

A hybrid modelling framework for the estimation of dynamic origin–destination flows[☆]

Sakitha Kumarage, Mehmet Yildirimoglu ^{*}, Zuduo Zheng

School of Civil Engineering, University of Queensland, Australia



ARTICLE INFO

Keywords:

Dynamic OD estimation
Macroscopic fundamental diagram
Large-scale traffic networks
Non-linear optimization

ABSTRACT

The dynamic origin–destination flow estimation (DODE) problem requires scalable methods for large scale traffic networks and consistent techniques for capturing both uncongested and congested traffic conditions. Despite numerous efforts on incorporating multifold data sources and developing manifold mathematical models, the DODE problem remains a challenging problem in terms of both scalability and consistency. To fill this gap, we propose a novel hybrid DODE framework that integrates region-level (macro) and centroid-level (micro) traffic dynamics. The region-level traffic flows are described by the macroscopic fundamental diagram, while the centroid-level traffic flows are represented by the linear mapping of origin–destination flows onto link counts. This hybrid approach enables us to (i) incorporate region-level traffic measures into the problem, addressing scalability issues arising in large scale traffic networks, and (ii) capture non-linear behaviour of traffic in the regional context, enhancing consistency of the estimation results with respect to traffic conditions. The proposed methodology is experimented in a large-scale traffic network, which is benchmarked for DODE problems. The results indicate an outstanding performance of the hybrid DODE particularly in congested traffic conditions and highlight the effectiveness of aggregated (regional) traffic models in enhancing DODE methods with minimal computational burden.

1. Introduction

Understanding the spatial–temporal variation of demand is a vital requirement in mobility management. In fact, obtaining a reliable estimate on time-varying Origin–Destination (OD) flows (or dynamic demand flows) in a traffic network is a crucial first step in developing an accurate traffic model, which can be used to evaluate various policies and develop traffic management strategies. Traditionally, demographic surveys and/or mathematical models were used to estimate prior OD flows, (see Ortízar and Willumsen (2011) for details). However, such methods are often considered erroneous because, surveys may be outdated and expensive to conduct, and prevailing mathematical models may not be advanced enough to tackle the inherent complexity of the underlying demand patterns. Hence, early works of Robillard (1975), Bell (1983), Fisk and Boyce (1983), Zuylen and Willumsen (1980) proposed the estimation of posterior OD flows (in the static context) based on prior OD estimates and traffic counts (indirect measurements) that can easily be collected from induction-based loop detectors. Later, these static OD estimation methods which consider OD flows in a single time horizon were extended to dynamic OD estimation (DODE) to estimate OD flows in multiple time

[☆] Funding: This research was partly funded by iMOVE CRC and supported by the Cooperative Research Centres program, an Australian Government initiative. This research was also partly funded by the Australian Research Council (ARC) through Dr. Mehmet Yildirimoglu's Discovery Early Career Researcher Award (DECRA; DE220101320).

* Corresponding author.

E-mail addresses: s.kumarage@uqconnect.edu.au (S. Kumarage), m.yildirimoglu@uq.edu.au (M. Yildirimoglu), zduo.zheng@uq.edu.au (Z. Zheng).

intervals. A summary on early literature on OD estimation could be found in [Cascetta et al. \(2013\)](#). There are two main difficulties we are focusing on in this study that remain a challenge in the DODE problems;

(i) *Scalability* issues that arise from *indeterminateness* of OD flows and that exacerbate with increasing network size. The core of existing DODE methods is to find the share of trips passing through a link entering from any origin and headed towards any destination such that the resulting link counts would match the observed link counts. The observability of the network, interpreted as the ratio of the number of unknowns/dimensions (i.e. OD estimates) to the number of equations (i.e. link count observations), should be closer to unity to find a unique (or determinate) solution for this problem ([Marzano et al., 2009](#)). However, this ratio is larger than one even in small networks, which leads to indeterminateness of OD flows, and this is expected to increase with increasing network size. This leads to a broader solution space in large-scale networks because there could be numerous combinations of OD flows that can result in the same link counts. That means, the indeterminateness issues observed in small networks are only expected to exacerbate in large-scale networks. The solution space available for OD flows will increase exponentially with increasing network size and affect the quality of solutions for any OD estimation method. This broadening of solution space or the exacerbation of the indeterminateness issues is referred to as ‘scalability’ in our study. Several techniques, ranging from dimension reduction ([Djukic et al., 2012](#)) to quasi-dynamic approaches ([Cascetta et al., 2013](#)) and simulation-based optimization ([Osorio, 2019](#)), have been proposed to address the *scalability* issues.

(ii) *Consistency* of the estimation results in congested traffic conditions due to nonlinear relationship between traffic counts and OD flows. The existing DODE methods underline the importance of the accurate modelling of the relation between OD flows and traffic counts, which is highly nonlinear in congested traffic conditions. The accuracy of the estimated OD flows should be *consistent* to changing traffic conditions, and should not deviate with respect to whether the scenario is uncongested or congested. A significant part of the literature models the relation between OD flows and traffic counts as a linear approximation, which obviously is limited in capturing nonlinear changes going from uncongested to congested traffic conditions, and vice versa. The nonlinear nature of the DODE problem has been typically addressed by incorporating a dynamic traffic assignment (DTA) model as the lower-level in bi-level optimization formulation, where the system is expected to move towards the optimal solution over iterations ([Djukic, 2014](#)). While the employment of DTA models is commendable, such approaches being combined with general-purpose optimization algorithms (such as Simultaneous Perturbation Stochastic Approximation (SPSA)) has brought in computational inefficiencies ([Osorio, 2019](#)). Note that this aspect of the DODE problem will be referred to as consistency in the remainder of the paper.

Considering the existing challenges of DODE methods, this study propose a hybrid DODE framework that combines two traffic models which is scalable to large-scale networks and efficient in tracking congested traffic conditions. Hybrid DODE is designed as a bi-level arrangement with lower-level being a DTA model and upper-level being an optimization algorithm to derive OD flows. The OD estimation level (upper-level) is formulated as an optimization algorithm which incorporates centroid-level and region-level traffic models that estimate/approximate traffic dynamics in the DTA model (lower-level). One could argue that these two models serve as hierarchical meta-models for the optimization problem, and it is revealed that the development of optimization algorithms that are aware of traffic dynamics (simulation-based optimization) has been very successful in overcoming scalability and improving computational efficiency ([Osorio and Punzo, 2019](#); [Gu and Saberi, 2021](#); [Patwary et al., 2021](#)). The hybrid framework simultaneously captures both traditional granular modelling (link-level or centroid-level traffic dynamics) and aggregated modelling (region-level traffic dynamics). While the traditional granular modelling relies on linear mapping of OD flows onto the link flows through an assignment matrix, the region-level model builds on macroscopic fundamental diagram (MFD) to describe the traffic performance in urban areas. MFD has emerged as a crucial modelling tool in the last decade enabling the development of aggregated modelling and control approaches for large-scale traffic networks. An urban region (network) with roughly homogeneous accumulation can be modelled using the MFD, which provides a uni-modal, low-scatter, and demand-insensitive relationship between accumulation and trip completion flow ([Geroliminis and Daganzo, 2008](#)). The parsimonious and analytically tractable nature of MFD enables researchers to develop large-scale and efficient models for traffic control purposes (see [Kumarage et al., 2021](#) for a summary). However, this modelling breakthrough has not been well explored for demand estimation purposes yet. Essentially, the inclusion of regional representation provides additional valuable information within the DODE problem and enables an analytically tractable and nonlinear model (i.e. MFD) to be embedded in the optimization problem.

This study has two major contributions to the large body of DODE literature:

- Developing an optimization algorithm that incorporates a centroid level linear approximation and a region-level traffic model to tackle OD estimation problems. Additional constraints that describe the region-level traffic model are expected to reduce the solution space of the optimization problem, and thus enhance the scalability of the OD estimation procedure.
- Developing OD estimation problems that are consistent in both congested and uncongested traffic conditions by incorporation of region-level models that capture the non-linear behaviour of traffic at region level.

The paper is organized as follows: Section 2 outlines existing works on DODE problem and their limitations, Section 3 presents the hybrid DODE framework, Section 5 presents results on implementation of the proposed hybrid DODE framework in a large-scale traffic network which is benchmarked for DODE problems, and Section 6 concludes the paper with final remarks.

2. Prior works

The essential theory on static OD flow estimation problem builds upon error minimization between observed and estimated values of link counts and the difference between a priori and estimated OD flows. This static OD estimation problem is extended to dynamic (time-varying) scenarios in the work of [Cascetta et al. \(1993\)](#), by discretizing the time horizon into a finite number of

time steps (j). Then, time-varying link counts are measured and OD flows are estimated for each time step. However, when dealing with these time-varying link counts the complexity increases significantly as the demand at time step t_0 can affect link counts at any time step $t \geq t_0$. Hence, the traffic assignment becomes a dynamic problem where dynamic route choice and dynamic link costs must be incorporated, which makes the estimation problem a more complicated, computationally expensive and iterative process. The general formulation of a dynamic OD flow estimator $(d_{[1]}^*, d_{[J]}^*)$ can be presented as;

$$(d_{[1]}^*, d_{[J]}^*) = \operatorname{argmin}_{x_{[1]},..x_{[J]}} \left\{ z_1(x_{[1]},..x_{[J]}, \hat{d}_{[1]},..,\hat{d}_{[J]}) + z_2(F_{[1]},..F_{[J]}, \hat{F}_{[1]},..,\hat{F}_{[J]}) \right\} \quad (1)$$

Note, $(z_1(..))$ is the measure of the distance between estimated OD flows $(x_{[1]},..x_{[J]})$ and the a priori OD flows $(\hat{d}_{[1]},..,\hat{d}_{[J]})$, and $(z_2(..))$ is the measure of the distance between the link counts resulting from the estimated OD flows $(\hat{F}_{[1]},..,\hat{F}_{[J]})$ and the observed link counts $(F_{[1]},..F_{[J]})$.

Cascetta et al. (1993) proposed two estimators based on the above fundamental concept which could be categorized as; (1) simultaneous estimators (offline methods) focusing on the finite time horizon, and (2) sequential estimators (online methods) which target estimation in moving time horizons. Although it is possible to apply both estimators in offline estimation, seminal work of Ashok and Ben-Akiva (2000) revealed the reduced computational burden of sequential estimators in real-time/continuous-time applications. Our attention in this study is on offline methods but online methods could be a good future extension.

The offline DODE problem is formulated as a constrained optimization problem in most studies, where the objective function takes a functional form (penalty function) such as generalized least squares, entropy maximization, maximum likelihood, Bayes inference, normalized squared error (Cascetta et al., 2013; Antoniou et al., 2016). We can see a spectrum of algorithms being used to solve the offline DODE problem, which can be summarized into three broad families as; (i) gradient search algorithms such as steepest gradient solvers (Cremer and Keller, 1987), conjugate gradient solvers (Sherali and Park, 2001), quasi-dynamic GLS solvers (Cascetta et al., 2013), sparse least square solvers (LSQR algorithm) (Bierlaire and Crittin, 2004), quasi-newton solvers (Toledo and Kolechkina, 2013), non-linear interior point solvers (Shafiei et al., 2016); (ii) stochastic search algorithms such as box-algorithm (Darda, 2002), simultaneous perturbation stochastic approximation (SPSA) algorithm (Cipriani et al., 2011), genetic algorithms (Huang et al., 2013), evolutionary algorithms (Kattan and Abdulhai, 2006); and (iii) heuristic search methods such as fuzzy logic based algorithms (Gómez et al., 2015), artificial neural networks (Ou et al., 2019; Sun et al., 2019) and forward-backward algorithms (Ma et al., 2020). It should be emphasized that the use of general-purpose algorithms to solve the optimization problem is very common in OD estimation literature. However, it has numerous disadvantages such as scalability and efficiency. The fundamental cause for these drawbacks is the lack of understanding on the behaviour of traffic networks within these optimization problems, and they are designed with asymptotic convergence properties for a wide range of problems (Osorio, 2019).

Note that DODE and DTA are interdependent, where a reliable DODE will lead to an accurate DTA and vice versa. Hence, the DODE problem is often formulated as a bi-level framework where the first level accounts for DODE, and the second for DTA such that an iterative process between these two levels leads to a reliable DODE. This also allows the DODE framework to indirectly capture the nonlinear relationship between OD flows and link traffic counts. Although there are attempts to combine these two levels (as in Darda, 2002), such developments are limited due to their computational complexities. There are a variety of perspectives on bi-level DODE models. Several research have argued for the correctness of DTA models by establishing ways for obtaining assignment matrices for use in DTA through the use of network-wide data sources as seen in Krishnakumari et al. (2020), Ros-Roca et al. (2022), Nigro et al. (2018b). On the Other hand, Cipriani et al. (2011), Balakrishna et al. (2008), Nigro et al. (2018a), Qurashi et al. (2020) avoid the use of a bi-level framework by considering DTA as a black box and develop solutions and methods that are based on the SPSA algorithm. While Ros-Roca et al. (2020), Qurashi et al. (2020), Tympanianaki et al. (2018) have seen SPSA based algorithms to be efficient, Osorio (2019), Ma et al. (2020), Djukic et al. (2012) and Lu et al. (2013) have pointed out the SPSA based solvers' lack of computational efficiency.

Consistency of estimation results in the face of traffic congestion is a major challenge for existing DODE methods. A number of techniques have been proposed in literature to enhance consistency of DODE problems. Firstly, having accurate a priori OD estimates that produce traffic patterns with the same congestion states as observed in reality are important for DODE to perform well in congested traffic conditions (Frederix et al., 2013). However, research efforts find that a priori OD flows are frequently biased and carry noise (Marzano et al., 2009). Secondly, capturing traffic congestion within the DODE is essential. Initial works of Cascetta et al. (1993) introduce demand vs link count interaction as a linear mapping, while Frederix et al. (2013) reveals the unsuitability of such linear models in congested conditions as they fail to capture spatial-temporal effects of queues, delays, spillbacks and rerouting effects which are non-linear and non-convex. There has been attempts to embed non-linear traffic assignment/dynamics in DODE such as algorithms based on kinetic wave theory (Frederix et al., 2011; Lu et al., 2013) and algorithms based on simulation meta-models (Osorio, 2019; Osorio and Punzo, 2019; Gu and Saberi, 2021; Patwary et al., 2021). A sequential decomposition approach proposed by Hu et al. (2017) is another technique where the first stage preserves a simple linear structure and the second stage builds on a demand-capacity-volume relationship at a congested road segment and utilizes shockwave theory to capture the differences between true demand and volume output. Nevertheless, the second stage in the proposed model is tested only with few links, and extending this in large-scale networks can significantly complicate the estimation procedure. While these attempts are comprehensive in capturing traffic dynamics, incorporation of such microscopic models/details into the optimization problem increase the number of inputs within the DODE, resulting in incompatibilities in obtaining optimal solutions, as objective functions frequently become non-convex and non-differentiable with numerous local minima. Such attempts also increase the computational complexities in large scale traffic networks (Zhang et al., 2018). Alternatively, providing more information on traffic states to DODE via different

forms of traffic data such as link densities, travel times, vehicle trajectories, vehicle (re) identification data, mobile phone data (see Krishnakumari et al., 2020 for a summary) has been seen as a remedy to capture congested traffic conditions.

Scalability of DODE problem is another concern, Marzano et al. (2009), Krishnakumari et al. (2020) reveal that the scalability of existing DODE methods is limited to spatial extents where unknown-equation ratio is preserved around unity. However, the DODE problem is a high-dimension problem in large scale networks where unknown-equation ratio is far from unity, resulting in broader solution spaces and leading to indeterminate solutions. We find two main approaches to address scalability issue in literature; (i) decomposition of a large scale network DODE into series of sub-level DODEs, such as hierarchical decomposition (Frederix et al., 2013), meta-model based decomposition (Osorio, 2019) and infrastructure based decomposition (Lou and Yin, 2010) and (ii) dimensionality reduction of decision variables by referring to structural similarity of network (Behara et al., 2020), activity-based demand structure (Qurashi et al., 2020), principal component analysis (Djukic et al., 2012; Prakash et al., 2018), and congestion patterns (Andrienko et al., 2017). While existing works have focused on confining the solution space by reducing the number of decision variables through decomposition or dimensionality reduction, we aim to limit the solution space for decision variables through constraints that relate region-level and centroid-level traffic measures in this work. This is a promising direction that has the potential to address the consistency issues, scalability issues and computational difficulties observed when using DODE for medium to large scale traffic networks and congested traffic conditions.

MFD is the main modelling tool utilized in the region-level traffic model proposed in the study. Estimation of MFD for a region is a crucial task for the region-level model. While MFD estimation is usually done with loop detectors (Leclercq and Geroliminis, 2013), there are several studies aiming to improve MFD estimations considering the optimal sensor locations (Saffari et al., 2020; Ortigosa et al., 2015; Zockaie et al., 2018), or alternative data sources such as taxi GPS data (Huang et al., 2019), floating car data (Knoop et al., 2018), mobile sensor data (Paipuri et al., 2020), and combined data sources (Ambühl and Menendez, 2016; Saffari et al., 2022). MFD-based modelling has been used for numerous applications such as perimeter control (Aboudolas and Geroliminis, 2013), route guidance (Yildirimoglu et al., 2018), demand re-distribution (Yildirimoglu and Ramezani, 2020; Kumarage et al., 2021) and work schedule design (Yildirimoglu et al., 2021). However, there has been less attention on integrating MFD models into demand estimation. To this end, Dantsuji et al. (2022) introduce the region-level DODE using MFD based traffic models. The study focuses only on the estimation of region-level OD flows where the estimated region-level ODs are distributed to centroid-level ODs in proportion with the *a priori* centroid-level OD flows. The lack of centroid-level path choices in DODE is a major drawback in Dantsuji et al. (2022). On the other hand, this study proposes a hybrid framework capturing both region-level and centroid-level interactions; the nonlinear relation between OD flows and traffic conditions is captured at a regional context and link-level path choices are incorporated at a centroid-level formulation.

3. Methodology

The hybrid DODE problem presented in this study introduces the hybrid modelling of traffic dynamics. This is accomplished by incorporating a centroid-level linear approximation and a region-level traffic model which operates simultaneously. Fig. 1 shows a conceptual diagram of the hybrid DODE framework. We have hybrid DODE as the upper level and traffic assignment as the lower-level. Note that there are three traffic models in this framework. The traffic assignment level is a simulation based model which operates closer to real traffic networks, while there are two other traffic models being used in the OD estimation level. These two models are analytical models based on the hierarchy of operation. The centroid-level is modelled as a linear traffic approximation model which maps the OD flows onto link counts via an assignment matrix. The accumulation based MFD model is used at the region-level which defines traffic dynamics based on conservation of vehicle flows at the regional context. The centroid-level demand ($q(t)$) is the main decision variable in the proposed framework that links these three models. The process starts with measurements from the traffic network (e.g., link counts, regional accumulations) and *a priori* OD data that might be retrieved from travel surveys, etc. Then, we simulate the *a priori* OD matrix in the traffic assignment level and record simulation results such as link counts, assignment matrices, region accumulations, regional route choice (split ratios), regional trip lengths. Next, the hybrid DODE model is used to estimate ODs where we combine the centroid-level linear approximation and the region-level traffic model. Note that this hybrid approach involves modelling of traffic at different granular levels (i.e., centroid-level and region-level) and requires a strong agreement between the two at the optimal solution point. Finally, the estimated ODs from the hybrid DODE are fed back to the traffic assignment level. The iterations between the traffic assignment level and the hybrid DODE level is continued until the observed and simulated traffic conditions (counts) converge, which is equivalent to existing bi-level OD estimation problems.

We propose a DTA model for the lower level as it has advantages for OD estimation problems in modelling the complex relationship between OD flows and link counts (Antoniou et al., 2016). The upper-level or OD estimation level is formulated as an optimization problem which can be tackled with a myriad of solution algorithms. We propose a gradient search algorithm for the OD estimation level in this study and we find stochastic search methods or heuristic search methods could be a good future direction. The proposed hybrid DODE framework is formulated by incorporating a centroid-level linear approximation and a region-level traffic model. The centroid-level model simply builds on a linear approximation approach between demand and link counts, while we use an MFD-based approach for region-level traffic modelling. Hence, the scope of this study is limited to urban networks where an MFD could be defined. It should be stressed that by incorporating centroid-level and region-level modelling, it is possible to define specific constraints that guarantee a certain level of accordance between the two levels.

The next sections will go into greater details concerning model development and interactions. The region-level model based on the MFD traffic dynamics is explained in Section 3.1 while the linear approximation method used to define the centroid-level is explained in Section 3.2. The combined formulation of OD estimation with centroid-level approximation and region-level model is presented in Section 3.3. A summary of the variables used in the hybrid DODE framework is given in Table 1.

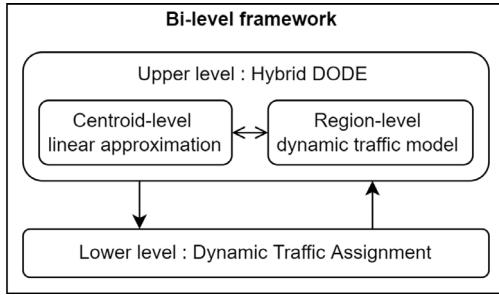


Fig. 1. Bi-level formulation of hybrid DODE.

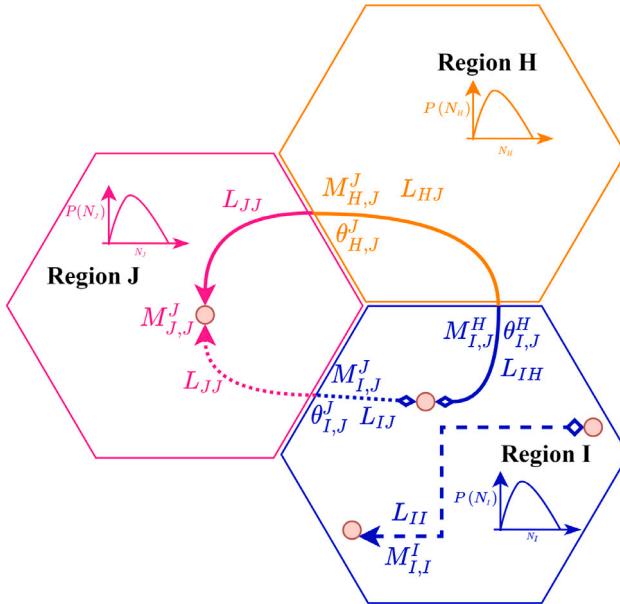


Fig. 2. Multi-region MFD dynamics.

3.1. Region-level modelling

The region-level traffic model incorporated in this study focus on defining specific functions that relate region-level OD flows on to regional accumulations. The multi-region accumulation-based MFD model, which is built upon the conservation of vehicle flows in a multi-region traffic network is utilized as the region-level traffic model.

Let an urban network \mathcal{R} be partitioned into R regions such that $\mathcal{R} = \{1, 2, \dots, I, H, J, \dots, R\}$, and \mathcal{G}_I be the number of neighbouring regions to region I . Let the $Q_{I,J}(t)$ be the demand inflow from region I to region J in current time step t . Let $N_I(t)$ be the total number of vehicles (accumulation) currently in region I , and $N_{I,J}(t)$ be accumulation currently in region I with final destination region J such that $N_I = \sum_{J=1}^R N_{I,J}$. Regional route choice arises when dealing with multi-region traffic networks. While some studies omit regional route choice due to network topology when just a single evident route exists across regions (see Sirmatel and Geroliminis (2018)). However, we provide a generic framework as such network topologies are not always available. Fig. 2 illustrates two regional paths observed between region I and region J . One trajectory (solid-line) travels from region I to H and reaches region J , while the other trajectory (dotted-line) travels from region I to J directly. We adopt the region split ratio parameter to identify the relative preference of transfer flows in multi-region configuration. Here, the split ratio parameter $\theta_{I,J}^H$ ($I, J \in \mathcal{R}, H \in \mathcal{G}_I$) is the proportion of vehicles choosing to go the neighbouring region H among the vehicles currently in region I with final destination region J . Hence, $\sum_{H \in \mathcal{G}_I} \theta_{I,J}^H = 1$. Following above definitions, mass conservation equations for an R -region MFD network are:

$$\dot{N}_{I,J}(t) = Q_{I,J}(t) + \sum_{H \in \mathcal{G}_I} M_{H,J}^I - \sum_{H \in \mathcal{G}_I} M_{I,J}^H \quad (2a)$$

$$\dot{N}_{I,I}(t) = Q_{I,I}(t) - M_{I,I}^I + \sum_{H \in \mathcal{G}_I} M_{H,I}^I - \sum_{H \in \mathcal{G}_I} M_{I,H}^H \quad (2b)$$

Table 1
Nomenclature.

Notation	Units	Description
n		Number of OD pairs
m		Number of traffic detectors
R		Number of macroscopic regions in urban network
$N_I(t)$	[veh]	Accumulation in region I at time interval t
$N_I^0(t)$	[veh]	Observed accumulation in region I at time interval t
$K_I(t)$	[veh/km]	Average density in region I at time interval t
$K_I^O(t)$	[veh/km]	Observed average density in region I at time interval t
L_I	[km]	Roadway length in region I
$q^k(\lambda)$	[veh/h]	OD flow between OD pair k departing at time step λ
$\hat{f}_I(t)$	[veh/h]	Estimated link count in time step t
$f_I^o(t)$	[veh/h]	Observed link count in time step t
$\hat{\alpha}_d^k(\lambda t)$	[] ∈ [0, 1]	Estimated fraction of OD flow (q^k) departing in time step λ , and contributing to link count at detector d at time step t
$\hat{\alpha}_d^l(\lambda t)$	[] ∈ {0, 1}	Path incidence variable which defines whether a vehicle which started the journey in time step λ and following the route d being able to pass through detector l at time step t ,
$\hat{p}_d^k(\lambda)$	[] ∈ [0, 1]	Probability of selecting OD route d between OD pair k in departure time step λ
\mathcal{R}		Urban network with R number of regions
a_I, b_I, c_I		Estimated MFD parameters for reg
\mathcal{G}_I		Set of neighbouring regions to region I
$P_I(\cdot)$	[veh · m/s]	Production in region I
$Q_{I,J}(t)$	[veh/s]	Region level demand inflow from region I to region J
$N_{I,J}(t)$	[veh]	Accumulation of vehicles currently in region I with final destination region J
$\theta_{I,J}^H$	[] ∈ [0, 1]	Ratio of flow transferring from region I to neighbouring region H and, with final destination region J
$M_{I,I}^I(t)$	[veh/s]	Exit flow in region I
$M_{I,J}^I(t)$	[veh/s]	Transfer flow from region I to next region H and final destination region J
L_{IH}	[m]	Average trip length for vehicles transferring from region I to neighbouring region H
L_{II}	[m]	Average trip length for internal trips in region I
$N_I^{J\text{am}}$	[veh]	Jam accumulation of region I
Δ_{IJ}^k	[] ∈ {0, 1}	Binary variable set to one if origin centroid and destination centroid of OD pair k are found in origin region I and destination region J
α_{lb}		Lower bound for regional accumulations given by MFD ($N^M(t_c)$)
α_{ub}		Upper bound for regional accumulations given by MFD ($N^M(t_c)$)
β_{lb}		Lower bound for a priori OD estimates (q^0)
β_{ub}		Upper bound for a priori OD estimates (q^0)
Vector notation	Dimension	Description
$q^g(t)$	∈ ℝ ^{n×1}	Centroid-level ground truth OD vector in time step t
$q^0(t)$	∈ ℝ ^{n×1}	Centroid-level perturbed OD vector in time step t
$q(t)$	∈ ℝ ^{n×1}	Centroid-level estimated OD vector in time step t
$f(t)$	∈ ℝ ^{m×1}	Link count vector in time step t
$\hat{f}(t)$	∈ ℝ ^{m×1}	Estimated link count vector in time step t
$\hat{A}(t)$	∈ ℝ ^{m×n}	Approximated assignment matrix
τ	∈ ℝ ^{n×1}	Random error and noise vector used for perturb ODs
$Q(t_c)$	∈ ℝ ^{R×R×1}	Region level demand flow vector in time step t_c
Δ	∈ ℝ ^{R×R×n}	Mapping vector of demand flows from centroid-level to region-level
$N^M(t_c)$	∈ ℝ ^{R×R×1}	Accumulation vector in time step t_c given by MFDs
$N^O(t_c)$	∈ ℝ ^{R×R×1}	Observed accumulation vector in time step t_c
$L(t_c)$	∈ ℝ ^{R×R×1}	Average trip length vector in time step t_c
$\theta(t_c)$	∈ ℝ ^{R×R×R}	Split ratios vector in time step t_c

where $M_{I,I}^I(t)$ (veh/s) is the exit flow (internal trip completion rate) which account for the completed trips in region I , whereas $M_{I,J}^I(t)$ (veh/s) is the transfer flow rate of vehicles from region I , with destination region J and next transfer region H . Fig. 2 provides an illustration on how transfer flows and exits flows are related to within region (single region) trips and inter-region trips in multi-region MFD dynamics. We calculate exit flows and transfer flows by distributing regional outflows (i.e. $P(.)$ /L(.)) as follows;

$$M_{I,J}^H(t) = \theta_{I,J}^H \cdot \frac{N_{I,J}(t)}{N_I(t)} \cdot \frac{P_I(N_I(t))}{L_{IH}(t)} \quad (3a)$$

$$M_{I,I}^I(t) = \frac{N_{I,I}(t)}{N_I(t)} \cdot \frac{P_I(N_I(t))}{L_{II}(t)} \quad (3b)$$

Here, L_{II} is the average trip length for within region (internal) trips and L_{IH} is the average trip length (external) for trips that are transferring from region I to neighbouring region H (see Fig. 2 for an illustration). Empirical experiments conclude that an asymmetric uni-modal, right-skewed curve could be approximated as a low scatter MFD generated in a homogeneous traffic

network. In this study, we prefer a third degree polynomial to fit MFDs in the form of; $P_I(N_I) = a_I \cdot N_I^3 + b_I \cdot N_I^2 + c_I \cdot N_I$ where, $P_I(\cdot)$ is the regional production (veh.m/s), N_I is the accumulation (veh) in region I r and, a_I, b_I, c_I are estimated MFD parameters. It should be noted that the functional form of the MFD is independent of the proposed DODE methodology, one could use extended MFD functional forms as suggested in Ramezani et al. (2015), Zheng and Geroliminis (2020), and it is also possible to use the vehicle outflow as in Sirmatel and Geroliminis (2021) or average network speed as in Arnott and Buli (2018) to define MFDs. The traffic states of this region-level model are based on third order functions (MFD) which are non-convex (concave) at the operation range. On the other hand, the traffic dynamics are defined based on the conservation of vehicles flows within the multi-region network. Hence, the resulting model is a non-convex non-linear but differentiable analytical model which could be incorporated into the hybrid DODE.

3.2. Centroid-level modelling

While the aggregate-level traffic dynamics are explained by multi-region MFD model, granular traffic dynamics are approximated by the centroid-level model in the framework. Here, we use the traditional approach which relates time-dependent OD flows to time-dependent link counts using a linear assignment method as elaborated in Cascetta et al. (1993). Consider an urban network with n number of OD pairs and m number of detectors which are sufficient in numbers and well placed in the network. Let $q(t) = [q^1(t) \dots q^n(t)] \in \mathbb{R}^{1 \times n}$ be the centroid-level demand vector in time step t , and let $f(t) = [f_1(t) \dots f_m(t)] \in \mathbb{R}^{1 \times m}$ be the link count vector in time step t . Then, we propose an estimation model for link counts as; $f(t) = \hat{f}(t) + e_t$, where e_t is the estimation error and $\hat{f}(t)$ is the estimated link counts which is formulated as a linear mapping between OD flows ($q(\cdot)$) and estimated assignment matrix (mapping matrix) ($\hat{A}(\cdot|t)$) as follows;

$$\hat{f}(t) = \sum_{\lambda=t-\eta:t} \hat{A}(\lambda|t) \times q(\lambda) \quad (4)$$

Note that the calculated link counts in current time step t are made up of OD flows leaving (or starting the journey) during the current time step or in previous time steps and reaching the detector l in current time step. In formulation, we consider η number of previous time steps such that link counts at time step t being contributed by OD flows from $t - \eta : t$ previous time steps. Note that η complies with the travel times in the network between the considered OD pair. The estimated assignment matrix $\hat{A}(\lambda|t) = [\hat{a}_1^{\lambda} \dots \hat{a}_m^{\lambda}] \in \mathbb{R}^{m \times n}$ defines the fraction of OD flows which started the journey in time step λ , and contributing to the link counts in time step t . The assignment matrix is derived from path choice matrices and path incident matrices, which are extracted from the DTA model (lower-level). Each element in assignment matrix ($\hat{a}_l^k(\lambda|t)$) which gives the estimated fraction of OD flow (q^k) departing in time step λ , and contributing to link count at detector l in time step t , is calculated by: $\hat{a}_l^k(\lambda|t) = \sum_{d \in D} \hat{a}_d^{\lambda}(\lambda|t) \times \hat{p}_d^k(\lambda)$. Here, $\hat{a}_d^{\lambda}(\lambda|t)$ is the path incidence variable which defines whether a vehicle that started the journey in time step λ and following the route d being able to pass through detector l at current time step t , and $\hat{p}_d^k(\lambda)$ defines the probability of selecting OD route d between OD pair k in departure time step λ .

The centroid-level linear approximation enables us to embed an analytical model that describes link-level traffic flows in the DODE problem. Nevertheless, the linear mapping approach, as the name implies, cannot capture the nonlinear relation that arises in congested traffic conditions. The effectiveness of the linear approximation deteriorates with increase in traffic congestion. Hence, we incorporate region-level modelling with the ability to capture nonlinear traffic dynamics in the regional context, and aim to reach consistent OD estimates for both congested and uncongested conditions over the iterations. Incorporating a more detailed model such as a link level meta-model (Osorio, 2019) instead of a linear approximation could be an alluring future direction.

3.3. Hybrid DODE formulation

Here, we propose a novel approach to DODE problem. The conventional DODE problem is designed to estimate optimal OD flows that are close enough to existing a priori OD flows while producing link counts that are sufficiently close to observed (ground-truth) link counts. However, in medium to large-scale traffic networks, this optimization criterion faces scalability issues, because the indeterminate nature of the problem formulation leads to sub-optimal outputs as the solution space increases with the number of OD pairs. While existing research has focused on techniques to restrict solution space by decreasing decision variables (through decomposition or dimensionality reduction), we propose limiting the solution space by introducing specific constraints that tie the optimum solution to regional traffic conditions.

Furthermore, when the network is significantly congested, sub-optimal solutions are inevitable because conventional methods lack information (knowledge) on the propagation of traffic congestion in the network and the non-linear behaviour of traffic in congested conditions (Frederix et al., 2013), which challenge the consistency of reliability in OD estimates. Incorporation of more information (about congestion) via accurate a priori ODs, non-linear assignment models, and more data sources such as link densities, travel times, and GPS trajectories have been proposed as solutions in the literature to address this consistency issue. The hybrid DODE framework presented in this study, on the other hand, overcomes such consistency issues by adding the region-level model into the DODE framework, which captures the non-linear relationship between regional OD flows and regional traffic conditions. The problem is formulated as follows:

$$\underset{q}{\text{minimize}} \sum_{t=0:T} \left\{ \sum_{l=1:m} \left(\frac{f_l^0(t) - \hat{f}_l(t)}{f_l^0(t)} \right)^2 \right\} \quad (5a)$$

subject to for $t = 0 \dots \Delta t \dots T$

$$\hat{f}(t) = h(\hat{A}(\cdot|t), q, \eta) \quad (5b)$$

$$q^0(t) * \beta_{lb} \leq q(t) \leq q^0(t) * \beta_{ub} \quad (5c)$$

for $t_c = t \dots \delta t_c \dots (t + \Delta t - \delta t_c)$

$$Q(t_c) = \Delta \cdot q(t) \quad , \quad (5d)$$

$$N^M(t_c + 1) = g \{ Q(t_c), N^M(t_c), L(t_c), \theta(t_c) \} , \quad (5e)$$

$$N^O(t_c) * \alpha_{lb} \leq N^M(t_c) \leq N^O(t_c) * \alpha_{ub} \quad (5f)$$

The objective function given in Eq. (5a) targets to minimize the normalized squared error (NSE) between the ground-truth link counts ($f_l^0(t)$) and calculated/estimated link counts ($\hat{f}_l(t)$) for all the traffic detectors ($l \in [1 : m]$), in all the time steps ($t \in (0 \dots \Delta t \dots T]$). Note that the NSE is an algebraic rather than an absolute error measurement capable of balancing the instabilities in link counts (caused in assignment phase) and capture the improvements that can be detected at network level when the estimations move along a descent direction. However, it is possible to use different penalty function combinations as an improvement to the proposed study considering several factors such as prioritizing critical links, weighted based on the importance of road hierarchy and availability of data sources. The framework that we develop here is not bound by the choice of objective function, any GOF function that calculates the relative or absolute errors in the link counts can be readily adapted in the proposed framework. $q = \{q_k(t) \forall t, k\}$ is the decision variable of the objective function where, $q_k(t)$ represents the flow between (centroid-level) origin–destination pair k in time step t .

Eq. (5b) defines the centroid-level traffic approximation, where the estimated link counts ($\hat{f}(t) = \{\hat{f}_l(t), \forall l, t\}$) are derived by the function $h(\cdot)$, which takes assignment matrix ($A(\cdot|t)$), centroid-level demand (q), and number of previous time steps under consideration (η) as inputs and solves the linear system given in Eq. (4). Eq. (5c) generates constraints for the decision variable q where the OD estimates are limited to a boundary defined by the a priori OD estimates q^0 where, β_{lb}, β_{ub} are non-negative constants.¹

Eq. (5d) computes the regional demand by summing up the centroid-level demand flows which contribute to the regional demand flow ($Q(t_c) = \{Q_{IJ}(t_c) \forall I, J\}$). The relation between centroids and regions are defined by $\Delta = \{\Delta_{IJ}^k \forall I, J, k\}$. Here, Δ_{IJ}^k is set as a binary variable such that the value is equal to 1, if the origin centroid and destination centroid of OD pair k are found in origin region I and the destination region J , respectively, otherwise equal to 0. It is important to note the difference of the time steps in the centroid-level (t) and region-level (t_c) formulations. The time step in the region-level formulation is finer than the time step in the centroid-level, because t_c should be lesser than the free-flow travel time of shortest distance average trip length ($\min \{L_{IH}(t), \forall I, H\}$), in each region to capture transfer flows and exit flows. Nonetheless, the centroid level OD estimates and measurements are captured in coarser intervals (e.g.: 15-min, 30-min) in practice. Hence, we formulate the DODE problem to capture the difference of time-steps by assuming that the $q(t)$ is uniformly distributed within time-step t allowing us to obtain $Q(t_c)$ for finer time steps of t_c . Notably, the interdependence between the centroid- and region-levels specified by Eq. (5d) ensures that any change in demand at the centroid level should satisfy both the centroid- and region-level traffic dynamics. For example, assume regional OD flow $Q_{I,J}$ is generated by aggregation of centroid level OD flows q_a, q_b, \dots, q_g . When $Q_{I,J}$ is adjusted by the region-level model, the centroid level model has to adjust q_a, q_b, \dots, q_g . These constraints brought by the region-level ease the indeterminateness in the centroid level by providing a guidance on adjustment of centroid-level OD flows.

Eq. (5)e defines the dynamics of the multi-region accumulation-based model using the function $g(\cdot)$, which is the compact form of the accumulation-based MFD model presented in Section 3.1. Function $g(\cdot)$ gives the accumulation in the next time step ($N^M(t_c + 1) = \{N_{I,J}^M(t_c + 1), \forall I, J\}$) by taking current accumulation states ($N^M(t_c) = \{N_{I,J}^M(t_c), \forall I, J\}$), vector of demand flows ($Q(t_c) = \{Q_{IJ}(t_c) \forall I, J\}$), average trip length ($L(t_c)$) and split ratios/route choice variables ($\theta(t_c) = \{\theta_{I,J}^H(t_c) \forall I, J, H\}$) as inputs. The function $g(\cdot)$ is nonlinear, as the MFD-based traffic dynamics are nonlinear. Additionally, the function $g(\cdot)$ solves ordinary differential equations given in Eqs. (2a)–(2b) using the Runge–Kutta method by temporal discretization of the prediction interval into a step size significantly smaller than t_c . Note that t_c should be short enough to assume the demand within the time step does not vary, which enables us to estimate the accumulation in next time step using the four steps of the Runge–Kutta method.

Eq. (5f) defines the bounds (α_{lb}, α_{ub}) for the regional accumulations given by MFD dynamics ($N^M(t_c)$) with respect to the ground-truth accumulations ($N^O(t_c)$) taking α_{lb}, α_{ub} as non-negative constants.² Note that ground-truth accumulation values reflect the ground truth scenario as discussed earlier in the test-bed description. This constraint together with the constraint in Eq. (5d) helps to bound the solution space of q such that the optimizer is guided to find an optimal solution that complies with regional traffic conditions.

Eqs. (5d)–(5f) describe the MFD-based region-level traffic model. MFD essentially describes the relation between average flow and average density in urban regions, which is mathematically expressed as a third order polynomial function in this study and

¹ The selection of β_{ub}, β_{lb} is dependent on the reliability of *a priori* OD matrices. Comparing trip generations and trip attractions for a sample of traffic analysis zones or centroids (through boundary flows and screen-line checks) in a network could provide further information on the reliability of *a priori* matrices and allow the practitioners to identify the confidence level in the prior OD estimates.

² The selection of α_{ub}, α_{lb} is dependent on the reliability of observed region accumulations. The initial values for α_{ub}, α_{lb} could be established considering the measurement error of detectors (or estimation error if different data sources were used to estimate region accumulation). However, these values might require fine-tuning through a trial-and-error procedure.

many others in the literature. Further, this relation is useful to develop a dynamic regional traffic model that describes the flow of vehicles between regions and within regions (i.e., Eqs. (5d)–(5f)). This regional model describes the traffic propagation in the network inherently and nonlinearly, while the centroid-level linear model approximates the relation between vehicle flows and demand. For example, increasing demand for a centroid-level OD pair will increase link counts along the chosen path without changing the travel time or affecting traffic propagation across links. This is because the centroid-level approximation assumes the assignment matrix is constant in a given iteration. On the other hand, increasing the demand between two regions will increase the vehicle accumulation along the regional path (i.e., regions which the regional OD route follows) but also will change the travel times (in a nonlinear fashion) and affect the traffic propagation across the regions. In other words, the relation between demand and traffic counts is defined inherently in the nonlinear MFD-based traffic model, while it is defined with respect to exogenous travel times (from the last iteration) in the centroid-level approximation. Therefore, the regional traffic model (Eqs. (5)d–(5)f) is capable of assisting the centroid-level approximation (Eq. (5)b) in capturing the nonlinear nature of traffic. Note that this interaction cannot be fully accomplished in one single iteration where the assignment matrix at the centroid-level is constant. It is important to emphasize that several iterations are needed to expose the full effect of interaction between the two modelling levels.

Compared to the conventional methods, we bring in additional information to the optimization problem via region-level model, which allow to capture the non-linear behaviour of traffic dynamics in regional context and eliminate consistency issues faced in deriving optimal solutions. Further, having combined contribution of centroid-level and region-level within the optimization problem is expected to solve scalability issues as the optimal solution for centroid-level is found adhering to traffic conditions of region-level. However, the complex formulation of the problem leads to categorizing this optimization problem as a non-linear, non-convex optimization problem and we use iterative non-linear solvers (such as interior-point solvers) to derive optimal solutions.

4. Implementation framework

Following the conceptual framework presented in Section 3 we present the modelling procedure that should be followed to implement the hybrid DODE in this section. We propose a four stage modelling framework developed by referring the benchmark study on DODE by Antoniou et al. (2016), which integrates all modelling components of Fig. 1 within four layers such as experimental set-up, test-bed, hybrid DODE and performance evaluation layers as shown in Fig. 3. Following sections will elaborate on each layer.

4.1. Experimental setup

The experimental setup gathers preliminary modelling components which must be performed before running the hybrid DODE, namely ground truth OD, ground-truth link counts, observed region accumulations, and macroscopic components such as partitioning of the network and MFD estimations. Note that the modelling components in this layer (connected by the *red dashed* lines in Figure the 3) will be static throughout the hybrid DODE experiments and will not be updated through out the iterations. Essential components of the experimental set-up are elaborated below;

- *Ground-truth OD matrices:* This is the main reference element in the proposed hybrid DODE. Although a ground-truth OD is not accessible in practice, it is important to compare the estimation results of hybrid DODE to evaluate the estimation reliability. Note that we will be using the ground truth OD matrices produced by the Antoniou et al. (2016) along with benchmark network infrastructure in this study.
- *Observed link counts:* We use link-level traffic count data collected from traffic detectors in the network to derive traffic observations since they remain the most commonly available type of sensors. It is important to emphasize that the ground truth OD matrix has been fed to the mesoscopic simulator to derive the observed link counts, which emulate the empirical measurements collected from the traffic networks. We acknowledge that traffic counts may correspond to traffic states in different regimes (i.e., congested or free-flow), and they cannot be used as a sole indicator of congestion. Therefore, we consider other traffic measurements (i.e., region accumulations) to complement this information.
- *Observed region accumulations:* The regional modelling requires regional accumulations as equivalent to link densities in the link-level traffic modelling. This requires the average density values in the regions, which can be collected again from loop detectors located at various links in the region. The observed accumulation $N_I^O(t)$ of a region I could be estimated by; $N_I^O(t) = K_I^O(t) \cdot L_I$, where $K_I^O(t)$ is the observed average density in region I and L_I is the roadway length in region I . The observed average density is the mean of density measurements from all (detector-equipped) links l in region I , $K_I^O(t) = 1/m_I \sum_{l \in I} k_l^o(t)$, where m_I is the number of (detector-equipped) links in region I . Note that region accumulation is a proxy for average density in a region; therefore, the average congestion-level in a region can be captured by the region accumulation value.
- *Partitioning:* For the regional modelling purposes, the network must be partitioned into multiple regions. Partitioning of urban networks based on spatial correlations is studied in the works of Ji et al. (2014), Saeedmanesh and Geroliminis (2016), and a practical trial-and-error process based partitioning approach is presented recently in Mariotte et al. (2020), which is adopted and extended in this study. Briefly, the urban network is partitioned into regions by a trial-and-error method which sequentially prioritizes: (i) keeping natural borders from the network topology when they exist, (ii) ensuring compact reservoir shapes, (iii) avoiding reservoir borders too close to major arterial roads and crossroads, (iv) ensuring a good definition of low-scattered MFD, (v) ensuring fewer number (10–15) of centroids are included in each region, and (vi) ensuring links/nodes connected to an OD node is within the same reservoir.

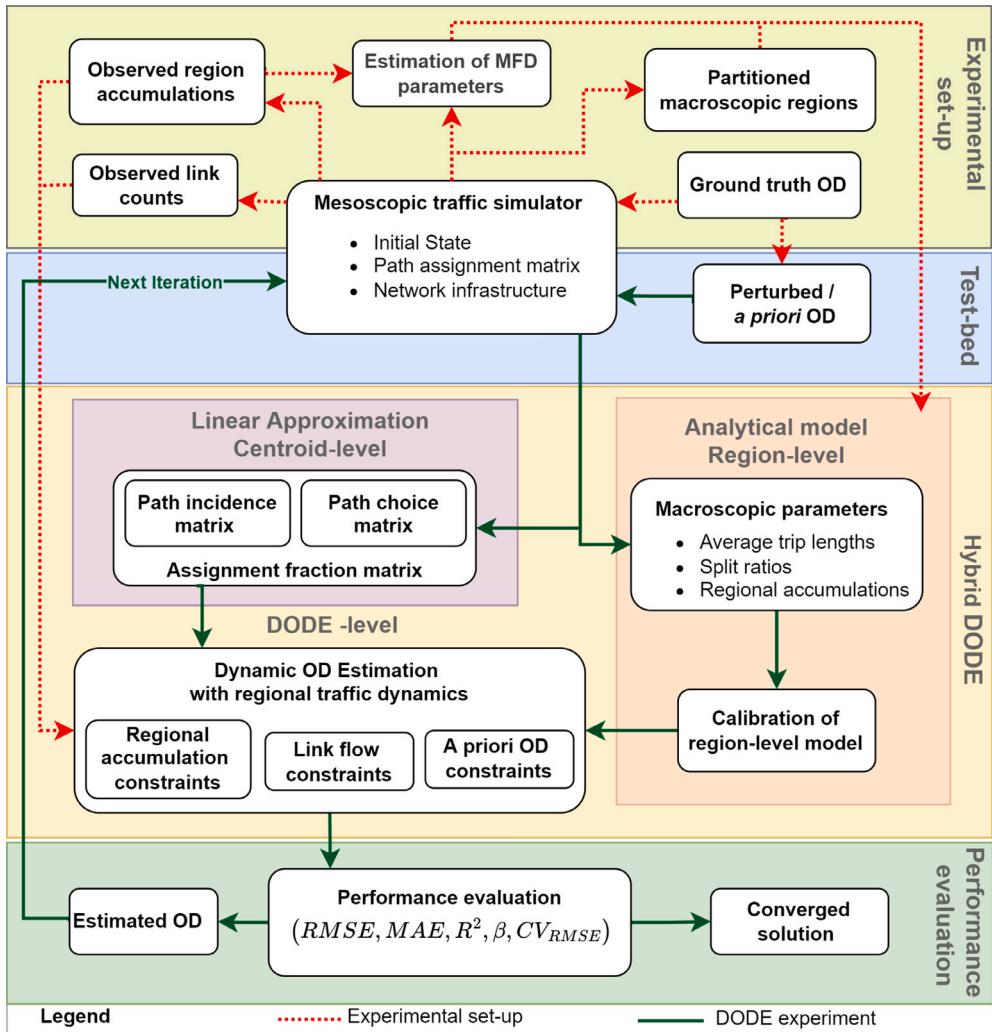


Fig. 3. Modelling framework of hybrid DODE method.

- **Estimation of MFD parameters:** A crucial input in the regional model is the functional form of MFDs for each one of the partitioned regions. In this paper, we first make use of the loop detector data that can be easily collected from the mesoscopic simulator (in the experimental setup) to derive average density (or accumulation) and average flow (or production) values in the partitioned regions. Note that the simulator in the experimental setup represents the observed traffic conditions, and the same data could be collected from existing loop detectors in a real-world application. Second, we consider a third degree polynomial function to represent the functional form for the relation between the production [veh.km/h] and the accumulation [veh] in the regions. This functional form for MFDs is kept unchanged over the iterations, as MFDs have been considered demand-independent to a large extent.

4.2. Test-bed

The test-bed layer establishes the traffic simulation experiment used for the DODE and the *a priori* OD matrices. Upon establishing the preliminary modelling components in the experimental set up, the hybrid DODE could be initiated with the *a priori* OD matrices being fed into the mesoscopic simulator. Then, the hybrid DODE will run iteratively completing the modelling components in test-bed, hybrid DODE and GOF (connected by green solid lines in Fig. 3) until a converged solution is received. Below are the essential components of the test-bed layer.

- **Mesoscopic simulator:** Traffic loading mechanism is implemented through a mesoscopic traffic simulator with the capability of evaluation up to individual vehicle-level. The selection of mesoscopic simulator instead of a microscopic approach reduces the simulation burden, and the selection of open source or commercial software tool is not an inherent component of the

proposed methodology. As a good modelling approach, it is recommended to use data by averaging several replication runs of the mesoscopic simulator in order to eliminate stochastic bias.

- *A priori OD flows:* A priori OD estimates are an input to the framework, which should be representative of the initial OD flow estimations. While a priori OD estimates (or seed matrix) may be obtained from surveys in reality, here, we use a perturbation function to derive them. The perturbation function in the form of; $q^0 = q^g \times \tau$ gives the perturbed OD matrix (q^0) by transforming the ground truth OD matrix (q^g) with an error vector τ generated by random uniform, a random Gaussian, scaled up/down error terms. This perturbation formula has been adopted from [Antoniou et al. \(2016\)](#), and we keep them as they are in order to produce comparable scenarios. However, they can be adjusted in future studies to include more noise in the prior demand flows.

4.3. Hybrid DODE

The third layer of the modelling framework is the hybrid DODE. Here, we have three allotments: region-level, centroid-level and DODE-level. The centroid-level relates to the components in the traditional (conventional) DODE problem given in Section 3.2. The region-level relates to the generation of macroscopic parameters that contribute to regional modelling of traffic dynamics as given in Section 3.1, and the calibration of region-level model to DTA model as explained in following Section 4.3.1. Finally, the DODE-level is the component where we incorporate hybrid DODE formulation explained in Section 3.3 that combine both region and centroid-level traffic models.

Note that centroid-level and region level are analytical approximations to the DTA in the test-bed layer (or lower level of [Fig. 1](#)). Hence, input parameters to these models could be extracted from the DTA model if possible or incorporate observed data when necessary. The parameters required for the hybrid DODE layer could be summarized as follows:

- *Assignment fraction matrix:* The assignment matrix establishes the relation between OD flows and link counts by defining the fraction of OD flows which start their journey in a previous time step, and contributing to link counts in the current time step. The traditional methods to derive the assignment matrix are based on deriving a route choice model considering utility (travel cost) maximization principles ([Ashok and Ben-Akiva, 2000](#)). In this study we derive assignment matrices from a dynamic user equilibrium experiment using the DTA model. The use of data-driven methods (vehicle location data, vehicle trajectory, etc.) to derive assignment matrices, as demonstrated in [Nigro et al. \(2018b\)](#), [Krishnakumari et al. \(2020\)](#), [Ros-Roca et al. \(2022\)](#) are possible alternatives that are beyond the scope of this study.
- *Macroscopic parameters:* The multi-region MFD dynamics defined in Section 3.1 require the estimation of parameters, such as MFD coefficients ($a_I, b_I, c_I \forall I \in \mathcal{R}$) for all regions, average trip lengths ($L_{II}, L_{IH}, \forall I \in \mathcal{R}, H \in \mathcal{G}_I$) and split ratios ($\theta_{I,J}^H, \forall I, J \in R, H \in \mathcal{G}_I$) for the simulation time horizon. Numerous studies have utilized manifolds of data sources including loop detectors, GPS trajectories, simulated data, and mobile phones to estimate these variables ([Saeedmanesh and Geroliminis, 2016](#); [Leclercq et al., 2014](#); [Ambühl et al., 2021](#)). However, the assortment of these estimates taken from several data sources include measurement and estimation errors which result in accrual of total error within the MFD model. Nonetheless, in the proposed framework, we can take full advantage of simulation data to derive macroscopic parameters, rather than relying on limited observations from particular sensors in real world settings. In other words, we can make use of vehicle trajectory data to estimate the macroscopic parameters that are required in the region-level model. The individual vehicle paths given as a sequence of sections (mesoscopic trajectories) extracted from the simulation experiments are used to derive initial values of macroscopic parameters ($\theta_{I,J}^H, L_{II}, L_{IH}$). A spatial-temporal grouping of vehicle trajectories (mesoscopic level) based on the partitioned regions is needed to produce time dependent split ratios and average trip lengths, as described in [Yildirimoglu and Geroliminis \(2016\)](#). Note that the estimation of the macroscopic parameters from the simulator data is repeated at every iteration, as the updated OD flows can trigger new conditions, see [Fig. 3](#).

4.3.1. Calibration of region-level modelling

We propose a calibration process of MFD parameters prior to the DODE problem such that the MFD model outputs coincide with outputs from the mesoscopic simulator. Similar attempt is taken in [Sirmatel et al. \(2021\)](#) which adequately estimated MFD parameters ($a_I, b_I, c_I, L_{II}, \theta_{IJ}^H$) using simulation (traffic states) results. We present an optimization-based calibration mechanism that aims to minimize the mismatch between the regional accumulation (N_I) observed from DTA model (simulation) and the multi-region MFD model. We formulate the optimization problem as follows;

$$\underset{\mathbf{L}}{\text{minimize}} \sum_{t_c=1:T_c} \left\{ \sum_{I=1:R} (N_I^S(t_c) - N_I^M(t_c))^2 \right\} \quad (6a)$$

subject to for $t_c = 0 \dots \delta t_c \dots T$,

$$\mathbf{L}^0(t_c) \cdot \gamma_{lb} \leq \mathbf{L}(t_c) \leq \mathbf{L}^0(t_c) \cdot \gamma_{ub}, \quad (6b)$$

$$\mathbf{N}^M(t_c + 1) = g(\mathbf{Q}(t_c), \mathbf{N}^M(t_c), \mathbf{L}(t_c), \boldsymbol{\theta}(t_c)), \quad (6c)$$

$$\mathbf{0} \leq \mathbf{N}^M(t_c) \leq \mathbf{N}^{Jam}, \quad (6d)$$

$$\gamma_{ub}, \gamma_{lb} \in \mathbb{R}^+, \quad t_c \in T_c \quad (6e)$$

Eq. (6a) targets to minimize the squared difference between regional accumulation values given by the mesoscopic simulator ($N_I^S(t_c)$) and regional accumulation values given by the region-level model ($N_I^M(t_c)$) by considering $L = \{L_{IH}(t_c), \forall I, H, t_c\}$ as decision variables. Note that the output of the mesoscopic simulator are obviously link-level measures; however, it is straightforward to derive the regional accumulation values from them.

Eq. (6b) constrains the average trip lengths $L(t_c) = \{L_{IH}(t_c), \forall I, H\}$ within a boundary defined around the initial average trip lengths ($L^0(t_c) = \{L_{IH}^0(t_c), \forall I, H\}$), which are extracted from vehicle trajectories in the mesoscopic simulator. Note that $L_{IH}(t_c)$ represents the average distance traversed by vehicles going from region I ($\in \mathcal{R}$) to H ($\in \mathcal{G}_I$) departing at time t_c , and this can be directly calculated considering all vehicle trajectories available from the mesoscopic simulator. Here, γ_{ub} and γ_{lb} are the non-negative constants which have to be selected such that the resulting optimal average trip length values are limited to a truncated boundary around the initial value $L_{IH}^0(t_c)$. We choose L as the decision variables of the optimization framework, because the trajectory-based average trip lengths may not always match outflow-based calculations (Yildirimoglu and Geroliminis, 2016).

Eq. (6c) defines the MFD traffic dynamics using $g(\cdot)$, which is the accumulation-based MFD model used for region-level modelling given in Section 3.1. Similar to Eq. (5e), function $g(\cdot)$ gives the accumulation in the next time step ($N^M(t_c + 1) = \{N_{I,J}^M(t_c + 1), \forall I, J\}$) by taking current accumulation states ($N^M(t_c) = \{N_{I,J}^M(t_c), \forall I, J\}$), vector of demand flows ($Q(t_c) = \{Q_{IJ}(t_c) \forall I, J\}$), average trip length ($L(t_c)$) and split ratios/route choice variables ($\theta(t_c) = \{\theta_{I,J}^H(t_c) \forall I, J, H\}$) as inputs. Further, $g(\cdot)$ solves the ordinary differential equations given in Eqs. (2a)–(2b) by temporal discretization of input parameters. Note that average trip lengths $L(t_c)$ in this equation are the output of the calibration procedure so that the mismatch between the aggregate model and the mesoscopic simulator is minimized.

The Eq. (6d) is included in the optimization problem to maintain the non-negativity in the regional accumulations and limit the regional accumulations below the jam accumulation ($N^{\text{jam}} = \{N_I^{\text{jam}} \forall I\}$) level. The resulting calibration problem is a non-linear, non-convex problem for which we have to use non-linear solvers such as interior-point algorithms to derive local optimal points.

4.4. Performance evaluation

The fourth layer of the proposed framework focuses on evaluating the performance of the hybrid DODE method. We consider several goodness of fit (Gof) measures for both link counts and OD flow values. Antoniou et al. (2016) provides a brief summary on recommended GoF measures and Buisson et al. (2014), Toledo and Koutsopoulos (2004), Ciuffo (2012), Hollander and Liu (2008) provide further details on the selection of GoF tests relevant for comparison and evaluation of traffic simulation models. On that note, we propose the mean absolute error (MAE) given in Eq. (7a) and root mean squared error (RMSE) given in Eq. (7b) of link counts. The choice of RMSE allows for a higher sensitivity to outliers when evaluating the error, whereas MAE provides an average error measurement with equivalent sensitivity to all values.

$$MAE = \frac{1}{m \times T} \left(\sum_{l=1}^m \sum_{t=1}^T |f_l(t) - \hat{f}_l(t)| \right) \quad (7a)$$

$$RMSE = \sqrt{\frac{1}{m \times T} \sum_{l=1}^m \sum_{t=1}^T (f_l(t) - \hat{f}_l(t))^2} \quad (7b)$$

In OD estimation, the ability to observe a strong match between estimated ODs and ground truth ODs (linearity), as well as the validity of unbiasedness, is critical. Hence, we use R^2 to measure the overall fit between the estimated and ground truth ODs. Following Antoniou et al. (2016), the slope ($\hat{\beta}_1$) of the linear regression estimator (i.e., $y = \hat{\beta}_0 + \hat{\beta}_1 x$) between ground truth and estimated OD flows is also investigated. This is an important measure that characterizes over/underestimation of OD flows. If $\hat{\beta}_1$ is closer to 1, it means the overall fit is impartial. While a complex statistical test may be used to demonstrate this agreement, selecting $\hat{\beta}_1$ is straightforward and practical. In addition, we report the computational times for reference.

In addition to the common GOF measurements, we propose specific measurements to further investigate the scalability and consistency aspects. Maximum possible relative error (MPRE) is a measure proposed by Yang et al. (1991) to investigate the theoretical reliability of an estimated OD matrix. While the solution space of DODE is unexposed (oblivious due to non-observability of all links), the MPRE can be used as a relative measure on the solution space's upper bound. As a result, a lower MPRE indicates a more constrained solution space. (see Yang et al., 1991 for more details).

We utilize the following reliability index (Re) based on MPRE to assess the estimation's reliability:

$$Re = \frac{1}{1 + \sqrt{\frac{\sum_{k=1}^n \gamma_k^2}{n}}} \quad (8)$$

Here, $\sqrt{\frac{\sum_{k=1}^n \gamma_k^2}{n}}$ is the MPRE and γ_k is defined as: $\gamma_k = \frac{q_k^g(t) - q_k(t)}{q_k(t)}$ where $q_k^g(t)$ is the ground-truth OD flow and $q_k(t)$ is estimated OD flow. It should be emphasized that Re index varies between 1 and 0 where it will be 1 when MPRE is 0 and vice versa. Thus, DODE method which can give Re measures closer to one are considered as more reliable estimates compared to rest.

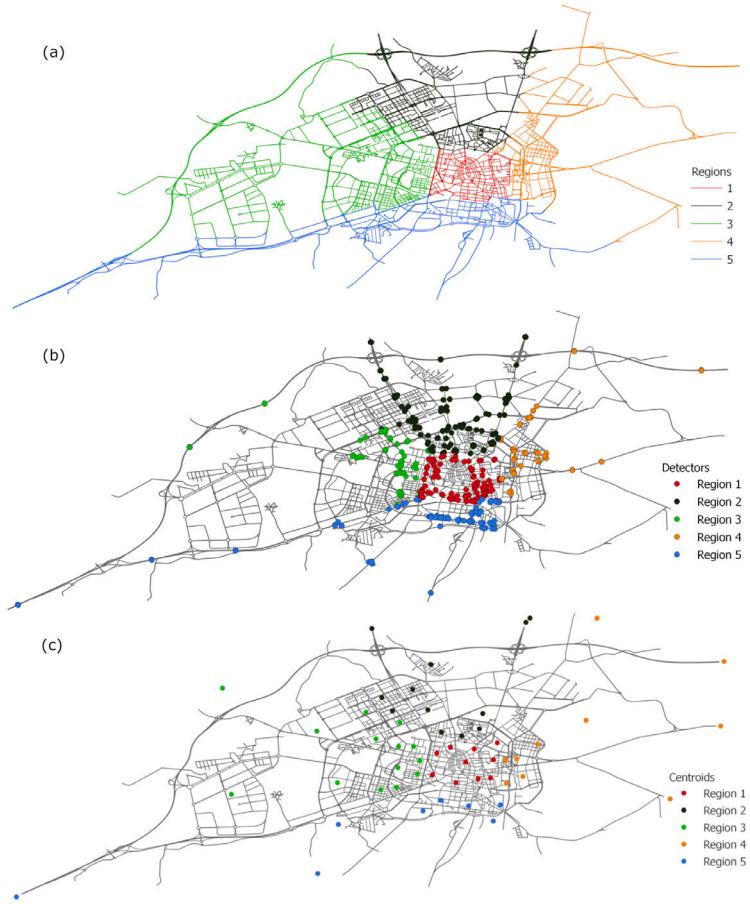


Fig. 4. Traffic network of Vitoria, Spain (a) regional representation, (b) traffic detectors (c) origin–destination centroids.

We propose the coefficient of variation of the RMSE (CV_{RMSE}) to measure the relative closeness of the predictions to the actual values which is given by:

$$CV_{RMSE} = \frac{\sqrt{\frac{1}{n \times T} \sum_{\forall n} \sum_{\forall t} (q^g(t) - q(t))^2}}{\frac{1}{n \times T} \sum_{\forall n} \sum_{\forall t} (q(t))} \quad (9)$$

This measure could be used to empirically evaluate the consistency of the DODE models by comparing the estimation results across uncongested and congested scenarios.

5. Case study and results

We evaluate the proposed DODE model in a large-scale traffic network with both congested and uncongested traffic scenarios. The traffic network of Vitoria, Spain which is adopted in [Antoniou et al. \(2016\)](#) as a benchmark network for OD estimation, and utilized in [Cantelmo et al. \(2020\)](#), [Qurashi et al. \(2020\)](#) is considered in this study.

[Fig. 4](#) presents three layers of the traffic network in Vitoria, Spain, which includes 2884 nodes and 5799 links. The first layer, [Fig. 4-\(a\)](#) provides the partitioned regions identified for multi-region MFD dynamics. We have identified 5 regions ($\mathcal{R} = [1, 2, 3, 4, 5]$) that demonstrate a homogeneous traffic condition in each region. Note that the regional configuration was selected following the partitioning criteria described in Section 4.2, which is an extension of the approach proposed by [Mariotte et al. \(2020\)](#). [Fig. 4-\(b\)](#) shows the spatial distribution of 395 detectors available in the network, and [Fig. 4-\(c\)](#) shows the map of 57 origin–destination centroids considered in the DODE. The links, detectors and centroids are colour coded based on the regional configuration.

A mesoscopic simulation experiment of the network based on Aimsun Next 8.4 traffic simulator is used as the test-bed. The benchmark experiment established by [Antoniou et al. \(2016\)](#) is considered as the ground truth scenario, and observed (ground-truth) link counts, regional accumulations were obtained by assignment of the ground truth OD in the traffic simulator, as presented in [Fig. 3](#). Following the benchmark study, we use OD paths and path flow computations obtained from a dynamic user equilibrium (DUE) based experiment to obtain the path assignment matrix used for the centroid-level formulation.

Table 2
Experiment parameters.

Aimsun Next 8.4 simulation					
Simulation class	Mesoscopic				
Simulation time	1-h				
Simulation demand time step (t)	15-min				
Perturbation of demand					
Demand scenario	D7	$\tau_{D7} = 0.7 + 0.3 \times U(0, 1)$			
$q^0 = q^g \times \tau_D$	D10	$\tau_{D10} = 1.0 + 0.3 \times U(0, 1)$			
	D11	$\tau_{D11} = 1.1 + 0.3 \times U(0, 1)$			
MFD Estimates					
Region MFDs	1	2	3	4	5
a	2.40E-06	6.63E-07	1.73E-06	3.62E-06	7.15E-07
b	-1.02E-02	-6.42E-03	-1.25E-02	-1.86E-02	-7.13E-03
c	1.07E+01	1.54E+01	2.25E+01	2.38E+01	1.77E+01
Jam accumulation (veh)	1940	5200	3800	2380	4620
Network length (km)	70	181	270	146	190
Calibration of MFDs					
Time step (δt_c)	3-min	Centroid level parameters			
RK4 prediction time step	1-s	Assignment matrix			
γ_{lb}	0.7	η	Time step (Δt)		
γ_{ub}	1.3	Route update	15-min		
Hybrid OD estimation					
Centroid-level time step (t)	15-min	β_{lb}	0.75	α_{lb}	0.92
Region-level time step (t_c)	3-min	β_{ub}	1.25	α_{ub}	1.08

Table 2 presents a summary of the parameters used in the experiment. Although the total simulation time is indifferent, the discretized time steps of centroid-level and region-level models are not the same. The centroid-level demand flows and measurements are discretized into 15-minute intervals following the benchmark study of Antoniou et al. (2016) (t in Eqs. (5a) and (5f)), and the region-level demand flows, measurements and dynamic parameters (average trip length, split ratios) are reported with 3-minute intervals (t_c in Eqs. (5a)–(5f)). As mentioned previously, the time step in the regional model should be smaller than the free flow travel time in the region considering the minimum average trip length ($\min\{L_{IH}(t)\}$). Modelling components of the hybrid DODE framework have been developed using MATLAB scripts, and Aimsun Next 8.4–MATLAB interactions were handled using python scripts. The non-convex nonlinear optimization problem proposed in the framework is solved using the interior-point solver (IPOPT) (Wächter and Biegler, 2006) included in the optimization library of CasADi (Andersson, 2013).

In this article, we present DODE results of three demand scenarios, obtained by up-scaling (or down-scaling) the true demand (q_k^g) values using a random noise function ($q_k^0 = q_k^g \times \tau_D \quad \forall k \in n$). As presented in **Table 2**, the demand scenario D7 (down-scaled by 15% on average) is used to evaluate the proposed method in uncongested conditions and the results are compared with existing methods. The scenario D10 (up-scaled by 15% on average) and the scenario D11 (up-scaled by 25% on average) reveal the performance of the proposed method in moderately and highly congested traffic conditions, respectively. Note that the ground truth scenario remains fairly uncongested in all three applications. We test the capability of the algorithm in estimating demand flows under a variety of traffic conditions.

5.1. Centroid-level evaluation

The DODE procedure that we propose builds on an iterative process to derive the optimal solution, as presented in **Fig. 3**. A single iteration in this process is shown by the components which are connected by solid lines in **Fig. 3**. A single iteration includes a mesoscopic simulation experiment (with ten replications), Calculation of macroscopic parameters using simulation outputs, calculation of centroid-level assignment matrices, a run of MFD calibration, and a run of hybrid DODE. This iterative process enables the synchronous convergence in OD flows and link counts to ground truth/observed conditions, as the changes in demand flows trigger changes in the simulator results, and vice versa. **Fig. 5** affirms this synchronous convergence of OD flows and link counts for the three demand scenarios D7 (red), D10 (yellow), and D11 (green). In **Fig. 5-(a)** we observe that $\hat{\beta}_1$ approaches to 1 within few iterations and remain stable across all three scenarios. R^2 shown in **Fig. 5-(b)** does not show any significant deviations, albeit always very close to 1. Correspondingly, **Figs. 5-(c)** and **5-(d)** depict the convergence of MAE and RMSE in link counts resulting from the estimated OD flows. We can see that both MAE and RMSE reach their minimum over the first few iterations and converge to steady values. The star sign (\star) on each plot indicates a decision point to stop the DODE, where the change in error (MAE or RMSE) between three consecutive iterations are below 3%. The number of iterations required for convergence is at most 8 in the considered demand scenarios, which highlights the efficiency of hybrid DODE. Note that the existing DODE methods require a higher number of iterations (more than 50 iterations) to reach a stable solution (Antoniou et al., 2016; Osorio, 2019; Tympakianaki et al., 2018).

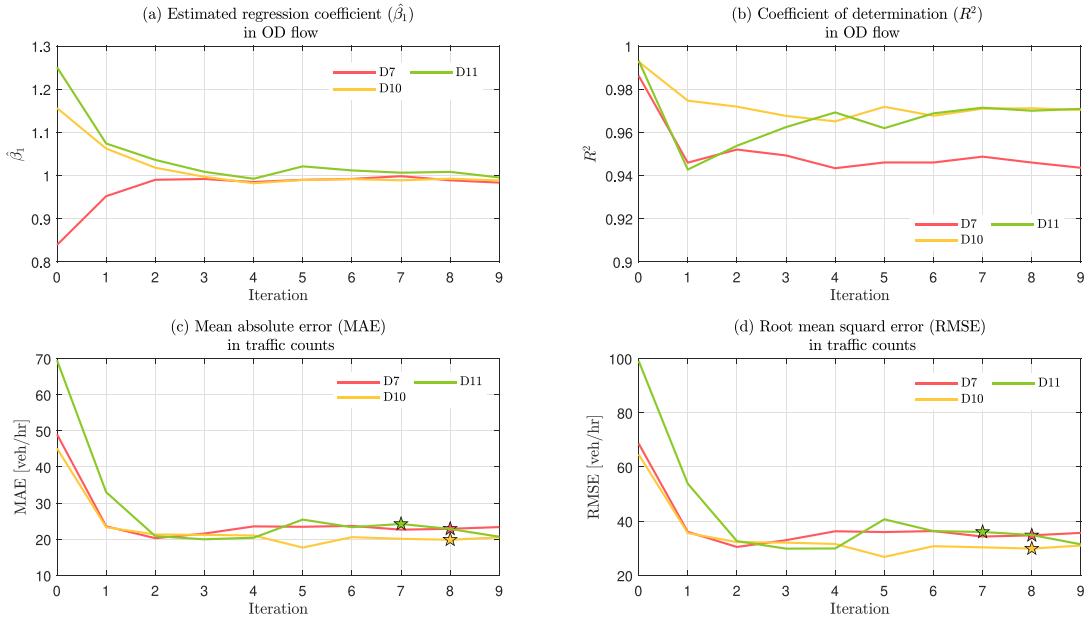


Fig. 5. Convergence of estimation error. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 6 presents the centroid-level or link-level results corresponding to the iteration marked by the star sign in **Fig. 5** for each demand scenario. The initial condition of perturbed vs ground truth OD flows for each demand scenario are plotted in **Fig. 6-(a)**, (b) and (c). Note that the coefficient of the regressor (β_1) connotes the average perturbation in each demand scenario. **Fig. 6-(d)**, (e) and (f) present the scatter plot of ground truth OD flows vs estimated/optimized OD flows which result from the hybrid DODE procedure (corresponding to the iteration marked by the star sign in **Fig. 5**). The β_1 values observed in regression lines rejects the null hypothesis $\beta \neq 1$ at all significant levels, and also we observe a regression $R^2 \geq 0.95$ at all scenarios. Similarly, **Fig. 6-(g), (h)** and (i) illustrates the agreement of ground-truth link counts with simulated link counts resulting from the estimated OD flows. Overall, the results indicate a significant agreement between ground truth OD flows and estimated OD flows, and between simulated link counts and ground-truth link counts.

While the results show that the algorithm performs quite well in scenarios ranging from uncongested to highly congested, it is important to emphasize the capabilities of the proposed hybrid DODE in estimating OD flows in congested traffic scenarios, as revealed in the demand scenarios *D10* and *D11*. The mapping between OD flows and link counts tends to be linear when traffic conditions are uncongested and traffic speeds are close to free flow speed. Existing methods are capable of producing accurate OD estimates considering the linear relation. However, such methods fail to perform well when the network is congested. The hybrid DODE method we introduce, captures this non-linear behaviour of traffic propagation in congested regions via MFD traffic dynamics, and guides the DODE to an optimal solution over the iterations as we see in *D10* and *D11*. A comparison with the baseline methods is presented in Section 5.3 below.

Although we see a very good match between ground-truth link counts and simulated link counts, it is important to understand the distribution of error across traffic detectors. **Fig. 7** exhibits the cumulative distribution of error across traffic detectors over iterations. A significant level of error is observed among detectors at the first iteration, where we see about 40% of detectors demonstrating an error of more than 50 veh/h in *D7* and *D10* scenarios. The error diminishes after the implementation of DODE where we see only about 10% of the detectors having such an error. Similarly, in *D11*, 65% of the detectors demonstrate an error of more than 50 veh/h, and that share has curbed down to 10% after running the DODE. We see that the final error values are quite small for most detectors, but there exists about 5% of detectors with high error measures (irregularities) which cause an increase in the aggregate measurements of MAE and RMSE.

5.2. Region-level evaluation

This section investigates the region-level components of the hybrid DODE framework given in **Fig. 3**. We focus on the calibration process and the convergence of region-level traffic outputs, which are in fact intermediate steps towards producing the final centroid-level estimates presented in the previous section. We run the calibration process at every iteration before the hybrid DODE as shown in **Fig. 3**. The calibration process aims to minimize the mismatch between the region-level traffic conditions observed in the mesoscopic model and the MFD model. This is a crucial step before employing the region-level model within the hybrid DODE framework, as it guarantees a certain level of agreement between region-level and centroid-level modelling. **Fig. 8** demonstrates the

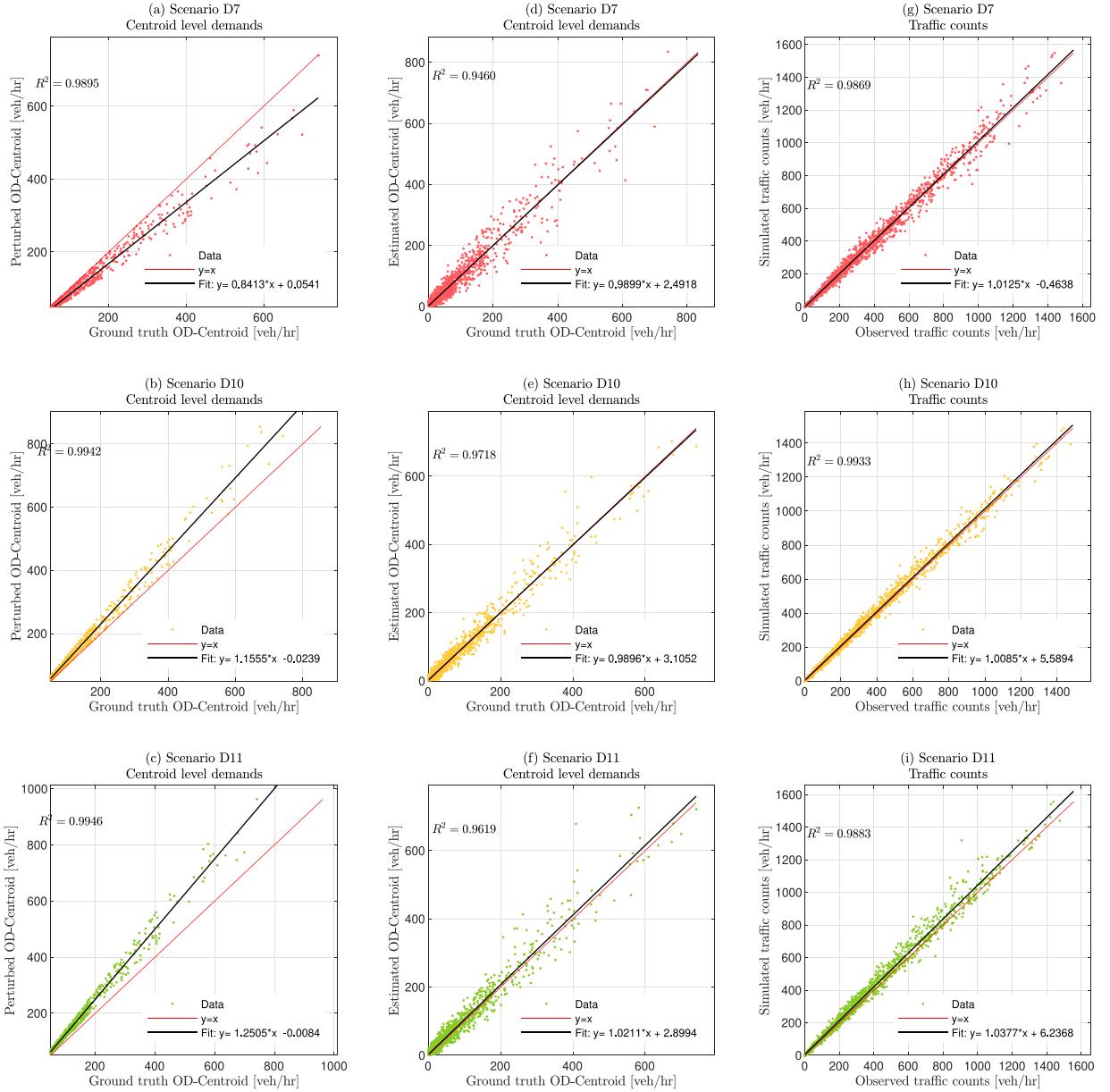


Fig. 6. Estimation results of the proposed hybrid DODE.

calibration process where it presents accumulation curves of five regions in the first iteration in the D10 demand scenario. Each sub-graph shows the accumulation curve before calibration, after calibration, and accumulation curve obtained from mesoscopic simulation. We can see that the calibration process is successful in solving the optimization problem in Eqs. (6a)–(6d) and producing an optimal set of $L_{IH}(t)$ such that the accumulation curve after calibration is in resemblance of the accumulations observed via the mesoscopic simulator. Similar results have also been obtained for other scenarios D7 and D11.

The calibration component ensures that traffic dynamics from the analytical model (multi-region MFD dynamics) used in DODE are analogous to the traffic conditions observed in the mesoscopic simulator. Importance of calibration is clear by observing the disparity in the accumulation curves before and after accumulation. The critical accumulation level is the accumulation level reached by a region before it becomes congested, at which point the region production is maximized. For example, the accumulation curve in region-3 which stays below the critical accumulation level before calibration exceeds the critical accumulation conditions upon calibration. Hence, deviance of calibration and feeding initial values ($L_{IH}^0(t)$) instead of calibrated values ($L_{IH}(t)$) may lead to a significant mismatch within the hybrid DODE framework and may eventually result in sub-optimal (or incompatible) outputs. Note that $L_{IH}^0(t)$ is measured from vehicle trajectories observed in the mesoscopic simulation. However, as it is expected to have a certain

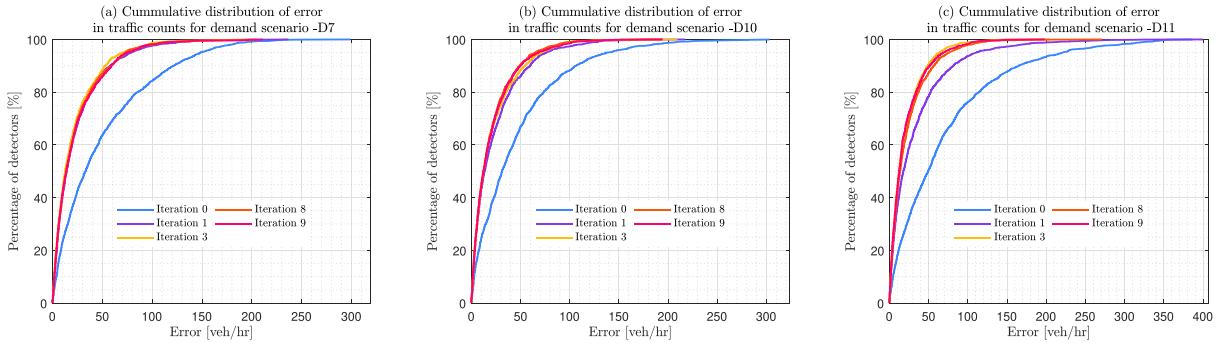


Fig. 7. Convergence of estimation error.

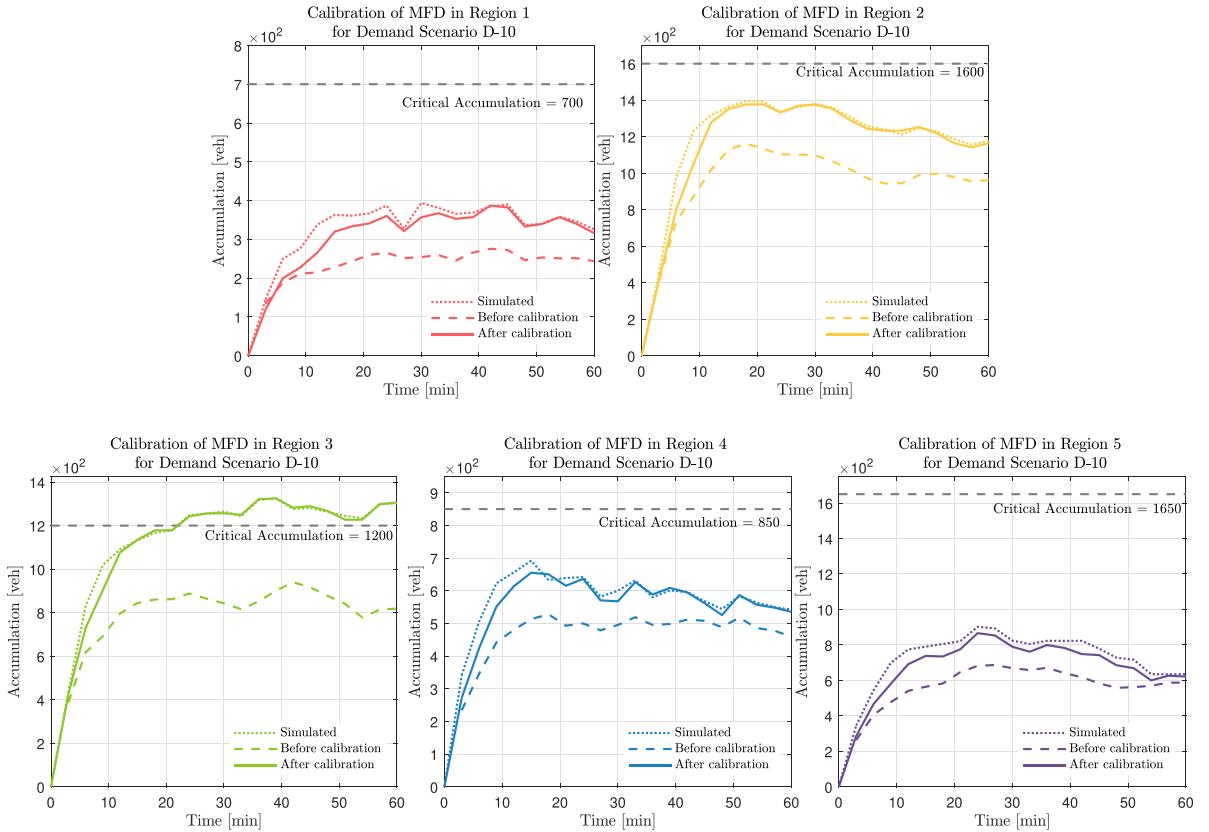


Fig. 8. Calibration of regional traffic dynamics in demand scenario D10.

disagreement between trajectory-based trip lengths and outflow based calculations (Yildirimoglu and Geroliminis, 2016), they are considered as the decision variables in the calibration process (Eqs. (6a)–(6d)).

The performance of the proposed hybrid DODE is highly dependent on the regional traffic dynamics. Hence, it is important to understand the behaviour of MFD dynamics within the hybrid DODE. Eq. (5e) defines the MFD dynamics in the DODE problem, and the outcome of this constraint over iterations for demand scenario D11 is shown in Fig. 9. Five graphs in Fig. 9 show the accumulation curves from the mesoscopic simulations for each region in iteration 0 (—), 1 (—), 2 (—), 5 (—), and 9 (—) along with the ground-truth accumulation curve (····) from the mesoscopic simulator with the ground truth demand. The scenario D11 is a 25% over-estimate of the ground truth OD with random noise. Consequently, iteration-0 results in a significant level of congestion in the network, where we can see region-2 reaching the critical accumulation level, region-3 going beyond the critical accumulation level and other regions demonstrating uncongested traffic conditions. Identification of regions with homogeneous traffic conditions has allowed us to observe spatial variations in traffic dynamics, which indicates the high accuracy of the partitioning step.

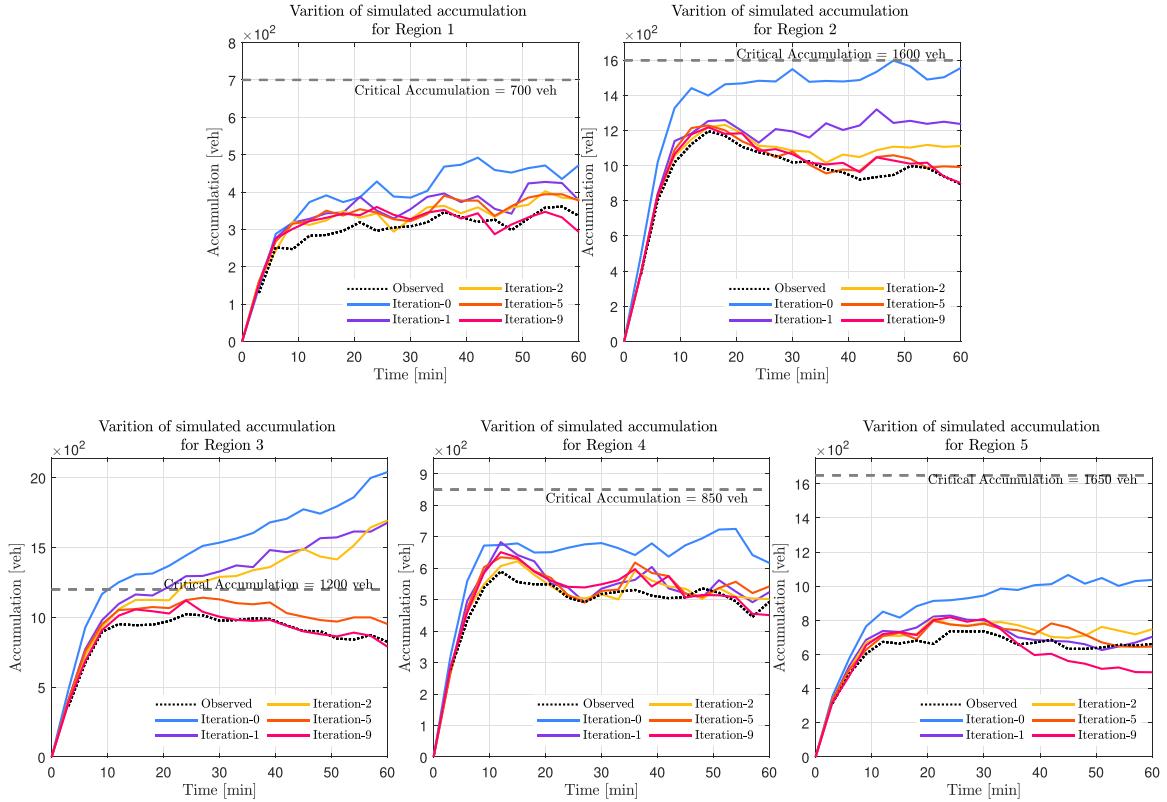


Fig. 9. Convergence of regional traffic dynamics in demand scenario D11. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In Fig. 9, we see a convergence of the accumulation curves over iterations. There exist contrasting deviations between the observed (ground-truth) accumulation curves and simulated accumulation curves at iteration-0 for all five regions, but especially in region-2 and region-3. However, the mismatch between accumulation curves diminishes over iterations, and the two curves eventually become quite similar. It takes several iterations for the mismatch between the initial measurements (i.e. iteration-0) and the observed conditions to diminish, mainly because the proposed hybrid framework still includes a centroid-level formulation which is a linear approximation. Nevertheless, over the iterations, the region-level model captures the inherent nonlinear relations within the system, and produces a strong match with the observed traffic conditions. This highlights the importance of the region-level traffic model in capturing nonlinear characteristics of traffic congestion, and ultimately producing consistent OD estimations in congested scenarios.

Note that we do not expect an identical match between simulated and ground-truth accumulation curves, as regional model is incorporated to the optimization problem as an inequality constraint (Eq. (5f)). Our main target is to optimize centroid-level OD flows that result in ground-truth link counts in the network. We consider the resulting regional accumulation values as an intermediary result or a by-product of the hybrid framework, which helps narrow down the solution space. Note that regional ODs are a linear combination of centroid-level ODs via Eq. (5c). Therefore, manipulating regional ODs via MFD dynamics disrupts the centroid-level DODE and enables centroid-level OD flows to satisfy both region-level and centroid-level formulations. Thus, the synchronous effect of these constraints enables the centroid-level estimations to capture the traffic dynamics observed at the region-level and eventually guide the optimization problem in the direction of the observed conditions in the large-scale traffic network.

This approach closely aligns with the existing works that focus on decomposition methods to address the scalability issue in DODE, as highlighted in Section 2. However, our approach differs in the sense that it includes a physics-based upper-level model that describes traffic dynamics in the regional context rather than relying on statistical approaches that yield purely mathematical changes in the problem formulation and solution space. This allows us to explore the interrelated (dependent) behaviour of the regions which is not taken into consideration in existing works.

Overall, the integration of centroid-level and region-level modelling produces promising results and demonstrates its potential in overcoming consistency and scalability issues in large-scale applications.

Table 3
Comparison with LSQR.

Demand scenario	OD Flow				Link counts (veh/h)			
	β_1		R^2		MAE		RMSE	
	Hybrid	LSQR	Hybrid	LSQR	Hybrid	LSQR	Hybrid	LSQR
D7	1.005	0.913	0.948	0.964	23.364	12.692	35.68	18.176
D10	1.004	1.089	0.972	0.974	21.048	20.616	31.588	29.736
D11	1.021	1.150	0.962	0.972	20.852	29.952	32.664	43.076

5.3. Comparison of hybrid DODE

In this subsection, we compare the hybrid DODE model with existing methods. The benchmark platform (Antoniou et al., 2016) recommends several solvers for DODE such as sparse least squares solver (LSQR) (Bierlaire and Crittin, 2004), SPSA and improved versions of SPSA (Cipriani et al., 2011; Nigro et al., 2018a). While LSQR considers an assignment matrix in formulation, the SPSA disregards it and considers the loading mechanism as a black box. In this section, we directly compare the results of hybrid DODE with the results of the LSQR method developed by Paige and Saunders (1982) and regarded as one of the best performing algorithms in the recent literature (Antoniou et al., 2016). Furthermore, we compare the findings in relation to SPSA from the benchmarking study of Antoniou et al. (2016).

The comparison is summarized in Table 3, which shows β_1 , R^2 for OD flows, and MAE, RMSE for link counts. Table 3 shows that, in the D11 scenario, the hybrid DODE method outperforms the LSQR algorithm in terms of both link count (MAE and RMSE) and OD flow (β_1) estimations. Note that both algorithms produce fairly high R^2 values in all scenarios. This implies that LSQR fails to correct the overestimation in *a priori* OD flows and link counts, while the hybrid DODE captures the overestimation issue and corrects it to a large extent. The hybrid DODE is the clear winner in this congested scenario (D11), as it performs better than LSQR in terms of both link counts and OD flows. In the less congested scenarios (D7 and D10), Table 3 shows that LSQR performs similarly or slightly better in terms of link counts (MAE and RMSE); however, it fails to capture under/overestimation in OD flows. Particularly, the slope β_1 produced by LSQR never gets as close to 1 as the hybrid DODE model. This implies that LSQR might be running into indeterminateness issues, where it is able to create a decent match between the observed and estimated link counts at the expense of causing a mismatch with the ground-truth OD flows. Obviously, the ground-truth OD flows are not observable within DODE framework, but this mismatch in OD flows is an artefact caused by the lack of constraints in LSQR. The hybrid DODE includes constraints that represents the bounds on centroid-level OD flows and the region-level traffic interactions, which in turn allows the algorithm produce better OD estimation results than LSQR.

Overall, the hybrid DODE outperforms LSQR in terms of OD estimation. LSQR is a generic optimization algorithm which can be used to solve an $F \cdot x = b$ system where matrix F is large and sparse. The algorithm involves only a centroid level approximation where the assignment matrix ($\hat{A}(\lambda|t)$) is incorporated when deriving the matrix F . Note that the optimization problem solved in LSQR has no constraints; LSQR is an unconstrained optimization algorithm. Although LSQR performs similarly to the hybrid DODE in link count estimations (MAE and RMSE) in some scenarios, it always fails to capture under/overestimation in OD flows, as seen in Table 3.

DODE formulation in Eq. (1) contains traffic assignment as a linear mapping as in Equation. (4). Therefore, the optimization problem has limited information about traffic dynamics and fails to diagnose deviations in traffic dynamics resulting due to over/under estimated OD flows. In contrast, the hybrid DODE method includes the region-level MFD dynamics in addition to centroid-level mapping formulation. The incorporation of MFD dynamics enables the hybrid DODE method to perform well in both congested and uncongested traffic conditions. This implies that the nonlinear modelling at the region-level guides the solution in the right direction with respect to OD flows.

While the hybrid DODE method has significant improvements compared to other existing methods which employ traffic assignment matrices, it also demonstrates significant efficacy compared to methods that consider assignment matrix as a black box. The results of DODE for demand scenario D7 with improved versions of SPSA algorithm were given in method 2 (SPSA-TD), method 3 (SPSA-AD-PI) and method 4 (SPSA-CG-TR) of Antoniou et al. (2016). The reported results are: $\beta_1^{\text{method 2}} = 0.89$, $\beta_1^{\text{method 3}} = 0.93$, $\beta_1^{\text{method 4}} = 1.04$, and all three methods report R^2 greater than 0.9. The results from the hybrid DODE in Table 3 with respect to the scenario D7 are better than these three SPSA-based DODE methods. The benchmark study does not report error measurements for link counts but states that the simulated link counts gives an insignificant correlation when compared with ground-truth link counts (Antoniou et al., 2016).

The computational complexity is a crucial matter in DODE as existing methods such as SPSA based solvers report significantly high computational demand. Computational efficiency is reported in CPU computational time in this study following the benchmark of Antoniou et al. (2016). The reported average computational times for a single iteration of hybrid DODE in an Intel(R) Core(TM) i7-7700 CPU with a 3.60 GHz clock time are; calibration = 185.5 s, hybrid DODE = 422.2 s and mesoscopic simulation = 2133.7 s. Unfortunately, Antoniou et al. (2016) does not report computational times, but one can assume that they must be significantly high, as SPSA based methods need multiple runs of the mesoscopic simulation within a single iteration and require high number of iterations (more than 50) to converge.

5.4. Empirical evidence of scalability and consistency

The results presented in previous sections indicate that combined modelling of centroid-level and region-level demonstrates the potential of hybrid DODE in overcoming consistency and scalability issues in large-scale applications. This section will empirically investigate the importance of incorporating MFD based traffic dynamics in DODE to ameliorate scalability and consistency issues. We will incorporate more sensitive evaluation methods to gain a better understanding of the scalability and consistency of the system, which the summary metrics (R^2 , $\hat{\beta}_1$, RMSE, MAE) utilized in the previous section seem unable to capture.

5.4.1. Empirical evidence of scalability

The target of traditional DODE methods is finding OD flows closer to *a priori* OD flows that minimize the mismatch between ground-truth link counts and simulated link counts. In addition to this requirement, the hybrid DODE presented in this study should estimate OD flows by constraining the mismatch between simulated and observed region accumulations. This additional constraint or requirement allows the framework to shrink the feasible solution space (for OD flows) and lead the solution towards ground-truth OD flows. We can validate this claim by comparing the hybrid DODE presented in Eq. (5a) with a traditional DODE without region-level constraints (Eq. (5d) to (5f)) as follows:

$$\underset{\mathbf{q}}{\text{minimize}} \sum_{t=0:T} \left\{ \sum_{l=1:m} \left(\frac{f_l^0(t) - \hat{f}_l(t)}{f_l^0(t)} \right)^2 \right\} \quad (10a)$$

subject to for $t = 0 \dots \Delta t \dots T$,

$$\hat{f}(t) = h(\hat{\mathbf{A}}(\cdot|t), \mathbf{q}, \eta) \quad (10b)$$

$$\mathbf{q}^0(t) * \beta_{lb} \leq \mathbf{q}(t) \leq \mathbf{q}^0(t) * \beta_{ub} \quad (10c)$$

Note that variables in the optimization problem given in Eq. (10a) to Eq. (10c) have the same meanings of Eq. (5a) to Eq. (5c). This formulation is similar to existing DODE methods and we will refer to it as non-hybrid DODE model. The proposed non-hybrid DODE is a quadratic program where the feasibility and uniqueness of the solution is dependent on the assignment matrix ($\hat{\mathbf{A}}(\cdot|t)$). The assignment matrix being strictly increasing (Hessian matrix of $\mathbf{A}^T \mathbf{A}$ being positive semi-definite) is sufficient for the quadratic program to result in a unique solution (Cascetta, 2001), which requires all vehicles that enter the network being observed, in other words having all the link flows can be observed. However, large-scale traffic networks are far from such ideal conditions where all links being observed. Hence, the proposed non-hybrid DODE method becomes a non-convex problem (similar to the hybrid DODE method) where feasibility of a locally optimal point is dependent on the analytical properties of constraints and ability to identify a descent direction (approximate the Hessian matrix). We use the non-convex quadratic program solver in Gurobi (Gurobi Optimization, 2022) to solve the non-hybrid model (it is more suitable than an interior point solver). Further, we compare the results with those from the LSQR algorithm. Note that LSQR is a sparse least squares solver (linear system solver based on bidiagonalization process) which is analytically equivalent to a conjugate gradient solver (Paige and Saunders, 1982). It is important to emphasize here that the explicit bounds on OD flows (as described in Eq. (10c)) are not taken into account in LSQR algorithm as it is an unconstrained optimization solver.

Fig. 10 demonstrates the centroid-level and region-level DODE results for the hybrid, non-hybrid and LSQR DODE methods highlighted above. We use the D10 scenario where the *a priori* OD flows are overestimated by 15% and result in moderately congested traffic conditions. Fig. 10-(a), (d) and (g) present scatter plots between ground-truth OD flows vs estimated OD flows in three models. We can see that the overestimation issue has not been fully addressed in either non-hybrid or LSQR method, while the hybrid DODE produces a much better match in terms of overall agreement in OD flows. Fig. 10-(b), (e) and (h) shows the agreement between observed traffic counts and simulated traffic counts corresponding to three models. A significant dispersion in the link counts is observed in the non-hybrid method, whereas the hybrid model and LSQR demonstrate a strong agreement between observed and simulated link counts. Note that the resulting objective function value in the LSQR and the hybrid model are fairly close (see the value of J in Fig. 10-(b) and (h)), indicating the presence of locally optimal points with similar performance. Fig. 10-(c), (f) and (i) shows the scatter between aggregated region-level OD flows where we see over-estimation not being corrected in either non-hybrid or LSQR DODE methods. Notably, the non-hybrid and the LSQR methods do not use region-level OD flows as input variables. Overall, there are two important observations from the analysis above: (i) although the solver indicates an optimal solution is found, the non-hybrid DODE model performs poorly without constraints specific to the region-level model, (ii) the LSQR algorithm performs similarly to the hybrid DODE model in terms of link counts (and the corresponding objective function), but fails to produce a strong agreement in OD flows. These two observations indicate the existence of many local minima in the solution space. The incorporation of the region-level model as additional constraints (in the hybrid DODE) shrinks the feasible solution space (for OD flows) and pushes the solution towards the ground truth where both OD flows and link counts show a strong agreement.

While Fig. 10 provides a direct comparison of OD flows and link counts across the three methods, there are other measures proposed in the literature that further describe the quality of estimated OD flows. The Re index proposed by Yang et al. (1991) based on the MPRE will be used to assess the estimation's reliability (See Section 4.4).

Fig. 11-(a) illustrates the cumulative distribution of the Re as a function of the percentage of OD pairs. Note that OD pairs are sorted here in the ascending order of γ_k . Fig. 11-(b) and (c) gives the cumulative distributions of RMSE and MAE as function of percentage of link counts respectively (sorted in the ascending order of absolute error). Fig. 11-(a) shows that the Re index drops

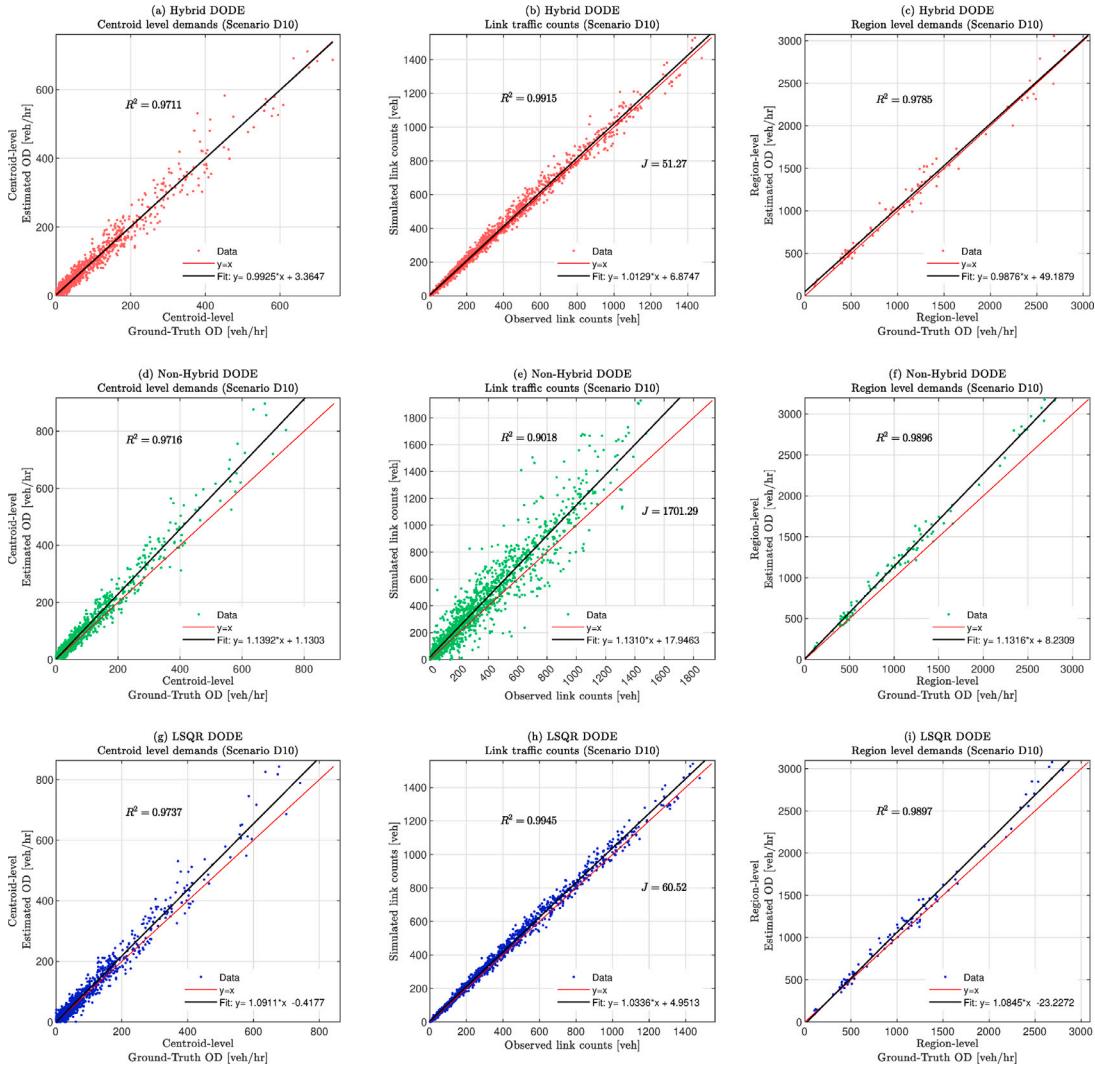


Fig. 10. DODE results given by hybrid, non-hybrid and LSQR solver methods.

to a relatively low value in LSQR method compared to other two methods, while we see LSQR reporting low RMSE and MAE values. Conversely, the non-hybrid DODE method reports higher Re index values close to the hybrid DODE, while RMSE and MAE distributions indicate that it fails in minimizing the error between link counts. On the other hand, the hybrid DODE model produces high Re values indicating highly reliable results with respect to OD flows while maintaining low errors in link count estimates. Note that Fig. 10 does not exhibit a large difference in estimation performance (of OD flows) between LSQR and other methods, while the reliability metric presented in Fig. 11-(a) shows a significant contrast as it is a relative measure and highly sensitive to deviations.

It is clear that LSQR is likely to converge to lesser reliable OD estimates as it solves the problem via a series of approximations to the linear system (Paige and Saunders, 1982). Note that the formulation used in the LSQR algorithm does not have specific bounds on OD flows (see Eq. (10c)) which significantly affects the estimation reliability (of OD flows) albeit producing a strong match in link counts. On the other hand, the non-hybrid DODE method shows a similar estimation reliability in OD flows because of constraints, but fails in minimizing link count error. The error in link counts in non-hybrid DODE emphasizes the significance of a region-level model. By including a region-level traffic model, the hybrid DODE technique improves both estimation reliability of OD flows and link count errors, making it more suitable for OD estimation problems where the number of unknowns largely exceeds the number of observations. This analysis highlights the varying nature and quality of (local) optimal solutions to the DODE problem. In particular, it shows that the proposed hybrid framework is able to produce high-quality estimates for both OD flows and links counts, while other methods struggle in one way or another. This is another important piece of empirical evidence showing the existence of local optimal solutions to the OD estimation problem and the importance of reducing the solution space to address the scalability issues.

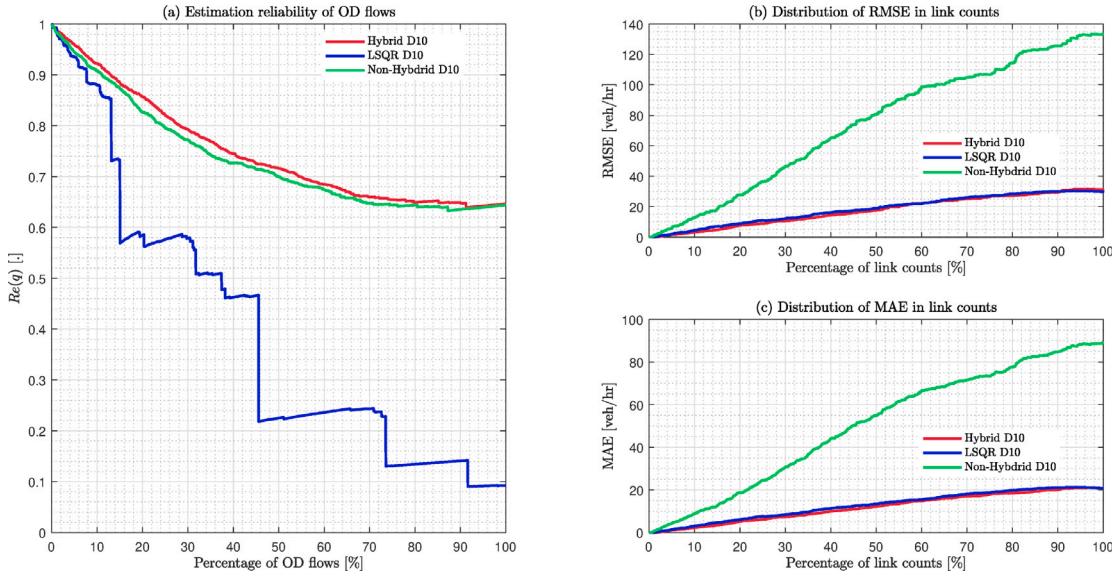


Fig. 11. DODE results given by hybrid, non-hybrid and LSQR solver methods.

Table 4
 CV_{RMSE} for DODE scenarios.

Scenario	Prior OD	Hybrid		Non-Hybrid		LSQR	
		CV_{RMSE}	%	CV_{RMSE}	%	CV_{RMSE}	%
D7	0.3322	0.2959	11%	0.3258	2%	0.3695	-11%
D10	0.3283	0.2767	16%	0.3281	0%	0.4102	-25%
D11	0.4917	0.3126	36%	0.4103	17%	0.3861	21%

5.4.2. Empirical evidence of consistency

The consistency in this study refers to the ability of the hybrid DODE method to operate similarly in both uncongested and congested traffic conditions and estimate OD flows accurately. The inclusion of region-level traffic model allows the hybrid DODE algorithm to capture how the traffic conditions change in the region-level context. Note that the region level model is an MFD based traffic model which is capable of estimating the regional accumulations when demand profile, route choice parameters and network structure (MFD estimates and region configuration) are given as inputs. In this study, we used MFD-based modelling to determine the demand profile that is accountable for the observed traffic conditions as indicated by regional accumulations (via detectors). It should be noted that the relationship between demand and traffic counts is defined inherently in the nonlinear MFD-based traffic model (region-level), which allows it to capture both congested and uncongested traffic conditions, making it more beneficial for OD estimation in congested traffic conditions. The centroid-level, on the other hand, lacks this feature because the demand to link count relation is defined with reference to exogenous travel times (from the last iteration). As a result, it is a linear approximation without a defined analytical function for state changes.

We can investigate the consistency of the DODE models by comparing the estimation results across uncongested and congested scenarios, namely D7 and D11. Here we refer to the CV_{RMSE} given in Eq. (9). The CV_{RMSE} evaluates the relative closeness of the predicted OD flows to the ground truth OD flows. The DODE method with the smaller CV_{RMSE} is better when compared (Cascetta et al., 2013). Table 4 presents the CV_{RMSE} value for the three demand scenarios D7, D10 and D11. We would like to compare the results obtained for the hybrid DODE with the non-hybrid DODE and LSQR DODE methods. It could be seen that CV_{RMSE} improved by 11% in D7, 16% in D10 and 36% for D11 in estimates by hybrid DODE. However, the non-hybrid and LSQR report either lower improvements or further worsening in CV_{RMSE} value which simply indicate the unsuitability of the non-hybrid and LSQR methods for OD estimation. However, we understand that having few outliers in the estimated ODs (high estimation error for few OD pairs) given by non-hybrid and LSQR methods may reduce the CV_{RMSE} as the measure is highly sensitive to outliers. Therefore, we look at the distribution of CV_{RMSE} against the total OD flow.

Fig. 12 illustrates the cumulative distribution of the CV_{RMSE} as a function of the percentage of overall demand ($q(t)$) for D7 and D11 demand scenarios. We present the CV_{RMSE} distributions given by hybrid DODE method (Red), LSQR method (Blue) and non-hybrid DODE(Green) for comparison purposes. Solid lines show the CV_{RMSE} distribution for demand scenario D7 where the DODE starts with uncongested traffic conditions as the *a priori* demand profile is 15% underestimate of the ground truth. We can see that the 80% of total OD flows hybrid DODE method demonstrate CV_{RMSE} value less than 0.20 (20%) which is acceptable for practical purposes according to Cascetta et al. (2013). However, LSQR and non-hybrid DODE methods report about 70% of total OD

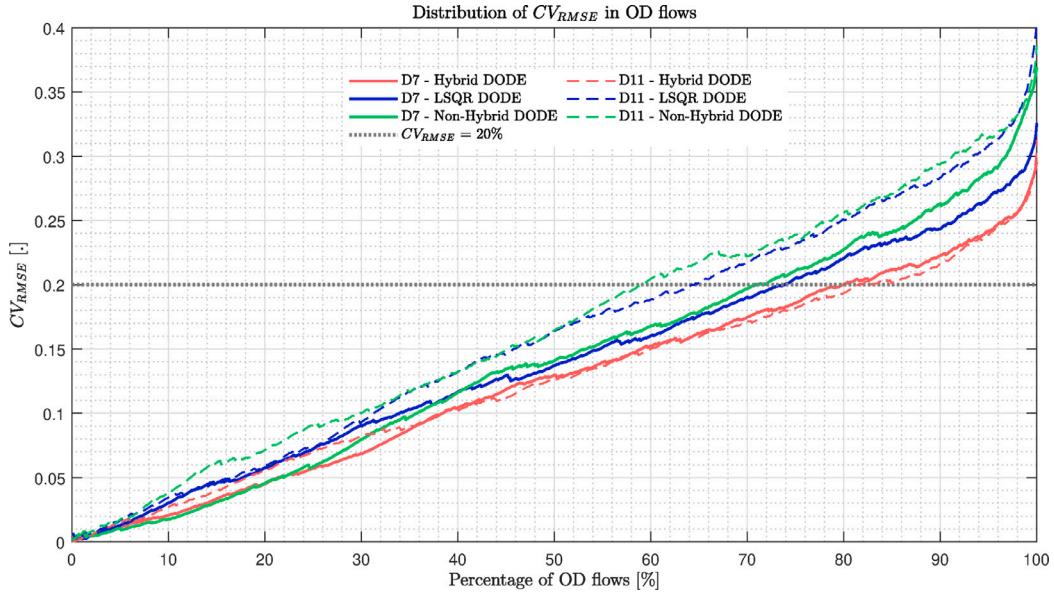


Fig. 12. Distribution of the CV_{RMSE} as a function of the considered percentage of overall demand volume.

flows with CV_{RMSE} values less than 0.20. Hence, it is clear that hybrid DODE has a higher relative closeness of predicted OD flows to the ground truth OD flows compared to LSQR and non-hybrid DODE. Similarly, dashed lines show the CV_{RMSE} distribution for demand scenario D11 where the DODE starts with congested traffic conditions as the *a priori* demand profile is 25% overestimate of the ground truth. We can see that the percentage of OD flows with CV_{RMSE} less than 0.20 drops to 60% in non-hybrid DODE and to 65% in LSQR DODE, whereas it stays at 80% in the hybrid DODE method. The change in traffic conditions cause both LSQR DODE and non-hybrid DODE methods to perform rather poorly in terms of the relative closeness of predicted OD flows, please see the shift between blue/green solid and dashed lines in Fig. 12. However, the hybrid DODE performs similarly in both uncongested and congested traffic conditions considering the relative closeness of OD flows (please see the similarity between red solid and dashed curves in Fig. 12), which indicates the consistency of DODE method for different traffic conditions. It is evident that the hybrid DODE method, which captures nonlinear nature of traffic through the region-level constraints, produces *consistent* results across different traffic conditions compared to LSQR DODE and non-hybrid DODE.

5.5. Evaluation of other measures of congestion

The proposed hybrid DODE formulation builds on the optimization problem presented in Eq. (5a)–(5f). This problem considers two key measures to be matched; link counts and region accumulations. Link counts alone can be misleading as indicator of congestion, because the same link counts could correspond to different traffic regimes (i.e. free-flow and congested). On the other hand, region accumulation is a proxy for average density in a region and it can fully capture the average congestion-level in a region.

The use of region accumulations in the proposed framework ensures similarity in the congestion level at the regional scale; however, we acknowledge that this does not guarantee a one-to-one match at the link-level. There are various studies that consider additional traffic measures (e.g., link densities, travel times, vehicle trajectories) in the objective function formulation in order to achieve a better match with respect to the link-level congestion (Antoniou et al., 2016). However, these models mostly bypass the complex relationship between the OD flows and traffic flow measurements by treating the assignment model as a black box (Balakrishna et al., 2008; Cipriani et al., 2011). On the other hand, we formulate this relationship building on an analytical formulation that consists of centroid-level linear assignment and region-level traffic model, see the upper-level in Fig. 1. Link densities or travel times cannot be added to the proposed optimization formulation, because they are not inherent outputs of the two analytical models of the upper-level. This can be considered a limitation of the proposed model. However, the core contribution of this paper is to incorporate the region-level traffic modelling into centroid-level OD estimation problem. The additional constraints brought by the region-level modelling ease the indeterminateness in the centroid level by providing a guidance on adjustment of centroid-level OD flows. The same approach could potentially be used with other OD estimation models that inherently produce additional traffic measurements at the link-level. This can be a future research direction.

While these additional traffic measures cannot be added to the optimization problem in Eq. (5a)–(5f), they can still be produced as part of the bi-level framework, please see Fig. 1. The lower-level here is a DTA model with a mesoscopic simulator, which is implemented over the iterations with the estimated OD flows, and the upper-level is the hybrid DODE formulation. Fig. 3 provides a more detailed overview of the proposed bi-level mechanism. The upper-level (i.e., hybrid DODE) requires various measurements to

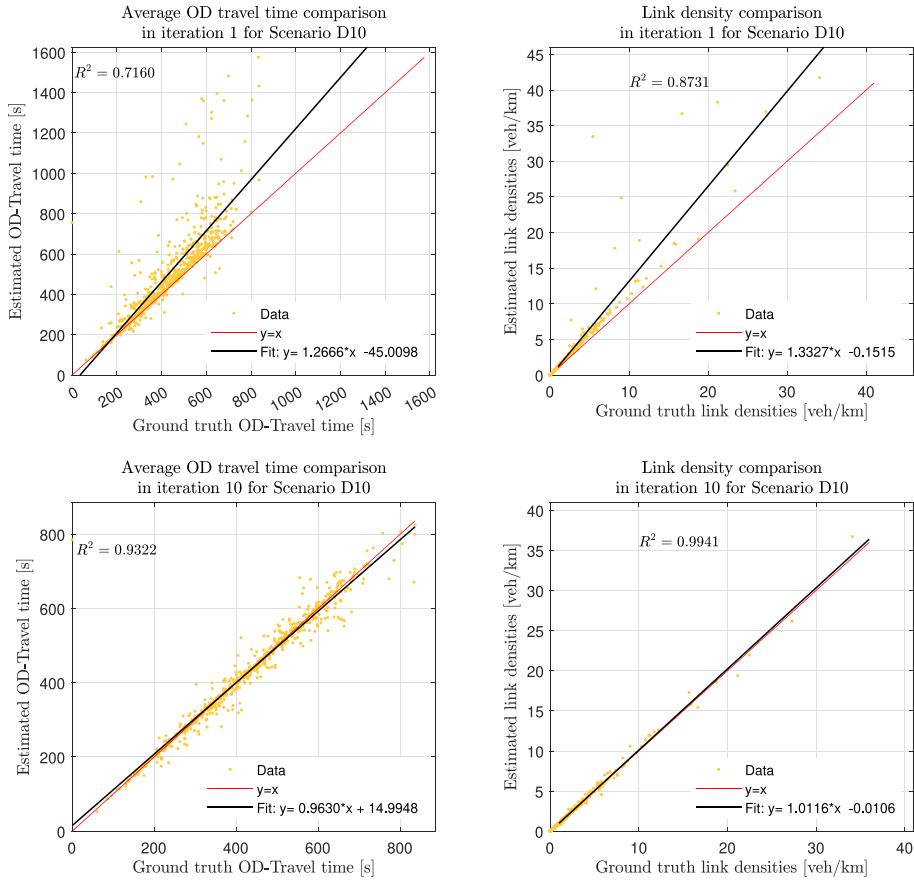


Fig. 13. Estimation results of OD travel times and link densities in the scenario D10.

be collected from the lower-level (i.e., test-bed) at each iteration: assignment matrix, regional accumulations, regional route choice, average trip lengths as depicted in Fig. 3. Additionally, the lower-level simulator can produce other measures of traffic including travel times, link densities, etc. Although these measures of congestion are not directly being used in the upper-level due to reasons described above, they still influence the DODE mechanism over the iterations via the assignment matrix, regional route choices, etc. For instance, the time-dependent assignment matrix captures the travel times across the network and the changing route choices in response to congestion.

Fig. 13 shows the estimation results of OD travel times and link densities produced by the mesoscopic simulator in iteration 1 and iteration 10. Iteration 1 represents the initial conditions that are estimated with a priori OD matrix, and iteration 10 represents the final results from the DODE procedure. Note that this figure includes all links with detectors, all OD pairs across all time intervals in the simulation model. The travel times represent the average travel times across OD pairs irrespective of the chosen paths. Fig. 13 clearly depicts a mismatch between simulated output and ground-truth measurements in iteration 1, which results in regression slope to be significantly different from 1. Additionally, the scatter is evident for both travel times and link densities, resulting in low R^2 values. The proposed DODE mechanism is able to significantly improve the results over the iterations, which leads to lesser scatter and higher accuracy as depicted in Fig. 13. The results in iteration 10 clearly show an agreement between observed and simulated values in terms of both travel times and link densities. We do not observe any measurements that is significantly apart from the proportionality line, which would represent a mismatch in the traffic regime (uncongested vs congested) despite having similar link traffic counts as depicted in Fig. 10. Note that Fig. 13 presents the results for the D10 scenario, but we observe a fairly similar performance across all scenarios. These results indicate that although additional congestion measures cannot be directly added to the proposed hybrid DODE formulation, the use of regional accumulation values as well as the iterative bi-level framework helps the model produce traffic conditions that are consistent with the ground-truth observations.

6. Conclusion

This study proposes a novel hybrid DODE method that harvests the synchronous modelling of region-level and centroid-level traffic dynamics. The proposed DODE method is formulated as an offline (simultaneous) OD estimation problem that aims to

minimize the mismatch between estimated and ground-truth link counts. Regional traffic dynamics are represented by a multi-region MFD model, while centroid-level traffic dynamics are represented by a linear mapping between demand and link counts via a dynamic assignment matrix. The hybrid DODE is formulated as a non-convex, non-linear optimization problem, and implemented in a large-scale traffic network, which is benchmarked for OD estimation in the literature (Antoniou et al., 2016).

Results demonstrate the potential of the hybrid DODE method in addressing several limitations observed in the existing OD estimation methods. The combined modelling of centroid-level and region-level in the hybrid DODE model allows us to define specific constraints that bound the solution space of OD estimation problem and derive optimal solutions adhering to region-level traffic conditions. By doing so, the scalability issues suffered by the existing OD estimation methods when they are applied to mid to large-scale traffic networks are avoided. Moreover, the inclusion of MFD model in the hybrid DODE model facilitates the recognition of congestion dynamics within the OD estimation problem, which enables us to determine the non-linear relationship between the regional OD flows and regional traffic dynamics. This feature makes the OD estimates given by hybrid DODE model being consistent in both uncongested and congested traffic conditions, which cannot be achieved by the existing OD estimation methods.

Compared to the conventional methods, we bring in additional information to the optimization problem via region-level traffic observations and MFD-based regional traffic model. There are two potential difficulties associated with the proposed region-level approach. (i) Partitioning of a large-scale network into regions is a crucial requirement to ensure homogeneity of traffic conditions, tractability of outflows, preservation of reservoir shape and invariant distribution of link level ODs (see Section 4.1 for more details). This paper does not propose a methodology for network partitioning (this is beyond the scope of this work), but Section 4.1 lists the criteria that one should consider when partitioning the network for OD estimation purposes. (ii) Limitations on the number and distribution of detectors can make it difficult to collect region-level traffic observations and estimate MFD parameters. An adequate coverage of detectors in the network is required to ensure reliable measurements on observed region accumulations. Nonetheless, this is a common requirement for DODE problems, where link count observations are also crucial. In addition, hybrid DODE formulation may require extra precaution, such as specifying constraints boundaries ($\beta_{ub}, \beta_{lb}, \alpha_{ub}, \alpha_{lb}$), to ensure the proposed method is implemented effectively. The offline DODE formulation could be extended to online (sequential) DODE by introducing a state-space model with dynamic controls, which could be an alluring future extension.

CRediT authorship contribution statement

Sakitha Kumarage: Conceptualization, Methodology, Formal analysis, Investigation, Writing. **Mehmet Yildirimoglu:** Conceptualization, Methodology, Formal analysis, Investigation, Writing. **Zuduo Zheng:** Methodology, Formal analysis, Investigation, Writing.

Acknowledgements

This research is funded by iMOVE CRC and supported by the Cooperative Research Centres program, an Australian Government initiative. This research is also partly funded by the Australian Research Council (ARC) through Dr. Mehmet Yildirimoglu's Discovery Early Career Researcher Award (DECRA; DE220101320). Methods and software tools for the proposed Hybrid DODE framework is accessible via the following GitHub repository : https://github.com/sakithakumarage/hybrid_DODE.

References

- Aboudolas, K., Geroliminis, N., 2013. Perimeter and boundary flow control in multi-reservoir heterogeneous networks. *Transp. Res. B* 55, 265–281. <http://dx.doi.org/10.1016/j.trb.2013.07.003>.
- Ambühl, L., Loder, A., Leclercq, L., Menendez, M., 2021. Disentangling the city traffic rhythms: A longitudinal analysis of MFD patterns over a year. *Transp. Res. C* 126, 103065. <http://dx.doi.org/10.1016/j.trc.2021.103065>.
- Ambühl, L., Menendez, M., 2016. Data fusion algorithm for macroscopic fundamental diagram estimation. *Transp. Res. C* 71, 184–197.
- Andersson, J., 2013. A General-Purpose Software Framework for Dynamic Optimization (Ph.D. thesis). Ku Leuven, p. 186, URL: <https://lirias.kuleuven.be/handle/123456789/418048>.
- Andrienko, G., Andrienko, N., Fuchs, G., Wood, J., 2017. Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data. *IEEE Trans. Vis. Comput. Graphics* 23 (9), 2120–2136. <http://dx.doi.org/10.1109/TVCG.2016.2616404>.
- Antoniou, C., Barceló, J., Breen, M., Bullejos, M., Casas, J., Cipriani, E., Ciuffo, B., Djukic, T., Hoogendoorn, S., Marzano, V., Montero, L., Nigro, M., Perarnau, J., Punzo, V., Toledo, T., van Lint, H., 2016. Towards a generic benchmarking platform for origin-destination flows estimation/updating algorithms: Design, demonstration and validation. *Transp. Res. C* <http://dx.doi.org/10.1016/j.trc.2015.08.009>.
- Arnott, R., Buli, J., 2018. Solving for equilibrium in the basic bathtub model. *Transp. Res. B* 109, 150–175. <http://dx.doi.org/10.1016/j.trb.2017.12.003>.
- Ashok, K., Ben-Akiva, M.E., 2000. Alternative approaches for real-time estimation and prediction of time-dependent origin-destination flows. *Transp. Sci.* 34 (1), 21–36. <http://dx.doi.org/10.1287/trsc.34.1.21.12282>.
- Balakrishna, R., Ben-Akiva, M., Koutsopoulos, H., 2008. Time-dependent origin-destination estimation without assignment matrices. In: Second International Symposium of Transport Simulation. ISTS06. Lausanne, Switzerland. 4–6 September 2006, EPFL Press.
- Behara, K.N., Bhaskar, A., Chung, E., 2020. A novel approach for the structural comparison of origin-destination matrices: Levenshtein distance. *Transp. Res. C* 111, 513–530. <http://dx.doi.org/10.1016/j.trc.2020.01.005>.
- Bell, M., 1983. The estimation of an origin-destination matrix from traffic counts. *Transp. Sci.* 17, 198–217. <http://dx.doi.org/10.1287/trsc.17.2.198>.
- Bierlaire, M., Crittin, F., 2004. An efficient algorithm for real-time estimation and prediction of dynamic OD tables. *Oper. Res.* 52 (1), 116–127. <http://dx.doi.org/10.1287/opre.1030.0071>.
- Buisson, C., Daamen, W., Punzo, V., Wagner, P., Montanino, M., Ciuffo, B., 2014. Calibration and validation principles. In: Traffic Simulation and Data: Validation Methods and Applications. CRC Press, pp. 89–118.
- Cantelmo, G., Qurashi, M., Prakash, A.A., Antoniou, C., Viti, F., 2020. Incorporating trip chaining within online demand estimation. *Transp. Res. B* 132, 171–187. <http://dx.doi.org/10.1016/j.trb.2019.05.010>.

- Cascetta, E., 2001. Transportation Systems Engineering: Theory and Methods. Springer US, <http://dx.doi.org/10.1007/978-1-4757-6873-2>.
- Cascetta, E., Inaudi, D., Marquis, G., 1993. Dynamic estimators of origin-destination matrices using traffic counts. *Transp. Sci.* 27 (4), 363–373. <http://dx.doi.org/10.1287/trsc.27.4.363>.
- Cascetta, E., Papola, A., Marzano, V., Simonelli, F., Vitiello, I., 2013. Quasi-dynamic estimation of O-D flows from traffic counts: Formulation, statistical validation and performance analysis on real data. *Transp. Res. B* 55, 171–187. <http://dx.doi.org/10.1016/j.trb.2013.06.007>.
- Cipriani, E., Florian, M., Mahut, M., Nigro, M., 2011. A gradient approximation approach for adjusting temporal origin-destination matrices. *Transp. Res. C* 19 (2), 270–282. <http://dx.doi.org/10.1016/j.trc.2010.05.013>.
- Ciuffo, B., 2012. *The Calibration of Traffic Simulation Models : Report on the Assessment of Different Goodness of Fit Measures and Optimization Algorithms*. Publications Office, Luxembourg.
- Cremer, M., Keller, H., 1987. A new class of dynamic methods for the identification of origin-destination flows. *Transp. Res. B* 21 (2), 117–132. [http://dx.doi.org/10.1016/0191-2615\(87\)90011-7](http://dx.doi.org/10.1016/0191-2615(87)90011-7).
- Dantsuji, T., Hoang, N.H., Zheng, N., Vu, H.L., 2022. A novel metamodel-based framework for large-scale dynamic origin-destination demand calibration. *Transp. Res. C* 136, 103545. <http://dx.doi.org/10.1016/j.trc.2021.103545>.
- Darda, D., 2002. *Joint Calibration of a Microscopic Traffic Simulator and Estimation of Origin-Destination Flows* (Ph.D. thesis).
- Djukic, T., 2014. *Dynamic OD Demand Estimation and Prediction for Dynamic Traffic Management*. TRAIL research school.
- Djukic, T., Lint, J.W.C.V., Hoogendoorn, S.P., 2012. Application of principal component analysis to predict dynamic origin-destination matrices. *Transp. Res. Rec.: J. Transp. Res. Board* 2283 (1), 81–89. <http://dx.doi.org/10.3141/2283-09>.
- Fisk, C.S., Boyce, D.E., 1983. A note on trip matrix estimation from link traffic count data. *Transp. Res. B* 17, 245–250. [http://dx.doi.org/10.1016/0191-2615\(83\)90018-8](http://dx.doi.org/10.1016/0191-2615(83)90018-8).
- Frederix, R., Viti, F., Corthout, R., Tampère, C.M.J., 2011. New gradient approximation method for dynamic origin-destination matrix estimation on congested networks. *Transp. Res. Rec.: J. Transp. Res. Board* 2263 (1), 19–25. <http://dx.doi.org/10.3141/2263-03>.
- Frederix, R., Viti, F., Himpe, W.W.E., Tampère, C.M.J., 2013. Dynamic origin-destination matrix estimation on large-scale congested networks using a hierarchical decomposition scheme. *J. Intell. Transp. Syst.* 18 (1), 51–66. <http://dx.doi.org/10.1080/15472450.2013.773249>.
- Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transp. Res. B* 42, 759–770. <http://dx.doi.org/10.1016/j.trb.2008.02.002>.
- Gómez, P., Menéndez, M., Mérida-Casermeiro, E., 2015. Evaluation of trade-offs between two data sources for the accurate estimation of origin-destination matrices. *Transportmetrica B* 3 (3), 222–245. <http://dx.doi.org/10.1080/21680566.2015.1025892>.
- Gu, Z., Saberi, M., 2021. Simulation-based optimization of toll pricing in large-scale urban networks using the network fundamental diagram: A cross-comparison of methods. *Transp. Res. C* 122, 102894. <http://dx.doi.org/10.1016/j.trc.2020.102894>.
- Gurobi Optimization, L., 2022. Gurobi Optimizer Reference Manual. URL: <https://www.gurobi.com>.
- Hollander, Y., Liu, R., 2008. The principles of calibrating traffic microsimulation models. *Transportation* 35 (3), 347–362. <http://dx.doi.org/10.1007/s11116-007-9156-2>.
- Hu, X., Chiu, Y.-C., Villalobos, J.A., Nava, E., 2017. A sequential decomposition framework and method for calibrating dynamic origin-destination demand in a congested network. *IEEE Trans. Intell. Transp. Syst.* 18 (10), 2790–2797.
- Huang, S., Sadek, A.W., Guo, L., 2013. Computational-based approach to estimating travel demand in large-scale microscopic traffic simulation models. *J. Comput. Civ. Eng.* 27 (1), 78–86. [http://dx.doi.org/10.1061/\(asce\)cp.1943-5487.0000202](http://dx.doi.org/10.1061/(asce)cp.1943-5487.0000202).
- Huang, C., Zheng, N., Zhang, J., 2019. Investigation of bimodal macroscopic fundamental diagrams in large-scale urban networks: Empirical study with GPS data for Shenzhen city. *Transp. Res. Rec.: J. Transp. Res. Board* 2673 (6), 114–128. <http://dx.doi.org/10.1177/0361198119843472>.
- Ji, Y., Luo, J., Geroliminis, N., 2014. Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems. *Transp. Res. Rec.: J. Transp. Res. Board* 2422 (1), 1–11. <http://dx.doi.org/10.3141/2422-01>.
- Kattan, L., Abdulhai, B., 2006. Noniterative approach to dynamic traffic origin-destination estimation with parallel evolutionary algorithms. *Transp. Res. Rec.: J. Transp. Res. Board* 1964 (1), 201–210. <http://dx.doi.org/10.1177/036119810619640012>.
- Knoop, V.L., van Erp, P.B., Leclercq, L., Hoogendoorn, S.P., 2018. Empirical MFDs using Google traffic data. In: 2018 21st International Conference on Intelligent Transportation Systems. ITSC, IEEE, <http://dx.doi.org/10.1109/itsc.2018.8570005>.
- Krishnakumari, P., van Lint, H., Djukic, T., Cats, O., 2020. A data driven method for OD matrix estimation. *Transp. Res. C* 113, 38–56. <http://dx.doi.org/10.1016/j.trc.2019.05.014>.
- Kumarage, S., Yildirimoglu, M., Ramezani, M., Zheng, Z., 2021. Schedule-constrained demand management in two-region urban networks. *Transp. Sci.* <http://dx.doi.org/10.1287/trsc.2021.1052>.
- Leclercq, L., Chiabaut, N., Trinquier, B., 2014. Macroscopic fundamental diagrams: A cross-comparison of estimation methods. *Transp. Res. B* 62, 1–12. <http://dx.doi.org/10.1016/j.trb.2014.01.007>.
- Leclercq, L., Geroliminis, N., 2013. Estimating MFDs in simple networks with route choice. *Procedia - Soc. Behav. Sci.* 80, 99–118. <http://dx.doi.org/10.1016/j.sbspro.2013.05.008>.
- Lou, Y., Yin, Y., 2010. A decomposition scheme for estimating dynamic origin-destination flows on actuation-controlled signalized arterials. *Transp. Res. C* 18 (5), 643–655. <http://dx.doi.org/10.1016/j.trc.2009.06.005>.
- Lu, C.-C., Zhou, X., Zhang, K., 2013. Dynamic origin-destination demand flow estimation under congested traffic conditions. *Transp. Res. C* 34, 16–37. <http://dx.doi.org/10.1016/j.trc.2013.05.006>.
- Ma, W., Pi, X., Qian, S., 2020. Estimating multi-class dynamic origin-destination demand through a forward-backward algorithm on computational graphs. *Transp. Res. C* 119, 102747. <http://dx.doi.org/10.1016/j.trc.2020.102747>.
- Mariotte, G., Leclercq, L., Batista, S., Krug, J., Paipuri, M., 2020. Calibration and validation of multi-reservoir MFD models: A case study in Lyon. *Transp. Res. B* 136, 62–86. <http://dx.doi.org/10.1016/j.trb.2020.03.006>.
- Marzano, V., Papola, A., Simonelli, F., 2009. Limits and perspectives of effective O-D matrix correction using traffic counts. *Transp. Res. C* 17 (2), 120–132. <http://dx.doi.org/10.1016/j.trc.2008.09.001>.
- MATLAB, Version 9.9.0.1467703 (R2020b). The MathWorks Inc., Natick, Massachusetts.**
- Nigro, M., Abdelfatah, A., Cipriani, E., Colombaroni, C., Fusco, G., Gemma, A., 2018a. Dynamic O-D demand estimation: Application of SPSA AD-PI method in conjunction with different assignment strategies. *J. Adv. Transp.* 2018, 1–18. <http://dx.doi.org/10.1155/2018/2085625>.
- Nigro, M., Cipriani, E., del Giudice, A., 2018b. Exploiting floating car data for time-dependent origin-destination matrices estimation. *J. Intell. Transp. Syst.* 22 (2), 159–174. <http://dx.doi.org/10.1080/15472450.2017.1421462>.
- Ortigosa, J., Menendez, M., Tapia, H., 2015. Study on the number and location of measurement points for an MFD perimeter control scheme: A case study of zurich. *EURO J. Transp. Logist.* 3 (3), 245–266. <http://dx.doi.org/10.1007/s13676-013-0034-0>.
- Ortúzar, J.D., Willumsen, L.G., 2011. Modelling transport. *Modell. Transp.* <http://dx.doi.org/10.1002/9781119993308>.
- Osorio, C., 2019. Dynamic origin-destination matrix calibration for large-scale network simulators. *Transp. Res. C* 98, 186–206. <http://dx.doi.org/10.1016/j.trc.2018.09.023>.
- Osorio, C., Punzo, V., 2019. Efficient calibration of microscopic car-following models for large-scale stochastic network simulators. *Transp. Res. B* 119, 156–173. <http://dx.doi.org/10.1016/j.trb.2018.09.005>.

- Ou, J., Lu, J., Xia, J., An, C., Lu, Z., 2019. Learn, assign, and search: Real-time estimation of dynamic origin-destination flows using machine learning algorithms. *IEEE Access* 7, 26967–26983. <http://dx.doi.org/10.1109/access.2019.2901289>.
- Paige, C.C., Saunders, M.A., 1982. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Softw.* 8 (1), 43–71.
- Pajipuri, M., Xu, Y., González, M.C., Leclercq, L., 2020. Estimating MFDs, trip lengths and path flow distributions in a multi-region setting using mobile phone data. *Transp. Res. C* 118, 102709. <http://dx.doi.org/10.1016/j.trc.2020.102709>.
- Patwary, A., Huang, W., Lo, H.K., 2021. Metamodel-based calibration of large-scale multimodal microscopic traffic simulation. *Transp. Res. C* 124, 102859. <http://dx.doi.org/10.1016/j.trc.2020.102859>.
- Prakash, A.A., Seshadri, R., Antoniou, C., Pereira, F.C., Ben-Akiva, M., 2018. Improving scalability of generic online calibration for real-time dynamic traffic assignment systems. *Transp. Res. Rec.: J. Transp. Res. Board* 2672 (48), 79–92. <http://dx.doi.org/10.1177/0361198118791360>.
- Qurashi, M., Ma, T., Chanotakis, E., Antoniou, C., 2020. PC-SPSA: Employing dimensionality reduction to limit SPSA search noise in DTA model calibration. *IEEE Trans. Intell. Transp. Syst.* 21 (4), 1635–1645. <http://dx.doi.org/10.1109/tits.2019.2915273>.
- Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. *Transp. Res. B* 74, 1–19. <http://dx.doi.org/10.1016/j.trb.2014.12.010>.
- Robillard, P., 1975. Estimating the O-D matrix from observed link volumes. *Transp. Res.* 9 (2–3), 123–128. [http://dx.doi.org/10.1016/0041-1647\(75\)90049-0](http://dx.doi.org/10.1016/0041-1647(75)90049-0).
- Ros-Roca, X., Montero, L., Barceló, J., 2020. Investigating the quality of spieß-like and SPSA approaches for dynamic OD matrix estimation. *Transportmetrica A* 17 (3), 235–257. <http://dx.doi.org/10.1080/23249935.2020.1722282>.
- Ros-Roca, X., Montero, L., Barceló, J., Nökel, K., Gentile, G., 2022. A practical approach to assignment-free dynamic origin–destination matrix estimation problem. *Transp. Res. C* 134, 103477. <http://dx.doi.org/10.1016/j.trc.2021.103477>.
- Saeedmanesh, M., Geroliminis, N., 2016. Clustering of heterogeneous networks with directional flows based on ‘snake’ similarities. *Transp. Res. B* 91, 250–269. <http://dx.doi.org/10.1016/j.trb.2016.05.008>.
- Saffari, E., Yildirimoglu, M., Hickman, M., 2020. A methodology for identifying critical links and estimating macroscopic fundamental diagram in large-scale urban networks. *Transp. Res. C* 119, 102743. <http://dx.doi.org/10.1016/j.trc.2020.102743>.
- Saffari, E., Yildirimoglu, M., Hickman, M., 2022. Data fusion for estimating macroscopic fundamental diagram in large-scale urban networks. *Transp. Res. C* 137, 103555. <http://dx.doi.org/10.1016/j.trc.2022.103555>.
- Shafiei, S., Saberi, M., Sarvi, M., 2016. Application of an exact gradient method to estimate dynamic origin–destination demand for melbourne network. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems. ITSC, IEEE, <http://dx.doi.org/10.1109/itsc.2016.7795870>.
- Sherali, H.D., Park, T., 2001. Estimation of dynamic origin–destination trip tables for a general network. *Transp. Res. B* 35 (3), 217–235. [http://dx.doi.org/10.1016/s0191-2615\(99\)00048-x](http://dx.doi.org/10.1016/s0191-2615(99)00048-x).
- Sirmatel, I.I., Geroliminis, N., 2018. Economic model predictive control of large-scale urban road networks via perimeter control and regional route guidance. *IEEE Trans. Intell. Transp. Syst.* 19 (4), 1112–1121. <http://dx.doi.org/10.1109/tits.2017.2716541>.
- Sirmatel, I.I., Geroliminis, N., 2021. Stabilization of city-scale road traffic networks via macroscopic fundamental diagram-based model predictive perimeter control. *Control Eng. Pract.* 109, 104750. <http://dx.doi.org/10.1016/j.conengprac.2021.104750>.
- Sirmatel, I.I., Tsitsikas, D., Kouvellas, A., Geroliminis, N., 2021. Modeling, estimation, and control in large-scale urban road networks with remaining travel distance dynamics. *Transp. Res. C* 128, 103157. <http://dx.doi.org/10.1016/j.trc.2021.103157>.
- Sun, J., Guo, J., Wu, X., Zhu, Q., Wu, D., Xian, K., Zhou, X., 2019. Analyzing the impact of traffic congestion mitigation: From an explainable neural network learning framework to marginal effect analyses. *Sensors* 19 (10), 2254.
- Toledo, T., Kolechkina, T., 2013. Estimation of dynamic origin–destination matrices using linear assignment matrix approximations. *IEEE Trans. Intell. Transp. Syst.* 14 (2), 618–626. <http://dx.doi.org/10.1109/tits.2012.2226211>.
- Toledo, T., Koutsopoulos, H.N., 2004. Statistical validation of traffic simulation models. *Transp. Res. Rec.: J. Transp. Res. Board* 1876 (1), 142–150. <http://dx.doi.org/10.3141/1876-15>.
- Tympanianaki, A., Koutsopoulos, H.N., Jenelius, E., 2018. Robust SPSA algorithms for dynamic OD matrix estimation. *Procedia Comput. Sci.* 130, 57–64. <http://dx.doi.org/10.1016/j.procs.2018.04.012>.
- Wächter, A., Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.* 106 (1), 25–57. <http://dx.doi.org/10.1007/s10107-004-0559-y>.
- Yang, H., Iida, Y., Sasaki, T., 1991. An analysis of the reliability of an origin–destination trip matrix estimated from traffic counts. *Transp. Res. B* 25 (5), 351–363. [http://dx.doi.org/10.1016/0191-2615\(91\)90028-H](http://dx.doi.org/10.1016/0191-2615(91)90028-H), URL: <https://www.sciencedirect.com/science/article/pii/019126159100028H>.
- Yildirimoglu, M., Geroliminis, N., 2016. Observing and reconstructing aggregated dynamic route choice patterns for large-scale mixed urban/freeway networks. In: *Transportation Research Board 95th Annual Meeting*. 16-5186.
- Yildirimoglu, M., Ramezani, M., 2020. Demand management with limited cooperation among travellers: A doubly dynamic approach. *Transp. Res. B* 132, 267–284. <http://dx.doi.org/10.1016/j.trb.2019.02.012>.
- Yildirimoglu, M., Ramezani, M., Amirgholy, M., 2021. Staggered work schedules for congestion mitigation: A morning commute problem. *Transp. Res. C* 132, 103391. <http://dx.doi.org/10.1016/j.trc.2021.103391>.
- Yildirimoglu, M., Sirmatel, I.I., Geroliminis, N., 2018. Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance. *Transp. Res. B* 118, 106–123. <http://dx.doi.org/10.1016/j.trb.2018.10.007>.
- Zhang, H., Seshadri, R., Prakash, A.A., Antoniou, C., Pereira, F.C., Ben-Akiva, M., 2018. Towards dynamic Bayesian networks: State augmentation for online calibration of DTA systems. In: 2018 21st International Conference on Intelligent Transportation Systems. ITSC, IEEE, <http://dx.doi.org/10.1109/itsc.2018.8569926>.
- Zheng, N., Geroliminis, N., 2020. Area-based equitable pricing strategies for multimodal urban networks with heterogeneous users. *Transp. Res. A* 136, 357–374. <http://dx.doi.org/10.1016/j.tra.2020.04.009>.
- Zockaie, A., Saberi, M., Saedi, R., 2018. A resource allocation problem to estimate network fundamental diagram in heterogeneous networks: Optimal locating of fixed measurement points and sampling of probe trajectories. *Transp. Res. C* 86, 245–262. <http://dx.doi.org/10.1016/j.trc.2017.11.017>.
- Zuylen, H.J.V., Willumsen, L.G., 1980. The most likely trip matrix estimated from traffic counts. *Transp. Res. B* 14, 281–293. [http://dx.doi.org/10.1016/0191-2615\(80\)90008-9](http://dx.doi.org/10.1016/0191-2615(80)90008-9).