, T. V.; Ukkusuri, S. V. 2011. Approximation Techniques for Transportation Network Design Problem under Demand Uncertainty, *Journal of Computing in Civil Engineering* 25(4): 316–329. [http://dx.doi.org/10.1061/\(ASCE\)CP.1943-5487.0000091](http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000091)
- Snickars, F.; Weibull, J. W. 1977. A minimum information principle: theory and practice, *Regional Science and Urban Economics* 7(1–2): 137–168. [http://dx.doi.org/10.1016/0166-0462\(77\)90021-7](http://dx.doi.org/10.1016/0166-0462(77)90021-7)
- Timms, P. 2001. A philosophical context for methods to estimate origin–destination trip matrices using link counts, *Transport Reviews* 21(3): 269–301. <http://dx.doi.org/10.1080/713868144>
- Ullah, M. S.; Molakatalla, U.; Morcoima-Black, R.; Mohideen, A. Z. 2011. *Travel Demand Modeling for the Small and Medium Sized MPOs in Illinois*. Research Report ICT-11-091, Illinois Center for Transportation, Illinois, USA. 123 p. Available from Internet: <http://ict.illinois.edu/publications/report%20files/FHWA-ICT-11-091.pdf>
- Van Zuylen, H. J.; Willumsen, L. G. 1980. The most likely trip matrix estimated from traffic counts, *Transportation Research Part B: Methodological* 14(3): 281–293. [http://dx.doi.org/10.1016/0191-2615\(80\)90008-9](http://dx.doi.org/10.1016/0191-2615(80)90008-9)
- Veenstra, S. A.; Thomas, T.; Tutert, S. I. A. 2010. Trip distribution for limited destinations: a case study for grocery shopping trips in the Netherlands, *Transportation* 37(4): 663–676. <http://dx.doi.org/10.1007/s11116-010-9274-0>
- Vitoşoğlu, Y. 2006. *Türkiye’de demiryolu ağırlıklı kombine yük taşımacılığı olanaklarının araştırılması*. İstanbul Teknik Üniversitesi. 277 y. (in Turkish).
- Wang, Y.-P.; Friedrich, B. 2009. Improving matrix estimation pertaining to detailed traffic information and sophisticated traffic state, in *Transportation Research Board 88th Annual Meeting Compendium of Papers DVD*, 11–15 January, 2009, Washington D.C. 15 p. (DVD).
- Wilson, A. G. 1970. Inter-regional commodity flows: entropy maximizing approaches, *Geographical Analysis* 2(3): 255–282. <http://dx.doi.org/10.1111/j.1538-4632.1970.tb00859.x>