IET Intelligent Transport Systems

Research Article

Methodology for O-D matrix estimation using the revealed paths of floating car data on large-scale networks

ISSN 1751-956X
Received on 16th October 2019
Revised 16th July 2020
Accepted on 28th August 2020
doi: 10.1049/iet-its.2019.0684
www.ietdl.org

Anna Mitra[1] ✉, Alessandro Attanasi[2], Lorenzo Meschini[2], Guido Gentile[1]
[1]*Sapienza University - DICEA, Via Eudossiana 18, Rome, Italy*
[2]*PTV SISTeMA, Via Bonghi 11b, Roma, Italy*
✉ *E-mail: anna.mitra@uniroma1.it*

Abstract: The increasing availability of historical floating car data (FCD) represents a relevant chance to improve the accuracy of model-based traffic forecasting systems. A more precise estimation of origin–destination (O-D) matrices is a critical issue for the successful application of traffic assignment models. The authors developed a methodology for obtaining demand matrices without any prior information, but just starting from a data set of vehicle trajectories, and without using any assignment model, as traditional correction approaches do. Several steps are considered. A data-driven approach is applied to determine both observed departure shares from origins to destinations and static assignment matrices. Then the O-D matrix estimation problem is formulated as a scaling problem of the observed FCD demand and carried out using as inputs: a set of traffic counts, the FCD revealed assignment matrix and the observed departure shares as an a-priori matrix. Four different optimisation solutions are proposed. The methodology was successfully tested on the network of Turin. The results highlight the concrete opportunity to perform a data-driven methodology that, independently from the reliability of the reference demand, minimises manual and specialised effort to build and calibrate the transportation demand models.

1 Introduction

The origin–destination (O-D) demand, referred to as the O-D matrix, is one essential input for transportation analysis, particularly for simulation models, like static or dynamic traffic assignment. Typically, a seed O-D matrix is estimated from a large-scale survey of mobility (direct estimation) or from production and gravity models (indirect estimation). Then, this is corrected (or adjusted) to match traffic counts.

This second step has received a lot of interest in the last decades. In the literature, the O-D matrix estimation has been initially studied, assuming static conditions for a simulation process that considers time-independent traffic flows. In this case, an average O-D demand is estimated for a specific temporal window, under the assumption of steady-state flows on the network. If instead the demand varies during the observation interval with peaks and queues, we have to consider a dynamic model.

Passing from the static to the dynamic context, the demand adjustment problem increases its complexity due to the role of time, which adds one-dimensional on link flows and one on O-D flows, to reflect the congestion conditions on the network. This is known in the literature as a dynamic O-D estimation problem.

Both static and dynamic versions face the under-specification issue, i.e. in theory, multiple solutions may exist which map O-D flows to link counts; in practice, the problem is degenerate, and no exact solution exists.

According to Cascetta and Nguyen [1], most popular O-D estimation models that utilise link flow counts can be formulated as a bi-level optimisation problem. In static traffic modelling contexts, depending on the form of the upper problem objective function, several methods were developed based on Bayesian statistics (Maher [2], Tebaldi and West [3] and Li [4]), generalised least squares (GLS) (Bell [5], Cascetta [6], Cascetta *et al.* [7] and Cascetta *et al.* [8]), maximum likelihood (ML) models (Spiess [9] and Cascetta and Nguyen [1]).

Marzano *et al.* [10] and Cascetta *et al.* [8] concluded that a satisfactory estimation of the O-D matrix, regardless of the quality of the prior estimate, can be obtained only when the number of equations (that equals the number of link measurements) is comparable to the number of unknowns (that equals the number of positive O-D flows). Both Marzano *et al.* [10] and Cascetta *et al.* [8] demonstrated the importance of a good seed matrix. Frederix *et al.* [11] underlined how these methods are very sensitive to prior information as well as to the chosen methodology.

Regarding the dynamic O-D matrix estimation, generally, the research contributions are categorised into offline and online contexts. The offline problem for congested networks is an extension of the static version and is thus usually formulated as a bi-level optimisation problem (Tavana [12], Van der Zijpp and Lindveld [13], Zhou *et al.* [14] and Zhou and Mahmassani [15]). In the upper level, time-dependent O-D matrices are corrected in order to replicate the observations, which can include speeds or travel times, while in the lower level, a dynamic traffic assignment (DTA) model is used to load the demand onto the network according to a user equilibrium or fixed-route choices (dynamic network loading). However, the resulting optimisation is a highly underdetermined, non-linear and non-convex problem (Antoniou *et al.* [16]). To solve this difficult problem, the most widely used algorithm in the literature is the simultaneous perturbation stochastic approximation (SPSA) (Balakrishna and Koutsopoulos [17], Cipriani *et al.* [18] and Frederix *et al.* [19]). To improve the performance of SPSA, various authors proposed modifications to reduce the problem dimensionality and its non-linearity or tested other derivative-free algorithms. The examples have been reported by Antoniou and co-authors [20], Lu *et al.* [20], Cantelmo *et al.* [21], Tympakianaki *et al.* [22], Cantelmo *et al.* [23], Kostic *et al.* [24, 25] and Ros-Roca *et al.* [26, 27].

Another category of dynamic O-D matrix estimation model assumes a form of evolution in the O-D flows across time slices, thus imposing a reduction in the number of unknowns. In this respect, a previous work by Cascetta *et al.* [8] proposed an application of a GLS estimator to offline quasi-dynamic contexts, while Marzano *et al.* [28] applied this approach to a Kalman filter estimator.

An emerging research field concerns the use of Big Data to describe users' mobility behaviours. Collecting data from strategically positioned sensors, from social networks or from other devices to detect the position, allows in principle to obtain complete information about the main traffic features, in order to evaluate and manage scenarios, assessing external effects such as safety risks and fuel consumptions (Astarita *et al.* [29], Astarita and Giofré [30] and Astarita *et al.* [31]), regulating traffic signal systems in real time (Astarita *et al.* [32, 33]) or representing the evolution of the rerouting phenomena in time and space when the information about an event spreads onto the network (Kucharski and Gentile [34, 35]).

Several researchers have tried to integrate additional information into the O-D demand estimation problem in terms of speed and density, turning fractions [36, 37] travel time and route choice [38–41].

More recent approaches, include as new data sources, automatic vehicle identification (AVI) point-to-point detection sensor systems (Avi-tags [15, 42, 43], license plate recognition data [36, 38, 44], bluetooth data [45–47], radio-frequency identification [48]) or AVI area-wide systems, such as GPS-based, cell phone tracking or floating car data (FCD) [49–54].

With regard to the FCD, Cipriani *et al.* [55] and Carrese *et al.* [40] show the benefit gained in the dynamic O-D matrix estimation using the information on the O-D zone of individually trip, their route choices and route travel times.

FCD has been adopted by Ásmundsdóttir [56], Ásmundsdóttir *et al.* [57] and Zhao *et al.* [58]. The first two contributions used FCD to obtain a-priori matrices and to correct them in the dynamic case also using information from the different historical databases. In addition, the FCD is used to analyse route choices and trip lengths. Zhao *et al.* [58] developed a practical approach to estimate time-varying O-D demand matrices incorporating both FCD and remote traffic microwave sensors (RTMSs) data. The RTMS data is used to obtain the static O-D demands. Then, the authors show a method to gain time-varying splitting rates combining FCD and RTMS data and subsequently apply them to control static O-D demands in order to derive the time-varying O-D matrices.

Yamamoto *et al.* [59] and Cao *et al.* [60] used observed link speed from FCD to estimate link flows, based on a Bayesian inference approach. Cao *et al.* [60], in particular, proposed a two-step framework incorporate probe vehicle data: in the first phase, link flows are estimated based only on observed link speed and a macroscopic speed–density relationship; in the second phase, a bi-level GLS estimator is formulated to estimate O-D flows.

Instead, in Vogt *et al.* [61], FCD trajectories are used to determine the number of turns at the intersections on the one hand and to determine the a-priori matrix on the other hand. The turning volume is applied in the information minimisation model to increase the quality of the estimated O-D demand.

A recent approach by Yang *et al.* [62] proposed a method that corrects an a-priori O-D matrix obtained by aggregating the origin and destination zones of each probe vehicle trace, based on link counts and probe ratios. An important key concept proposed in Yang *et al.* [62] is the estimation of the assignment matrix directly from the sampled trajectory, under specific modelling assumptions, avoiding sophisticated traffic assignment computations.

In recent work, Krishnakumari *et al.* [41] suggested a data-driven method to estimate dynamic O-D matrices introducing an empirical estimate of assignment matrices, obtained assuming that the proportion of each O-D flow and the paths is inversely proportional to the observed travel time along the paths.

A major challenge of conventional O-D demand estimation models is the computation of traffic assignment, facing the issue of solving (dynamic) user equilibrium problems with a sufficient level of precision through rapidly convergent and stable algorithms [63, 64].

The present research finds, therefore, its motivations in the potential of FCD and in the opportunity to process a big amount of information by specialised tools. Our main goal is to estimate a base O-D demand for a large-scale network, using only a data set of observed vehicle trajectories and traffic counts.

The proposed approach considers the revealed O-D matrix obtained from the FCD trajectories as the seed of the entire estimation process, even if it is known that the penetration rate of FCDs could be poor, not homogeneous among different couples, and anyhow not known in advance. In other words, a critical point we must never forget is that the FCD sample can be biased and thus not all mobility patterns are captured in the right proportions, timing and spatial configurations. However, this does not mean that it is impossible to use such data in a meaningful way.

The recent literature has highlighted how the inclusion of FCD samples in the O-D demand estimation problem positively impacts the reliability of the result [62, 65], also proving that travel times and route choice probabilities derived from FCD are effective when added as traffic measurements in the optimisation problem [47, 66].

Our target is to produce an O-D matrix by using FCD and a training set of traffic counts to scale the result. Our approach is shown in this paper with reference to the static case, but it can be generalised to the dynamic case. Indeed, the O-D matrices obtained for a sequence of time intervals can represent a reliable and sound seed to be subsequently refined by a dynamic O-D estimation model, as proposed by Yang and Rakha [67].

More specifically, we extracted from the FCD seed matrices profiled for each hour and revealed O-D paths for the same intervals. For each of the 24-hourly time slices, we can take the revealed O-D matrix and load it on the corresponding hourly paths using their specific shares and so replacing any assignment model with the observed assignment map. This way, we are assuming that those paths are representative of the mobility under stationary conditions in the considered hourly time slice, so neglecting the fact that trips along the network may intersect several intervals.

Also, as for the O-D seed matrix, the observed route choices can be biased because the FCD sample can be just a subset not distributed on the network in space a time accordingly to the real overall mobility distribution.

It is worth to underline again that the proposed methodology is applied in a static setting. Even if this is a strong limitation and the definitive solution to the O-D matrix calibration must be dynamic, we decided to reduce the complexity of the proposed methodology, leaving out the dynamicity in order to focus on the key point for which we can simplify the calibration process removing the equilibrium loop and replacing it with observed path choices.

The present research treats the O-D matrix estimation as a scaling problem of the observed FCD demand. Different optimisation-based solutions are performed in order to find the best set of multiplier factors. The achieved results are analysed and discussed, both in terms of the capacity of the model to fit traffic measurements used during the training phase, and its performance against a validation set.

The paper is organised as follows. Section 2 presents the general framework. Section 3 provides a discussion related to problem formulations and optimisation algorithms. Section 4 shows the experimental results. Section 5 presents our concluding remarks.

General framework: The proposed general framework consists of three different steps, depicted in Fig. 1: data-driven modelling, assignment matrix estimation and O-D matrix correction. The first two steps are preparatory to the last part, which, instead, represents the focus of this research. Anyhow a brief description of the first two steps is reported for a better understanding of the entire process.

1.1 Data-driven modelling

A data-driven approach is used to get the maximum amount of information from GPS trajectories (a collection of sequences of GPS points). The entire process is fully described under the USA patent number ‘US 10,636,294 B1’. Here we report some main points of the process that devises a dedicated paper in the future.

This process requires as inputs a graph, defined as usual in terms of nodes and directed links connecting them, a zoning and the definition of both day-types, to group all together days for which the average traffic behaviour is similar, and a time window resolution for profiling the output results of the process itself.

The preliminary step is to map-match each single raw GPS trajectory on the transportation graph in order to reconstruct the most probable path on the graph. This is achieved by specialised map-matching algorithms capable of handling high and low sampling rate GPS trajectories [68–70]. The map-matched trajectories are associated with a specific departure zone at a given

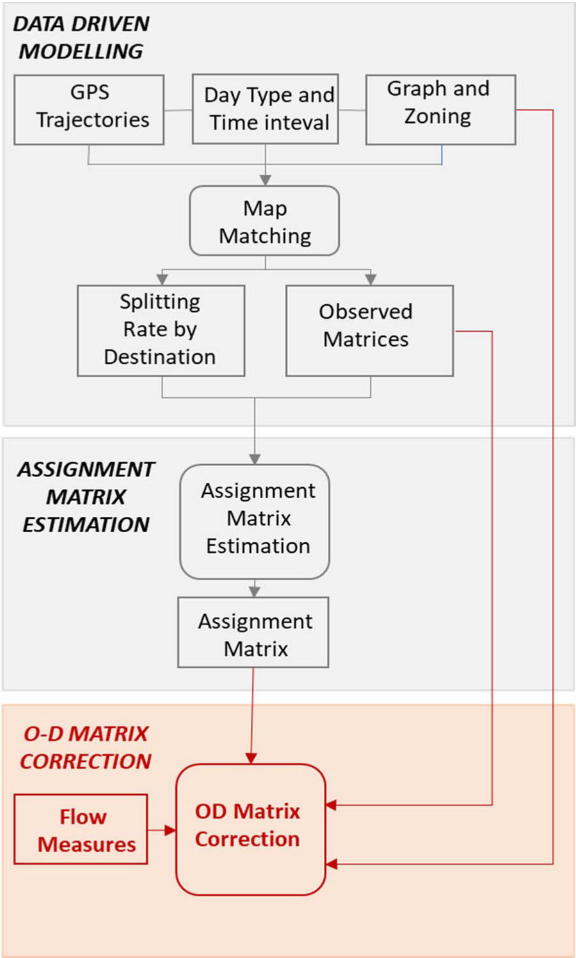

Fig. 1 *General framework*

time of day and to a given destination zone. They allow to compute turn probabilities by destination at every intersection to describe route choice behaviour better and thus reconstruct aggregated paths over the network by using the local route choice destination-based. The specific algorithm used to reconstruct the most used paths connecting an origin zone to a destination zone at a given departure time will be detailed in a future paper.

All outputs are profiled by day-types, transportation modes and time interval resolution. The detailed description of the individual outputs will be described in the next sections.

1.1.1 Observed O-D matrices: Given zoning, each trajectory of a probe vehicle is associated with an origin zone and a destination zone, if it intercepts at least one zone of the study area. Therefore, the departure rate from origin to destination can be reconstructed aggregating data by day-types and within daytime accordingly to the time interval resolution. The contribution of each FCD trajectory to an element of the observed O-D matrix is based on the entry time of the first GPS point into the origin zone.

1.1.2 Turn probabilities and destination-based turn probabilities: One of the most valuable information that can be extracted from FCDs is the turn probabilities, i.e. at a given intersection, the percentage of people taking a manoeuvre. The turn probabilities represent the people route choice behaviour, and usually, they are an output of an equilibrium assignment. The data-driven approach aims to provide turn probabilities as they are effectively observed through the sample.

More informative results on local route choices can be produced if, for each trajectory,, the destination zone is saved, indeed; in such a situation, it is possible to create destination-based turn probabilities. These ones allow the reconstruction of the aggregated explicit paths, as will be described in Section 1.1.4.

1.1.3 Attraction and generation shares: The turn probabilities corresponding to the exit flow from the network will be called attraction shares and they are used to create connectors through which the demand flows exit the network. As flows are admitted disappearing, symmetrically, they are created too. For all links, where there is at least one starting trajectory, a share can be assigned, i.e. the ratio between the total number of trajectories starting on the link itself and the total number of trajectories originated in the considered zone.

Generation and attraction shares are used in conjunction with the turn probabilities for the explicit path reconstruction, but they could be used in a transportation model having in mind the idea to remove the connectors and to load the demand matrices on the links where generation shares are not zero, and making the flows going out of the network in correspondence of the attraction shares. In this way, it is possible to get an implicit definition of the connectors to spread all over the network.

1.1.4 Explicit paths: Route choice behaviour can be expressed either by local choices at intersections as turn probabilities or by explicit paths, having a weight representing the probability to choose it.

For each O-D couple, the data-driven approach reconstructs the most used paths from the destination-based turn probabilities, via an explorative algorithm that from the origin reach the destination following the observed manoeuvres. The algorithm assigns all the paths a probability to be chosen, and the paths with the higher probability above a desired threshold are returned.

1.2 Assignment matrix estimation

The second step of the entire procedure is aimed at the definition of an observed assignment matrix expressing formally the fraction of demand using each arc.

As stated earlier, FCDs processing allows us to determine the turn probabilities at intersections by the destination and the generation and attraction shares in traffic zones. These local outputs are referred to predefined time intervals and allow to reconstruct aggregated path shares over the network.

Let $\boldsymbol{d}$ denote the unknown O-D vector containing $I=U\times T$ elements, where U is the number of O-D pairs and T represents the number of (discretised) departure time periods. Let $\boldsymbol{Y}$ be the vector with traffic counts containing $\Omega=A\times T$ elements, with A is the number of sensors and T is the number of time periods in which traffic counts are collected.

According to Cascetta [71], the relationship between the O-D flows and the link flows is described by the assignment matrix $\boldsymbol{M}$, which has dimensions $\Omega\times I$. This matrix can be further subdivided into a crossing fraction matrix $\boldsymbol{B}$ with dimensions $I\times K$ (K is the number of route flows) and a route fraction matrix $\boldsymbol{P}$, which has dimensions $K\times I$. The elements of crossing fraction matrix $\boldsymbol{B}$ express the proportion of a route flow that passes a link, thus describing the spatial–temporal propagation of the route flows throughout the network. The elements of route fraction matrix $\boldsymbol{P}$ express the proportion of an O-D flow choosing a certain route. There is a dependence of the crossing fraction matrix $\boldsymbol{B}$ on the O-D flow and, moreover, a dependence of the route fraction matrix $\boldsymbol{P}$ on the O-D flows. Let $\boldsymbol{\Delta}$ be the link-paths incidence matrix, using matrix notation; the observed assignment matrix can be estimated as

$$\boldsymbol{M} = \boldsymbol{\Delta} \times \boldsymbol{P} \quad (1)$$

The presented methodology assumes that the whole journey between an O-D pair is started and completed in the same time period. Under this assumption, the crossing fraction matrix $\boldsymbol{B}$ is equal to the link-path incidence matrix and is not a function of the demand flows. In other words, it is accepted the approximation of computing such path shares separately with reference to each time interval, neglecting the fact that trips along the network may intersect several time intervals. Our research is addressing the matrix scaling considering this quasi-static assumption, but it can be generalised to a fully dynamic case too. However, most important, while in standard formulations of the demand calibration problem, the path probabilities and the arc-path shares are given by the route choice as obtained by an assignment model, here we take the explicit paths reconstructed via the FCD probes as input.

1.3 O-D matrix correction

The last step of the methodology focuses on estimating O-D flows by combining flow measures observed an assignment matrix and a seed demand.

The classical formulation of an O-D matrix correction problem, which is usually referred to as demand adjustment in the literature, can be formulated as a minimisation problem and the objective function can be formally expressed as the sum of two functions z_1 and z_2, considered as different 'performance measures': z_1 measures the 'distance' of the unknown demand from the reference demand and z_2 measures the 'distance' of the flows obtained by assigning the unknown demand to the network from the traffic measures.

Compared to the classical formulation, we propose a model where the unknown demand is a function of design variables that represent the 'scaling factors' of the reference demand, as defined in Section 2. We search the optimum values of such variables in order to define a new demand vector, which, once assigned to the network, produces the vector flows closest to the vector counts by do not alter too much the reference demand.

We assume that the reference demand corresponds to the observed O-D matrix, described previously in Sections 1.1, and the assignment relationship is replaced by an observed assignment matrix, described in Sections 1.2.

The proposed model is ascribed to a least-squares problem and the general formulation is, thus, expressed as

$$\underset{x \geq 0}{\mathrm{Min}}\, \varphi(\boldsymbol{x}) = \frac{1}{2} \cdot \sum_{i=1}^{m} z_i(\boldsymbol{x})^2 \tag{2}$$

where the performance measure $\boldsymbol{z} = (z_1(\boldsymbol{x}), \ldots, z_m(\boldsymbol{x}))$ represents the vector of m function often called 'residuals' and $\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$ is the vector of n design variables that represent the scaling components of demand.

2 Problem formulation

Consider a road network represented by a direct graph $\langle N, A \rangle$, where N is the set of nodes and $A \subseteq N \times N$ is the set of links. Some of the nodes $o \subseteq N$ are trip origins and some $d \subseteq N$ are trip destinations. The main variables of the proposed model are

d^u_{odt} demand flow of class $u \in U$ users departing during time interval $t \in T$, from origin $o \in N$ to destination $d \in N$.
q^v_{at} traffic variable of type $v \in V$ (e.g. flow, speed, density) at arc $a \in A$ during interval $t \in T$.
$\bar{q}^v_{at}$ traffic measurement $(v, a, t) \in \Omega \subseteq J = V \times A \times T$.
x_p value of the generic demand parameters $p \in P$ (e.g. generation, attraction etc.)
λ the relative weight of reference demand.

Time is discretised into a set T of intervals; because in our quasi-static approach, all the elaborations are referred to a fixed time interval and they do not either depend from or affect other time intervals, we will omit the explicit time interval to simplify the notation.

The assignment operator $q(\cdot)$ loads on the network a vector of demand components $\boldsymbol{d}$ to obtain a vector of traffic variables q:

$$\boldsymbol{q} = q(\boldsymbol{d}) \tag{3}$$

Usually, not all arcs have detected measurements, then the relevant components of the assignment operator are obtained by projection on the measurement set Ω:

$$\boldsymbol{q}_\Omega = q_\Omega(\boldsymbol{d}) \tag{4}$$

In this context, the objective of adjustment is not the demand itself, but rather its aggregation in some form; therefore, the single components $\boldsymbol{d} \in \Re^{+I}$ are obtained from the parameters $\boldsymbol{x} \in \Re^{+P}$ through a demand operator $d(\cdot)$

$$\boldsymbol{d} = d(\boldsymbol{x}) \tag{5}$$

The adjustment of the demand $\boldsymbol{d}$ aims at minimising the distance between the traffic provided by the assignment operator $q_\Omega(\boldsymbol{d})$ and the measurements $\bar{\boldsymbol{q}}_\Omega$ of a subset of arcs, considering a certain weight λ of the reference demand $\bar{\boldsymbol{d}}$. The adjustment is obtained through the scaling demand parameters by solving the following optimisation problem:

$$\underset{x \geq 0}{\mathrm{Min}}\left(\varphi(\boldsymbol{x}) = z(q_\Omega(d(\boldsymbol{x})), \bar{\boldsymbol{q}}_\Omega) + \lambda \cdot z\left(d(\boldsymbol{x}), \bar{\boldsymbol{d}}\right)\right) \tag{6}$$

We assume that the demand operator $d(\cdot)$ is defined as a function of the design variables, namely

$$d_{\mathrm{od}} = \gamma \cdot \boldsymbol{\alpha}_{\mathrm{o}} \cdot \boldsymbol{\beta}_{\mathrm{d}} \cdot \bar{d}_{\mathrm{od}} \tag{7}$$

where γ is a 'constant factor' and represents a homogeneous scaling factor among all O-D pairs; $\boldsymbol{\alpha}_{\mathrm{o}}$ and $\boldsymbol{\beta}_{\mathrm{d}}$, are the 'attraction factor' vector and the 'generation factor' vector, respectively; $\bar{d}_{\mathrm{od}}$ is the reference demand obtained by processing FCDs, as described in Section 1.1.1.

The assignment operator $q(\cdot)$ is considered to be the observed assignment matrix $\boldsymbol{M}$, described in Sections 1.2.

Hence the traffic variables q_Ω can be expressed as

$$q_\Omega(\boldsymbol{d}) = \boldsymbol{M} \cdot \boldsymbol{d} \tag{8}$$

and the O-D matrix correction problem (6) is formulated as

$$\mathrm{Min}\varphi(\boldsymbol{x}) = \frac{1}{2} \cdot \| \boldsymbol{M} \cdot d(\boldsymbol{x}) - \bar{\boldsymbol{q}}_\Omega \|^2 + \lambda \cdot \frac{1}{2} \cdot \| d(\boldsymbol{x}) - \bar{\boldsymbol{d}} \|^2 \tag{9}$$

Moreover, the gradient of the objective function can be calculated as the product of the transposed Jacobian of residuals and the residuals themselves

$$\nabla \varphi = \nabla z \cdot z = \boldsymbol{J}^T \cdot z \tag{10}$$

The problem has been solved by generating a sequence of feasible solutions $\{\boldsymbol{x}^k\}$, where at each iteration k, the new iterate $\boldsymbol{x}^{k+1}$ is obtained, starting from the current iterate $\boldsymbol{x}^k$, by taking a step $\alpha^k \in$ (0,1] along a feasible $(\boldsymbol{x}^k + \Delta \boldsymbol{x}^k \geq \boldsymbol{0})$ descent direction $\Delta \boldsymbol{x}^k$:

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \cdot \Delta \boldsymbol{x}^k \tag{11}$$

In order to find an optimal solution for the problems (9), we have settled for the Newton method among the different classes of optimisation algorithms. In the Newton method, one considers the second-order Taylor expansion of $\varphi(\boldsymbol{x})$ at the current $\boldsymbol{x}^k$ and its gradient

$$\begin{aligned} \varphi\left(\boldsymbol{x}^k + \Delta \boldsymbol{x}\right) &\simeq \varphi\left(\boldsymbol{x}^k\right) + \Delta \boldsymbol{x}^T \cdot \nabla \varphi\left(\boldsymbol{x}^k\right) \\ &+ \frac{1}{2} \cdot \Delta \boldsymbol{x}^T \cdot \nabla^2 \varphi\left(\boldsymbol{x}^k\right) \cdot \Delta \boldsymbol{x}, \end{aligned} \tag{12}$$

$$\nabla \varphi\left(\boldsymbol{x}^k + \Delta \boldsymbol{x}\right) \simeq \nabla \varphi\left(\boldsymbol{x}^k\right) + \nabla^2 \varphi\left(\boldsymbol{x}^k\right) \cdot \Delta \boldsymbol{x} \tag{13}$$

The minimisation of the quadratic expansion is attained where its gradient is null, which provides the descent direction $\Delta \boldsymbol{x}^k$ of the Newton algorithm

$$\nabla^2\varphi(\boldsymbol{x}^k)\cdot\Delta\boldsymbol{x}^k = -\nabla\varphi(\boldsymbol{x}^k) \quad (14)$$

To obtain the result in terms of $\Delta\boldsymbol{x}^k$ one must solve the above system of linear equations. In theory, computing the inverse of the Hessian, we have

$$\Delta\boldsymbol{x}^k = -\left(\nabla^2\varphi(\boldsymbol{x}^k)\right)^{-1}\cdot\nabla\varphi(\boldsymbol{x}^k) \quad (15)$$

In practice, inversion is never a good idea, because the dimension $n\times n$ of the Hessian may be large, and the matrix could be not invertible. Instead, the linear system is solved via an iterative method.

In the case of least squares, the Newton descent direction is obtained by solving

$$\left(\boldsymbol{J}^{kT}\cdot\boldsymbol{J}^k + \sum_{i=1}^{m}\boldsymbol{H}_i^k\cdot z_i(\boldsymbol{x}^k)\right)\cdot\Delta\boldsymbol{x}^k = -\boldsymbol{J}^{kT}\cdot z(\boldsymbol{x}^k) \quad (16)$$

Clearly, if the residuals are linear functions, then the solution is attained in one iteration for Step 1. If the Hessian is positive definite, then the shift of the Newton algorithm is a descent direction:

$$\nabla\varphi(\boldsymbol{x}^k)^T\cdot\Delta\boldsymbol{x}^k = -\nabla\varphi(\boldsymbol{x}^k)^T\cdot\left(\nabla^2\varphi(\boldsymbol{x}^k)\right)^{-1}\cdot\nabla\varphi(\boldsymbol{x}^k) < 0, \quad (17)$$

$$\nabla\varphi(\boldsymbol{x}^k) \neq \boldsymbol{0} \quad (18)$$

In the next subsections, we focus on four alternative models from the general context by assuming for each instance different design variables of the optimisation problem (9). The purpose of the first three approaches is optimising the 'scaling variables' γ, $\boldsymbol{\alpha}_{\rm o}$ and $\boldsymbol{\beta}_{\rm o}$ instead, the last proposed model carries out the correction of the single O-D pairs.

2.1 Demand scaling using a constant factor γ

The first formulation is aimed at finding the optimum constant factor γ in order to perform a 'global scaling' of the reference matrix. The problem formulation referred to as DSCF (stands for demand scaling using a constant factor) can be formulated as

$$\begin{aligned}&\min_{\gamma\geq 0}\varphi\\&\boldsymbol{\alpha}_{\rm o} = \boldsymbol{\beta}_{\rm d} = 1\end{aligned} \quad (19)$$

The formulations (19) constitute a quadratic problem, and the closed-form solution is expressed as

$$\gamma * = \frac{\sum_{a\in\bar{A}}\bar{q}_a\cdot\dfrac{\partial q_a}{\partial\gamma}}{\sum_{a\in\bar{A}}\left(\dfrac{\partial q_a}{\partial\gamma}\right)^2} \quad (20)$$

2.2 Demand scaling using different attraction and generation factors $\boldsymbol{\alpha}_o$ and $\boldsymbol{\beta}_d$

The second formulation proposed in this study provides a re-proportioning method of O-D matrix, carried out using different factors for each O-D pair.

The design variables $\boldsymbol{\alpha}_{\rm o}$ and $\boldsymbol{\beta}_{\rm o}$ are obtained by two different optimisation-based approaches. The first approach considers the minimisation of objective function respect to $\boldsymbol{\alpha}_{\rm o}$ and $\boldsymbol{\beta}_{\rm o}$ jointly; on the second one, instead, the objective function is minimised respect to one of the variables, maintaining the other constant, according to an iterative process.

(i) $\boldsymbol{\alpha}_{\rm o}$ *and* $\boldsymbol{\beta}_{\rm o}$ *joint optimisation (ABJO)*: the problem formulation can be expressed as

Fig. 2 *Test site: road network of Turin*

$$\begin{aligned}&\min_{\boldsymbol{\alpha}_{\rm o},\boldsymbol{\beta}_{\rm d}\geq 0}\varphi\\&\boldsymbol{\alpha}_{\rm o},\boldsymbol{\beta}_{\rm d}\geq 0\\&\gamma\geq 0\end{aligned} \quad (21)$$

(ii) $\boldsymbol{\alpha}_{\rm o}$ *and* $\boldsymbol{\beta}_{\rm o}$ *iterative optimisation (ABIO)*: the problem formulation can be expressed as

$$\begin{aligned}&\min_{\boldsymbol{\alpha}_{\rm o}(\boldsymbol{\beta}_{\rm d}={\rm cos}t)}\varphi\\&\min_{\boldsymbol{\beta}_{\rm d}(\boldsymbol{\alpha}_{\rm o}={\rm cos}t)}\varphi\\&\boldsymbol{\alpha}_{\rm o},\boldsymbol{\beta}_{\rm d}\geq 0\\&\gamma\geq 0\end{aligned} \quad (22)$$

2.3 Single O-D pairs optimisation (SPO)

The latter problem formulation is not different than the classical O-D estimation problem, where the $N\times N$ elements of the O-D matrix are the variables to be optimised. Thus, the optimisation formulation is given as

$$\begin{aligned}&\min_{d\geq 0}\varphi\\&\boldsymbol{\alpha}_{\rm o} = \boldsymbol{\beta}_{\rm d} = 1\\&\gamma\geq 0\end{aligned} \quad (23)$$

3 Numerical example

To evaluate the effectiveness of the proposed models, the road network of Turin was selected as a test site.

The network with its count locations is presented in Fig. 2. It is a large-scale network divided into 438 zones. The number of links is 96,420 connecting 6532 nodes. The referenced time interval includes the morning period of 1 h (07:00–08:00 am).

Traffic data are obtained through 2.064.656 records, referred to the year 2016, afterwards aggregated by working days and referenced time interval. The total number of used count locations was 1203 (red triangles in Fig. 2), of which the 80% of these measurements were used in calibration procedure and whereas the remaining 20% were used in the validation process; specifically, 963 links have been used for the calibration procedure and 240 for validation.

About the FCD dataset, roughly 10^9 GPS points, referred to the year 2016, have been processed and aggregated by working days and referenced time interval (07:00–08:00 am). The FCDs processing outputs 402,628 observed assignment paths and one observed O-D matrix for the just specified day aggregation and time of the day. Note that the FCD data are totally different from the count data being the last ones coming from loop detectors.

For each O-D pair, the proportion of observed versus total vehicle population within the same time interval is low, estimated

Table 1 RMSE comparison among all optimisation procedures in both calibration and validation cases, and varying the parameter $\boldsymbol{\lambda}$

Problem OPT.	Λ	$RMSE_{CAL}$	$RMSE_{VAL}$
ABJO	0	138	360
ABIO	0	93	1460
SPO	0	49	176
ABJO	10^{-5}	141	311
ABIO	10^{-5}	103	644
SPO	10^{-5}	49	161
ABJO	10^{-4}	146	238
ABIO	10^{-4}	121	226
SPO	10^{-4}	52	159
ABJO	10^{-3}	166	179
ABIO	10^{-3}	151	181
SPO	10^{-3}	64	154
ABJO	10^{-1}	267	192
ABIO	10^{-1}	258	195
SPO	10^{-1}	150	152
ABJO	10^{-0}	335	203
ABIO	10^{-0}	—	—
SPO	10^{-0}	242	178

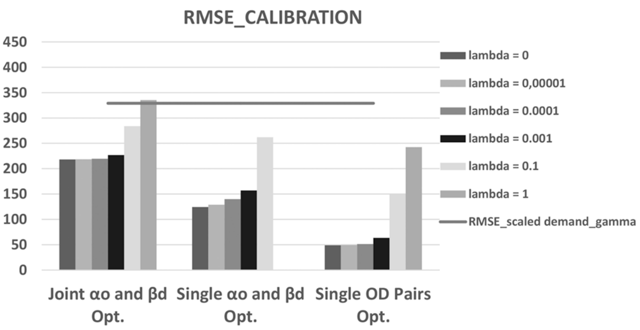

Fig. 3 *RMSE for the calibration results set*

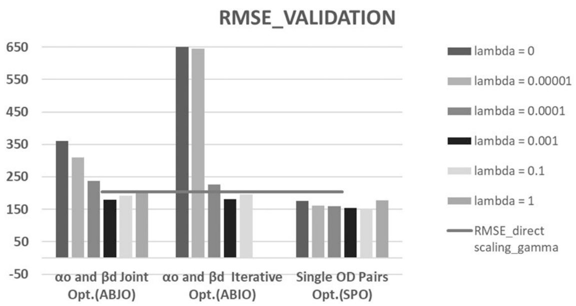

Fig. 4 *RMSE for the validation results set*

to be around 2% maximum of the total mobility. Due to the low penetration rate of FCD probes, the observed O-D matrix from FCD trajectories does not comprise sufficient information to serve as a good a-priori matrix for allowing a convergence of the optimisation procedures; therefore, it is necessary to scale up it. For this purpose, the estimated demand carried out from the 'direct scaling', using a constant factor γ^*, according to the formulations (19) and (20), represents not only the first crude estimation but also provides a reasonable starting point for the optimisation algorithm of the ABJO, ABIO and SPO problem formulations. In other words, the ABJO, ABIO and SPO formulations represent a further adjustment of the 'direct-scaled' O-D matrix and, therefore, the optimisation problems (21)–(23) have been formulated for $\gamma = \gamma *$.

A sensitivity analysis is performed against the λ parameter, varying it from 0 to 1, i.e. moving from being free to destroy at the maximum extent the structure of the seed matrix up to be as much conservative as possible, leaving the seed matrix structure roughly unchanged. The value used for λ will be the following: (10^{-5}; 10^{-4}; 10^{-3}; 10^{-1}; 1).

In order to evaluate the goodness of the models, we will use it as a key performance indicator (KPI), the RMSE between the flow values assigned on the links via the assignment map we observed thorough the FCD and the flow data measured on the count sections. The performance indicators are evaluated for both the calibration set and the validation set, and they are expressed as

$$\mathrm{RMSE_{CAL}} = \sqrt{\frac{\sum_{a \in \Omega'} (q_{\mathrm{od}} - \bar{q}_{\mathrm{od}})^2}{|\Omega'|}} \tag{24}$$

where a belongs to the sample $\Omega' \in \Omega$ representing the set of links used for the calibration procedure, and

$$\mathrm{RMSE_{VAL}} = \sqrt{\frac{\sum_{a \in \Omega''} (q_{\mathrm{od}} - \bar{q}_{\mathrm{od}})^2}{|\Omega''|}} \tag{25}$$

where a belongs to the sample $\Omega'' \in \Omega$ representing the set of links used for the validation procedure.

Such KPIs are reported in Table 1 for each optimisation problem formulations and for each value of the parameter λ.

Figs. 3 and 4 summarise the results obtained.

The horizontal lines depicted in Figs. 3 and 4 show the RMSE obtained just scaling up the observed O-D matrix with the constant factor γ^*.

3.1 Results interpretation

Looking at the KPI results reported in Table 1 and Figs. 3 and 4, it is evident that the performances of all models, regardless of the value of λ, are better in the calibration phase compared to the validation one. This is an expected output in general because all the models are optimised explicitly against the calibration data set and so their robustness with respect another data set is not granted at all, especially if, from a statistical point of view, the calibration data contains only a sub-sampling of the variety of the features of the entire data population. All the optimisation procedures improve the simple O-D overall scaling, being the SPO the best one. This remark is valid for each level of the λ parameter. Moreover, the sensitivity analysis with respect to λ shows that the lower the value, the better it is, but moving on the validation data set, such statement must be changed. Indeed, while for the calibration case, the optimisation models tend to overfit the data. On the validation set, the mitigation of such overfitting is achieved by fine-tuning the λ value reaching a good tradeoff where the validation is the best possible even if the calibration is not perfect. The best value of λ in such an experiment is equal to 10^{-3}. It is interesting to note that while for the ABJO and ABIO models, the RMSE for the calibration and validation at the λ=10^{-3} are roughly the same, for the SPO, the KPI in calibration is around three times better than the validation. Despite the SPO results are anyhow better than the ABJO and ABIO ones, this situation is an indicator of the fact that the SPO approach is less stable than the other two; and this can be imputed to the fact that the number of unknowns over which the SPO method is working is much higher than the other cases and so the optimisation space of the problem is less constrained and so able to accommodate in a more wider way the observational data.

4 Conclusion

This paper account for the development of an O-D matrix estimation methodology using as main input data the vehicle trajectories are coming from FCDs. The proposed methodology does not require the availability of a-priori matrix and an assignment model.

Four alternative optimisation approaches, with a different number of design variables, were tested and compared. The first main observation is that a direct scaling of the observed O-D matrix, using a constant factor for the whole matrix, is an essential precondition for all optimisation problems. The results indicate that every optimisation problem tends to overfit the training data, but it's possible to improve the simple overall scaling estimation by applying a further optimisation step.

The best performance is achieved by estimating the single O-D pairs; nevertheless, the different approaches tested led to similar results. Finally, regardless of the optimisation procedure, the best solution is obtained considering an intermediate value for λ in the objective function, i.e. it is convenient to search a final solution that does not yield a significant distortion from the seed matrix.

As for future research, the proposed approach can be generalised to a dynamic setting, and a comparative analysis against classical assignment-based methods is expected. More tests are also needed on other real data sets and planned for future research activities.

5 References

[1] Cascetta, E., Nguyen, S.: 'A unified framework for estimating or updating origin/destination matrices from traffic counts', *Transp. Res. B*, 1988, **22**, pp. 437–455

[2] Maher, M.J.: 'Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach', *Transp. Res. B*, 1983, **17**, pp. 435–447

[3] Tebaldi, C., West, M.: 'Bayesian inference on network traffic using link count data', *J. Am. Stat. Assoc.*, 1998, **93**, p. 576

[4] Li, B.: 'Bayesian inference for origin-destination matrices of transport networks using the EM algorithm', *Technometrics*, 2005, **47**, pp. 399–408

[5] Bell, M.G. H.: 'The estimation of an origin-destination matrix from traffic counts', *Transp. Sci.*, 1983, **17**, pp. 198–217

[6] Cascetta, E.: 'Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator', *Transp. Res. B*, 1984, **18**, pp. 289–299

[7] Cascetta, E., Inaudi, D., Marquis, G.: 'Dynamic estimators of origin-destination matrices using traffic counts', *Transp. Sci.*, 1993, **27**, pp. 363–373

[8] Cascetta, E., Papola, A., Marzano, V., *et al.*: 'Quasi-dynamic estimation of o–d flows from traffic counts: formulation, statistical validation and performance analysis on real data', *Transp. Res. B Methodol.*, 2013, **55**, pp. 171–187

[9] Spiess, H.: 'A maximum likelihood model for estimating origin-destination matrices', *Transp. Res. B*, 1987, **21**, pp. 395–412

[10] Marzano, V., Papola, A., Simonelli, F.: 'Limits and perspectives of effective O-D matrix correction using traffic counts', *Transp. Res. C Emerg. Technol.*, 2009, **17**, pp. 120–132

[11] Frederix, R., Viti, F., Tampère, C.M. J.: 'Dynamic origin–destination estimation in congested networks: theoretical findings and implications in practice', *Transp. A Transp. Sci.*, 2013, **9**, (6), pp. 494–513

[12] Tavana, H.: 'Internally-consistent estimation of dynamic network origin–destination flows from intelligent transportation systems data using bi-level optimization', 2001, Ph.D. Dissertation, University of Texas, Austin, USA

[13] Van der Zijpp, N.J., Lindveld, C.D.R.: 'Estimation of origin-destination demand for dynamic assignment with simultaneous route and departure time choice', *Transp. Res. Rec.*, 2001, **1771**, pp. 75–82

[14] Zhou, X., Qin, X., Mahmassani, H.S.: 'Dynamic origin-destination demand estimation with multiday link traffic counts for planning applications', *Transp. Res. Rec. J. Transp. Res. Board*, 2003, **1831**, (1), pp. 30–38

[15] Zhou, X., Mahmassani, H.S.: 'Dynamic origin–destination demand estimation using automatic vehicle identification data', *IEEE Trans. Intell. Transp. Syst.*, 2006, **7**, (1), pp. 105–114

[16] Antoniou, C., Barceló, J., Breen, M., *et al.* 'Towards a generic benchmarking platform for origin-destination flows estimation/updating algorithms: design, demonstration and validation', *Transp. Res. C Emerg. Technol.*, 2016, **66**, pp. 79–98 doi: 10.1016/j.trc.2015.08.009

[17] Balakrishna, R., Koutsopoulos, H.N.: 'Incorporating within-day transitions in simultaneous offline estimation of dynamic origin-destination flows without assignment matrices', *Transp. Res. Rec.*, 2008, **2085**, (1), pp. 31–38, https://journals.sagepub.com/toc/trra/2085/1

[18] Cipriani, E., Florian, M., Mahut, M., *et al.*: 'A gradient approximation approach for adjusting temporal origin-destination matrices', *Transp. Res. C Emerg. Technol.*, 2011, **19**, pp. 270–282

[19] Frederix, R., Viti, F., Corthout, R., *et al.*: 'New gradient approximation method for dynamic origin-destination matrix estimation on congested networks', *Transp. Res. Rec.*, 2011, **2263**, (1), pp. 1543–1548, https://journals.sagepub.com/toc/trra/2263/1

[20] Lu, L., Xu, Y., Antoniou, C., *et al.*: 'An enhanced SPSA algorithm for the calibration of dynamic traffic assignment models', *Transp. Res. C Emerg. Technol.*, 2015, **51**, pp. 149–166

[21] Cantelmo, G., Cipriani, E., Gemma, A., *et al.*: 'An adaptive Bi-level gradient procedure for the estimation of dynamic traffic demand', *IEEE Trans. Intell. Transp. Syst.*, 2014, **15**, pp. 1348–1361

[22] Tympakianaki, A., Koutsopoulos, H.N., Jenelius, E.: 'c-SPSA: cluster-wise simultaneous perturbation stochastic approximation algorithm and its application to dynamic origin-destination matrix estimation', *Transp. Res. C Emerg. Technol.*, 2015, **55**, pp. 231–245

[23] Cantelmo, G., Viti, F., Cipriani, E., *et al.*: 'A utility-based dynamic demand estimation model that explicitly accounts for activity scheduling and duration', *Transp. Res. A Policy Pract.*, 2018, **114**, pp. 303–320

[24] Kostic, B., Meschini, L., Gentile, G.: 'Calibration of the demand structure for dynamic traffic assignment using flow and speed data: exploiting the advantage of distributed computing in derivative-free optimization algorithms', *Transp. Res. Procedia*, 2017, **27**, pp. 993–1000

[25] Kostic, B., Annunziata, A., Gentile, G., *et al.*: 'A sequential approach to time-dependent demand calibration: application, validation and practical implications for large-scale networks'. 2017 5th IEEE Int. Conf. on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Naples, Italy, 2017, pp. 362–367

[26] Ros-Roca, X., Montero, L., Schneck, A., *et al.*: 'Investigating the performance of SPSA in simulation-optimization approaches to transportation problems', *Transp. Res. Procedia*, 2018, **34**, pp. 83–90

[27] Ros-Roca, X., Montero, L., Barceló, J.: 'Investigating the quality of Spiess-like and SPSA approaches for dynamic OD matrix estimation', *Transp. A Transp. Sci.*, 2020, pp. 1–23, https://doi.org/10.1080/23249935.2020.1722282

[28] Marzano, V., Papola, A., Simonelli, F., *et al.*: 'A Kalman filter for quasi-dynamic o-d flow estimation/updating', *IEEE Trans. Intell. Transp. Syst.*, 2018, **19**, (11), pp. 3604–3612

[29] Astarita, V., Guido, G., Mongelli, D., *et al.*: 'A co-operative methodology to estimate car fuel consumption by using smartphone sensors', *Transport*, 2015, **30**, pp. 307–311

[30] Astarita, V., Giofré, V.P.: 'From traffic conflict simulation to traffic crash simulation: introducing traffic safety indicators based on the explicit simulation of potential driver errors', *Simul. Model. Pract. Theory*, 2019, **94**, pp. 215–236

[31] Astarita, V., Festa, D.C., Giofrè, V.P., *et al.*: 'Bluetooth portal-based system to measure the performance of building emergency evacuation plans and drills', *IET Intell. Transp. Syst.*, 2018, **12**, pp. 294–300

[32] Astarita, V., Giofrè, V.P., Guido, G., *et al.*: 'A single intersection cooperative-competitive paradigm in real time traffic signal settings based on floating car data', *Energies*, 2019, **12**, (3), p. 409

[33] Astarita, V., Giofrè, V.P., Guido, G., *et al.*: 'The use of adaptive traffic signal systems based on floating car data', *Wirel. Commun. Mob. Comput.*, 2017, **2017**, p. 4617451

[34] Kucharski, R., Gentile, G.: 'Simulation of rerouting phenomena in dynamic traffic assignment with the information comply model', *Transp. Res. B Methodol.*, 2019, **126**, pp. 414–441

[35] Kucharski, R., Gentile, G.: 'Observing rerouting phenomena in dynamic traffic networks'. 2015 Int. Conf. on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Budapest, Hungary, 2015, pp. 140–147

[36] Van Der Zijpp, N.J.: 'Dynamic origin-destination matrix estimation from traffic counts and automated vehicle identification data', *Transp. Res. Rec.*, 1997, **1607**, pp. 87–94

[37] Mishalani, R.G., Coifman, B., Gopalakrishna, D.: 'Evaluating real-time origin-destination flow estimation using remote sensing-based surveillance data'. Proc. of the Int. Conf. on Applications of Advanced Technologies in Transportation Engineering, Boston, Massachusetts, 2002

[38] Asakura, Y., Hato, E., Kashiwadani, M.: 'Origin-destination matrices estimation model using automatic vehicle identification data and its application to the Han-Shin expressway network', *Transportation (Amst)*, 2000, **27**, (4), pp. 419–438

[39] Sun, J., Feng, Y.: 'A novel OD estimation method based on automatic vehicle identification data', *Commun. Comput. Inf. Sci.*, 2011, **135**, pp. 461–470

[40] Carrese, S., Cipriani, E., Mannini, L., *et al.*: 'Dynamic demand estimation and prediction for traffic urban networks adopting new data sources', *Transp. Res. C Emerg. Technol.*, 2017, **81**, pp. 83–98

[41] Krishnakumari, P., van Lint, H., Djukic, T., *et al.*: 'A data driven method for OD matrix estimation', *Transp. Res. C Emerg. Technol.*, 2019, **38**, pp. 139–159

[42] Dixon, M.P., Rilett, L.R.: 'Real-time OD estimation using automatic vehicle identification and traffic count data', *Comput. Civ. Infrastruct. Eng.*, 2002, **17**, (1), pp. 7–21

[43] Antoniou, C., Ben-Akiva, M., Koutsopoulos, H.N.: 'Dynamic traffic demand prediction using conventional and emerging data sources', *IEE Proc. - Intell. Transp. Syst.*, 2006, **153**, (1), p. 97

[44] Rostami Nasab, M., Shafahi, Y.: 'Estimation of origin–destination matrices using link counts and partial path data', *Transportation (Amst)*, 2019, pp. 1–28, https://doi.org/10.1007/s11116-019-09999-1

[45] Barceló, J., Montero, L., Bullejos, M., *et al.*: 'A Kalman filter approach for exploiting bluetooth traffic data when estimating time-dependent od matrices', *J. Intell. Transp. Syst. Technol. Plann., Oper.*, 2013, **17**, pp. 123–141

[46] Michau, G., Pustelnik, N., Borgnat, P., *et al.*: 'Combining traffic counts and bluetooth data for link-origin-destination matrix estimation in large urban networks: the Brisbane case study: a preprint', 2019, pp. 1–15, https://arxiv.org/https://arxiv.org/abs/1907.07495

[47] Behara, K., Bhaskar, A., Chung, E.: 'A novel methodology to assimilate sub-path flows in bi-level OD matrix estimation process', *IEEE Trans. Intell. Transp. Syst.*, 2020, pp. 1–11, https://doi.org/10.1109/TITS.2020.2998475

[48] Guo, J., Liu, Y., Li, X., *et al.*: 'Enhanced least square based dynamic OD matrix estimation using radio frequency identification data', *Math. Comput. Simul.*, 2019, **155**, pp. 27–40

[49] Iqbal, M.S., Choudhury, C.F., Wang, P., *et al.*: 'Development of origin-destination matrices using mobile phone call data', *Transp. Res. C Emerg. Technol.*, 2014, **40**, pp. 63–74

[50] Calabrese, F., Di Lorenzo, G., Liu, L., *et al.*: 'Estimating origin-destination flows using mobile phone location data', *IEEE Pervasive Comput.*, 2011, **10**, pp. 36–44

[51] Ge, Q., Fukuda, D.: 'Updating origin–destination matrices with aggregated data of GPS traces', *Transp. Res. C Emerg. Technol.*, 2016, **69**, pp. 291–312

[52] Sohn, K., Kim, D.: 'Dynamic origin-destination flow estimation using cellular communication system', *IEEE Trans. Veh. Technol.*, 2008, **57**, pp. 2703–2713

[53] Alexander, L., Jiang, S., Murga, M., *et al.*: 'Origin-destination trips by purpose and time of day inferred from mobile phone data', *Transp. Res. C Emerg. Technol.*, 2015, **58**, pp. 240–250

[54] Zin, T.A., Kyaing, K., Lwin, K., *et al.*: 'Estimation of originating-destination trips in Yangon by using big data source', *J. Disaster Res.*, 2018, **13**, (1), pp. 6–13

[55] Cipriani, E., Del Giudice, A., Marialisa, N., *et al.*: 'The impact of route choice modeling on dynamic OD estimation'. 2015 IEEE 18th Int. Conf. on Intelligent Transportation Systems, Las Palmas de Gran Canaria, Spain, 2015, pp. 1483–1488

[56] Ásmundsdóttir, R.: 'Dynamic OD matrix estimation using floating car data', MSc thesis, Civil Engineering Delft University of Technology, 2008

[57] Ásmundsdóttir, R., Chen, Y., van Zuylen, H.J.: 'Dynamic origin–destination matrix estimation using probe vehicle data as a priori information', 2010, **144**, pp. 89–108

[58] Zhao, H., Yu, L., Guo, J.F., *et al.*: 'Estimation of time-varying OD demands incorporating FCD and RTMS data', *J. Transp. Syst. Eng. Inf. Technol.*, 2010, **10**, pp. 72–80

[59] Yamamoto, T., Miwa, T., Takeshita, T., *et al.*: 'Updating dynamic origin-destination matrices using observed link travel speed by probe vehicles', in Lam, W., Wong, S., Lo, H. (Eds.): '*Transportation and Traffic Theory 2009: Golden Jubilee*, (Springer, Boston, MA, 2009), p. 723, https://doi.org/10.1007/978-1-4419-0820-9_35

[60] Cao, P., Miwa, T., Yamamoto, T., *et al.*: 'Bilevel generalized least squares estimation of dynamic origin–destination matrix for urban network with probe vehicle data', *Transp. Res. Rec. J. Transp. Res. Board*, 2013, **2333**, (1), pp. 66–73

[61] Vogt, S., Fourati, W., Schendzielorz, T., *et al.*: 'Estimation of origin-destination matrices by fusing detector data and floating car data', *Transp. Res. Procedia*, 2019, **37**, pp. 473–480

[62] Yang, X., Lu, Y., Hao, W.: 'Origin-destination estimation using probe vehicle trajectory and link counts', *J. Adv. Transp.*, 2017, **2017**, pp. 1–18

[63] Gentile, G.: 'Solving a dynamic user equilibrium model based on splitting rates with gradient projection algorithms', *Transp. Res. B Methodol.*, 2016, **92**, pp. 120–147

[64] Gentilea, G.: 'New formulations of the stochastic user equilibrium with logit route choice as an extension of the deterministic model', *Transp. Sci.*, 2018, **52**, pp. 1531–1547

[65] Djukic, T., Barcelo Bugeda, J., Bullejos, M., *et al.*: 'Advanced traffic data for dynamic od demand estimation: the state of the art and benchmark study'. TRB 94th Annual Meeting Compendium of Papers, Washington, D.C., 2015, pp. 1–16

[66] Nigro, M., Cipriani, E., del Giudice, A.: 'Exploiting floating car data for time-dependent origin–destination matrices estimation', *J. Intell. Transp. Syst.*, 2018, **22**, (2), pp. 159–174

[67] Yang, H., Rakha, H.: 'A novel approach for estimation of dynamic from static origin–destination matrices', *Transp. Lett.*, 2019, **11**, (4), pp. 219–228

[68] Hart, P.E., Nilsson, N.J., Raphael, B.: 'A formal basis for the heuristic determination of minimum cost paths', *IEEE Trans. Syst. Sci. Cybern.*, 1968, **4**, pp. 100–107

[69] Marchal, F., Hackney, J.K., Axhausen, K.W.: 'Efficient map-matching of large GPS data sets – tests on a speed monitoring experiment in Zurich', arbeitsbericht verkehrs und raumplanung. Technical report, UNAM, **244**, 2004

[70] Quddus, M.A., Ochieng, W.Y., Noland, R.B.: 'Current map-matching algorithms for transport applications: state-of-the art and future research directions', *Transp. Res. C Emerg. Technol.*, 2007, **15**, pp. 312–328

[71] Cascetta, E.: *Transportation systems analysis. Models and applications*, 2009, 2nd ed., doi: 10.1007/978-0-387-75857-2