Advanced traffic data for dynamic OD demand estimation: The state of the art and benchmark study

Tamara Djukic
Delft University of Technology, Department of Transport and Planning, The Netherlands
+31 15 278 1723 email:t.djukic@tudelft.nl
(Corresponding author)

Jaume Barcelò
Universitat Politcnica de Catalunya BarcelonaTECH, inLab FIB, Spain
jaume.barcelo@upc.edu

Manuel Bullejos
Universitat Politcnica de Catalunya BarcelonaTECH, inLab FIB, Spain
manuel.bullejos@upc.edu

Lidia Montero
Universitat Politcnica de Catalunya BarcelonaTECH, Departament dEstadstica i Investigaci Operativa, Spain
lidia.montero@tudelft.nl

Ernesto Cipriani
Universit Degli Studi Roma Tre, Department of Civil Engineering, Italy
ernesto.cipriani@uniroma3.it

Hans van Lint
Delft University of Technology, Department of Transport and Planning, The Netherlands
j.w.c.vanLint@tudelft.nl

Serge P. Hoogendoorn
Delft University of Technology, Department of Transport and Planning, The Netherlands
s.p.hoogendoorn@tudelft.nl

November 14, 2014

Word count:
Number of words in abstract 153
Number of words in text (including abstract, first page and excl.ref) 5526
Number of figures and tables 6 * 250 = 1500
Total 7179

Submitted to the 94rd Annual Meeting of the Transportation Research Board, 11-15 January 2015, Washington D.C.
ABSTRACT

In this paper, the use of advanced traffic data is discussed to contribute to the ongoing debate about their applications in dynamic OD estimation. This is done by discussing the advantages and disadvantages of traffic data with support of the findings of a benchmark study. The benchmark framework is designed to assess the performance of the dynamic OD estimation methods using different traffic data. Results show that despite the use of traffic condition data to identify traffic regime, the use of unreliable prior OD demand has a strong influence on estimation ability. The greatest estimation occurs when the prior OD demand information is aligned with the real traffic state or omitted and using information from AVI measurements to establish accurate and meaningful values of OD demand. A common feature observed by methods in this paper indicates that advanced traffic data require more research attention and new techniques to turn them into usable information.
INTRODUCTION

With advanced real-time simulation tools that are able to integrate data from many different sensors both traffic operators and travelers can be provided with up-to-date and even projected travel and traffic information. Thorough calibration and validation procedures with sufficient data and granularity are critical in establishing the credibility of simulation tools for these different (real-time and planning) purposes. However, unreliability and lack of knowledge about the projected OD demand makes prediction with advanced simulation models simply impossible, regardless of how well these models have been calibrated. In this respect, the estimation of dynamic OD demand has received a lot of attention in the last decades.

Dynamic OD demand estimation methods have been proposed since the 1980s and since then a great evolution of these models have taken place. It started with the use of link traffic counts at intersection (1) that were translated into a practical generic optimization problem by (2). This OD demand problem formulation has received a lot of attention in the literature and has been continuously improved and extended. However, the methods and tools used in this area are still largely based on data that have been around from the early 1980s and onwards, such as traffic counts and questionnaire data. As a result these methods are, to use a euphemism, “assumption rich and data poor”.

In the last decade the amount of empirical traffic data becoming available for both on-line and off-line use has increased, particularly in terms of the wide range of sensor technologies developed and applied to collect traffic data. New emerging data collection methods, such as large scale revealed travel itineraries and route patterns from GPS, Bluetooth and WiFi scanners, and cameras to name just a few, offer tremendous opportunities to extract more detailed and more valuable information about (origin-destination) demand patterns and travel behavior, than ever before. The literature intensively explores application of advanced traffic data for dynamic OD demand estimation, but a lot is yet to be done to extract sensible and valid information from these new data sources.

In this paper, we present detailed literature review of different traffic data sources used in dynamic OD demand estimation. In addition, each of these traffic data has been independently applied to estimate dynamic OD demand, but to the authors’ knowledge, a deeper comparative analysis of their advances has not yet been performed. Although many methods have been proposed to solve the dynamic OD estimation problem using more traffic data, much fewer efforts have been reported to actually evaluate and cross-compare these methods under different circumstances (e.g. different network structures, different sets of data available in different qualities) (3), (4). In most cases where new OD estimation method is proposed, also a sensitivity analysis to demonstrate different properties of the method (5), (6), (7) is performed. The purpose of this paper is to contribute to a better understanding of traffic data advances by presenting a comparative analysis of the advantages and disadvantages of four least square dynamic OD demand estimation methods using different traffic data.

The next section introduces the traffic data used in OD demand estimation process. The benchmark study is based on urban large-scale network and focuses on the suitability of each method to estimate dynamic OD demand under different input data. Further, discussion about their advantages and disadvantages from points of view of theory and application is presented. Finally, conclusions and recommendations for future research are provided.

TRAFFIC DATA FOR ESTIMATION OF DYNAMIC OD DEMAND

Since OD matrices are often not directly observable, they have to be estimated from any available relevant data. Sensors typically measure traffic characteristics, which are the result of not just OD demand, but also of route choice and of traffic operations. Although data from traffic sensors come in many forms and qualities, they can essentially be subdivided into three categories depending on the source of information on OD flows and traffic operations. The types of input data that are used in literature for dynamic OD estimation and prediction are subdivided as depicted in Figure 1(a): (1) OD flow data; (2) link flow data; and (3) traffic condition data. The first type of input data, OD flow data, represent direct observations of OD flows obtained
from surveys or probe vehicles. The second type of input data, link flow data, are determined by the travel
behaviour process. This process describes travel choices: when to depart, which mode to use, which route
to choose. The third category of input data, traffic condition data over network, is determined by traffic
operations. These data describe traffic state on a network: travel speeds, travel times, densities, etc. The
sources of input data and their application in dynamic OD demand estimation and prediction process are
discussed below.

(a) Classification of input data for dynamic OD demand estimation.

(b) Examples of some traffic data and their sources.

FIGURE 1 Types of input data used in dynamic OD demand estimation.

To illustrate on a very simple network example, suppose there are four OD flows going from A to
C, A to D, B to C and B to D (see Figure 1(b)). Figure 1(b) gives some examples of different traffic sensors
and the spatio-temporal semantics of the traffic variables, which can be observed with those sensors. We
will use this figure as a reference to explain the most important sources of OD flow data, and their features
for dynamic OD demand estimation.

OD flow data
Observations of OD flow data are rare. The practical and theoretical limitations of survey OD generation
techniques have led to an exploration of how such data could be derived from equipped vehicles with in-
vehicle traffic sensors which act as probes by transmitting their origin and intended trip destination when
they initiate a trip.

Automatic vehicle location (AVL) data are receiving attention for their potential to provide a large sample of OD flow data. The observation of OD flows from in-vehicle traffic sensors (e.g., GPS and GSM) allows the detection of vehicles in multiple locations as they traverse the network. This feature makes the re-identification and tracking of these probe vehicles possible, which in turn may (under certain conditions) provide information on particular OD pairs (e.g., OD pair AC in Figure 2). In an ideal case, if data on OD flows are collected from all vehicles equipped with in-vehicle traffic sensors, full information on OD flows over all OD pairs can be extracted. Today, probe vehicles constitute only a fraction of the total number of vehicles in a network. Several models have been developed for the estimation of OD flows using AVL data ((8), (9), (10)). (9) introduced the notion of direct measurements for the incorporation of AVL data into the solution of the OD estimation and prediction problem.

Automatic vehicle identification (AVI) data represent another OD flow data source of growing importance for estimating dynamic OD demand flows. The observation of OD flows from AVI sensors (e.g., electronic-toll collection devices, infrared cameras, Bluetooth, WiFi, etc.) depends on: a) the location of these traffic sensors on a network, as depicted in Figure 1(b) and b) the sample of tagged vehicles. In an ideal case, if cameras are located on links connected to origin and/or destination nodes on a network, they can provide under some assumptions total demand that departs from origin B or arrives at destination C. If only a subset of vehicles is equipped with transponder tags or only a subset of vehicles is correctly identified by the AVI readers, then these OD flow data need to be explicitly considered in order to infer OD flows over all OD pairs. Several models have been developed for the estimation of OD flows using AVI data ((11), (12), (13), (14)). In brief, these models require estimating the sample rate (either market penetration rates or identification rates) so as to relate the AVI samples to the OD demand. The estimation of sample rates, however, is a difficult problem in its own right, as these rates are essentially time-dependent and location-dependent random variables. Moreover, the inclusion of sample rates in the OD demand estimation problem could dramatically increase the number of unknown variables and impact the reliability of OD demand estimates. To circumvent primary difficulties associated with estimating sample rates, (15) developed an OD demand estimation model using partially observed AVI data.

### Link flow data

Traffic link flow data collected from loop detectors at specific locations on a network are the most common type of input data used in dynamic OD demand estimation. The traffic link flow data could either be collected in the middle of a roadway segment, at entry or exit ramps on highways, or across a screen-line in an urban area. The number and position of loop detectors on an urban or highway network plays an important role, since traffic link flow data from these detectors can provide different information on OD flows. In an ideal case, if link flow data are collected on road segments belonging exclusively to routes used to serve one particular OD pair, they can provide information on OD volume for that particular OD pair. In addition, if loop detectors are located on links connected to origin or destination nodes on a network, they can provide under some assumptions total demand that departs from origin B or arrives at destination C, as represented in Figure 1(b). The traffic link flow data observed by loop detectors located on links between nodes 1 and 2 in Figure 1(b) are comprised of contributions from several OD flows (i.e., OD pairs: AC, AD, BC, BD). Thus, such link flow data require adequate specification of relation and mapping with OD flows. This procedure describes the most critical issue in OD matrix estimation, that is the relationship of the observed link flow data and traffic condition data with the unobserved OD flows.

### Traffic condition data

Apart from traffic link flow data, loop detectors are able to detect speeds and turn fractions at bifurcations in the network. The available speed or derived density measurements can help to identify whether traffic link flow data represents a congested or uncongested traffic state on a network. As such, they can facilitate
correct interpretation of traffic link flow data, and identification of OD flows that need to be adjusted, and in which direction. The simplest approach to including this type of input data is to include speed or density measurements in the goal function of the dynamic OD estimation problem ((16), (17), (18)). Turn fractions data collected at bifurcations in the network may provide constraints on the route choice patterns ((19)).

New technologies for probe vehicle re-identification and tracking (e.g. AVI systems and AVL systems) might provide traffic condition data, such as partial point-to-point travel times, route choice fractions, vehicle paths, and turning fractions. The data may come from cameras that capture and compare vehicle plates or from floating car data which may report the vehicle’s location at certain intervals to construct trajectories, as is depicted in Figure 1(b). The difficulty for the OD demand model formulation is to define the relationship between traffic flow data and OD flows. The identification of trajectories or link travel times can help to identify or estimate route flows. Therefore, they provide constraints on the traffic conditions resulting from assigning the OD flows to the network. Estimating OD matrices only from link flow data can be rather challenging given the indeterminate relation between link flow observations and route flows ((20)). Hence, many researchers have tried to integrate traffic condition data into the dynamic OD demand estimation and prediction problem. Examples include: speed and density data (e.g., (21), ((16), (18)); turning fractions (e.g.,((19), (22)); travel times (e.g., (17), (16)); and route flows (e.g., (23), (14)).

BENCHMARK STUDY

In this section we will discuss the overall benchmark study and provide some more detail on the components. First, design and implementation of benchmark platform will be briefly described. Then, generation of input scenarios, varying in terms of network topology, traffic conditions, and data availability is provided.

Overview of benchmark platform

The benchmark platform used in this benchmark study has been developed within European Union COST Action MULTITUDE project (24). The main goal of this platform was to ensure equal testing conditions for various OD demand estimation methods that would support fair comparison and an understanding of their relative merits. The benchmark platform consists of two main elements:

- **Traffic simulator**: In this benchmark study we use the mesoscopic version of the Aimsun simulation model (25) as the common traffic model. The mesoscopic model with default set of parameters was used because it is substantially faster than the microscopic one.

- **OD demand estimation algorithms**: This element refers to selection and implementation of a single or multiple OD demand estimation algorithms to be compared. Note that more information on selected dynamic OD estimation methods is given in following section.

For more detail description of the workings of benchmark platform, we refer to (3).

Case study

A key requirement for the task of evaluating an OD demand estimation algorithms, and for comparison of multiple ones, is to test the performance under a range of different conditions and scenarios and to ensure that these conditions are consistent across algorithms. For that purpose, in this benchmark study, we consider the following input scenarios:

1. We will test on large size network from Vitoria, Spain, with route choice.
2. We will consider different scenarios in terms of dynamic prior OD matrices, varying bias and random errors.
3. We will consider different scenarios in terms of data availability (i.e. the number and location of sensors and the type of surveillance information).
Network topology

Prior to methods evaluation, we define Vitoria network that consists of 57 centroids, 3249 OD pairs with a 600km road network, 2800 intersections and 389 detectors presented with black dots in Figure 2(a). This network was chosen because of the availability and quality of the empirical detector data on network, and because a calibrated OD matrix was available in the mesoscopic version of the Aimsun (25). This network resembles a reasonable sized real-life network, and is representative for congested road networks, as found in many large urban areas. The true link flow on detectors is derived from assignment of true OD matrix in Aimsun for one hour peak-afternoon period reflecting the congested state at the network. The simulation period is divided in 15 minutes time intervals with additional warm-up time interval, $T = 5$. The trips between some of the OD pairs are not completed within one time interval due to congestion on network or the distance between OD pairs resulting in 4 lagged time intervals and very sparse assignment matrices.

(a) Loop detector sensor layout.
(b) AVI sensor layout.

FIGURE 2 The Vitoria network, Basque Country, Spain

OD flow scenarios

To estimate the dynamic OD matrix for a specific day and time period $t$, information on OD flows given by prior OD matrix $\tilde{x}_{ij,t}$ turns out to be an important source of information. Generally, the dynamic prior OD matrix provides the base OD matrix which is matched and scaled on the basis of additional information (e.g. link flow data and traffic condition data) using different methods. The demand level is a key element affecting the performance of dynamic OD estimation methods (24). We can simulate different prior OD demand patterns which capture various demand levels by randomly perturbing each entry in the "true" OD demand matrix and for each departure time interval, $t$.

The experimental design considers the following three prior OD demand scenarios:

1. **Low demand scenario (D7)**: This scenario addresses situations where the prior OD demand might be a result of OD demand generated from out of date surveys. The low prior OD demand pattern is generated for 85% of the "true" OD demand level with random fluctuations over each OD pair and departure time interval in range of +/- 15%, that is

$$x_{ij,t}^{LD} = x_{ij,t} \times [0.7 + 0.3 \times \alpha_{ij,t}] \quad \alpha_{ij,t} \sim U(0, 1)$$

2. **Random demand scenario (RD)**: This scenario is based on the assumption that the prior OD matrix is the best estimate of the mean of the dynamic OD matrices. Any survey or off-line OD estimation procedure will utilize data from several days, inherently smoothing out any day to day variation present in
the flows. In this scenario, the prior OD demand pattern is generated for 95% of the “true” OD demand level and varied by adding uniformly random components in range of +/- 15%, representing the difference between the smoothed historical OD demand estimates and the particular daily realization:

\[ x_{ij,t}^{RD} = x_{ij,t} \times [0.8 + 0.3 \times \alpha_{ij,t}] \quad \alpha_{ij,t} \sim U(0, 1) \]  

(2)

3. **High demand scenario (D9):** This scenario addresses situations where the prior OD demand reflects travel demand in peak-hours, when congestion occurs on network. The prior OD demand pattern is generated for 105% of the “true” OD demand level and varied by adding uniformly random components in range of +/- 15%, that is

\[ x_{ij,t}^{HD} = x_{ij,t} \times [0.9 + 0.3 \times \alpha_{ij,t}] \quad \alpha_{ij,t} \sim U(0, 1) \]  

(3)

**Link flow and traffic condition data scenarios**

Sensors located on the Vitoria network can be divided into two main groups: loop detectors and AVL sensors. Loop detector sensors might produce local flows, densities, occupancies, etc. related to all vehicles at the detected loop. AVL sensors usually provide automatic signature identification for a subset of the vehicles; i.e., WiFi antennas to catch Bluetooth devices in discovery mode.

Traffic data are collected from 389 loop detectors and also 50 AVL detectors located using the layout models in (26). Almost 90% of the trips are collected twice at least in the peak-afternoon demand scenario, which account for 95% of the number of OD pairs and 86% of the most likely used paths identified in a DUE assignment with the “true” prior OD matrix. The procedure proposed by (26) returns simulated travel times on these predefined and stored routes. Figure 2(b) shows Vitoria’s network and subnetwork covered by AVI sensor layout.

**SELECTION OF OD DEMAND ESTIMATION METHODS**

In this section we make a choice of dynamic OD demand estimation methods used within today’s dynamic traffic management systems for the benchmark study. Since the main goal of the study is to evaluate the expected improvements due to implementation of richer and more varied traffic data, in this benchmark study we will focus on dynamic OD estimation methods that share same performance measure, i.e. least square error measure. In addition, one of the key requirements for successful benchmark study is to ensure good understanding and experience with various dynamic OD estimation methods that would support fair comparison.

First we provide definitions that will be used further in formulation of dynamic OD demand methods. The traffic demand between origin node \( o \) and destination node \( d \) is stored in the origin-destination (OD) matrix, \( x \). \( I \) is the set of all OD pairs and the vector \( x = \{ x_i | i \in I \} \) is the OD demand. The historical or prior OD matrix \( \tilde{x} = \{ \tilde{x}_i | i \in I \} \) is a matrix defined in OD flow scenarios that needs to be updated. \( y = \{ y_l | l \in L \} \) are the link flow data. Link flow data and traffic condition data (e.g., speed, density, occupancy) are available on links \( \hat{L} \subseteq L \). Thus, the observed link flows on those links are denoted as \( \hat{y} = \{ \hat{y}_l | l \in \hat{L} \} \) and observed traffic condition data are denoted as \( \hat{c} = \{ \hat{c}_l | l \in \hat{L} \} \). Additional traffic condition data, such as travel times, collected from AVI sensors available on links \( \tilde{L} \subseteq L \) are denoted as \( \tilde{z} = \{ \tilde{z}_l | l \in \tilde{L} \} \). The study period has \( T \) time steps, and is divided in time intervals \( t, t = 1, 2, ..., T \).

The generic formulation of dynamic OD estimation methods considered in this benchmark study,
combining observed link flow data and traffic condition data, can be expressed as

$$\hat{x} = \arg\min_{x \in S} [f_1(x, \bar{x}) + f_2(y(x), \bar{y}) + f_3(c(x), \bar{c}) + f_4(z(x), \bar{z})]$$

subject to

$$y = \text{assign}(x),$$

$$x \geq 0,$$

$$y \geq 0$$

where $x$ is the unknown OD demand vector $\hat{x} = [\hat{x}_1, ..., \hat{x}_T]$, for time intervals $t \in 1, 2, ..., T$. The four functions $f_1$, $f_2$, $f_3$ and $f_4$ expresses the performance as a function of different error measures. An intuitive interpretation of the problem given in (4) is that it searches the vector $\hat{x}$ that is closest to the a priori estimate $\bar{x}$, and, once it is assigned to the network produces the traffic data $y(x)$, $c(x)$ and $z(x)$ closest to their observed values. At each iteration step or time interval, $t$, $y(x)$, $c(x)$ and $z(x)$, could be extracted from inputs of the AIMSUN traffic simulator and could be calculated using traffic assignment of the DUE simulation (see subsection Benchmark platform). The set of constraints depends on application of the problem as well as the desired level of accuracy, and it can include non-negativity constraints, initial condition constraints, lower and upper bound constraints to avoid infeasible solutions and restrict search space, etc.

Traveler’s route choice or traffic assignment rules are often obtained by optimizing an objective function, which can be explicitly included in the set of constraints. This formulation results in a bi-level optimization and represents the solution framework for considered OD estimation methods in this benchmark study. The functional form of the four functions $f_1$, $f_2$, $f_3$ and $f_4$ for estimators considered in this benchmark study is given by least square formulation. Although, the use of least square approach to formulate the dynamic OD demand estimation model has been originally proposed by (27), many authors build-on their modeling frameworks by exploiting different traffic data. Since different traffic condition data contain very diverse values normalized least square functions are applied. The selection of normalized least square objective function indicates that considered methods belong to a common ”family” and ensures to get a better grasp of the algorithms performance and improvements due to application of richer traffic data on OD flows. Table 1 presents the main properties of selected dynamic OD estimation methods.

### Table 1: Properties of Selected Dynamic OD Estimation Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Input data</th>
<th>Objective function</th>
<th>Solution algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>+ +</td>
<td>least square (LS)</td>
<td>LSQR</td>
</tr>
<tr>
<td>Method 2</td>
<td>+ + +</td>
<td>normalized LS</td>
<td>SPSA AD-PI</td>
</tr>
<tr>
<td>Method 3</td>
<td>+ + + +</td>
<td>normalized LS</td>
<td>SPSA AD</td>
</tr>
<tr>
<td>Method 4</td>
<td>+ +</td>
<td>normalized LS</td>
<td>SPSA CG-TR</td>
</tr>
</tbody>
</table>

**Method 1: The LSQR method**

The least square approach to formulate the dynamic OD demand estimation model given in Eqn.(4) is used by (6). They build-on their modeling framework by exploiting (28) proposal of using deviation of OD flows as state variables and deviations of link flows. The main properties of the model are given as follows:

- input data: prior OD flow and link flow data
- solution approach: LSQR algorithm

To estimate dynamic OD demand by solving Eqn.(4) given by least square functions $f_1$ and $f_2$, Bierlaire (6) proposed the LSQR solution algorithm to get the computational performance required for very large networks. LSQR is an iterative method for solving the least square problem, analytically equivalent to a
conjugate gradient method, based on bi-diagonalization procedures (29). Key properties of LSQR approach are that assignment matrix (very sparse in large scale networks) does not need to be explicitly constructed or stored, only multiplications with vectors need to be implemented. This feature is attractive for large sparse problems, which is the network case in Figure 2. For more detail explanation of this algorithm we refer to the paper (6).

Method 2: The SPSA AD-PI method

Cipriani (30) formulate the dynamic OD demand estimation model by adding traffic condition data, i.e., densities, providing additional information on traffic regime. The main properties of the model are given as follows:

- input data: prior OD flow, link flow data and density data
- solution approach: SPSA AD-PI algorithm

The solution approach to solve dynamic OD demand problem given by normalized least square functions $f_1$, $f_2$ and $f_3$ in Eqn.(4), is modified SPSA (Simultaneous Perturbation Stochastic Approximation) algorithm proposed by (30). Different variants of the SPSA algorithm have been proposed in (30), (16), where the off-line dynamic OD demand estimation problem is formulated as a bi-level nonlinear optimization program and solved with an assignment-matrix-free method. The authors proposed solution approach that is modification of the gradient-based path search optimization method (SPSA) dealing with the Asymmetric Design (AD) for gradient computation and the Polynomial Interpolation (PI) of the objective function (4) for the linear optimization. SPSA AD-PI permits to reduce the computational efforts with respect to the usual gradient-based methods, that is a basic issue to deal with a simultaneous demand estimation for on-line applications. For more detail explanation of this algorithm we refer to the paper (16).

Method 3: The BiLevel-DUE method

An improvement of the previous Method 2, proposed in (16) has been studied assuming the availability of travel times between Bluetooth sensors along the main paths connecting them in the network (Figure 2(b)). The previous research reported in (31) has proved that a suitable Bluetooth sensor layout allows the identification of the paths between sensors and therefore the measurement of the associated travel times. Consequently, to implement the proposed method, the lower level DUE conducted with AimsunMeso needs to generate also the simulated travel time estimates from Bluetooth antennas along the corresponding paths.

The main properties of the model are given as follows:

- input data: prior OD flow, link flow data, density data and travel time data
- solution approach: SPSA AD algorithm

Thus, the dynamic OD estimation problem is defined by normalized least square functions $f_1$, $f_2$, $f_3$ and $f_4$ in Eqn.(4) and solved by modified SPSA AD-PI approach used in Method 2.

Method 4: The Enhanced BiLevel-DUE method

The computational experience showed that prior OD flow information had a twofold negative influence avoiding the estimated matrix to move away from the prior matrix on one hand, and a high computational cost on the other hand. (32) proposed framework by excluding information on OD flow data given in Eqn.(4). The main properties of the model are given as follows:

- input data: link flow data, occupancy data and travel time data
- solution approach: SPSA CG-TR

Thus, the dynamic OD estimation problem is defined by normalized least square functions $f_2$, $f_3$ and $f_4$ in Eqn.(4). This case study considers solution approach for given OD estimation problem to reduce computational time of the experiments. First, the computation of the approximated average gradient that could be enhanced using a conjugate gradient strategy as suggested in (33). It is known that conjugate directions permit to reach faster the solution than using the basic gradient method. Second, the use of a trust region
scheme is included as in (34). The main idea of trust region is to set implicitly at each iteration, a neighborhood around the current solution. Avoiding replications of matrices outside of the trust region is essential to reduce the computational burden. For more detail explanation of this algorithm we refer to the paper (32).

RESULTS
The performance of Method 1 using prior OD demand information and link flow data is presented in Figure 3. The estimation ability of the Method 1 demonstrates good performance, since no traffic condition data has been included in estimation process. This result can be explained by definition of state variables in Method 1, i.e., deviation of OD flows captures spatial and temporal deviations between prior and real OD flows. Although, the Method 1 shows no significant differences between considered scenarios when estimating OD demand (Figure 3(a) and 3(c)) for low (D7) and high (D9) demand level, link flow results indicate slightly worse estimates (Figure 3(b) and 3(d)). We could infer from the results that even a good estimates of OD demand can produce different link flow results, which is a proof of under-determinedness of OD demand estimation problem.

FIGURE 3 Method 1 results: $R^2$ for prior demand scenario D7 and D9

In line with findings described in literature and from Method 1, traffic condition data should improve the estimation ability of dynamic OD estimation algorithms, especially in congested networks such as one considered in this case study. Figure 4(a) and 4(b) provides an overview of Method 2 considering a prior OD demand lower then the real one (scenario D7). Information from traffic condition data, i.e., densities, has
the potential to influence the improvements in OD demand estimation from prior OD matrix, but this is not always the case. Results indicate that the highest estimation accuracy of Method 2 is observed for estimated OD flows (Figure 4(a)) and the lowest is observed for estimated link flows and densities (Figure 4(b)). These results show that including information on traffic conditions, despite its importance, may not suffice: while density allows to capture correct traffic regime at link level, its contribution at area level lowers for increasing network size and complexity because many OD flows combinations generate same link solution. Moreover, experiments show that SPSA algorithm is largely affected by a set of parameters related to its stochasticity and accuracy of assignment phase. Thus, appropriate refinement of values of such parameters has been adopted for Method 3, where different random seeds and objective function specifications have been used.

When travel time information is included in estimation process, the progress of Method 3 is improved over all prior demand scenarios, especially when prior OD information is close to the real traffic demand (Figure 4(c) and (d)). In addition, if travel time information is not included, the worst performance occurs when prior demand is lower then real one. Figure 4(c)(d) demonstrates improvement in estimation accuracy when travel time information is included in estimation process. These results imply the necessity of establishing new techniques to extract valuable information from AVI and AVL sensors.

Since results indicate strong dependency on the demand level of the prior OD, Figure 5 illustrates...
performance of Method 4 without prior OD information. When prior OD information is not included in ob-
jective function, solution approach without defined trust region needs more iterations to converge. However,
when solution approach based on conjugate gradient and trust region techniques is applied, computation
time is decreased. Figure 5 demonstrates that estimation accuracy increase for both OD demand and link
flows, when prior OD demand information is not provided.

![Objective Function D7](image1)

![Real vs Estimated OD Pairs D7](image2)

![Objective Function D9](image3)

![Real vs Estimated OD Pairs D9](image4)

**FIGURE 5** Solution approaches for Method 4: a) OD flow term evolution, b) real vs estimated total
OD flows.

Consequently, results obtained using Method 4 are improved and also a significant reduction in
the computational time is achieved. This is very important feature for on-line applications. For example,
Method 2 requires 40 dynamic equilibrium assignments for each iteration, resulting in 7.5 minutes for each
assignment on the Vitoria network, a total amount of 5 hours per iteration was needed.

**CONCLUSIONS**

In this paper, results show that despite the potential of information from advanced traffic data to improve
OD demand estimation, the information captured by these data are not fully explored by the available es-
timation procedures. Traffic condition data may help to correctly interpret the traffic link flow data, and to
identify which OD flows need to be adjusted, and in which direction. However, the main issue underling
the OD estimation methods, is spatial and temporal OD pattern given by prior OD matrix, especially in
congested networks. It is possible to infer from the results that even a good estimates of OD demand can
produce different link flow and traffic condition data, which is a consequence of under-determination of
OD estimation problem. In addition, the computational experiments presented in this paper prove the ro-
bustness and quality of the OD estimates exploiting AVI measurements. The computational performance of
the Enhanced Bilevel DUE method without prior OD information and using gradients and trust region has
been substantially increased by significantly reducing the number of function evaluations and the number of
iterations, converging faster in this way to better demand estimates. These OD estimation methods provide
effective tools for off-line pre-processing of prior OD data for on-line applications.

This paper did not intend to claim the superiority of one type of traffic data over the other but
was intended to show the potential of different types of traffic data for dynamic OD estimation. The use
of advanced traffic data to model dynamic OD demand is relatively new, and the literature still show a lack of empirical experiments to validate their use for dynamic OD estimation clear. The benchmark study presented here indicates that advanced traffic data require more research efforts and new techniques to turn them into usable information.

ACKNOWLEDGMENT

This research is partly funded by the ITS Edulab, a collaboration between TUDelft and Rijkswaterstaat. Also, this research is supported by the EU COST Action TU0903 MULTITUDE Methods and tools for supporting the Use of validation and evaluation of Traffic simulation models project and AIMSUN - Transport Simulation Systems.

REFERENCES


