

Statistical models for the estimation of the origin-destination matrix from traffic counts

Anselmo Ramalho Pitombeira Neto¹, Francisco Moraes de Oliveira Neto²,
Carlos Felipe Grangeiro Loureiro³✉

¹Departamento de Engenharia de Produção, UFC, anselmo.pitombeira@ufc.br

²Departamento de Engenharia de Transportes, UFC, moraes@det.ufc.br

³Departamento de Engenharia de Transportes, UFC, felipe@det.ufc.br

Recebido:

16 de março de 2017

Aceito para publicação:

28 de julho de 2017

Publicado:

30 de dezembro de 2017

Editor de área:

Cira Souza Pitombo

Keywords:

Origin-destination matrix,
Transportation demand,
Statistical models.

Palavras-chaves:

Matriz origem-destino,
Demanda por transportes,
Modelos estatísticos.

DOI:10.14295/transportes.v25i4.1344

ABSTRACT

In transportation planning, one of the first steps is to estimate the travel demand. The final product of the estimation process is an origin-destination (OD) matrix, whose entries correspond to the number of trips between pairs of origin-destination zones in a study region. In this paper, we review the main statistical models proposed in the literature for the estimation of the OD matrix based on traffic counts. Unlike reconstruction models, statistical models do not aim at estimating the exact OD matrix corresponding to observed traffic volumes, but they rather aim at estimating the parameters of a statistical model of the population of OD matrices. Initially we define the estimation problem, emphasizing its underspecified nature, which has led to the development of several models based on different approaches. We describe static models whose parameters are estimated by means of maximum likelihood, the method of moments, and Bayesian inference. We also describe some recent dynamic models. Following that, we discuss research questions related to the underspecification problem, model assumptions and the estimation of the route choice matrix, and indicate promising research directions.

RESUMO

No planejamento dos transportes, um dos primeiros passos é estimar a demanda por viagens. O produto final do processo de estimação é uma matriz origem-destino (OD), cujas entradas correspondem ao número de viagens entre pares de zonas de origem-destino em uma região de estudo. Neste artigo, revisamos os principais modelos estatísticos propostos na literatura para a estimação da matriz OD com base em contagens de tráfego. Ao contrário dos modelos de reconstrução, os modelos estatísticos não visam a estimar a matriz OD exata correspondente aos volumes de tráfego observados, mas sim a estimar os parâmetros de um modelo estatístico da população de matrizes OD. Inicialmente, define-se o problema da estimação, enfatizando sua natureza subespecificada, o que levou ao desenvolvimento de vários modelos baseados em diferentes abordagens. Descrevem-se modelos estáticos cujos parâmetros são estimados por meio da máxima verossimilhança, do método dos momentos e da inferência bayesiana. Descrevem-se também alguns modelos dinâmicos recentes. Em seguida, discutem-se questões de pesquisa relacionadas ao problema da subespecificação, às premissas adotadas nos modelos e à estimativa da matriz de escolha de rotas, e indicam-se direções de pesquisa promissoras.



1. INTRODUCTION

In transportation planning, one of the first steps is to estimate travel demand. Generally, demand is measured in terms of trip flows between zones in a geographic region. The final product of the estimation process is a so-called *origin-destination matrix* (OD matrix, for short), whose entries correspond to the number of trips between pairs of zones in a reference time period.

Traditionally, OD matrices have been estimated through direct methods, such as home-based surveys,

road-side interviews and license plate automatic recognition. These methods collect sample data on the number of trips performed daily, their origins and their destinations. Such data can be compiled and several statistics may be computed, such as the mean, standard deviation and confidence intervals. However, these direct methods require large samples to achieve a target statistical significance, which may be technically or economically infeasible (Cascetta, 2009).

Another way of estimating OD matrices is by using trip generation and distribution models. In this approach, social and economic data are used to estimate the number of trips produced and attracted by each zone. A gravity-type model is applied in order to distribute the generated trips between zones (Ortúzar and Willumsen, 2011). These models are in general applied in long term planning studies, which require the prediction of OD matrices for nonexistent or developing geographic regions. This approach involves the acquisition of large amounts of data through surveys, with high accompanying costs, which precludes its use in short term applications, such as traffic management systems and public transit operation.

In the 1970s, researchers started developing alternative mathematical models whose objective was to obtain an OD matrix from indirect data on trip patterns (Robillard, 1975). The main sources of indirect data were traffic volumes observed on links of the transportation network (also called *traffic counts*). The development of traffic monitoring systems opened up the possibility of acquiring data on traffic volumes in an automated way at low costs. In road networks this acquisition takes place by means of sensors installed on the roads, and in transit networks data on traffic of passengers can be acquired by means of electronic ticketing.

The rationale of these alternative models is to estimate OD flows through a mathematical model which relates traffic volumes on links of the transportation network to OD flows between zones. The OD matrix so obtained is called a *synthetic* OD matrix, since it is not estimated by direct observation of trips (e.g., by directly sampling OD trips), but as an output of a model which uses indirect data on travel demand. The usefulness of such a model is evident: the transportation demand patterns in a part or in a whole region may be, in theory, traced to a finer time scale of days or hours, or even in real time. This is a great improvement over household surveys, which are typically carried out once in a decade, a time period during which the demand pattern may have changed considerably.

In the literature, a distinction is made between *reconstruction* and *estimation*, as pointed out by Spiess (1987), Lo *et al.* (1996), Hazelton (2001), Timms (2001) and Carvalho (2014). We define as reconstruction of the OD matrix the attempt to recover the "exact" OD matrix which produced an observed vector of link volumes in a given time period. On the other hand, we refer to estimation of the OD matrix when we intend to estimate the parameters of a statistical model which describes a "population" of OD matrices. Statistical models assume that OD matrices are random variables, with a particular realization in a given time period, from which we can compute mean OD flows, variances and higher moments which may be valuable.

Oliveira Neto *et al.* (2016) review the main reconstruction models proposed in the literature. They treated the phenomenon that generates the OD flows as either probabilistic or deterministic. In the case of probabilistic OD flows, they assume that the phenomenon is stationary, that is, with constant parameters over time. They review the main approaches based on generalized least squares, maximum entropy and Bayesian inference, and discuss the main characteristics and drawbacks of such methods. The authors also propose research questions related to the underspecification problem, quality of prior information, validity of probability models, application to real networks, equilibrium in congested networks, and the impact of route choice modeling.

Aiming to expand the review of Oliveira Neto *et al.* (2016), this paper focus on statistical models for estimation of the OD matrix. Unlike reconstruction models, statistical models treat OD flows as random variables, and the main interest lies on the estimation of parameters of the model, such as mean OD flows, rather than reconstructing a realized OD matrix. Considering the availability of traffic data from real-time monitoring systems, it is believed that a large spatio-temporal dataset of traffic volumes might

be suitable for estimating mean OD flows in large networks. The main objectives of this paper are to describe the main features and limitations of the models; to discuss issues and research questions related to the problem and to the models; and to suggest future research directions.

The remainder of this paper is divided up into the following sections: in section 2 we describe the problem of estimation of the OD matrix based on traffic counts; in sections 3 we describe the main statistical models; in section 4 we discuss issues and research questions; finally, we draw concluding remarks and point to future research in section 5.

2. THE OD MATRIX ESTIMATION PROBLEM

Let $(\mathcal{N}, \mathcal{A})$ be a transportation network, in which \mathcal{N} is a set of *nodes* and \mathcal{A} is a set of *directed links*. Typically for road networks the links and nodes correspond to road segments and intersections between road segments respectively. The transportation network connects *zones* of a certain geographic region (e.g., a city), which "produce and attract" trips, so that there is also a set of zones, denoted by \mathcal{I} . A *trip* is a movement of a user (person, freight, or vehicle) between an *origin zone* and a *destination zone* (referred to simply as an *OD pair*). All trips enter and exit the network through *centroid nodes*, which are terminal nodes located at zone centroids. Intrazonal trips are not taken into account, since their origin and destination centroid nodes coincide.

We denote by x_i the total flow of trips in an OD pair $i \in \mathcal{I}$ for a given time period. In applications, a time period may be, e.g., the morning peak hour or a whole business day. x_i is regarded as a random variable with a known density function probability mass function (if OD flows are treated as discrete variables). What is traditionally meant as an *OD matrix* is a two-dimensional array whose row indices identify origin zones, column indices identify destination zones, and the entries are the number of trips in an OD pair. As a matter of analytical convenience, in our notation the OD matrix is stretched out as a vector $\mathbf{x} \in \mathbb{R}_+^n$ (respectively \mathbb{Z}_+^n in the discrete case) for which $n = |\mathcal{I}|$ is the number of OD pairs.

For a given OD pair i , there is a set of *routes* connecting its origin and destination zones. A *route* is a simple path between a pair of nodes. Let \mathcal{K}_i be a set of routes associated with OD pair i . For a route $k \in \mathcal{K}_i$, we define y_{ik} as the flow of trips through route k in a reference time period. Let also $\mathbf{y} \in \mathbb{R}_+^r$ be the vector whose components are y_{ik} , and $r = \sum_{i \in \mathcal{I}} |\mathcal{K}_i|$ is the number of routes over all OD pairs. In general, only a reduced set of routes is considered in a model, since the full set of routes may be prohibitively large in real scale networks. (See Bekhor *et. al.*, 2006, for an evaluation of route choice set generation algorithms). We notice that y_{ik} are also random variables, whose relationship with x_i is given by the following equation:

$$y_{ik} = p_{ik} x_i \quad \forall k \in \mathcal{K}_i, \forall i \in \mathcal{I} \tag{1}$$

where p_{ik} for $k \in \mathcal{K}_i$ and $i \in \mathcal{I}$ gives the expected proportion (i.e., the probability) of users in OD pair i that chooses to follow route k . Notice that equation (1) implies that $\sum_{k \in \mathcal{K}_i} y_{ik} = x_i$ for all $i \in \mathcal{I}$.

The total count of trips which flow through a link $a \in \mathcal{A}$ for a given time period is denoted as the *traffic volume* z_a . Traffic volumes on all links are represented by a vector $\mathbf{z} \in \mathbb{R}_+^m$, and $m = |\mathcal{A}|$ is the number of links in the network. Flows on routes and traffic volumes are related by the following flow conservation equation:

$$z_a = \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}_i} \delta_{ak} y_{ik} \quad \forall a \in \mathcal{A} \tag{2}$$

In equation (2), δ_{ak} takes the value 1 if link a is part of route k and 0 otherwise. It synthesizes information on the topology of the network. Conveniently we will refer to equation (2) in matrix

notation, in which $\Delta = [\delta_{ak}]_{m \times r}$ is called the *link-path incidence matrix*. Notice that it is possible to establish a direct relationship between OD flows and traffic volumes through the route choice probabilities. Let $\mathbf{P} = [p_{ik}]_{r \times n}$ be a *route choice matrix*, from equations (1) and (2), in matrix notation:

$$\mathbf{z} = \Delta \mathbf{P} \mathbf{x} \tag{3}$$

The route choice matrix \mathbf{P} is commonly specified by means of a route choice model, which estimates the probability that a user chooses a route as a function of the travel time (or generalized cost) associated with the route. Deterministic route choice models assume that the users have perfect knowledge of costs and always choose the route with minimum cost. In contrast, stochastic models assume that *perceived* costs of the users are different from actual costs, so that they may choose routes which do not have minimum costs. The probabilities are estimated by means of discrete choice models (Ben-Akiva; Lerman, 1985).

The term $\mathbf{F} = \Delta \mathbf{P}$ is called the *assignment matrix*, through which the relationship between \mathbf{x} and \mathbf{z} may be directly expressed by:

$$\mathbf{z} = \mathbf{F} \mathbf{x} \tag{4}$$

After presenting the basic mathematical relationships among variables, we can finally state the estimation problem. Let $f(\mathbf{x}; \boldsymbol{\theta})$ be a probability distribution for the random OD matrices \mathbf{x} , where $\boldsymbol{\theta}$ is a vector of parameters, and $g(\mathbf{z}; \boldsymbol{\theta})$ the probability distribution of traffic volumes \mathbf{z} . Given a sample of observed traffic volumes $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(N)}$ taken on different days during the same reference time period (for example, the peak hour for different days) the problem is how to estimate parameters $\boldsymbol{\theta}$ (Hazelton, 2000). In the following section, we describe the main models proposed for solving this estimation problem according to different estimation principles from classic and Bayesian statistics.

3. STATISTICAL MODELS FOR OD MATRIX ESTIMATION

In the following subsections, we present statistical models whose parameters are estimated by maximum likelihood (section 3.1), the method of moments (section 3.2) and Bayesian inference (section 3.3). The models in these sections assume that the distribution probability of the OD matrices does not vary in time, and are thus termed *static* models. In section 3.4 we describe dynamic models, which allow the probability distribution of OD matrices to vary over time.

3.1 Models based on maximum likelihood

Estimation by maximum likelihood is performed through maximization of the likelihood function. Some authors (Spiess, 1987; Vardi, 1996; Tebaldi and West, 1998; Hazelton, 2000) assume that OD flows follow independent Poisson distributions with parameters $\theta_i = E[x_i]$. We denote by $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)^T$ with $\boldsymbol{\theta} \in \mathbb{R}^n$ the vector of mean OD flows for all OD pairs. It can be shown (Hazelton, 2000) that route flows $y_{ik}, \forall k \in \mathcal{K}_i, i \in \mathcal{I}$ are also independent random variables which follow Poisson distributions with expected value $E[y_{ik}] = p_{ik} \theta_i$, in which p_{ik} is the probability that a trip occurs in route $k \in \mathcal{K}_i$.

Moreover, as traffic volumes on links are sums of route flows according to the flow conservation equation (2), they also marginally follow Poisson distributions. Nevertheless, since some links share the same route flows, link volumes are not independent random variables, so that their covariances will be different from zero. Their joint distribution $g(\mathbf{z} | \boldsymbol{\theta})$ should be some complicated form of multivariate Poisson. Vardi (1996) and Hazelton (2000) proposed to approximate this distribution by a multivariate normal distribution:

$$g(\mathbf{z} | \boldsymbol{\theta}) = (2\pi)^{-m/2} |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{z} - \mathbf{F} \boldsymbol{\theta})^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} (\mathbf{z} - \mathbf{F} \boldsymbol{\theta}) \right\} \tag{5}$$

where $E[\mathbf{z}] = \boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{F} \boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \Delta \text{diag}(\mathbf{P} \boldsymbol{\theta}) \Delta^T$, in which $\text{diag}(\cdot)$ denotes a

diagonal matrix and $|\cdot|$ denotes the determinant of a matrix.

A noteworthy fact about the multivariate normal in (5) is that both expected value μ and covariance matrix Σ are functions of the parameters θ . This is reminiscent of the Poisson univariate distribution, whose mean and variance are equal. It should also be noted in the density function defined by (5) that the matrix Δ should be formed only by independent rows, otherwise the covariance matrix $\Sigma(\theta)$ will be singular. This means that only the rows corresponding to non-redundant observed links should be included in Δ .

Another important issue related to (5) is how to set the route choice matrix \mathbf{P} . In uncongested networks, it could be estimated separately by means of a route choice model, since an assumption in uncongested networks is that route choices are independent of the OD flows. In case of congested networks, the matrix \mathbf{P} is dependent on the OD flows and a more appropriate approach would be to treat it as a parameter of the model, so that it could be jointly estimated with mean OD flows.

Given a sample of size N of traffic volumes vectors $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(N)}$, each observed in different days during the same reference time period and assumed independent, we define the *likelihood function* as the following:

$$\mathcal{L}(\theta) = \prod_{j=1}^N f(\mathbf{z}^{(j)} | \theta) \tag{6}$$

We can define as a maximum likelihood estimate of the mean OD matrix a maximizer of equation (6). It is though computationally more convenient to maximize the *log-likelihood* function, given by $\ell(\theta) = \log \mathcal{L}(\theta)$, which in our present case takes the following form:

$$\ell(\theta) = -\frac{N}{2} \log |\Sigma(\theta)| - \frac{1}{2} \sum_{j=1}^N (\mathbf{z}^{(j)} - \mathbf{F}\theta)^T \Sigma(\theta)^{-1} (\mathbf{z}^{(j)} - \mathbf{F}\theta) + c \tag{7}$$

In which $c = -(Nm/2) \log(2\pi)$ is constant with respect to θ . The maximum likelihood estimator is $\hat{\theta} = \operatorname{argmax}_{\theta \geq 0} \ell(\theta)$, which may be obtained by means of non-linear programming methods. A particular difficulty in maximizing equation (7) is that it may not be strictly concave, i.e., it may have multiple maximizers. This will require some form of regularization, for example the selection of $\hat{\theta}$ which is closest to a prior OD matrix or the use of a Bayesian approach as described in section 3.3.

In case the sample of observed volume vectors is large, Hazelton (2000) proposes to substitute the population covariance matrix by the sample covariance matrix S in the log-likelihood, resulting in the following approximation to the log-likelihood that is computationally more tractable:

$$\tilde{\ell}(\theta) = -\frac{1}{2} \sum_{j=1}^N (\mathbf{z}^{(j)} - \mathbf{F}\theta)^T \mathbf{S}^{-1} (\mathbf{z}^{(j)} - \mathbf{F}\theta) \tag{8}$$

Maximizing equation (8) for $\theta \geq 0$ can be interpreted as finding an estimate for which the sum of the weighted quadratic distance (\mathbf{S}^{-1} is the matrix of weights) between the expected volume vector $\mu = \mathbf{F}\theta$ and the observed volume vectors $\mathbf{z}^{(j)}, j = 1, 2, \dots, N$ is minimal. However, $\tilde{\ell}(\theta)$ is not strictly concave if Δ has more columns than rows, which occurs if there are more OD pairs than observed links, and we may also have to resort to a prior OD matrix.

3.2 Models based on the method of moments

The method of moments is a classical technique in point estimation. Its basic idea is to solve the equations obtained by equating population and sample moments. Let $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(N)}$ be a sample of traffic volumes vectors, each observed in different days during the same reference time period and assumed independent. Their sample mean $\bar{\mathbf{z}}$ and sample covariance matrix S are given respectively by:

$$\bar{\mathbf{z}} = \frac{1}{N} \sum_{j=1}^N \mathbf{z}^{(j)} \tag{9}$$

$$\mathbf{S} = \frac{1}{N-1} \sum_{j=1}^N (\mathbf{z}^{(j)} - \bar{\mathbf{z}})(\mathbf{z}^{(j)} - \bar{\mathbf{z}})^T \tag{10}$$

In theory, an estimate of the mean OD flows θ may be obtained by solving the linear system of equations resulting from equating population and sample moments with the additional constrain the $\theta \geq 0$, as given below:

$$\begin{aligned} \Delta \mathbf{P} \theta &= \bar{\mathbf{z}} \\ \Delta \text{diag}(\mathbf{P} \theta) \Delta^T &= \mathbf{S} \end{aligned} \tag{11}$$

The linear system given by (11) will have multiple solutions if its number of independent equations is less than the number of OD pairs. Moreover, it may even be inconsistent, i.e., it may have no $\theta \geq 0$ which satisfy all equations. The inconsistencies may arise from sampling error. To overcome this problem, Hazelton (2003) proposed to set an optimization model whose objective-function is the minimization of the distances between theoretical and sample moments:

$$\hat{\theta} = \underset{\theta \geq 0}{\text{argmin}} \{ \|\Delta \mathbf{P} \theta - \bar{\mathbf{z}}\| + \rho \|\text{vec}(\Delta \text{diag}(\mathbf{P} \theta) \Delta^T) - \text{vec}(\mathbf{S})\| \} \tag{12}$$

In which $\text{vec}(\cdot)$ is the vector concatenation of the columns of a matrix, $\|\cdot\|$ denotes a suitable distance measure (a common one is the Euclidian distance), $\rho \geq 0$ is a weighting factor. It is worth noting that the method of moments is not dependent on the Poisson assumption, and may be applied to more general models in which only a relationship between the mean vector and the covariances matrix is assumed.

3.3 Bayesian models

Models based on Bayesian inference define, in addition to the likelihood of the data, prior and posterior probability distributions for the OD matrix. The main model in this class was proposed by Tebaldi and West (1998). Following Vardi (1996), they assume that OD flows \mathbf{x} are discrete and follow independent Poisson distributions, whose mean value θ_i for OD pair i follows a gamma prior probability distribution. They also assume that only one route is available for each OD pair, so that the route choice matrix \mathbf{P} equals the identity matrix.

In their contribution, the authors assume one has a vector \mathbf{z} of observed traffic volumes in a single point in time. In Bayesian inference the unknown vector of parameters θ is treated as a random variable, for which we define a prior distribution $p(\theta)$ before link volume data \mathbf{z} are observed, and a posterior distribution $p(\mathbf{z} | \theta)$ updated via Bayes theorem after link volumes data are observed: (See Gelman *et al.*, 2014 for a comprehensive reference on Bayesian inference)

$$p(\theta | \mathbf{z}) = \frac{p(\mathbf{z} | \theta)p(\theta)}{p(\mathbf{z})} \tag{13}$$

with the following likelihood function, obtained by marginalizing \mathbf{x} in the joint probability distribution of \mathbf{x} and \mathbf{z} :

$$p(\mathbf{z} | \theta) = \sum_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}, \mathbf{z} | \theta) \tag{14}$$

in which $\mathcal{X} = \{\mathbf{x} : \Delta \mathbf{x} = \mathbf{z}\}$ is the set of OD flow vectors consistent with the observed link volume vector \mathbf{z} . (this implicitly assumes that all \mathbf{x} which do not satisfy $\Delta \mathbf{x} = \mathbf{z}$ have probability zero) The problem with the likelihood function (14) is that the evaluation of the sum requires the enumeration of all vectors $\mathbf{x} \in \mathcal{X}$, which is computationally infeasible for even moderately-sized networks. Their solution strategy

is to evaluate not the marginal posterior $p(\boldsymbol{\theta}|\mathbf{z})$, but the joint posterior distribution $p(\mathbf{x},\boldsymbol{\theta}|\mathbf{z})$, by noting that:

$$p(\mathbf{x},\boldsymbol{\theta}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x})p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{15}$$

In addition, they assume there are no observation errors in link volumes, so that $p(\mathbf{z}|\mathbf{x}) = I(\Delta\mathbf{x} = \mathbf{z})$ and I denotes the *indicator function*, $I(A) = 1$ if A is true and 0 otherwise. In order to sample from the joint posterior given by (15), the authors proposed a Gibbs sampler (Geman and Geman, 1984). Gibbs sampling is a type of Markov chain Monte Carlo method (Gelman *et al.*, 2014) which allows sampling from a joint distribution by sampling from conditional distributions.

The proposed Gibbs sampler iteratively samples from the conditional distributions $p(x_i | \mathbf{x}_{[-i]}, \boldsymbol{\theta}, \mathbf{z})$ and $p(\theta_i | \boldsymbol{\theta}_{[-i]}, \mathbf{x}, \mathbf{z})$, in which $\mathbf{x}_{[-i]}$ and $\boldsymbol{\theta}_{[-i]}$ denote the corresponding vectors with component i excluded. This sequence of conditional samples is a Markov chain, which converges (mixes) in the long run to the desired joint posterior distribution $p(\mathbf{x}, \boldsymbol{\theta} | \mathbf{z})$. We can draw samples from the marginal posteriors $p(\mathbf{x} | \mathbf{z})$ and $p(\boldsymbol{\theta} | \mathbf{z})$ by taking the corresponding values generated by the Markov chain. Then summaries such as sample mean OD flows and standard deviations can be calculated from the samples.

3.4 Dynamic models

All models discussed up to this point in this review take for granted that transportation systems are *stationary*. By stationary we mean parameters such as mean OD flows and variances do not vary over time. In this sense OD flows random pattern is static. However, in reality transportation systems are non-stationary, since social-economic conditions, infrastructure and other factors change over time. This implies mean OD flows, variances and possibly other parameters are *dynamic*. In order to predict future behavior of the system or to assess the impact of interventions in the system over time, we should be able to model the dynamic variation of the demand.

The types of dynamic models for OD matrix estimation may be classified in two broad classes, according to the time scale of the model: *within-day* and *day-to-day* models. Within-day models consider the time variation of the demand for a specified time period within a single day. The span of the time period under study may vary from a few minutes to the whole day. In contrast, day-to-day models are often concerned with the variation of the demand for a repeated reference time period (typically the peak hour) over a sequence of days. Within-day models are mainly approached from a reconstruction perspective via optimization models. For further reading on within-day dynamic models, we refer to the works of Willumsen (1984), Cremer and Keller (1987), Cascetta (1993) and Ashok and Ben-Akiva (2002). In this review, we focus on day-to-day models, which have been approached from a statistical estimation perspective.

In the day-to-day dynamic OD matrix estimation problem, we want to estimate a sequence of unobserved mean OD matrices given a sample of link volume vectors observed in a sequence of $t = 1, 2, \dots, T$ consecutive days on some links of the network. Hazelton (2008) proposes a dynamic model based on the multivariate normal distribution approximation to independent OD route flows following Poisson distributions. Let λ_{kt} be the mean OD flows on route k at time t , and $\boldsymbol{\lambda}_t = (\lambda_{1t}, \lambda_{2t}, \dots, \lambda_{rt})^T$ with $r = \sum_{i \in \mathcal{I}} |\mathcal{K}_i|$ the number of routes over all OD pairs. By using a normal approximation, the conditional distribution of link volumes \mathbf{z}_t given mean OD route flows $\boldsymbol{\lambda}_t$ is assumed to be $p(\mathbf{z}_t | \boldsymbol{\lambda}_t) = N(\Delta\boldsymbol{\lambda}_t, \phi\Delta\text{diag}(\boldsymbol{\lambda}_t)\Delta^T)$ and $\phi > 0$ is a scale factor which adjusts for link volumes which are not compliant with the Poisson assumption.

Hazelton (2008) further considers parsimonious parametrizations of the mean OD route flows. The idea is to represent the mean OD route flows $\boldsymbol{\lambda}_t$ as a function of a vector of parameters $\boldsymbol{\theta}$ which does

not change with time. A possible parsimonious model is to represent the vector λ_t as a linear model $\lambda_t = \lambda_0 + t\delta$, with a vector of parameters $\theta = (\lambda_0, \delta, \phi)$, so that we make inferences on the initial vector λ_0 , the time increment vector δ and the link volume scale factor ϕ . Another possibility is a weekday-weekend model, in which $\lambda_t = \lambda_0$ in weekdays and $\lambda_t = \gamma\lambda_0$ for weekends and γ is a demand adjustment factor.

Assuming that link volumes are independent over time, the likelihood function of the observed link volumes will be given by the following expression:

$$\mathcal{L}(\theta) = \prod_{t=1}^T (2\pi)^{-T/2} |\Sigma_t(\theta)|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{z}_t - \Delta\lambda_t(\theta))^T \Sigma_t(\theta)^{-1}(\mathbf{z}_t - \Delta\lambda_t(\theta))\right\} \quad (18)$$

Where $\Sigma_t(\theta) = \phi\Delta\text{diag}(\lambda_t)\Delta^T$. We can obtain a maximum likelihood estimator for θ by maximizing equation (18). Alternatively, we can obtain a Bayesian estimator by specifying a prior distribution on θ and evaluating the posterior distribution $p(\theta | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_T)$. This will require the use of Markov chain Monte Carlo methods, such as Gibbs sampling and Metropolis-Hastings. (Gelman *et al.*, 2014)

A limitation of this latter model is that it assumes independence of link volumes, which may not correspond to reality. In fact, it is expected that link volumes exhibit correlations over time due the nature of the decision process of users, who dynamically adapt to congestion conditions. Pitombeira-Neto and Loureiro (2016) proposed a dynamic linear model for the time evolution of OD flows. They assume that, in the short term, mean OD flows θ_t are *locally constant*. In other words, at time t mean OD flows should be equal to previous OD flows at time $t-1$ but shifted by some stochastic error:

$$\theta_t = \theta_{t-1} + \omega_t \quad (19)$$

In which $\omega_t \sim N(0, \mathbf{W}_t)$, N denotes the multivariate normal distribution and the covariance matrix \mathbf{W}_t determines the stochastic evolution of mean OD flows over time. Moreover, the vector of observed volumes \mathbf{z}_t at time t is related to the current mean OD vector θ_t through an *observation equation*:

$$\mathbf{z}_t = \mathbf{F}_t\theta_t + \nu_t \quad (20)$$

in which $\nu_t \sim N(0, \mathbf{V}_t)$ is the observation error at time t . It represents the variability of the observed volumes around the mean expected volumes given by $E[\mathbf{z}_t] = \mathbf{F}_t\theta_t$.

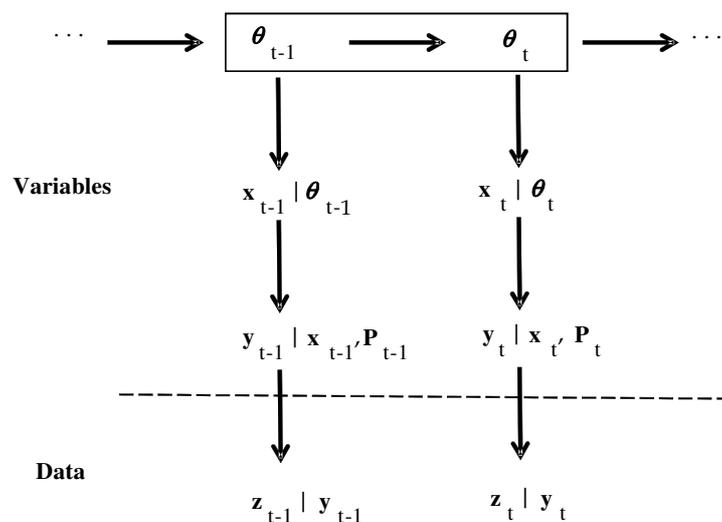


Figure 1: Dynamic model structure

Pitombeira-Neto and Loureiro (2016) notice that special attention should be given to the specification of the covariance matrix \mathbf{V}_t of the observed volumes. Due to the network structure, link volumes are correlated and this must be represented in the covariance matrix. They identify three sources of variability affecting link volumes: the generation of OD flows; the route choice process; and the counting of volumes on the links. The structure of the dynamic model may be represented as in Figure 1.

Given mean OD flows $\boldsymbol{\theta}_t$, they define the conditional distribution of realized OD flows $\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tm})^T$ as $p(\mathbf{x}_t | \boldsymbol{\theta}_t) = N(\boldsymbol{\theta}_t, \boldsymbol{\Sigma}_t^x)$, where $\boldsymbol{\Sigma}_t^x$ is the covariance matrix of the realized OD flows. Then, for each OD pair i they assume the vector of route flows $\mathbf{y}_{ii} = (y_{ii1}, y_{ii2}, \dots, y_{ii(n(i))})$, in which $n(i) = |\mathcal{K}_i|$, has a conditional probability distribution $p(\mathbf{y}_{ii} | x_{ii}, \mathbf{p}_{ii}) = N(x_{ii}\mathbf{p}_{ii}, \boldsymbol{\Sigma}_{ii}^y)$ with $\mathbf{p}_{ii} = (p_{ii1}, p_{ii2}, \dots, p_{ii(n(i))})^T$ the vector of route choice probabilities of OD pair $i \in \mathcal{I}$. $\boldsymbol{\Sigma}_{ii}^y$ has a multinomial-like structure given by:

$$\boldsymbol{\Sigma}_{ii}^y = x_{ii}(\text{diag}(\mathbf{p}_{ii}) - \mathbf{p}_{ii}\mathbf{p}_{ii}^T) \tag{21}$$

Let $\mathbf{y}_t = (\mathbf{y}_{t1}, \mathbf{y}_{t2}, \dots, \mathbf{y}_{tm})^T$. We specify the conditional probability density as $p(\mathbf{y}_t | \mathbf{x}_t) = N(\mathbf{P}_t\mathbf{x}_t, \boldsymbol{\Sigma}_t^y)$ with \mathbf{P}_t a block-diagonal matrix whose components are \mathbf{p}_{ii} and $\boldsymbol{\Sigma}_t^y$ a block-diagonal matrix whose components are $\boldsymbol{\Sigma}_{ii}^y$ for all $i \in \mathcal{I}$. The conditional distribution of link volumes \mathbf{z}_t given route flows \mathbf{y}_t is defined as $p(\mathbf{z}_t | \mathbf{y}_t) = N(\Delta\mathbf{y}_t, \boldsymbol{\Sigma}_t^z)$, where $\boldsymbol{\Sigma}_t^z$ is the covariance matrix related to measurement errors. Finally, it can be shown that the conditional density $p(\mathbf{z}_t | \boldsymbol{\theta}_t) = N(\mathbf{F}_t\boldsymbol{\theta}_t, \mathbf{V}_t)$ with $\mathbf{F}_t = \Delta\mathbf{P}_t$ and \mathbf{V}_t given by:

$$\mathbf{V}_t = \mathbf{F}_t\boldsymbol{\Sigma}_t^x\mathbf{F}_t^T + \Delta\overline{\boldsymbol{\Sigma}}_t^y\Delta^T + \boldsymbol{\Sigma}_t^z \tag{22}$$

in which $\overline{\boldsymbol{\Sigma}}_t^y$ is an estimate of $\boldsymbol{\Sigma}_t^y$ given by substituting a current estimate of \mathbf{x}_t in equation (21). Inference on mean OD flows $\boldsymbol{\theta}_t$ at current time t given a time series of traffic volumes $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_t$ is carried out by obtaining the conditional $p(\boldsymbol{\theta}_t | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_t)$ through Bayes theorem, which can be shown to be multivariate normal with mean \mathbf{m}_t and covariance matrix \mathbf{C}_t given respectively by the expressions below:

$$\begin{aligned} \mathbf{m}_t &= \overline{\mathbf{m}}_t + \mathbf{A}_t(\mathbf{z}_t - \mathbf{F}_t\overline{\mathbf{m}}_t) \\ \mathbf{C}_t &= \overline{\mathbf{C}}_t - \mathbf{A}_t(\mathbf{F}_t\overline{\mathbf{C}}_t\mathbf{F}_t^T + \mathbf{V}_t)\mathbf{A}_t^T \end{aligned} \tag{23}$$

with $\mathbf{A}_t = \overline{\mathbf{C}}_t\mathbf{F}_t^T(\mathbf{F}_t\overline{\mathbf{C}}_t\mathbf{F}_t^T + \mathbf{V}_t)^{-1}$. Equations (23) are computed recursively, one observation \mathbf{z}_t at a time, with the prior density $p(\boldsymbol{\theta}_t | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{t-1}) = N(\overline{\mathbf{m}}_t, \overline{\mathbf{C}}_t)$ and an initial prior $p(\boldsymbol{\theta}_0) = N(\mathbf{m}_0, \mathbf{C}_0)$. Through simulated experiments, Pitombeira-Neto and Loureiro (2016) show that this dynamic linear model is capable of approximating the unobserved mean OD flows as more data is gathered on traffic volumes with no prior information on OD flows.

4. DISCUSSION

In the following subsections, we discuss some research questions which are still not fully answered in the literature.

4.1 The underspecification problem

The underspecification problem refers to the fact that the linear relation between OD flows and traffic volumes, given by equation (4), has no unique solution. It occurs because real transportation networks

have more OD pairs than observed links, so that there will be many OD matrices consistent with observed traffic volumes. In statistical models, this problem may arise as unidentifiability of parameters, i.e., even under the hypothesis of an infinite sample of observed data, it would not be possible to uniquely determine the parameters of the statistical model. This hurdle is often overcome through some kind of *regularization*, i.e., some enrichment to the model which allows one to obtain unique parameter estimates.

In some models, such as the static ones presented in Sections 3.1 to 3.3, regularization takes the form of the additional assumption that OD flows follow independent Poisson distributions. Under such assumption and some more weak conditions on the link-path matrix, Vardi (1996) and Hazelton (2000) show that mean OD flows are identifiable. But this only means that mean OD flows may be uniquely determined if one has an infinite sample of observed traffic volumes, which gives no hint on the quality of the estimates with finite samples. This leads us to the following research question: *How to define the sample size of observed days of link volumes for a given level of confidence and a set of non-redundant links?*

In their model, described in Section 3.3, Tebaldi and West (1998) regularize the model by using prior information in the form of a prior OD matrix. This may be an old OD matrix obtained from house surveys. Nevertheless, the prior OD matrix will bias the estimate in its direction. In other words, the estimated OD matrix will be dependent on the prior OD matrix, but this dependence has not been properly investigated in the literature. Thus, an important research question is: *How sensitive is the estimated OD matrix to a prior OD matrix?*

Besides a prior OD matrix and link volumes, new models are beginning to incorporate other sources of information so as to obtain better estimates. Castillo *et al.* (2008) develop a model which uses plate scanning data, Parry and Hazelton (2012) propose a model which uses samples of route flows obtained from tracking vehicles in addition to link volumes, and Caceres *et al.* (2013) develop a methodology in order to estimate link volumes from mobile phone data.

4.2 Model assumptions

For analytical convenience, many researchers have assumed that OD flows follow independent Poisson distributions, with the implication that traffic volumes will also follow Poisson distributions with mean values equal to their variances. Nevertheless, Hazelton (2003) has argued that link volumes may be *overdispersed*, i.e., the variances may be greater than the mean values. This empirical finding poses the following research question: *Do we incur severe errors if we assume Poisson distributions in the presence of overdispersion?* In case link volumes depart considerably from the Poisson assumption, we should consider the use of more flexible probability distributions for OD flows which can accommodate the possibility of overdispersion or underdispersion.

Moreover, we should consider possible dependence regarding OD flows. As transportation demand has a spatial nature, it seems a plausible hypothesis that OD flows that share the same origin or the same destination may have covariances different from zero. In a recent study, Shao *et al.* (2014) proposed a model that estimates both the mean OD flows and their covariances from link counts. However, their approach greatly increases the number of parameters to be estimated, which demands a large amount of data in order to obtain estimates with good accuracy. This leads to the following question: *How to specify parsimonious models which can estimate OD flows covariances?*

4.3 Estimation of the route choice matrix

The route choice matrix gives the probabilities that a user chooses a specific route in an OD pair. It is a basic component in most models and it is generally estimated through discrete choice models. The most used model is the multinomial logit Ben-Akiva and Lerman (1985), though there are other models which consider correlation between paths in modeling route choices, such as the C-logit (Zhou *et al.*; 2012) and combinatorial paired logit (Chen *et al.*; 2014). In theory, the route choice matrix could also be jointly estimated along with mean OD flows from observed link volumes (Lo *et al.*, 1996; Hazelton, 2000), at

the price of greatly increasing the number of parameters to be estimated. This gives rise to the following research question: *Is it feasible to jointly estimate route choice probabilities and mean OD flows in real scale networks?* An alternative approach, pursued by Lo and Chan (2003) and Hazelton (2010), is to estimate the parameters of a route choice model jointly with the mean OD flows, instead of estimating directly the route choice probabilities. This reduces the number of parameters at the expense of adopting some assumptions on the choice behaviour of users according to the route choice model applied.

Another issue is that all models presented in this paper depend on an estimate of the matrix of route probabilities and consequently on the definition of a set of routes for each OD pair. It is known that in large networks, there is a large amount of paths connecting the same OD pair, and many of these paths are not even considered by the users. To find the routes actually used by drivers over time is not an easy task, especially in congested networks where drivers may exchange routes constantly. Besides, the estimated parameters of the route choice models are affected by the considered set of routes. So, *how sensitive is the estimated OD matrix to the route probability matrix or how sensitive is the estimated OD matrix to the traffic allocation matrix?*

5. CONCLUDING REMARKS

In this paper, we reviewed the main statistical models for OD matrix estimation based on link counts. It is worth noting that this paper focused on the estimation of mean OD matrix rather than reconstruction of a realized OD matrix that generated a given set of observed link volumes. Estimation methods rely on the main assumption that the trip patterns in a transportation network vary over time and we are interested in estimating a mean OD pattern. In contrast, reconstruction methods are applied for the case of deterministic OD flows or when we are interested in finding a certain OD matrix for a specific time reference.

The potential of these methods to support transportation operations and planning has been the main motivation for many years, more than four decades, of research efforts. It is believed that nowadays, with the possibility of acquiring data of traffic flows at low cost and in real time from traffic monitoring systems, we can finally understand the variation of traffic volumes on a real scale network and explore in what conditions reconstruction and estimation methods can be applied. As discussed here and in Oliveira Neto et. al (2016), all methods suffer from the similar limitations, related to the underspecification problem, model specification, and estimation of the route choice matrix, giving rise to open research questions.

This paper dealt specifically with the problem of estimating either static or dynamic OD flows, represented by models whose parameters are estimated by means of maximum likelihood, the method of moments or Bayesian inference. These models represent an advance in the past literature that focused on reconstruction of an approximately deterministic OD matrix. Although the problem of reconstructing and estimating OD matrix from traffic counts have attracted considerable attention in the past decades, there is still many gaps to be fulfilled so as to decide whether or not practitioners should invest in more sophisticated models in real scale transportation planning. On the list below we point to some research directions we think should be pursued in order to advance the theory and applications of the models for estimation of OD matrices:

To analyse the sensitivity of estimated OD matrices to prior OD matrices: Due to the underspecification problem, most models rely on a prior OD matrix. There is little research reported in the literature on the sensitivity of the estimated OD matrix to a prior matrix.

To assess the magnitude of the estimation error as the sample size of observed traffic volumes increases: In real networks, observed links form a small subset of all links in the network. Knowledge of how the size of this subset affects the estimation error can support the decision on how many links should be observed. Likewise, the dependence of the estimation error on the number of days for which one has observations on link volumes has not been fully addressed yet.

To investigate the validity of the Poisson assumption: Although some papers in the literature already

consider the relaxation of the Poisson assumption for OD flows, this assumption may be required for the sake of analytical or computational tractability. The trade-off between model complexity and tractability should be balanced in favor of the latter for the models to be applicable to real transportation networks.

To develop statistical models which jointly consider the probabilistic nature of OD flows and route choices: Most statistical models have assumed that the route choice matrix is previously known and independent of the OD matrix, which is a valid assumption for uncongested networks. However, in congested networks route choices are dependent on the OD matrix, in order that the route choice matrix has to be estimated along with OD flows. In deterministic models this has been done through equilibrium traffic assignment. Nevertheless, it is known that the classical definition of *stochastic user equilibrium*, due to Daganzo and Sheffi (1977), does not imply an equilibrium probability distribution on OD flows (Hazelton, 1998; Nakayama and Watling, 2014). Hence, the modelling of OD flows through a fully probabilistic model which does not assume an equilibrium state may provide new insights into the problem.

To develop models which take into account spatial correlation among OD flows: Most models proposed in the literature have assumed that OD flows are independent. Nevertheless, transportation demand has a strong spatial nature and it is expected that OD flows exhibit positive or negative spatial correlation. The explicit modeling of spatial correlation may add new sources of data and contribute to reducing the underspecification problem and estimation errors.

Finally, we point out that the recent development of dynamic models is a promising avenue for research. Dynamic models may mitigate the underspecification problem, as argued by Marzano *et al.* (2009). Pitombeira-Neto and Loureiro (2016) have recently shown through simulation studies that it is possible to estimate OD flows with no prior information if assignment matrices are known, and the accuracy of the estimates gets better as more observations on link volumes are gathered. Dynamic models also do not rely on the equilibrium assumption and allow the consideration of time dependence in OD flows and route flows. Moreover, the study of the time variation in OD flows may allow novel applications in transportation planning through the analysis of long term trends in OD demand.

ACKNOWLEDGEMENTS

The authors thank CNPq and CAPES for the financial support.

REFERENCES

- Ashok, K. and M. E. Ben-Akiva (2002) Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows. *Transportation Science*, v. 36, p. 184–198. DOI: 10.1287/trsc.36.2.184.56
- Bekhor, S.; M. E. Ben-Akiva and M. S. Ramming (2006) Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, v. 144, p. 235–247. DOI: 10.1007/s10479-006-0009-8
- Ben-Akiva, M. E. and S. R. Lerman (1985) *Discrete choice analysis: Theory and application to travel demand*. MIT Press, Cambridge, MA, USA.
- Caceres, N.; L. M. Romero and F. G. Benitez (2013) Inferring origin-destination trip matrices from aggregate volumes on groups of links: a case study using volumes inferred from mobile phone data. *Journal of Advanced Transportation*, v. 47, p. 650–666. DOI: 10.1002/atr.187
- Carvalho, L. E. (2014) A Bayesian statistical approach for inference on static origin-destination matrices in transportation studies. *Technometrics*, v. 56, p. 225–237. DOI: 10.1080/00401706.2013.826144
- Cascetta, E. (2009) *Transportation Systems Analysis: Models and Applications* (2nd ed.). Springer, New York, NY, USA.
- Cascetta, E. (1993) Dynamic estimators of origin-destination matrices using traffic counts. *Transportation Science*, v. 27, p. 363–373. DOI: 10.1287/trsc.27.4.363
- Castillo, E.; J. M. Menéndez and P. Jiménez (2008) Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. *Transportation Research Part B*, v. 42, p. 455–481. DOI: 10.1016/j.trb.2007.09.004
- Chen, A. S.; Ryu, X. Xu and K. Choi (2014) Computation and application of the paired combinatorial logit stochastic user equilibrium problem. *Computers and Operations Research*, v. 43, p. 68–77. DOI: 10.1016/j.cor.2013.08.022
- Cremer, M. and H. Keller (1987) A new class of dynamic methods for the identification of origin-destination flows. *Transportation Research Part B*, v. 21, p. 117–132. DOI: 10.1016/0191-2615(87)90011-7
- Daganzo, C. F. and Y. Sheffi (1977) On stochastic models of traffic assignment. *Transportation Science*, v. 11, n. 3, p. 253–274. DOI: 10.1287/trsc.11.3.253
- Gelman, A.; J. B. Carlin; H. S. Stern; D. B. Dunson; A. Vehtari and D. B. Rubin (2013) *Bayesian Data Analysis* (3rd ed.). Chapman & Hall, Boca Raton, FL, USA.

- Geman, S. and D. Geman (1984) Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 6, n. 6, p. 721–741. DOI: 10.1109/TPAMI.1984.4767596
- Hastings, W. K. (1970) Monte carlo sampling methods using markov chains and their applications. *Biometrika*, v. 57, n. 1, p. 97–109. DOI: 10.2307/2334940
- Hazelton, M. L. (1998) Some remarks on stochastic user equilibrium. *Transportation Research Part B*, v. 32, p. 101–108. DOI: 10.1016/S0191-2615(97)00015-5
- Hazelton, M. L. (2000) Estimation of origin-destination matrices from link flows on uncongested networks. *Transportation Research Part B*, v. 34, p. 549–566. DOI: 10.1016/S0191-2615(99)00037-5
- Hazelton, M. L. (2001) Inference for origin-destination matrices: estimation, prediction and reconstruction. *Transportation Research Part B*, v. 35, p. 667–676. DOI: 10.1016/S0191-2615(00)00009-6
- Hazelton, M. L. (2003) Some comments on origin-destination matrix estimation. *Transportation Research Part A*, v. 37, p. 811–822. DOI: 10.1016/S0965-8564(03)00044-2
- Hazelton, M. L. (2008) Statistical inference for time varying origin-destination matrices. *Transportation Research Part B*, v. 42, p. 542–552. DOI: 10.1016/j.trb.2007.11.003
- Hazelton, M. L. (2010) Bayesian inference for network-based models with a linear inverse structure. *Transportation Research Part B*, v. 44, p. 674–685. DOI: 10.1016/j.trb.2010.01.006
- Lo, H.-P.; N. Zhang and W. H. K. Lam (1996) Estimation of an origin-destination matrix with random link choice proportions: a statistical approach. *Transportation Research Part B*, v. 30, p. 309–324. DOI: 10.1016/0191-2615(95)00036-4
- Lo, H.-P. and C.-P. Chan (2003) Simultaneous estimation of an origin-destination matrix and link choice proportions using traffic counts. *Transportation Research Part A*, v. 37, p. 771–788. DOI: 10.1016/S0965-8564(03)00048-X
- Marzano, V.; A. Papola and F. Simonelli (2009) Limits and perspectives of effective O-D matrix correction using traffic counts. *Transportation Research Part C*, v. 17, p. 120–132. DOI: 10.1016/j.trc.2008.09.001
- Nakayama, S. and D. Watling (2014) Consistent formulation of network equilibrium with stochastic flows. *Transportation Research Part B*, v. 66, p. 50–69. DOI: 10.1016/j.trb.2014.03.007
- Oliveira Neto, F. M.; A. R. Pitombeira Neto; C. F. G. Loureiro and B. V. Bertoncini (2016) Discussão conceitual sobre métodos de reconstrução de matrizes origem-destino estáticas em redes de transportes. *Transportes*, v.24, n.1, p. 107-119. DOI: 10.14295/transportes.v24i1.1017
- Ortúzar, J. de D. and L. G. Willumsen (2011) *Modelling Transport* (4th ed.). Wiley, Chichester, UK.
- Parry, K. and M. L. Hazelton (2012) Estimation of origin-destination matrices from link count and sporadic routing data. *Transportation Research Part B*, v. 46, p. 175–188. DOI: 10.1016/j.trb.2011.09.009
- Parry, K. and M. L. Hazelton (2013) Bayesian inference for day-to-day dynamic traffic models. *Transportation Research Part B*, v. 50, p. 104–115. DOI: 10.1016/j.trb.2013.01.003
- Pitombeira-Neto, A. R. and C. F. G. Loureiro (2016) A dynamic linear model for the estimation of time-varying origin-destination matrices from link counts. *Journal of Advanced Transportation*, v. 50, n. 8, p. 2116–2129. DOI: 10.1002/atr.1449
- Robillard, P. (1975) Estimating the OD matrix from observed link volumes. *Transportation Research*, v. 9, p. 123–128. DOI: 10.1016/0041-1647(75)90049-0
- Shao, H.; W. H. K. Lam; A. Sumalee; A. Chen and M. L. Hazelton (2014) Estimation of mean and covariance of peak hour origin-destination demands from day-to-day traffic counts. *Transportation Research Part B*, v. 68, p. 52–75. DOI: 10.1016/j.trb.2014.06.002
- Spiess, H. (1987) A maximum likelihood model for estimating origin-destination matrices. *Transportation Research Part B*, v. 21, p. 395–412. DOI: 10.1016/0191-2615(87)90037-3
- Tebaldi, C. and M. West (1998) Bayesian inference on network traffic using link count data. *Journal of the American Statistical Association*, v. 93, p. 557–573. DOI: 10.1080/01621459.1998.10473707
- Timms, P. (2001) A philosophical context for methods to estimate origin-destination trip matrices using link counts. *Transport Reviews*, v. 21, p. 269–301. DOI: 10.1080/713868144
- Vardi, Y. (1996) Network tomography: Estimating source-destination traffic intensities from link data. *Journal of the American Statistical Association*, v. 91, p. 365–377. DOI: 10.2307/2291416
- Willumsen, L. G. (1981) Simplified transport models based on traffic counts. *Transportation*, v. 10, p. 257–258. DOI: 10.1007/BF00148462
- Willumsen, L. G. (1984) Estimating time dependent trip matrices from traffic counts. *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*, Delft, v. 1, p. 397–411.
- Zhou, Z.; A. Chen and S. Bekhor (2012) C-logit stochastic user equilibrium model: formulations and solution algorithm. *Transportmetrica*, v. 8, p. 17–41. DOI: 10.1080/18128600903489629