O–D matrix estimation based on data-driven network assignment

Nikolaos Tsanakas, David Gundlegård & Clas Rydergren

To cite this article: Nikolaos Tsanakas, David Gundlegård & Clas Rydergren (2022): O–D matrix estimation based on data-driven network assignment, Transportmetrica B: Transport Dynamics, DOI: 10.1080/21680566.2022.2080128

To link to this article: https://doi.org/10.1080/21680566.2022.2080128
O–D matrix estimation based on data-driven network assignment

Nikolaos Tsanakas 🏳️‍🌈, David Gundlegård 🏳️‍🌈 and Clas Rydergren 🏳️‍🌈

Department of Science and Technology, Linköping University, Norrköping, Sweden

ABSTRACT
Time-dependent Origin-Destination (OD) matrices are an essential input to transportation models. A cost-efficient and widely used approach for estimating OD matrices involves the exploitation of flow counts from stationary traffic detectors. This estimation approach is also referred to as assignment-based OD matrix estimation because, typically, Dynamic Traffic Assignment (DTA) models are used to map the OD matrix to the link flows. The conventional DTA establish a complex non-linear relationship between the demand, and the link flows, adding an inherent complexity to the OD matrix estimation problem. In this paper, attempting to exploit the growing availability of Floating-Car Data (FCD), we suggest a solution approach that is based on a Data-Driven Network Assignment (DDNA) mechanism. The DDNA utilises the FCD from probe vehicles to capture congestion effects, providing a linear mapping of the OD matrix to the link flow observations. We present the results of two synthetic-data experiments that serve as proof of concept, indicating that if FCD are available, the computationally costly DTA may not be necessary for solving the OD matrix estimation problem.

Abbreviations
APR: average penetration rate; DDNA: data driven network assignment; DDNL: data driven network loading; DODME: dynamic OD matrix estimation; DTA: dynamic traffic assignment; FCD: floating-car data; ITS: intelligent transportation systems; NNLSQ: non-negative least squares; OD: origin destination; ODME: OD matrix estimation; PV: probe vehicle; RUM: random utility model

1. Introduction
The growing demand for mobility in our modern societies has led to highly congested cities, revealing the necessity for more effective utilisation of the available infrastructure. By making appropriate decisions and interventions, transportation planners and policymakers can mitigate congestion and enhance a transportation system’s efficiency. Traffic models are essential tools commonly used in strategic and operational traffic planning that allow the evaluation of alternative strategies and interventions in the system. In a typical traffic modelling analysis, urban space is divided into traffic analysis zones according to certain socio-economic criteria. The number of trips from each origin to each destination zone is usually stored in a matrix called Origin–Destination (OD) matrix. OD matrices represent the demand for travelling and constitute a crucial input to traffic models and may affect the modelling quality significantly.

CONTACT Nikolaos Tsanakas 🏳️‍🌈 nikolaos.tsanakas@liu.se
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

Supplemental data for this article can be accessed here. https://doi.org/10.1080/21680566.2022.2080128

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.
However, OD matrices are very seldom directly observable, and thus, they need to be estimated, relying on information that implicitly evinces the demand. Due to the importance of OD matrices in traffic modelling, the underlying estimation mechanism has to be carefully constructed. Traditional OD Matrix Estimation (ODME) approaches are mainly based on survey and census data. Such approaches typically are time-consuming, while they can also be economically costly. Moreover, surveys might be outdated, yielding a poor quality of the output OD matrices (Lundgren and Peterson 2008). The inefficiency of the survey-based approaches has directed the research attention towards alternative inexpensive data sources. One such data source are the stationary traffic detectors, which have been already installed in most cities, and measure flows for other purposes, e.g. traffic monitoring or control. They can therefore provide updated information on traffic flows without additional costs. However, even though flow counts are vital indicators of the demand, they provide only indirect information of the OD matrices. Hence, a supplementary mechanism is required to map the OD matrices to the flow counts. Traffic assignment models are commonly employed to perform this mapping and define which OD pairs and to what extent they contribute to each flow count. For this reason, the ODME approach that utilises flow counts is also referred to as assignment-based ODME (Frederix, Viti, and Tampère 2013).

We can define traffic assignment as the process through which OD demand is distributed among the different routes that connect each OD pair, resulting in link flows. The assignment of the demand is typically performed under the assumption that users adhere to certain behavioural rules and choose their routes based on the (actual or perceived) travel cost of each alternative route. In particular, most traffic assignment models presume that route choices are performed in such a way that the system eventually reaches a stable condition called User Equilibrium (UE), where no user can further improve her/his travel cost (Wardrop 1952). While traffic assignment determines the link flows given a specific demand, the assignment-based ODME process inversely attempts to compute the OD matrices given observed link flows. Thus, we can define the ODME problem as the problem of finding the optimal OD matrix, which, once assigned to the network, results in link flows equal to the measured ones. The assignment-based ODME problem has attracted abundant research attention lately since its solution can provide an effective and low-cost estimation of OD matrices. However, solving the ODME problem is not a straightforward task because (i) the number of flow observations usually is significantly lower than the number of OD pairs, and (ii) the induced mapping of demand to link flows typically is non-linear.

The low ratio between the number of flow observations and the number of OD pairs yields an underdetermined system of equations. This implies that if a solution exists, it is not unique; we can find an infinite number of OD matrices that satisfy the observed flows. To restrict the number of solutions, further information is required, which typically has the form of a prior or target OD matrix (it can be estimated, for instance, by the conventional survey-based ODME approaches). Although such a prior matrix might be only a rough approximation of the ‘real’ OD matrix, it can provide vital structural information and mitigate the underdetermination. Therefore, solving the ODME problem typically involves the minimisation of a two-part error function (Ashok and Ben-Akiva 2002), where, besides minimising the deviations between the estimated and observed link flows, we also try to minimise the distances between the estimated and the prior demand.

Traffic assignment models, attempting to capture congestion effects, establish a complex non-linear relationship (in most cases, we cannot even express it in an analytical form) between the link flows and the OD trips. When demand gets higher, congestion increases, raising the costs and changing the route choices and their corresponding flow patterns. Typically, due to implicit link interactions, link flows are not separable (Frederix, Viti, and Tampère 2013) and the demand affects link flows in a manner that cannot be evaluated analytically. Therefore, this modelling setting commonly disturbs the mathematical tractability of the ODME problem, and properties such as convergence and solution uniqueness cannot be guaranteed.

The early solution approaches suggested by Van Zuylen and Willumsen (1980), Cascetta (1984), and Bell (1991) were developed under the assumption that route choices are made a priori and are
fixed (proportional traffic assignment). To this extent, a linear mapping of the OD matrix to the link flows is settled, and this mapping is typically represented by the so-called assignment matrix. Although proportional-assignment approaches provide formulations that allow efficient optimisation schemes, the modelling realism is highly questionable. Alternatively, Spiess (1990), Hai Yang (1995), Cascetta and Postorino (2001), and Lundgren and Peterson (2008) suggested solutions under the modelling hypothesis that route choices are contingent upon the demand. In this way, the assignment matrix depends on the demand and is endogenous as the demand is also the variable under estimation. This modelling setting expands the complexity, leading to a non-linear bilevel optimisation problem, which is iteratively solved using gradient approximations.

The limited performance of linear mapping schemes might be further aggravated if we desire to estimate a time-dependent OD matrix. In the Dynamic OD Matrix Estimation (DODME), apart from route choices, the assignment matrix should also reflect flow-propagation effects. Flow-propagation phenomena, such as spill-back, enhance the non-separability of link flows, and hence, the assumption of linear mapping might yield biased estimations. In this regard, the research attention is focussed on the bilevel formulations. However, in contrast to the static traffic assignment problem that can be solved analytically, solving the Dynamic Traffic Assignment (DTA) problem commonly requires computationally costly traffic simulations. Considering that several DTA simulations might be required to approximate the gradient (Frederix et al. 2011; Toledo and Kolechkina 2012), the bilevel approach extends the computational burden for large-scale networks and long analysis periods. Therefore, finding a solution approach that permits efficient optimisation schemes while at the same time sufficiently captures congestion effects has been proven to be a challenging task.

The latest advances in communication systems and the extended research on Intelligent Transportation Systems (ITS) has offered a great abundance of traffic-related data. Besides the conventional cross-sectional data from stationary detectors, the availability of Floating-Car Data (FCD) from probe vehicles or other mobile sources (GPS or cellular networks) is rapidly growing. Appropriate exploitations of such data may give new insights and provide effective solutions to the inherently complex DODME problem. However, in contrast to the flow counts, FCD might not reflect the total number of network users because of the usually low probe-vehicles penetration rates. FCD can, though, provide vital information on the users’ behaviour and system’s performance (Auberlet et al. 2014). In this respect, Sohn and Kim (2008), Xianfeng Yang, Lu, and Hao (2017), Ma and Qian (2018), Krishnakumari et al. (2020), Ros-Roca, Bugeda, and Mercadé (2020), and Mitra et al. (2020) suggest solution schemes where FCD play the key role of the external congestion indicators. In these data-driven approaches, the congestion level that defines the route choices and the flow propagation is independent of the demand (yielding separable link flows). For instance, travel time observations can be used to indicate the route costs or even flow-propagation phenomena, such as the formation or dissolution of the queues.

The methodological advantages of the data-driven solution schemes have been widely discussed by Xianfeng Yang, Lu, and Hao (2017) and Krishnakumari et al. (2020): The computationally expensive iterative DTA process is not required to solve the DODME problem. Instead, a one-shot Data Driven Network Assignment (DDNA) defines an exogenous assignment matrix, which can reflect the prevailing congestion although it does not depend on the demand. Undeniably, the quality of such an assignment matrix (i.e. its ability to reflect the underlying route choices and flow propagation) is crucial and may strongly affect the estimation’s accuracy. However, in most data-driven DODME estimations found in literature, developing the assignment matrix is a side-process, where some simplified and sometimes unrealistic assumptions are made (especially for the flow propagation). Instead, they focus on how the available field data can be exploited for providing a target (or seed) OD matrix (or production and generation patterns).

In this paper, although we highly acknowledge the importance of the FCD-driven target OD matrix generation, we focus on the exogenous assignment matrix itself, aiming at enhancing its quality. Under the hypothesis that FCD can sufficiently indicate the spatiotemporal variations of congestion, we suggest and evaluate a novel DDNA mechanism. Special attention is given to the flow propagation
sub-mechanism, where we identify the main limitations of the existing data-driven approaches: Usually, traffic flow is represented by point packets of vehicles having no physical length. This modelling setting may deteriorate the quality of the assignment matrix and lead to biased OD demand estimations. In this paper, we adopt and extend (to facilitate its use to the DODME) a more advanced data-driven flow-propagation mechanism, initially suggested by Tsanakas et al. (2021). Flow is also represented by vehicle-packets which, however, have an adjustable spatial length that dynamically adapts to the level of congestion. In this way, we are able to exploit even flow counts coming from congested parts of the network. Furthermore, we also propose a route choice estimation approach by fusing a random utility model with the available FCD.

The paper is structured as follows: In Section 2, we provide an overview of the DODME problem and its solution approaches. In Section 3, we present our method for constructing an exogenous assignment matrix and solving the DODME problem. In Section 4, we describe the experimental setup that aims to evaluate our approach, while in Section 5, we present the numerical results of this evaluation. In Section 6, we discuss the main findings and the limitations of this study. Finally, Section 7 presents the main conclusions of this study.

2. The DODME problem

2.1. Formulation

Consider a transportation network represented by a directed graph \( G(\mathcal{N}, \mathcal{A}) \), where \( \mathcal{N} \) is the set of nodes and \( \mathcal{A} \) is the set of links. Let \( o \) and \( d \) denote origin and destination nodes, respectively, with \( o \in \mathcal{O}, \mathcal{O} \subseteq \mathcal{N} \) and \( d \in \mathcal{D}, \mathcal{D} \subseteq \mathcal{N} \). Furthermore, let \( i = 1, 2, \ldots, I \) denote each OD pair \( o, d \), with \( I \) being the total number of OD pairs. The analysis period, \( \hat{T} \), is divided in \( H \) equal-length time periods, commonly called demand or assignment periods. Let each demand period be indexed by \( h \), with \( h = 1, 2, \ldots, H \), and have a length of \( T_h \), with \( T_h = \hat{T}/H \). Moreover, let \( x_{ih} \) denote the number of travellers in OD pair \( i \), whose departure time from the origin node lies within period \( h \). Then, the time-sliced demand is typically stored in vector \( \mathbf{x} = [x_1 \ldots x_h \ldots x_H]^T \), where \( x_h = [x_{1,h} \ldots x_{I,h} \ldots x_{I,h}] \). Consider that vehicle detectors are installed at a subset of links, \( B \subseteq \mathcal{A} \), providing vehicle counts for \( S \) equal-length counting periods indexed by \( s = 1, 2, \ldots, S \). Let the length of each counting period be \( T_s \) with \( T_s = \hat{T}/S \). The number of vehicles detected at link \( b \) during counting period \( s \) is denoted by \( \hat{y}_{b,s} \). Such vehicle counts are stored in vector \( \hat{\mathbf{y}} = [\hat{y}_1 \ldots \hat{y}_s \ldots \hat{y}_S]^T \), with \( \hat{y}_s = [\hat{y}_{b,1} \ldots \hat{y}_{b,s} \ldots \hat{y}_{b,S}] \).

A widely used formulation of the assignment-based DODME problem is given by (1) as

\[
\mathbf{x}^* = \arg\min_{\mathbf{x}} \left[ \xi_1 F_1(\mathbf{y}(\mathbf{x}), \hat{\mathbf{y}}) + \xi_2 F_2(\mathbf{x}, \hat{\mathbf{x}}) \right],
\]

s.t. \( \mathbf{y} = \mathbf{A}(\mathbf{x}) \mathbf{x} \),

\( \mathbf{x} \geq 0 \),

where \( \hat{\mathbf{x}} = [\hat{x}_1 \ldots \hat{x}_h \ldots \hat{x}_H]^T \) represents a prior demand vector, \( F_1, F_2 \) are functions that represent distance measures, and \( \mathbf{A} \) is the assignment matrix. The factors \( \xi_1, \xi_2 \) are weights, which, for instance, can reflect the relative uncertainty of the vehicle counts and the prior demand. The assignment matrix has dimensions of \( |B|S \times IH \) and maps the demand, \( \mathbf{x} \), to the link flows, \( \mathbf{y} \), by implicitly reflecting route choice and flow propagation effects. Therefore, we seek for the demand vector, \( \mathbf{x}^* \), that is the closest to the prior demand, \( \hat{\mathbf{x}} \), and, once assigned to the network, induces the link flows, \( \mathbf{y} \), that are the closest to the observed, \( \hat{\mathbf{y}} \). Formulation (1) corresponds to the simultaneous model proposed by Cascetta, Inaudi, and Marquis (1993), where the demand is simultaneously estimated for every demand period \( h \). Alternatively, one could use the sequential estimation approach (Cascetta, Inaudi, and Marquis 1993; Ashok and Ben-Akiva 2002), where the demand for each period is sequentially estimated. The sequential approach exhibits some computational advantages but may be less accurate.

We can further subdivide the assignment matrix (Frederix, Viti, and Tampère 2013; Mitra et al. 2020) by making the assumptions of constant route choices and uniform departures during each demand period.
period. Let \( P_i \) be a set containing the available routes that connect OD pair \( i \). The departure rate of the vehicles that depart during demand period \( h \) and follow the \( p \)th route, \( p \in P_i \), is denoted by \( f_{i,p,h} \) and is given as

\[
f_{i,p,h} = r_{i,p,h} \cdot x_{i,h} / T_h,
\]

where \( r_{i,p,h} \) connotes the portion of the total OD trips, \( x_{i,h} \), which are made through the \( p \)th route, \( p \in P_i \). Then, we can relate the induced flows at the location of the detectors to the route flows as

\[
y_b(t) = \sum_{i=1}^{l} \sum_{p \in P_i} \sum_{h=1}^{H} \hat{q}_{i,p,h}^b (t) \cdot f_{i,p,h}, \quad \forall b \in B,
\]

where \( \hat{q}_{i,p,h}^b (t) \) are the so-called crossing fractions, denoting the portion of route flow, \( f_{i,p,h} \), that passes by the detector’s location at link \( b \) and at time instant \( t \). Note that \( \hat{q}_{i,p,h}^b (t) \) is zero if link \( b \) does not belong to path \( p \) and if \( h \cdot T_h > t \). By applying (2), we are able to incorporate the OD trips in (3) as

\[
y_b(t) = \sum_{i=1}^{l} \sum_{p \in P_i} \sum_{h=1}^{H} \hat{q}_{i,p,h}^b (t) \cdot r_{i,p,h} \cdot x_{i,h} / T_h, \quad \forall b \in B,
\]

and derive an explicit relationship between the continuous-time link flows and the OD trips. As \( \hat{y}_{b,s} \) denotes the total number of vehicles crossed the detector at link \( b \) during counting period \( s \), we can cumulate the continuous-time flows and obtain the corresponding modelled counts, \( y_{b,s} \), as

\[
y_{b,s} = \sum_{i=1}^{l} \sum_{p \in P_i} \sum_{h=1}^{H} q_{i,p,h}^{b,s} \cdot r_{i,p,h} \cdot x_{i,h}, \quad \forall b \in B, \quad s = 1, 2, \ldots, S,
\]

with,

\[
q_{i,p,h}^{b,s} = \frac{1}{T_h} \int_{(s-1)T_s}^{sT_s} \hat{q}_{i,p,h}^b (\omega) \, d\omega.
\]

To acquire the matrix form of (5), we should first define a global route choice set as \( \hat{P} = \bigcup_i P_i \), \( i = 1, 2, \ldots, l \). Note that each route \( p \in \hat{P} \) connects a unique OD pair. We can reformulate Equation (5) using matrix notation as

\[
y = QRx, \quad \text{with } QR = A.
\]

\( R \) is a matrix with dimensions \(|\hat{P}|H \times lH \) that reflects the route choices. It can be defined as \({}^2\)

\[
R = \text{diag}(r) \left( I_H \otimes \hat{\delta} \right),
\]

where \( r \) is the so-called route choice vector with \( r = [r_1 \ldots r_h \ldots r_l] \), \( r_h = [r_{1,h} \ldots r_{p,h} \ldots r_{l,h}] \) and \( \hat{\delta} \) is an incidence matrix with dimensions \(|\hat{P}| \times l \), with its entries, \( \hat{\delta}_{p,i} \), taking the value of 1 if route \( p \in \hat{P} \) connects OD pair \( i \) and 0 otherwise. Accordingly, \( Q \) is a matrix with dimensions \(|B|S \times |\hat{P}|H \) that reflects flow-propagation effects and it is given by

\[
Q = \begin{bmatrix}
\hat{Q}_{1,1} & \cdots & \hat{Q}_{1,H} \\
\vdots & \ddots & \vdots \\
\hat{Q}_{S,1} & \cdots & \hat{Q}_{S,H}
\end{bmatrix}, \quad \hat{Q}_{s,h} = \begin{bmatrix}
q_{1,h}^{1,s} & \cdots & q_{1,H}^{1,s} \\
\vdots & \ddots & \vdots \\
q_{|B|,s}^{1,h} & \cdots & q_{|B|,s}^{1,H}
\end{bmatrix},
\]

with \( \hat{Q}_{s,h} = 0 \) for each \( h, s \) such that \( h \cdot T_h > s \cdot T_s \).
2.2. Conventional solution approaches

Depending on the underlying hypothesis on the nature of the assignment matrix, most solution approaches in the literature can be classified into two categories (Lundgren and Peterson 2008): In the first category, the assignment matrix, \( A \), does not depend on the demand vector, \( x \), while in the second, the assignment matrix is given as a function of the demand vector, i.e. \( A = A(x) = R(x)Q(x) \).

Even though the presumption of a fixed assignment matrix might be unrealistic for congested networks, this modelling setting provides some important mathematical properties: the mapping of OD demand to link flows described by constraint (1b) is linear, and by considering convex distance measures \( F_1, F_2 \) (e.g. Euclidean distances), the DODME problem is solved using standard techniques for convex programming (Van Zuylen and Willumsen 1980; Bell 1991; Cascetta, Inaudi, and Marquis 1993; Bierlaire and Crittin 2004).

On the other hand, if route choices and flow propagation are subject to congestion, it is almost impossible to derive an analytical expression of constraint (1b). Namely, we are unable to explicitly define how link flows vary with OD trips variations because of the complex link interactions incorporated in DTA models. In principle, computing the assignment matrix usually constitutes an optimisation problem in itself (e.g. a dynamic UE problem). In this context, DODME is treated as a bilevel programming problem and it is iteratively solved: solving the upper-level problem (1a) defines a current solution \( x' \), which once assigned to the network – usually using simulation-based DTA – provides an updated estimate of the link flows, \( y \). As the objective function cannot be evaluated analytically, the gradient that indicates the descent direction is approximated either by assuming that the assignment matrix is locally constant at \( x' \) (Spiess 1990; Hai Yang 1995; Lundgren and Peterson 2008) or by using finite-differences schemes at the neighbourhood of \( x' \) (Frederix et al. 2011; Toledo and Kolechkina 2012). Note that the latter approach requires several evaluations of the objective function using costly simulations, although more efficient gradient-approximation techniques have been suggested (Cipriani et al. 2011; Antoniou et al. 2015; Tympakianaki, Koutsopoulos, and Jenelius 2015). The bilevel DODME has attracted wide research attention, and in literature one may find numerous solution approaches. A more extensive description of such solution approaches lies beyond the scope of this paper. From the paper’s perspective, it is important to note that the endogenous modelling of route choice and flow propagation introduces non-convex constraints, disturbing convexity and deteriorating the deployment efficiency.

2.3. Incorporating data from alternative sources

Conventional approaches for estimating OD matrices commonly utilise data from only stationary detectors, i.e. traffic counts. This is reasonable since stationary detectors were the dominant source of traffic-related data when several of the seminal solution approaches (Van Zuylen and Willumsen 1980; Spiess 1990) were developed. However, since the time of Van Zuylen and Willumsen (1980) and Spiess (1990), the development of communication and positioning systems has provided a vast amount of traffic-related data from various sources. Lately, many researchers investigate the possibility of exploiting such data to enhance the DODME solution approaches. Consider now that apart from the traffic counts, data from alternatives sources (e.g. observed travel times from Bluetooth, vehicle trajectories from GPS) are available during the same analysis period, forming the vector of observations, \( \hat{z} \). A common approach (Barceló et al. 2013; Nigro, Cipriani, and del Giudice 2018) for exploiting this supplementary information is the addition of the term, \( \zeta_3 F_3(z(x), \hat{z}) \), to the objective function (1a). \( \zeta_3 \) is a weighting factor and \( F_3(z(x), \hat{z}) \) signifies the distance between the observed, \( \hat{z} \), and the induced values, \( z(x) \), obtained by assigning demand \( x \) to the network. The additional term increases the number of equations, partly alleviating the underdetermination and improving the quality of the solution. However, similar to the traffic counts, \( y \), a function (or better a process) is required to map the demand vector \( x \) to the observed values \( \hat{z} \). Such mapping is typically provided by simulation-based DTA and, thus, solving the DODME faces gradient approximation challenges that may lead to
poor deployment efficiency. Therefore, although this modelling setting may improve the accuracy, congestion is still an endogenous variable (it depends on the demand), retaining complexity and hindering any mathematical tractability.

Alternatively, we may use observations \( \hat{\mathbf{z}} \) as congestion indicators and obtain an analytical relationship between the OD trips and the link flows. This approach yields an empirical or exogenous assignment matrix, and the DODME problem can be reformulated as

\[
\mathbf{x}^* = \arg\min_x \left[ \zeta_1 \mathbf{F}_1 (\mathbf{y}(\mathbf{x}, \hat{\mathbf{z}}), \mathbf{\hat{y}}) + \zeta_2 \mathbf{F}_2 (\mathbf{x}, \hat{\mathbf{z}}) \right],
\]

s.t. \[
\mathbf{y} = \mathbf{A}(\mathbf{\hat{z}}) \mathbf{x} = \mathbf{Q}(\mathbf{\hat{z}}) \mathbf{R}(\mathbf{\hat{z}}) \mathbf{x},
\]

\[
\mathbf{x} \geq 0,
\]

The modelled link flows, \( \mathbf{y}(\mathbf{x}, \hat{\mathbf{z}}) \), are obtained by a DDNA mechanism which, in analogy to the conventional DTA, maps the OD trips to the link flows. In the data-driven case, however, such mapping is linear since congestion is captured by external variables. Therefore, this modelling setting can establish convexity (for convex distance functions). We should highlight here that the exogenous assignment matrix does not provide a generic OD demand–link flow mapping, but it instead regards a specific analysis period, during which the flow counts, \( \mathbf{\hat{y}} \), and the FCD, \( \hat{\mathbf{z}} \), are collected. In other words, the assignment matrix is case-specific, and we may not use the same assignment matrix for different analysis periods.

Sohn and Kim (2008), Xianfeng Yang, Lu, and Hao (2017), Ma and Qian (2018), Krishnakumari et al. (2020), Ros-Roca, Bugeda, and Mercadé (2020), Mitra et al. (2020), and Ros-Roca et al. (2022) follow this exogenous assignment matrix approach. Depending on the presumed type of the available data, they suggest different methods for obtaining the route choice and loading matrices. Although all the aforementioned studies use an exogenous assignment matrix, not all of them adopt formulation (10). For instance, Ros-Roca et al. (2022), Ros-Roca, Bugeda, and Mercadé (2020), and Mitra et al. (2020) treat the DODME problem as a scaling problem, while Krishnakumari et al. (2020) use production and attraction patterns instead of a prior OD matrix to decrease the ratio between the number of unknowns and the number of equations. However, what is important from this paper’s perspective, is the method used for constructing the exogenous assignment matrix.

Ma and Qian (2018), Krishnakumari et al. (2020), and Ros-Roca, Bugeda, and Mercadé (2020) estimated average realised link and/or route travel times by processing FCD from Probe Vehicles (PVs) or other sources. Let \( \bar{\tau}_{i,p,h}(\hat{\mathbf{z}}) \) denote the estimated travel time for the \( p \)-th route, \( p \in \mathcal{P}_i \), of OD pair \( i \) for the vehicles departing during the demand period \( h \), given the observations \( \mathbf{\hat{z}} \). Then, assuming that users choose their route attempting to maximise their perceived utility (in terms of travel times), the route choice proportions can be computed by a Random Utility Model (RUM)

\[
\rho_{i,p,h}^{\text{RUM}} = \text{prob} \left( \bar{\tau}_{i,p,h}(\hat{\mathbf{z}}) + \xi_{i,p,h} \leq \bar{\tau}_{i,p',h}(\hat{\mathbf{z}}) + \xi_{i,p',h}, \forall p' \in \mathcal{P}_i \right), \quad \forall p \in \mathcal{P}_i,
\]

where \( \xi_{i,p,h} \) is a random variable that represents the perception error. Ma and Qian (2018) use a simple MultiNomial Logit (MNL) structure, while Krishnakumari et al. (2020), Ros-Roca, Bugeda, and Mercadé (2020), and Ros-Roca et al. (2022) use modifications of the MNL by adding route size and route commonality penalisation factors, respectively. However, such modelling of route choices is based on behavioural assumptions, whose realism may be questionable (Nigro, Cipriani, and del Giudice 2018). Equation (11) implies that (experienced) travel time is the only feature affecting the route choices. Therefore, Equation (11) may not always be able to describe the inherently complex phenomenon of route choice.

Alternatively, Sohn and Kim (2008), Xianfeng Yang, Lu, and Hao (2017), and Mitra et al. (2020) estimated the route choice proportions by explicitly considering the FCD and scaling up individual choices of the PVs. This approach relies on the assumption that either entire trips (from origin to destination) are monitored for a subset of the vehicles or that turning fractions are available for each node of the network. Let \( w_{i,h} \) be the observed number of PVs of OD pair \( i \) departed from their origins during period
Furthermore, let $w_{i,p,h}$ be the number of PVs that select the $p$th route of OD pair $i$, $h$. Then, we can approximate the entries of the route choice vector as

$$r_{i,p,h}^{FCD} = \frac{w_{i,p,h}}{w_{i,h}}. \quad (12)$$

This approach may be meaningful for relatively high penetration rates of PVs. However, if the penetration rate is low, the consideration of random individual choices might yield biased estimations.

Xianfeng Yang, Lu, and Hao (2017) develop a similar PV-based mechanism for determining the flow propagation. Alternatively, Ros-Roca, Bugeda, and Mercadé (2020) and Ros-Roca et al. (2022) utilise the estimated (realised) link travel times to capture the flow propagation and compute the entries of the loading matrix as

$$q_{i,p,h}^{b,s} = \begin{cases} 1 & \text{if } (s - 1)T_s < \tilde{\tau}_{i,p,h}^b \leq sT_s, \quad \forall \ b \in B, \ p \in \bar{P} | \delta_{b,p} > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

where $\delta_{b,p}$ is an incidence indicator taking the value of 1 if link $b$ belongs to route $p$ and 0 otherwise. $\tilde{\tau}_{i,p,h}^b$ is the average travel time that the vehicles of OD pair $i$ departed during the $h$th demand period need to reach link $b$ from their origin. Equation (13) implies that traffic flow is treated as a point packet, where every vehicle of each packet is simultaneously propagated through the network (it strictly requires that $T_h = T_s$). We will refer to this approach as Point-Packet Network Loading (PPNL).

Although this discrete-time modelling structure partly captures congestion as it is based on realised travel times, the assumption of point packets might deteriorate the quality of the assignment matrix especially for long demand periods (Ashok and Ben-Akiva 2002). Krishnakumari et al. (2020) follow a similar PPNL approach considering, though, Equation (13) only if the links of route $p$ upstream of link $b$ are not congested. Mitra et al. (2020), who mainly focus on route choice estimation rather than on network loading, suggest a quasi-static flow-propagation approach. They assume that the departure time from the origin and the arrival time to the destination lies within the same demand period for every trip performed during the analysis period. This modelling setting implies that for each demand period, the loading matrix is replaced by the link-route incidence matrix $\delta \ (|B| \times |\bar{P}|)$, with $\delta = \{\delta_{b,p} | b \in B, p \in \bar{P}\}$. The Quasi-Static Network Loading (QSNL) matrix (which also requires $T_h = T_s$) is given by

$$Q^{QSNL} = I_H \otimes \delta. \quad (14)$$

The QSNL can also be considered as a point-packet approach because it yields a binary network loading matrix.

Instead, Sohn and Kim (2008) and Ma and Qian (2018) adopt the more sophisticated continuous packet network loading mechanism suggested by Cascetta, Inaudi, and Marquis (1993) and Ashok and Ben-Akiva (2002): The packets have physical length and travel time is uniformly distributed among the first and last vehicle of the packet. We can determine the travel times of each packet’s first and the last vehicle by defining a continuous-time travel time function based on the available observations. In this manner, the packets can ‘stretch’ or ‘squeeze’ while flow is propagated through the network and the strict binary nature of crossing fractions is relaxed. Undeniably, the continuous packet modelling assumption is more realistic than the point-packet assumption. It exhibits, however, some methodological limitations since an arbitrary travel time function might result in travel time–flow propagation inconsistencies, FIFO violations and even negative flows. Moreover, the network-loading approaches suggested by Sohn and Kim (2008) and Ma and Qian (2018) are developed and evaluated from a highway-network perspective, and their route-based structure might not be adequate for general networks.
3. The proposed approach

In this paper, we suggest a novel approach for constructing the exogenous assignment matrix. The methodological advances of our approach aim at addressing the limitations of the existing data-driven DODME techniques:

- The proposed flow-propagation mechanism is based on the Data-Driven Network Loading (DDNL) suggested by Tsanakas et al. (2021). DDNL considers a continuous packet modelling setting that ensures vehicle conservation, FIFO, positive flows and travel time to flow propagation consistency. Furthermore, the number and size of the packets are not fixed but rather adapt to the associated congestion level. Moreover, DDNL exhibits a link–node model structure instead of a route-based modelling structure. Namely, DDNL considers some network topology attributes, such as the incoming and the outgoing links at each node, which limit the number of operations (the operations are performed simultaneously for all the routes traversing the same node) and encourage the application to large-scale networks.

- For the estimation of route choice, we fuse the choices derived by a model-driven estimation (using a RUM similar to Ma and Qian 2018; Krishnakumari et al. 2020; Ros-Roca, Bugeda, and Mercadé 2020) with the corresponding choices obtained by a pure data-driven estimation (scaling observed individual choices as suggested by Sohn and Kim 2008; Xianfeng Yang, Lu, and Hao 2017). The route choice vector is expressed as a weighted average of the model-driven and the data-driven route choice estimations. We use an adaptive weight that adjusts to the underlying data availability level: a higher weight is given to the model-driven estimations if the data availability level is low and vice versa.

Before giving a more detailed description of the data-driven network-loading and route choice mechanisms, let us highlight the main assumptions of our approach. We assume that:

- The set or routes actually used, \( \hat{P} \), is known.
- Route choices are constant during each demand period.
- Vehicles uniformly depart from their origins, yielding piecewise constant inflow rates for the first link of each route.
- Stationary detectors are installed at a subset of links \( B \subseteq A \) (they provide the cross-sectional data that form vector \( \hat{y} \)).
- A subset of the vehicles (referred to as PVs) is equipped with a sensing device providing link travel times throughout the entire trip from their origin to their destination (they provide the FCD that form vector \( \hat{z} \)).

3.1. Flow propagation

Let \( U_a(t) \) and \( V_a(t) \) be the cumulative number of vehicles that have entered or exited link \( a \) by time \( t \), with

\[
U_a(t) = \int_0^t u(\omega) \, d\omega, \quad V_a(t) = \int_0^t v(\omega) \, d\omega, \tag{15}
\]

where \( u_a(t) \) and \( v_a(t) \) are the inflow and outflow rates of link \( a \) at time \( t \), respectively. Furthermore, let \( \tau_a(t) \) be the time that vehicles departing from the upstream node of link \( a \) at time instant \( t \) need to reach the downstream node of the same link. Since no vehicle can appear or disappear while traversing link \( a \), the cumulative outflow at \( \tau_a(t) \) time units after \( t \) equals the cumulative inflow at \( t \),

\[
V_a(t + \tau_a(t)) = U_a(t). \tag{16}
\]
By differentiating both sides of (16) with respect to time, we can obtain a relationship between the inflow and outflow rates,

\[ v_a(t + \tau_a(t)) = \frac{u_a(t)}{1 + \frac{d\tau_a(t)}{dt}}, \]

(17)

describing how flow propagates within link \( a \).

Traditionally, travel time, \( \tau(t) \), is given as a function of the congestion level at time instant \( t \), which implicitly or explicitly depends on the inflow and outflow rates. Instead, Tsanakas et al. (2021) suggest that the estimation of travel time can be based on the FCD, \( \hat{z} \), and be exogenous to the flow propagation. To this end, they propose a piecewise linear approximation of the travel time function using constrained segmented regression where the available travel time observations play the role of the independent variables. Consider that for each link, \( a \in A \), travel time observations, \( \hat{\tau}_a, j = 1, 2, \ldots, J \), are available from the \( J \) PVs traversing link \( a \) during the analysis period at discrete time instances \( \hat{t}_a \), (in principle, \( \hat{\tau}_a, j \) is the observed travel time from the \( j \)th PV which crosses the upstream end of link \( a \) at \( \hat{t}_a \)). Given the data set, \( \{\hat{t}_a, j, \hat{\tau}_a, j\}_{j=1}^J \), we employ a sliding window algorithm to define the number of linear segments for each link, \( a \), as well as the starting time, \( \theta_a,k \), the slope, \( \alpha_a,k \), and the intercept, \( \beta_a,k \), for each linear segment of link \( a \). At each iteration, the algorithm slides the estimation window forward along the time dimension and includes a new data point in the estimation of the parameters. A transition of a linear segment occurs when the estimation error exceeds a predefined threshold. Therefore, the number of linear segments (and the number of the break points) is not fixed, but it depend on each link’s congestion level. The segmented regression algorithm is constrained in such a way to ensure feasible and positive outflows, i.e. \( d\tau(t)/dt > -1 \). A complete description of the constrained segmented regression algorithm is given by Tsanakas et al. (2021).

The \( k \)th linear segment of the estimated piecewise linear travel time function, which spans over \( [\theta_k, \theta_{k+1}) \), is given by

\[ \tau(t) = \tau_k(t) = \alpha_k t + \beta_k, \quad \text{if } \theta_k \leq t < \theta_{k+1}. \]

(18)

The collection of the linear travel time intervals forms a partition, \( \mathcal{K} \), of the analysis period \( [0, T) \),

\[ \mathcal{K} = \{\theta_k\}, \quad k = 1, 2, \ldots, |\mathcal{K}|, \]

(19)

with

\[ 0 = \theta_1 < \theta_2 < \cdots < \theta_K = \hat{T}. \]

(20)

Note that the number and the parameters of the linear segments are link-specific, but we omit the link index here for notational convenience. For the links where no travel time observations are available, \( |\mathcal{K}| \) is set to one, \( \alpha \) to zero and \( \beta \) equals the free-flow travel time.

A piecewise linear travel time function together with piecewise constant inflows lead to piecewise constant outflows and consequently, to piecewise constant inflows for the proceeding links. The collection of time intervals, during which (i) the inflow of link \( a \) is constant, and (ii) the travel time of link \( a \) is described by a single linear segment, also forms a partition, \( \mathcal{K}_a \), of the analysis period,

\[ \mathcal{K}_a = \{t_j\}, \quad j = 1, 2, \ldots, |\mathcal{K}_a|, \]

(21)

with

\[ 0 = t_1 < t_2 < \cdots < t_J = T \quad \forall \ a \in A. \]

(22)

Therefore, for every link \( a \in A \), the continuous-time inflows, \( u(t) \), are constant between two discrete events, i.e.

\[ u(t) = u_j, \quad \text{if } t_j \leq t < t_{j+1}, \]

(23)
This modelling setting yields continuous packets of vehicles, where travel time linearly evolves between the two extreme points, $t_j, t_{j+1}$. Finally, using (17) we can compute the link outflows as

$$v(t) = \frac{u_j}{1 + \alpha_k} = v_j, \quad \text{if } (1 + \alpha_k)t_j + \beta_k \leq t < (1 + \alpha_k)t_{j+1} + \beta_k,$$

and $\theta_k \leq t_j < t_{j+1} \leq \theta_{k+1}, \quad (24)$

which are also piecewise constant. Accordingly, the collection of time intervals, $\tilde{K}$, during which the outflow of link $a$ is constant, is given by

$$\tilde{K} = \{(1 + \alpha_k)t_j + \beta_k\}. \quad (25)$$

Note that $\tilde{K}$ and $\tilde{K} | (|\tilde{K}| = |\tilde{K}|)$ are also link-specific, but for simplicity we skip the link index.

Consider now that $A^-_n$ and $A^+_n$ are the sets of the incoming and outgoing links to a node $n \in N$. Let $\phi_{a,a'}(t)$ be the turning proportions at $t$, denoting the part of the outflow, $v_a(t), a \in A^-_n$, that turns to link $a' \in A^+_n$. Such turning proportions are exogenous to the flow-propagation mechanism, and they are also piecewise constant if we assume fixed route choices during each demand period. The inflow of each outgoing link is given by

$$u_{a'}(t) = \sum_{a \in A^-_n} \phi_{a,a'}(t) \cdot v_a(t), \quad \forall a' \in A^+_n, \quad (26)$$

which is constant between $t_{j,a'}$ and $t_{j+1,a'}$.

Therefore, as the travel time function and the turning proportions are exogenous, defining the inflow partitions is enough for propagating the flow through the network (we can start from the inflows of each route’s first link, which are also exogenous and known a priori). For a more detailed description of the travel time function estimation and the data-driven flow propagation, we refer to Tsanakas et al. (2021).

The assumption of constant route choices during each demand period allows us to disaggregate the inflows and outflows as

$$u_a(t) = \sum_{i=1}^I \sum_{p \in P} \sum_{h=1}^H q^a_{i,p,h}(t) \cdot r_{i,p,h} \cdot x_{i,h}/T_h, \quad (27)$$

$$v_a(t) = \sum_{i=1}^I \sum_{p \in P} \sum_{h=1}^H q^a_{i,p,h}(t) \cdot r_{i,p,h} \cdot x_{i,h}/T_h, \quad (28)$$

where $q^a_{i,p,h}(t)$ and $q^a_{i,p,h}(t)$ are the dynamic crossing fractions of route flows, $f_{i,p,h} = r_{i,p,h} \cdot x_{i,h}/T_h$, on the upstream and downstream boundary of link $a$, respectively. The exogenous-congestion setting establishes an important property to such crossing fractions: their propagation is independent of the aggregated flows. This property becomes highly beneficial from a DODME perspective because it permits us to explicitly propagate the crossing fractions instead of the aggregated flows. We can perform such propagation by disaggregating the outflow Equation (24) as

$$\tilde{q}^a_{i,p,h}(t) = \frac{q^a_{i,p,h}}{1 + \alpha_k} = \tilde{q}^a_{i,p,h}, \quad \text{if } (1 + \alpha_k)t_j + \beta_k \leq t < (1 + \alpha_k)t_{j+1} + \beta_k,$$

and $\theta_k \leq t_j < t_{j+1} \leq \theta_{k+1}. \quad (29)$

The propagation of the crossing fractions exhibits the same properties as the propagation of the aggregated link flows. Piecewise constant crossing fractions on the upstream links’ boundaries combined with piecewise linear travel time function yield piecewise constant fractions on the downstream
boundaries. Therefore, similar to the aggregated flows, in order to propagate the crossing fractions we need only to find the partitions $K_a$ for each link $a \in \mathcal{A}$.

Figure 1(a) illustrates the piecewise constant crossing fractions on the upstream boundary of a hypothetical link and for a typical route $i, p, h$. For the same link, Figure 1(b) shows the estimated (given the travel time observations) piecewise linear travel time. Although the upstream crossing fractions take two distinct values (note that $q_{i,p,h}^1 = q_{i,p,h}^2$), the change of travel time function's slope at $\theta_2$ prompts a third packet of vehicles ($|\mathcal{K}_a| = 3$). Each one of those packets is associated with different crossing fractions on the downstream boundary since according to (29):

$$
\tilde{q}_{i,p,h}^1 = \frac{q_{i,p,h}^1}{1 + \alpha_1}, \quad \tilde{q}_{i,p,h}^2 = \frac{q_{i,p,h}^2}{1 + \alpha_2}, \quad \tilde{q}_{i,p,h}^3 = \frac{q_{i,p,h}^3}{1 + \alpha_2}.
$$

Note that as we do not explicitly propagate flows but fractions of OD trips, the number of vehicles at each packet is unknown and it depends on the route choice portion $r_{i,p,h}$ and the OD trips $x_{i,h}$. The propagation of packets aims at defining the crossing fractions at the detectors. To this end, we should first map the upstream crossing fractions, $q_{i,p,h}^b(t)$, to the counting location. Let $l_b, b \in B$ be the relative position of the detector along link $b$, with $0 \leq l_b < 1$. Assuming that speed is spatially constant throughout the link, we can compute the corresponding continuous-time crossing fractions at the detectors' location as

$$
\hat{q}_{i,p,h}^b(t) = \frac{q_{i,p,h}^b}{1 + \alpha_k \cdot l_b}, \quad \text{if } (1 + \alpha_k \cdot l_b)t_j + \beta_k \cdot l_b \leq t < (1 + \alpha_k \cdot l_b)t_{j+1} + \beta_k \cdot l_b,
$$

and $\theta_k \leq t_j < t_{j+1} \leq \theta_{k+1}$,

which are also piecewise constant. Figure 1(c) shows the location of a hypothetical detector. We also plot some representative vehicle trajectories (grey dotted lines) to show how the travel time difference between the first and the last vehicle of the packet (solid black lines) is uniformly distributed among the rest of the vehicles. Again, as we do not know the number of each packet’s vehicles, the grey dotted lines represent some hypothetical vehicles. However, the propagation of crossing fractions allows us to find the relative distances of the vehicles of a packet and the packet’s length while it crosses a detector – which, in principle, is the information we need for the estimation of the OD trips. Finally, Figure 1(d) depicts the piecewise constant crossing fractions at the detector’s location. To construct the elements of the network-loading matrix, $q_{i,p,h}^{b,s}$, we can cumulate the continuous-time crossing fractions, $\hat{q}_{i,p,h}^b(t)$, over each detection period, $s$, using (6).

In Section 1 of supplementary material, we include a detailed description of our approach for propagating the crossing fractions and computing the elements of the network-loading matrix. We propagate the flow using an event-based algorithm, which merges and sorts all the possible events that indicate a change in i) the piecewise constant upstream and downstream crossing fractions or ii) the slope of the travel time function for every link $a \in \mathcal{A}$. Under the hypothesis of uniform departures, we first define the crossing fractions on the upstream boundary of each route’s first link as

$$
q_{i,p,h}^a(t) = \begin{cases} 1 & \text{if } \delta_{a,p} > 0, \\ 0 & \text{otherwise} \end{cases}, \quad (h - 1)T_h \leq t < hT_h, \; \forall \; a \in \mathcal{A}_h^+, \; \forall \; n \in \mathcal{O}.
$$

Then, we propagate such disaggregated information through the network links by successively computing the crossing fractions on the downstream boundaries and on the upstream boundary of succeeding links. The crossing fractions, $q_{i,p,h}^a(t)$, on the downstream boundary of link $a$ are computed
via (29). The upstream crossing fractions, $q_{i,p,h}^{a'}(t)$, at each succeeding link of $a \in A_{p}^{-}$ can be defined as

$$q_{i,p,h}^{a'}(t) = \bar{q}_{i,p,h}^{a'}(t), \quad \forall \, a' \in A_{p}^+, \, p \in P_i \mid \delta_{a,p} \cdot \delta_{a',p} > 0. \quad (33)$$

The turning fractions are not required here because the crossing fractions carry route-disaggregated information.

### 3.2. Modelling the route choices

On the one hand, the underlying presumption on travellers behaviour of the RUM-based route choice estimations might not fully cover the ‘real’ choices. On the other hand, although the purely FCD-driven route choice estimations may be adequate for high penetration rates, scaling up observed individual choices may yield biased estimations for low penetration rates. In this regard, Sohn and Kim (2008)
suggest that RUM should be preferred when the Average Penetration Rate (APR) is less than a pre-
defined threshold. Let $\gamma_{i,h}$ be the APR of PVs for OD pair $i$ and demand period $h$, and $\gamma^*$ the threshold that defines the relative reliability of the FCD-driven route choices. Then, the route choice proportions, $r_{i,p,h}$, can be found as

$$r_{i,p,h} = \begin{cases} r_{i,p,h}^{\text{RUM}} & \text{if } \gamma_{i,h} \leq \gamma^*, \\ r_{i,p,h}^{\text{FCD}} & \text{otherwise} \end{cases}, \quad \forall p \in \mathcal{P}_i, \; i = 1, 2, \ldots, l, \; h = 1, 2, \ldots, H, \quad (34)$$

where $r_{i,p,h}^{\text{RUM}}$ is given by (11), and $r_{i,p,h}^{\text{FCD}}$ by (12).

In this paper, we propose the relaxation of the strict predefined threshold, $\gamma^*$, to achieve smoother transitions from the model-driven to the data-driven route choice estimation. The route choice proportions are given as Weighted Average (WA)

$$r_{i,p,h}^{\text{WA}} = (1 - v_{i,h}(\gamma_{i,h})) \cdot r_{i,p,h}^{\text{RUM}} + v_{i,h}(\gamma_{i,h}) \cdot r_{i,p,h}^{\text{FCD}}, \quad (35)$$

where the weight $v_{i,h}$ is given as a sigmoid function of the estimated APR $\gamma_{i,h}$

$$v_{i,h} = \frac{1}{2} \left( 1 + \tanh \left( \frac{\gamma^* - \gamma_{i,h}}{\Delta \gamma} \right) \right), \quad (36)$$

with $\Delta \gamma$ being the transition width. Hence, the contribution of individual observations to the estimation depends on the APR, while the RUM is employed to keep the route choices at ‘reasonable’ levels for both low and high penetration rates.

As the OD demand is unknown, we cannot explicitly calculate the OD pair-specific APR. However, assuming that the time instants when each PV crosses a detector are known, we can approximate the corresponding APRs. Let $\hat{\gamma}_{b,s}$ represent the fraction of detected PVs to the total vehicle count, $\hat{\gamma}_{b,s}$, for detector $b \in \mathcal{B}$ and during detection period $s \in \mathcal{S}$. We can then use the crossing fractions $q_{i,p,h}^{d,s}$ to map the detector-specific PV fractions to the OD pair-specific penetration rates as

$$\gamma_{i,h} = \frac{\sum_{p \in \mathcal{P}_i} \sum_{b \in \mathcal{B}} \sum_{s=1}^{S} q_{i,p,h}^{b,s} \cdot \hat{\gamma}_{b,s}}{|\mathcal{P}_i| \cdot \sum_{b \in \mathcal{B}} \sum_{s=1}^{S} q_{i,p,h}^{b,s}}, \quad (37)$$

Thus, we estimate the APR of each OD pair as the weighted average of the detector-specific PV fractions, $\hat{\gamma}_{b,s}$, accounting, though, only for the detectors at the routes connecting that OD pair. Equation (37) implies that the detector-specific PV fractions are uniformly distributed among the various route flows that cross a detector. We should highlight here that although this approach for estimating APRs captures temporal variations of data availability, it cannot reflect possible heterogeneities of APR among different OD pairs.

Finally, we also develop a method for estimating the experienced route travel times, which commonly constitute the explanatory variables in the RUM. By definition, the vehicle-average link travel time connotes the fraction of the total travel time spent on that link to the total number of vehicles traversing that link. Therefore, we can express the demand period-average link travel times as

$$\bar{\tau}_{i,p,h}^a = \frac{\int_0^T \left( q_{i,p,h}^a(t) \cdot r_{i,p,h} \cdot x_{i,h} \cdot \tau_a(t) \right) dt}{\int_0^T q_{i,p,h}^a(t) \cdot r_{i,p,h} \cdot x_{i,h} dt}, \quad a \in \mathcal{A}, \; p \in \mathcal{P}_i, \; i = 1, 2, \ldots, l, \; h = 1, 2, \ldots, H. \quad (38)$$

Since the crossing fractions, $q_{i,p,h}^a(t)$, are piecewise constant and the travel time function, $\tau_a(t)$, piecewise linear, we can compute the definite integrals in a straightforward manner. The estimation of
the travel times is embedded in the event-based algorithm described in Section 1 of supplementary material. Then, the experienced route travel time is given by

$$\hat{\tau}_{i,p,h} = \sum_{a \in A} \hat{\tau}_{i,p,h}^a.$$  \hfill (39)

By making the conventional assumption that the route choices follow a C-Logit model (Cascetta et al. 1996), we can estimate the probability of route $p$ as

$$p^{RUM}_{i,h,p} = \frac{\exp\left(\hat{\theta}(\hat{\tau}_{i,p,h} - \hat{\beta} \cdot CF_p)\right)}{\sum_{p' \in P_i} \exp\left(\hat{\theta}(\hat{\tau}_{i,p',h} - \hat{\beta} \cdot CF_{p'})\right)}.$$  \hfill (40)

where $CF_p$ and $CF_{p'}$ are the commonality factors of routes $p$ and $p'$, respectively, and $\hat{\theta}$, $\hat{\beta}$ are parameters to be estimated. The commonality factor can be defined as

$$CF_p = \ln \sum_{p' \in P_i} \left(\frac{L_{p,p'}}{\sqrt{L_p L_{p'}}}\right)^{\hat{\gamma}},$$  \hfill (41)

where $L_p$ and $L_{p'}$ are the lengths of routes $p$ and $p'$, respectively, $L_{p,p'}$ the common length between routes $p, p'$ and $\hat{\gamma}$ a parameter to be estimated (Prato 2009).

4. Experimental setup

To test the performance of our approach, we perform simulation-based experiments using the traffic simulator Aimsun (TSS n.d.) for two networks:

- a small grid toy network, and
- a larger corridor network from the city of Pasadena in Los Angeles, United States.

Aimsun is employed to assign the groundtruth demand, $\tilde{x}$, to the network, emulate the traffic flow propagation and generate synthetic data. We use Aimsun to synthesise the flow-counts, $\hat{y}$, and the FCD, $\tilde{z}$, that correspond to the groundtruth demand, $\tilde{x}$. Then, we consider different approaches for estimating the exogenous assignment matrix, $A(\tilde{z}) = Q(\tilde{z}) R(\tilde{z})$. In particular, we construct the network loading matrix based on three different methods: the QSNL, the PPNL and the DDNL, resulting in $Q^{QSNL}$, $Q^{PPNL}$ and $Q^{DDNL}$, respectively. Accordingly, we consider two different approaches for estimating the route choices: the RUM-based method and the fusion WA, leading to $R^{RUM}$ and $R^{WA}$, respectively. The assignment matrix is used to estimate the demand vector $x^*$, which is then compared to the groundtruth demand $\tilde{x}$. The flowchart illustrated in Figure 2 shows the experimental and evaluation framework used in this study. Each step is further described in the following subsections.

4.1. Solving the DODME problem

In our experiments, we follow formulation (10), assuming that a ‘reasonable’ guess, $\hat{x}$, of the real demand, $\tilde{x}$, is available a priori. For Euclidean distances, we can express (10) as a non-negative ordinary least squares (NNLSQ) problem

$$x^* = \arg \min_{x \geq 0} \left[ \zeta_1 \| \hat{y} - Ax \|_2^2 + \zeta_2 \| \hat{x} - x \|_2^2 \right].$$  \hfill (42)

Problem (42) can be solved using standard techniques for quadratic programming or NNLSQ solvers such as the active-set method proposed by Lawson and Hanson (1995). For high dimensions, the stochastic gradient descent method suggested by Ma and Qian (2018) could be used. Alternatively,
since the assignment matrix typically is very sparse (Bierlaire and Crittin 2004), one could exploit the algorithm for sparse least squares suggested by Paige and Saunders (1982). Generally, any NNLSQ solver can be applied – depending on the needs and the dimension of the problem – and the development of an NNLSQ solution algorithm lies beyond the scope of this paper. In our experiments, we deploy the *lsqin* solver of Matlab (MATLAB 2018), and in particular, the reflective Newton method, which is developed by Coleman and Li (1996) for box-constrained quadratic programming. This solver proved to be adequate for this paper’s needs.

### 4.2. Prior OD matrix

The structural similarity between the real OD matrix, \( \tilde{x} \), and the prior OD matrix, \( \hat{x} \), is a key factor that may significantly affect the estimation quality. Typically, for simulation-based experiments, where it is impossible to derive a prior matrix using conventional techniques (e.g. surveys), the prior matrix is obtained by adding some noise to the groundtruth matrix, \( \tilde{x} \). However, selecting the appropriate perturbations is a challenging task (Antoniou et al. 2016). The prior OD matrix should adequately reflect the possible errors that might be introduced by the conventional estimation approaches (e.g. outdated surveys), providing at the same time some kind of structural information. In this paper, we construct the prior matrix following the approach that Antoniou et al. (2016) suggest in their generic benchmarking platform for ODME approaches. We add random fluctuations to the groundtruth matrix, considering three scenarios:

- **a low demand scenario**, where we assume that the prior ‘guess’ underestimates the actual demand:
  \[
  \hat{x}_{ih}^L = \tilde{x}_{ih} (0.6 + 0.4\xi_{ih}) , \quad \xi_{ih} \sim U(0, 1),
  \]

- **a medium demand scenario**, where we assume that the mean value of the prior ‘guess’ corresponds to the mean of the groundtruth demand (with a wider fluctuation range than the high and low demand scenarios):
  \[
  \hat{x}_{ih}^M = \tilde{x}_{ih} (0.6 + 0.8\xi_{ih}) , \quad \xi_{ih} \sim U(0, 1),
  \]

- **and a high demand scenario**, where we assume that the prior ‘guess’ overestimates the actual demand:
  \[
  \hat{x}_{ih}^H = \tilde{x}_{ih} (1 + 0.4\xi_{ih}) , \quad \xi_{ih} \sim U(0, 1).
  \]
4.3. FCD availability scenarios

We assume that a portion of the simulated vehicles (virtual PVs) is equipped with a sensing device able to provide the vehicle’s travel time for each link of its route. Such link travel times constitute synthetic ‘observations’ that form the feature vector for the estimation of the piecewise linear function. Moreover, the link-specific travel times reveal the route taken by each PV, allowing us to compute the FCD-driven route choice vector. Furthermore, we match each PV to the detectors to construct the detector-specific PV fractions. To examine the effect of the APR, we consider seven different levels of APR: 0.1%, 0.25%, 0.5%, 1%, 2.5%, 5%, 10%. Since vehicle sampling is a random process and might influence the results (especially for low penetration rates), we run 20 replications for each APR and prior-demand scenario (constructing the prior matrix also involves random fluctuations). To construct the sigmoid function that determines the weights of the model-based and data-driven route choice, we consider three different sets for the values of the penetration rate threshold, \( \gamma^* \), and transition width, \( \Delta \gamma : [\gamma^*, \Delta \gamma] = [1%, 0.5%], [\gamma^*, \Delta \gamma] = [5%, 3%], \) and \( [\gamma^*, \Delta \gamma] = [8%, 4%] \).

4.4. Detection layout

Selecting the number and location of the detectors to generate the synthetic flow counts, \( \hat{y} \), is a difficult task as the detection layout strongly influences the estimation quality (Hai Yang and Zhou 1998; Castillo et al. 2008; Barceló et al. 2012). In this paper, we determine the detection layout according to the location rules proposed by Hai Yang and Zhou (1998): under a certain number of detectors, the detection layout should maximise the number of OD pairs covered and the flow intercepted, yielding a binary linear programming problem. To solve this problem, we use Cplex solver for binary integer programming problems (cplexbip, IBM 2019). Given the estimated detection layout, we synthesise the flow counts by making the assumption of error-free measurements.

4.5. Evaluating the quality of the assignment matrix

First, aiming at isolating the effect of the loading matrix, we assume that a mechanism which ‘perfectly’ models route choices is available. Let \( \hat{R} \) be such a perfect route choice matrix (we constructed this matrix considering the route choices of every simulated vehicle). To evaluate our propagation technique, we compare the assignment matrix \( A = Q^{DDNL} \hat{R} \) to the assignment matrices \( A = Q^{QSNL} \hat{R} \) and \( A = Q^{PPNL} \hat{R} \). Next, to evaluate the route choice mechanism, we compute and compare two different assignment matrices, which are derived by considering two different route choice matrices \( R^{RUM} \) and \( R^{WA} \), while keeping the same loading matrix \( Q^{DDNL} \). Therefore, in total, we consider five alternative approaches to construct the assignment matrix (see Table 1).

Ideally, assigning the groundtruth demand vector, \( \hat{x} \), to the network, under a ‘perfect’ assignment matrix (which accurately captures all the underlying flow propagation and route choices effects)
should satisfy the detected flows, \( \hat{y} \). Thus, the deviations between the modelled flows, \( y = A\hat{x} \), and the detected flows, \( \hat{y} \), indicate the quality of the corresponding assignment matrix, \( A \). As a first evaluation stage, we compare the modelled to the detected flows for each of the five estimation approaches presented in Table 1. We use a root mean square error (RMSE), a mean absolute error (MAE) as performance measures:

\[
RMSE_y = \sqrt{\frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} \sum_{s=1}^{5} (y_{b,s} - \hat{y}_{b,s})^2},
\]

\[
MAE_y = \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} \sum_{s=1}^{5} |y_{b,s} - \hat{y}_{b,s}|.
\]

As a second evaluation stage, we examine the effect of the assignment matrix on the DODME. We solve the NNLSQ Problem (42) for the five alternative assignment matrix estimation approaches, and we evaluate the estimated OD demand based on the same performance measures,

\[
RMSE_x = \sqrt{\frac{1}{I \cdot H} \sum_{i=1}^{I} \sum_{h=1}^{H} (x_{i,h} - \tilde{x}_{i,h})^2},
\]

\[
MAE_x = \frac{1}{I \cdot H} \sum_{i=1}^{I} \sum_{h=1}^{H} |x_{i,h} - \tilde{x}_{i,h}|,
\]

and the total demand deviation (TDD) (Chen et al. 2012),

\[
TDD = \frac{\sum_{i=1}^{I} \sum_{h=1}^{H} x_{i,h} - \sum_{i=1}^{I} \sum_{h=1}^{H} \hat{x}_{i,h}}{\sum_{i=1}^{I} \sum_{h=1}^{H} \hat{x}_{i,h}} \cdot 100.
\]

Finally, for each estimation approach, we compute the \( RMSE_y \) again, but for the link flows \( y \) induced by assigning the estimated demand, i.e. \( y = A\hat{x}^* \). This measure indicates the distance between the link flows caused by the estimated demand and the groundtruth link flows.

5. Results

5.1. Network type 1: grid network

Table 2 shows the main simulation parameters for the first simulation-based experiment performed for the grid network illustrated in Figure 3. The network consists of 126 links, 30 nodes and 122 OD pairs. Note that the centroids in the south attract/generate trips only from/to the centroids in the north, and the same applies for the west–east centroids. In this first experiment, we set the number of detectors to 12 (they yield 100% OD pair coverage and 99.03% flow coverage). Figure 3 depicts their optimal location.

The three-hour analysis period is divided into 18 equal-length demand periods (10 min each). Figure 4 depicts the total demand per demand period. We construct a highly fluctuating demand profile to test how the various approaches behave in abrupt traffic conditions changes. The within demand period variations are also high, as we randomly distribute the total demand differences among the OD pairs. Figure 4 also illustrates the average link travel time profile, which indicates that the demand perturbations yield congestion in the first half of the analysis period. The central links, where most of the OD trips pass through, face severe congestion, and queues propagate to several upstream links.
The demand illustrated in Figure 4 is then assigned to the network using Aimsun. First, we deploy the hybrid (meso-micro) simulation module of Aimsun (Barceló and Casas 2005) to find the dynamic UE routes via the method of successive averages. Then, we run a one-shot simulation using the stochastic route choice module of Aimsun, where a part of the users follow the UE routes (we assume that they represent the experienced network users) while a part of the users follow a RUM (we assume that they have some kind of information on the prevailing traffic conditions before their trip). Furthermore, a portion of the network users are able to continuously evaluate their cost during the trip and change their choice en route. We add a high degree of stochasticity in the route choice to reflect the inherent complexity of the real route choices. Aimsun is also employed to generate the global route choice set that contains the routes actually used for each OD pair. In this first experiment, we consider 393 routes for the 122 OD pairs ($|\mathcal{P}| = 393, I = 122$).

For each APR level, we randomly sample the corresponding percentage of simulated vehicles 20 times to synthesise 20 FCD sets. Then, for each FCD set, we construct the network-loading and route choice matrices. Table 3 shows the number of events (as the mean of the 20 replications) for propagating the crossing fractions through the DDNL approach. The number of events, which also determines the number of vehicle packets, increases as the APR gets higher. High APRs yield a more detailed description of the travel time function, implying an increased number of linear segments. Such travel time function might enhance the accuracy but increases the computational burden. Regarding the route choice matrix, the parameters of the C-Logit model, $\hat{\theta}, \hat{\beta}$, which define the sensitivity of the route choices to travel times and the impact of the commonality factors, respectively, are set to their optimal values (such that the deviations between the estimated route choices, $\hat{R}^{RUM}$, and the groundtruth ones, $\tilde{R}$, are minimised). We assume that a sufficient amount of route choice observations from previous days are available, allowing us to ‘perfectly’ calibrate the C-Logit, considering, though, a single value of $\hat{\theta}, \hat{\beta}$ for all OD pairs. The parameter $\hat{\gamma}$ is set to one. Moreover, we assume that $\xi_1 = \xi_2 = 1$ in (42).
Table 3. Number of events for the propagation of the crossing fractions.

<table>
<thead>
<tr>
<th>APR (%)</th>
<th>Events ($\times 10^4$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>3.01</td>
</tr>
<tr>
<td>0.25</td>
<td>3.16</td>
</tr>
<tr>
<td>0.50</td>
<td>3.37</td>
</tr>
<tr>
<td>1.00</td>
<td>3.93</td>
</tr>
<tr>
<td>2.50</td>
<td>5.53</td>
</tr>
<tr>
<td>5.00</td>
<td>7.90</td>
</tr>
<tr>
<td>10.00</td>
<td>12.44</td>
</tr>
</tbody>
</table>

Table 4. Link flow differences for the assignment matrix estimation approaches QSNL, PPNL and DDNL ($y = Ax$).

<table>
<thead>
<tr>
<th>A = Q^{QSNL} R</th>
<th>$\text{RMSE}_y$ (veh)</th>
<th>$\text{MAE}_y$ (veh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APR (%)</td>
<td>0.10</td>
<td>15.81</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>9.85</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>15.77</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>15.83</td>
</tr>
<tr>
<td></td>
<td>2.50</td>
<td>15.63</td>
</tr>
<tr>
<td></td>
<td>5.00</td>
<td>15.24</td>
</tr>
<tr>
<td></td>
<td>10.00</td>
<td>14.35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A = Q^{PPNL} R</th>
<th>$\text{RMSE}_y$ (veh)</th>
<th>$\text{MAE}_y$ (veh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APR (%)</td>
<td>0.10</td>
<td>9.84</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>9.88</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>9.81</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>9.63</td>
</tr>
<tr>
<td></td>
<td>2.50</td>
<td>9.35</td>
</tr>
<tr>
<td></td>
<td>5.00</td>
<td>9.17</td>
</tr>
<tr>
<td></td>
<td>10.00</td>
<td>9.17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A = Q^{DDNL} R</th>
<th>$\text{RMSE}_y$ (veh)</th>
<th>$\text{MAE}_y$ (veh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APR (%)</td>
<td>0.10</td>
<td>8.33</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>8.19</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>7.79</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>7.17</td>
</tr>
<tr>
<td></td>
<td>2.50</td>
<td>5.81</td>
</tr>
<tr>
<td></td>
<td>5.00</td>
<td>5.29</td>
</tr>
<tr>
<td></td>
<td>10.00</td>
<td>4.90</td>
</tr>
</tbody>
</table>

5.1.1. Evaluation stage 1: link flow differences

Table 4 presents the flow RMSEs and MAEs (in terms of numbers of vehicles per detection period) as the mean of the 20 replications for the first three estimation approaches. Note that since we consider ‘perfect’ route choices, the values of Table 4 solely reflect the errors of the underlying network-loading mechanism. Furthermore, note that the QSNL approach does not depend on the APR because it is purely based on graph properties. The lowest APR (0.1%) corresponds to only 12 PVs for the entire analysis period, which apparently are not able to reflect the underlying congestion. Therefore, we can say that the results for 0.1% represent the case where congestion is neglected (propagation based on free-flow times).

In general, the DDNL seems to be more effective at capturing flow-propagation phenomena compared to the point-packet approaches. It results into the lowest RMSE and MAE for every APR level. For APRs less than 0.5%, the PPNL exhibits almost the same performance as the QSNL and the additive value of the realised travel times in propagation appears to be negligible. However, for both PPNL and DDNL, the performance measures seem to improve while APR gets higher. This is reasonable because greater APRs imply more accurate travel time estimates (more PVs are employed to indicate congestion).

Figure 5 shows how the detected flows, $\hat{y}$, relate to the estimated flows, $y = Ax$, for the estimation approaches PPNL and DDNL and for a random replication of 2.5% APR. The point-packet approach (Figure 5(a)) leads to high discrepancies, which are mainly caused by the discrete nature of the flow propagation. Such differences between the detected and the estimated flows seem to grow as link flow gets higher due to the (i) increased congestion and (ii) increased number of packets merged at links with high flow. Furthermore, we can notice some outliers at low flows, which may come from the inability of PPNL to capture congestion effects such as spill-back. On the other hand, DDNL (Figure 5(b)) exhibits an improved performance for both low and high flow levels due to the continuous nature of flow propagation and the variable packet length. We can still see some outliers in Figure 5(b) possibly because the 2.5% APR of PVs may not be able to provide sufficient information on the prevailing congestion level.

Table 5 presents the flow RMSEs and MAEs as the mean of the 20 replications for the estimation approaches RUM and WA. The APR also seems to have a great impact here, especially for the WA route choice estimation approach. The effect of APR is less strong (but not negligible) for the RUM-based approach. Again, we can assume here that the APR of 0.1% corresponds to estimations that neglect congestion. Therefore, for APR 0.1% the route choice is based on free-flow route travel times rather than experienced travel times. High APRs lead to more accurate estimations of experienced route travel times. Furthermore, Table 5 shows that our fusion technique can effectively balance the RUM with the FCD-based route choices since the errors for the WA approach are the lowest for almost every APR. Moreover, it seems that the second set of the sigmoid function parameters (5.0%, 3.0%) exhibits the best performance in total. The first set of parameters (1.0%, 0.5%) allows low PV fractions.
Figure 5. Flow (numbers of vehicles per detection period) scatter plot for an APR of 2.5% and for the estimation approaches (a) PPNL and (b) DDNL.

Table 5. Link flow differences for the assignment matrix estimation approaches RUM and WA ($y = Ax$).

<table>
<thead>
<tr>
<th>APR (%)</th>
<th>0.10</th>
<th>0.25</th>
<th>0.50</th>
<th>1.00</th>
<th>2.50</th>
<th>5.00</th>
<th>10.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A = Q^{DDNL}R^{RUM}$</td>
<td>RMSE$_y$ (veh)</td>
<td>11.54</td>
<td>11.63</td>
<td>11.05</td>
<td>10.76</td>
<td>10.22</td>
<td>9.82</td>
</tr>
<tr>
<td></td>
<td>MAE$_y$ (veh)</td>
<td>7.81</td>
<td>7.97</td>
<td>7.68</td>
<td>7.56</td>
<td>7.34</td>
<td>7.11</td>
</tr>
<tr>
<td>$A = Q^{DDNL}R^{WA}$</td>
<td>RMSE$_y$ (veh)</td>
<td>11.52</td>
<td>11.61</td>
<td>10.99</td>
<td>10.60</td>
<td>9.94</td>
<td>9.25</td>
</tr>
<tr>
<td></td>
<td>MAE$_y$ (veh)</td>
<td>7.80</td>
<td>7.95</td>
<td>7.65</td>
<td>7.43</td>
<td>7.05</td>
<td>6.63</td>
</tr>
<tr>
<td>$[\gamma^*, \Delta\gamma] = [1.0%, 0.5%]$</td>
<td>RMSE$_y$ (veh)</td>
<td>11.53</td>
<td>11.62</td>
<td>11.04</td>
<td>10.71</td>
<td>10.02</td>
<td>9.10</td>
</tr>
<tr>
<td></td>
<td>MAE$_y$ (veh)</td>
<td>7.80</td>
<td>7.96</td>
<td>7.67</td>
<td>7.53</td>
<td>7.18</td>
<td>6.53</td>
</tr>
<tr>
<td>$[\gamma^*, \Delta\gamma] = [5.0%, 3.0%]$</td>
<td>RMSE$_y$ (veh)</td>
<td>11.54</td>
<td>11.62</td>
<td>11.04</td>
<td>10.74</td>
<td>10.13</td>
<td>9.39</td>
</tr>
<tr>
<td></td>
<td>MAE$_y$ (veh)</td>
<td>7.80</td>
<td>7.97</td>
<td>7.68</td>
<td>7.55</td>
<td>7.27</td>
<td>6.77</td>
</tr>
</tbody>
</table>

Table 6 shows how the three alternative flow-propagation mechanisms affect the solution of the DODME problem for three indicative APR levels. We present the results for every APR level in Section 2 in supplementary material. As a consequence of the extended flow discrepancies, the QSNL and PPNL approaches result in higher errors than the DDNL approach. However, even such simplified network-loading settings (QSNL and PPNL) seem to adequately 'correct' the high and low prior demand in terms of total number of trips (under the hypothesis that route choice modelling is 'perfect'). The absolute TDDs for the QSNL, PPNL and DDNL are significantly lower than the prior demand TDD for both low and high demand scenarios. The DDNL approach also provides structural enhancements, leading to lower RMSE and MAE than the prior demand for every APR level and prior-demand scenario (except for the medium demand scenario where the variance of the groundtruth demand fluctuation is higher). The PPNL seems to improve the prior-demand structure (in terms of RMSE) only for APRs higher than 2.5% (for $\gamma_1 = \gamma_2$).

Table 7 presents the performance measures for the two different approaches for obtaining the route choice matrix. We can observe the same patterns as in Table 5: the APR strongly affects the performance of the WA method, while the impact is not so intense for the RUM-based method. The WA approach proposed in this paper exhibits a better overall performance in terms of OD trips as well. Although both approaches effectively reduce the prior-demand TDDs, they seem to deteriorate the structure of the prior demands for low APRs, leading to high RMSEs and MAEs especially for the...
Even though this is a real-life network and we will have a greater impact than the route choice. In this second experiment, we set the number of detectors to 350. Figure 6 depicts their optimal location that maximises the OD pairs and flow coverage (99.50% coverage of OD pairs and 99.49% coverage of flow). The network consists of 4317 links, 2004 nodes and 7944 OD pairs, which are connected by 12607 routes in total. There is only one route available for several OD pairs connected via the main corridor of the network (part of the freeway I-210). Hence, we expect that in this network, the network loading will have a greater impact than the route choice. In this second experiment, we set the number of detectors to 350. Figure 6 depicts their optimal location that maximises the OD pairs and flow coverage (99.50% coverage of OD pairs and 99.49% coverage of flow). Even though this is a real-life network and

## 5.2. Network type 2: pasadena corridor network

The morning peak (7:00-11:00) of a typical day is simulated using Aimsun for the Pasadena network illustrated in Figure 6. The four-hour analysis period is divided into 24 equal-length demand periods (10 minutes each). Table 8 shows the main parameters for the second simulation-based experiment. The network consists of 4317 links, 2004 nodes and 7944 OD pairs, which are connected by 12607 routes in total. There is only one route available for several OD pairs connected via the main corridor of the network (part of the freeway I-210). Hence, we expect that in this network, the network loading will have a greater impact than the route choice. In this second experiment, we set the number of detectors to 350. Figure 6 depicts their optimal location that maximises the OD pairs and flow coverage (99.50% coverage of OD pairs and 99.49% coverage of flow). Even though this is a real-life network and

### Table 6. OD trip differences for the assignment matrix estimation approaches QSNL, PPNL and DDNL.

<table>
<thead>
<tr>
<th>Prior demand, (\hat{x})</th>
<th>(A = Q^{QSNL}_R)</th>
<th>(A = Q^{PPNL}_R)</th>
<th>(A = Q^{DDNL}_R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE(_x) (veh)</td>
<td>1.46</td>
<td>1.47</td>
<td>1.46</td>
</tr>
<tr>
<td>TTD (%)</td>
<td>0.10</td>
<td>19.91</td>
<td>0.10</td>
</tr>
<tr>
<td>RMSE(_x) (veh)</td>
<td>1.14</td>
<td>1.14</td>
<td>1.14</td>
</tr>
<tr>
<td>APR (%)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.14</td>
</tr>
<tr>
<td>Scenario 1: (\hat{x}_1)</td>
<td>Scenario 2: (\hat{x}_2)</td>
<td>Scenario 3: (\hat{x}_3)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 7. OD trip differences for the assignment matrix estimation approaches RUM and WA.

<table>
<thead>
<tr>
<th>Prior demand, (\hat{x})</th>
<th>(A = Q^{QSNL}_R)</th>
<th>(A = Q^{PPNL}_R)</th>
<th>(A = Q^{DDNL}_R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE(_x) (veh)</td>
<td>1.46</td>
<td>1.47</td>
<td>1.46</td>
</tr>
<tr>
<td>TTD (%)</td>
<td>1.46</td>
<td>1.46</td>
<td>1.46</td>
</tr>
<tr>
<td>RMSE(_x) (veh)</td>
<td>1.14</td>
<td>1.14</td>
<td>1.14</td>
</tr>
<tr>
<td>APR (%)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.14</td>
</tr>
<tr>
<td>Scenario 1: (\hat{x}_1)</td>
<td>Scenario 2: (\hat{x}_2)</td>
<td>Scenario 3: (\hat{x}_3)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 8. Main parameters for the second simulation-based experiment.

<table>
<thead>
<tr>
<th>Scenario 1: (\hat{x}_1)</th>
<th>Scenario 2: (\hat{x}_2)</th>
<th>Scenario 3: (\hat{x}_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE(_x) (veh)</td>
<td>1.46</td>
<td>1.47</td>
</tr>
<tr>
<td>TTD (%)</td>
<td>0.10</td>
<td>19.91</td>
</tr>
<tr>
<td>RMSE(_x) (veh)</td>
<td>1.14</td>
<td>1.14</td>
</tr>
<tr>
<td>APR (%)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

### Medium prior-demand scenario. In particular, both RUM and WA yield satisfactory results when the APR is greater than 2.5% for the high and low prior-demand scenarios (again for \(\xi_1 = \xi_2\)).
Table 8. Simulation experiment 2 – Pasadena network.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation time, $T$</td>
<td>14,400 s</td>
</tr>
<tr>
<td>Demand period duration, $T_h$</td>
<td>600 s</td>
</tr>
<tr>
<td>Counting period duration, $T_s$</td>
<td>600 s</td>
</tr>
<tr>
<td>Total demand, $\sum_{i=1}^{N_l} \sum_{h=1}^{N_h} \tilde{x}_{i,h}$</td>
<td>419,505 veh</td>
</tr>
</tbody>
</table>

Table 9. Number of events for the propagation of the crossing fractions.

<table>
<thead>
<tr>
<th>APR (%)</th>
<th>Events ($\times 10^6$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>3.88</td>
</tr>
<tr>
<td>0.25</td>
<td>4.41</td>
</tr>
<tr>
<td>0.50</td>
<td>4.94</td>
</tr>
<tr>
<td>1.00</td>
<td>5.62</td>
</tr>
<tr>
<td>2.50</td>
<td>6.31</td>
</tr>
<tr>
<td>5.00</td>
<td>6.38</td>
</tr>
<tr>
<td>10.00</td>
<td>6.12</td>
</tr>
</tbody>
</table>

detectors are already installed to several links for other purposes (their number is greater than 350), we consider only a subset of the existing detectors. We want to test our approach under the hypothetical scenario of marginally covering the OD pairs. Figure 7 depicts the total demand and average link travel time per demand period. The demand is assigned to the network using Aimsun, under the same route choice settings applied to the first simulation experiment.

Similar to the first simulation experiment, we randomly sample the simulated vehicles to synthesise 20 FCD data sets for each APR level. Table 9 shows the average number of events when DDNL is used to propagate the crossing fractions.

5.2.1. Evaluation stage 1: link flow differences

Table 10 presents the flow RMSEs and MAEs as the mean of the 20 replications for the first three estimation approaches. The DDNL exhibits considerably lower errors than the QSNL and the PPNL also in this second experiment. For both PPNL and DDNL, increasing the penetration rate seems to have a
marginal effect for APRs greater than 2.5%. The lowest APR, 0.1%, corresponds to 419 vehicles, which might be enough to provide some congestion indications, especially for the OD pairs that are connected via the main corridor (they are associated with the high demand levels). Figure 8(a,b) show how the detected flows, $\hat{y}$, relate to the estimated flows, $y = A\tilde{x}$, for the PPNL and DDNL approach, respectively, and for a random replication of 2.5% APR.

Table 11 presents the flow RMSEs and MAEs as the mean of the 20 replications for the estimation approaches RUM and WA. We can notice approximately the same patterns as the first experiment: (i) The APR effect is less evident for the RUM compared to the WA approach, and (ii) the WA errors are lower than the RUM ones for every APR level. However, we observe a significant difference compared to the first experiment: the first set of the sigmoid function parameters (1.0%, 0.5%) exhibits the best performance for every APR. The second and third sets of parameters give low weights to the FCD-driven route choice when the penetration rates are less than 2.5%. However, an APR of 2.5% makes up around 10500 vehicles, which might provide essential route choice information, at least for the main OD pairs connected through the corridor. Finally, for 10% APR the impact of route choice errors is almost marginal since the WA errors are slightly higher than the errors associated with a ‘perfect’ route choice (see Table 10).

5.2.2. Evaluation stage 2: OD trip differences

Table 12 shows how estimation approaches QSNL, PPNL and DDNL impact the solution of the DODME problem. Similar to the first experiment, the DDNL estimation approach exhibits a remarkably better performance than the QSNL and PPNL approach. All three estimation approaches are able to effectively ‘correct’ the prior demand in terms of total number of trips irrespectively of the APR level. The DDNL seems to enhance also the structure of the prior demand (it improves the RMSEs for the high and low prior-demand scenarios). As in the first experiment, the RMSEs are higher for the medium prior-demand scenario possibly because of the wider and anisotropic fluctuations added to the groundtruth
Table 11. Link flow differences for the assignment matrix estimation approaches RUM and WA ($y = Ax$).

<table>
<thead>
<tr>
<th>APR (%)</th>
<th>0.10</th>
<th>0.25</th>
<th>0.50</th>
<th>1.00</th>
<th>2.50</th>
<th>5.00</th>
<th>10.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A = Q^{\text{DDNL}} R^\text{RUM}$</td>
<td>RMSE$_y$ (veh)</td>
<td>27.15</td>
<td>26.25</td>
<td>25.78</td>
<td>25.53</td>
<td>25.27</td>
<td>25.23</td>
</tr>
<tr>
<td></td>
<td>MAE$_y$ (veh)</td>
<td>12.41</td>
<td>12.06</td>
<td>11.89</td>
<td>11.79</td>
<td>11.68</td>
<td>11.63</td>
</tr>
<tr>
<td>$A = Q^{\text{DDNL}} R^\text{WA}$</td>
<td>RMSE$_y$ (veh)</td>
<td>27.14</td>
<td>26.19</td>
<td>25.60</td>
<td>24.76</td>
<td>24.57</td>
<td>24.11</td>
</tr>
<tr>
<td>$[\gamma^*, \Delta \gamma] = [1.0%, 0.5%]$</td>
<td>MAE$_y$ (veh)</td>
<td>12.40</td>
<td>12.04</td>
<td>11.83</td>
<td>11.79</td>
<td>11.68</td>
<td>11.63</td>
</tr>
<tr>
<td>$A = Q^{\text{DDNL}} R^\text{WA}$</td>
<td>RMSE$_y$ (veh)</td>
<td>27.13</td>
<td>26.22</td>
<td>25.73</td>
<td>25.41</td>
<td>24.97</td>
<td>24.41</td>
</tr>
<tr>
<td>$[\gamma^*, \Delta \gamma] = [5.0%, 3.0%]$</td>
<td>MAE$_y$ (veh)</td>
<td>12.40</td>
<td>12.05</td>
<td>11.85</td>
<td>11.71</td>
<td>11.44</td>
<td>11.33</td>
</tr>
<tr>
<td>$A = Q^{\text{DDNL}} R^\text{WA}$</td>
<td>RMSE$_y$ (veh)</td>
<td>27.14</td>
<td>26.22</td>
<td>25.73</td>
<td>25.41</td>
<td>24.97</td>
<td>24.41</td>
</tr>
<tr>
<td>$[\gamma^*, \Delta \gamma] = [8.0%, 4.0%]$</td>
<td>MAE$_y$ (veh)</td>
<td>12.41</td>
<td>12.06</td>
<td>11.87</td>
<td>11.75</td>
<td>11.58</td>
<td>11.33</td>
</tr>
</tbody>
</table>

Table 12. OD trip differences for the assignment matrix estimation approaches QSNL, PPNL and DDNL.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$\hat{x}^L$</th>
<th>$\hat{x}^H$</th>
<th>$\hat{x}^M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior demand, $\hat{x}$</td>
<td>RMSE$_x$ (veh)</td>
<td>1.39</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td>MAE$_x$ (veh)</td>
<td>0.55</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>TTD (%)</td>
<td>$-19.99$</td>
<td>$19.98$</td>
</tr>
<tr>
<td>$A = Q^{\text{QSNL}} R$</td>
<td>RMSE$_x$ (veh)</td>
<td>1.49</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>MAE$_x$ (veh)</td>
<td>0.63</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>TTD (%)</td>
<td>$-3.52$</td>
<td>$-3.80$</td>
</tr>
<tr>
<td>$y = Ax^*$</td>
<td>APR (%)</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>RMSE$_y$ (veh)</td>
<td>1.49</td>
<td>1.41</td>
</tr>
<tr>
<td>$A = Q^{\text{PPNL}} R$</td>
<td>MAE$_x$ (veh)</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>TTD (%)</td>
<td>$-3.33$</td>
<td>$-3.23$</td>
</tr>
<tr>
<td>$y = Ax$</td>
<td>RMSE$_y$ (veh)</td>
<td>1.32</td>
<td>1.31</td>
</tr>
<tr>
<td>$A = Q^{\text{DDNL}} R$</td>
<td>MAE$_x$ (veh)</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>TTD (%)</td>
<td>$-3.33$</td>
<td>$-3.23$</td>
</tr>
</tbody>
</table>

In this second experiment, we notice a completely different MAE pattern. The DDNL results in higher MAEs than the prior demand MAEs, even though the corresponding RMSEs are considerably lower (again for $\gamma_1 = \gamma_2$). Since RMSE heavily penalises large errors (Buisson et al. 2014), such discrepancies imply that our approach ‘corrects’ the prior demand for OD pairs with a high number of trips. In contrast, the prior demand exhibits a close match to the groundtruth demand for OD pairs with a low number of trips. This is reasonable considering how we construct the prior demand: increased fluctuations are added to high-demand OD pairs, while the number of trips for low-demand OD pairs remains close to the groundtruth ones.

Table 13 presents the performance measures in terms of OD trips for the estimation approaches RUM and WA. Both estimation approaches enhance the quality of the prior demand in terms of total OD trips (see TDD measure). Regarding the RMSEs, the APR seems to have a weak impact on the RUM-based estimation. Moreover, in accordance with the link flows errors, the first set of parameters (1%, 0.5%) exhibits the best performance also in terms of OD trips. While the simple RUM approach itself seems unable to improve the prior-demand structure (it marginally improves it for the high-demand scenario), the WA approach exhibits satisfactory performance for APR higher than 2.5%, reducing the corresponding RMSEs.

Figure 9(a) shows how the link flows, $y$, relate to the link counts, $\hat{y}$, if we assign the estimated demand, $x^*$, the groundtruth demand, $\tilde{x}$, and the prior demand, $\hat{x}^H$, to the network, under the assignment matrix $A = Q^{\text{DDNL}} R^\text{WA}$ (APR = 2.5%, $\gamma^* = 1\%$, $\Delta \gamma = 0.5\%$). Solving the optimisation problem (42) minimises the differences between the link counts and the link flows induced by the estimated demand. Such link flows, however, differ from the flows that the groundtruth demand would cause. These deviations represent the errors that an ‘imperfect’ assignment matrix introduces into the estimation process. Figure 9(b) shows how the estimated demand relates to the groundtruth demand, under
Table 13. OD trip differences for the assignment matrix estimation approaches RUM and WA.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>RMSE ( \hat{x} ) (veh)</th>
<th>MAE ( \hat{x} ) (veh)</th>
<th>TTD (%)</th>
<th>APR (%)</th>
<th>RMSE ( \hat{y} ) (veh)</th>
<th>MAE ( \hat{y} ) (veh)</th>
<th>TTD (%)</th>
<th>APR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.39</td>
<td>0.44</td>
<td>-19.99</td>
<td>0.10</td>
<td>1.42</td>
<td>0.64</td>
<td>-4.83</td>
<td>-4.37</td>
</tr>
<tr>
<td>2</td>
<td>1.37</td>
<td>0.44</td>
<td>19.98</td>
<td>0.10</td>
<td>1.37</td>
<td>0.63</td>
<td>1.16</td>
<td>1.14</td>
</tr>
<tr>
<td>3</td>
<td>1.39</td>
<td>0.44</td>
<td>-0.03</td>
<td>1.00</td>
<td>1.37</td>
<td>0.63</td>
<td>2.84</td>
<td>2.82</td>
</tr>
</tbody>
</table>

Figure 9. Results for a random replication of 2.5% APR, estimation approach WA and high-demand scenario: (a) real vs estimated link flows (b) real vs estimated OD trips.

The same estimation setting (APR = 2.5%, \( \gamma^* = 1\% \), \( \Delta_\gamma = 0.5\% \)) and for the high-demand scenario (to avoid confusion, we do not plot the prior demand but only its regression line). At first glance, the improvement of the prior demand might seem to be negligible as the difference between the regression coefficients is relatively low. However, Figure 9(a) shows that assigning the two demand vectors, \( \hat{x}^H \) and \( \hat{x}^* \), to the network yield significantly different patterns in terms of link flows.

The results presented so far in this section are obtained under the hypothesis that \( \gamma_1 = \gamma_2 \). Table 14 shows how the estimation is affected if we set different weights, \( \gamma_1, \gamma_2 \), for the two parts of the objective function in (42). Giving a higher weight to the link flows, i.e. \( \gamma_1 > \gamma_2 \), yields low deviations between the estimated and observed link flows. The estimation is also improved in terms of TDD because the contribution of the under- or overestimated prior demand is not so strong. On the other hand, giving greater weight to the prior demand, i.e. \( \gamma_1 < \gamma_2 \), can improve the MAEs, especially for low penetration rates.
Table 14. OD trip differences for the assignment matrix estimation approach WA and for γ∗ = 1%, Δγ = 0.5%.

<table>
<thead>
<tr>
<th>Prior demand, x</th>
<th>Scenario 1: xL</th>
<th>Scenario 2: xH</th>
<th>Scenario 3: xM</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSEx (veh)</td>
<td>1.40</td>
<td>1.38</td>
<td>1.39</td>
</tr>
<tr>
<td>MAEx (veh)</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>TTD (%)</td>
<td>−20.03</td>
<td>19.97</td>
<td>−0.03</td>
</tr>
<tr>
<td>APR (%)</td>
<td>0.10</td>
<td>0.250</td>
<td>0.10</td>
</tr>
<tr>
<td>RMSEx (veh)</td>
<td>1.71</td>
<td>1.38</td>
<td>1.39</td>
</tr>
<tr>
<td>MAEx (veh)</td>
<td>0.57</td>
<td>0.61</td>
<td>0.65</td>
</tr>
<tr>
<td>TTD (%)</td>
<td>−4.52</td>
<td>−3.38</td>
<td>−0.03</td>
</tr>
<tr>
<td>RMSEx (veh)</td>
<td>0.09</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>MAEx (veh)</td>
<td>1.17</td>
<td>1.15</td>
<td>1.32</td>
</tr>
<tr>
<td>TTD (%)</td>
<td>−4.12</td>
<td>−0.67</td>
<td>1.19</td>
</tr>
<tr>
<td>RMSEx (veh)</td>
<td>1.71</td>
<td>1.38</td>
<td>1.39</td>
</tr>
<tr>
<td>MAEx (veh)</td>
<td>0.45</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>TTD (%)</td>
<td>−1.79</td>
<td>−0.71</td>
<td>1.19</td>
</tr>
<tr>
<td>RMSEx (veh)</td>
<td>0.09</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>MAEx (veh)</td>
<td>1.17</td>
<td>1.15</td>
<td>1.32</td>
</tr>
<tr>
<td>TTD (%)</td>
<td>−1.32</td>
<td>−0.36</td>
<td>1.19</td>
</tr>
<tr>
<td>RMSEx (veh)</td>
<td>0.76</td>
<td>0.72</td>
<td>0.61</td>
</tr>
<tr>
<td>MAEx (veh)</td>
<td>0.57</td>
<td>0.61</td>
<td>0.65</td>
</tr>
<tr>
<td>TTD (%)</td>
<td>−0.71</td>
<td>−0.03</td>
<td>1.19</td>
</tr>
<tr>
<td>RMSEx (veh)</td>
<td>1.17</td>
<td>1.15</td>
<td>1.32</td>
</tr>
<tr>
<td>MAEx (veh)</td>
<td>0.45</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>TTD (%)</td>
<td>−1.32</td>
<td>−0.36</td>
<td>1.19</td>
</tr>
</tbody>
</table>

6. Discussion

The numerical results presented in the previous section indicate that our approach for constructing the assignment matrix can improve – under certain settings – the estimation accuracy. As we expected, the detailed description of crossing-fractions propagation can enhance the quality of the assignment matrix. Our approach seems to capture congestion phenomena effectively, even for low penetration rates of PVs. However, the minimum penetration rate required for indicating propagation effects depends on the prevailing congestion level. In terms of route choices, although the WA approaches lead to lower errors, the estimation results strongly rely on the APR. In our experiments, we try three different values for the APR threshold that defines the contribution of the FCD-based route choices. Even though such values are arbitrarily selected, the results show the potential of modelling the route choices using such a threshold. Our results also indicate that the optimal values of the APR threshold depend on the network structure and topology. One crucial factor that also affects the APR threshold is the (perceived) quality of the RUM, i.e. its ability to capture the real route choices.

Undeniably, the assignment matrix is a crucial aspect that affects the DODME, and, hence, it should be carefully constructed. However, the solution of the DODME problem might be influenced by numerous other parameters which are not examined in this paper: the quality of the prior demand, the detection layout and the form of the distance functions, F1, F2. DODME is a multidimensional problem, and it is almost impossible to account for every impacting factor and estimate its relative importance. Nonetheless, although it is difficult to draw clear conclusions, considering also that there is no indisputable OD matrix quality measure (Antoniou et al. 2016), our results indicate that the estimation performance is contingent upon the various approaches of constructing the exogenous assignment (at least in terms of RMSE, MAE and TDD).

The weights, ζ1, ζ2, are one more aspect that can have an effect on the estimation. Initially, we set equal weights, i.e. ζ1 = ζ2. Then, we test various sets of ζ1, ζ2 values (see Table 14), observing that the weights affect the estimation in a dissimilar manner for different APRs. This is reasonable since a more accurate assignment matrix (high APRs) should be given a higher weight ζ1 than a less accurate assignment matrix (low APRs). Therefore, there is actually some information, i.e. the estimated APRs, γi,h, that could implicitly indicate the weighting values. As the quality of the assignment matrix depends on the APR, we could consider that ζ1 is proportional to the estimated penetration rate. This is something we...
consider for future research. We should also bear in mind that the quality level of prior demand, which cannot always be known, plays a key role in selecting the weights.

Let us highlight here the main limitations of this study. Our method is evaluated through simulation-based experiments using error-free synthetic flow counts. However, in reality, the link counts, as well as the travel time observations, typically exhibit some noise. Furthermore, the way in which the traffic simulator defines the route choices might affect the evaluation results. Real-life route choices are inherently complex, and they may be misestimated by the traffic simulator. Despite these limitations, simulation-based experiments provide a solid ground for evaluating ODME approaches. Simulation permits adjusting different parameters, such as the APR and the detection layout, and more importantly, in contrast to field experiments, a groundtruth OD matrix is available.

Moreover, even if perfect information about the route choice and flow propagation is available, our modelling setting might still yield deviations between the estimated and observed link flows. Constant route choices, uniform departure rates and packet-based propagation are assumptions that may be unrealistic, although they establish an efficient modelling structure. Additionally, we assume that there is no congestion on the links where no FCD are available, which might not always be true. Furthermore, in our experiments, we do not study the effect of the route choice set size (number of routes considered for each OD pair) as well as of the route choice set generation approach. As the demand period-specific link travel times are known, we may use a route choice set generation algorithm similar to the one used by Ros-Roca, Bugeda, and Mercadé (2020). This is also something we consider for future research.

Finally, similar to any data-driven OD estimation method (Xianfeng Yang, Lu, and Hao 2017; Ma and Qian 2018; Krishnakumari et al. 2020), our DDNA is free of any supply constraints. Namely, we assign the demand to the network without considering any capacity or jam density constraints. Ideally, the travel time observations should prevent any supply violations, but actually, the travel time is just an approximation of the real travel time. While the estimation keeps the flows at a realistic level for the links with detectors, there is nothing to constrain the flows for the rest of the network links. This issue, which is also discussed by Ma and Qian (2018), could yield capacity violations especially when our prior guess overestimates the demand (as in the high-demand scenario). To test this effect, we assume that a detector is installed at each one of the 4317 links of the Pasadena network, but we use only the 350 of them for estimating the demand (under the high demand scenario and 2.5% APR). Figure 10 shows how the link flows, $y$, of every network link relate to ‘real’ link flows, $\hat{y}$, if we assign the estimated demand, $x^*$, and the prior demand, $\hat{x}^H$. Note that although our estimation is based on a subset of the links, it is able to ‘correct’ the flows for almost every link of the network. While the prior demand leads

Figure 10. Real link flows, $\hat{y}$, vs estimated link flows, $y$, for every network link if we assign the estimated demand, $x^*$, and prior demand, $\hat{x}^H$. 

\[ y = Ax^* \]
\[ y = 0.989x + 2.234 \]
\[ y = Ax^H \]
\[ y = 1.165x + 4.501 \]
\[ y = x \]
to capacity violations for 0.61% of the link-flow estimates, the ‘corrected’ demand causes violations for 0.12% of the link-flow estimates.

7. Conclusions and future directions

In this paper, we suggest and evaluate a data-driven approach for solving the DODME problem, where an iterative traffic assignment process is not required. The main rationale behind this method is the exploitation of the available FCD for defining the prevailing congestion level that governs the route choices and flow propagation. To this end, we develop a novel DDNA approach which in contrast to the conventional DTA does not depend on the demand, allowing the use of efficient solution schemes.

The core element of our method is an exogenous assignment matrix that is given as the product of a network loading and a route choice matrix. Given link-specific travel time observations, we construct a piecewise linear travel time function for each link of the network. Then, the travel time function is employed to define the propagation of the crossing fractions (network loading matrix) and the average travel time of each route. Assuming that the distribution of OD trips over the available routes is subject to the route travel times, we employ a RUM to estimate the route choice matrix. The route choice matrix estimation is also enhanced by fusing individual PV route choices.

Our approach is evaluated through two simulation-based experiments, where we estimate the OD matrix considering seven FCD availability scenarios under the same detection layout and prior demand matrix. The numerical results are encouraging, indicating that under appropriate processing, the FCD might reveal the prevailing congestion and enhance the estimation accuracy. In particular, the network loading matrix seems to adequately capture flow propagation phenomena for penetration rates higher than 2.5%. At the same time, the APR level appears to have a stronger effect on the performance of the route choice mechanism, which is satisfactory for APRs higher than 5%. Essentially, the estimation accuracy improves as the APR gets higher, implying that our approach effectively exploits the available FCD. This is promising considering the continuously growing availability of FCD.

Future research will involve the evaluation of our method using real-life, instead of synthetic, FCD data. Furthermore, we plan to investigate the possibility of incorporating our approach into the online ODME frameworks. In this study, we consider PV penetration rates lower than 10%, which might not be sufficient to reveal any structural information for the OD matrix, even though they partly indicate congestion. However, if higher penetration rates are assumed (Van Aerde et al. 1993; Eisenman and List 2004; Xianfeng Yang, Lu, and Hao 2017), the PVs may provide essential information for constructing an FCD-based OD matrix, which can replace the target matrix in our formulation. Developing approaches for constructing an FCD-based target OD matrix is something we also consider for future research. Moreover, FCD can be exploited for local link flow estimations as suggested by Cao et al. (2013). Such link flow estimates can increase the ratio between the number of equations and the number of unknowns. Furthermore, instead of assuming that the entire trips of PVs are monitored and explicitly considering them in route choice estimation, we could assume that only (macroscopic) turning fractions are available at each network node, similar to the approach suggested by Mitra et al. (2020). The latter assumption might be more realistic. The turning fractions can be embedded in the event-based algorithm allowing us to directly compute the assignment matrix. Generally, the DDNA is a promising approach that creates an abundance of future research directions within the area of DODME.

Notes

1. |S| denotes the cardinality of set S.
2. Symbol ⊗ denotes the Kronecker product, \( I_n \) is the identity matrix with size \( n \times n \), and \( \text{diag}(c_1, c_2, \ldots, c_n) \) denotes an \( n \times n \) square matrix with \( c_1, c_2, \ldots, c_n \) being the elements of its main diagonal while all non-diagonal entries are zeros.
3. \( \|a\|_2 \) denotes the Euclidean norm of vector \( a \).
Acknowledgments

We are thankful to Dr. Joakim Ekström from the University of Linköping for the valuable suggestions on earlier drafts of this manuscript. We would also like to acknowledge Dr. Anthony Patire from UC Berkeley for kindly providing the Aimsun model as well as detailed knowledge of the Pasadena network. Finally, we are grateful for the constructive comments from three anonymous referees.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Swedish Transport Administration [Dnr TRV2018/132473 and TRV2021/22404] and the Swedish Energy Agency [project number 46963-1].

ORCID

Nikolaos Tsanakas http://orcid.org/0000-0003-2767-9415
David Gundlegård http://orcid.org/0000-0002-5961-5136
Clas Rydergren http://orcid.org/0000-0001-6405-5914

References


