Modified bi-level framework for dynamic OD demand estimation in the congested networks

Tamara Djukic
Aimsun SL
Ronda Universitat 22B, 08007 Barcelona, Spain
Phone: +34 933 171 693; Email: tamara.djukic@aimsun.com
(Corresponding author)

David Masip
Aimsun SL
Ronda Universitat 22B, 08007 Barcelona, Spain
Email: david.masip@aimsun.com

Martijn Breen
Aimsun SL
Ronda Universitat 22B, 08007 Barcelona, Spain
Email: martijn.breen@aimsun.com

Josep Perarnau
Aimsun SL
Ronda Universitat 22B, 08007 Barcelona, Spain
Email: josep.perarnau@aimsun.com

Jordi Casas
Aimsun SL
Ronda Universitat 22B, 08007 Barcelona, Spain
Email: jordi.casas@aimsun.com

November 15, 2017

Word count:

Number of words in abstract	234
Number of words in text (including abstract, first page and ref)	5976
Number of figures and tables	6 * 250 = 1500
Total	7476

Submitted to the 97th Annual Meeting of the Transportation Research Board, 7-11 January 2018, Washington D.C. November 15, 2017

ABSTRACT

This paper presents a modified bi-level optimization framework to solve the high-dimensionality of a non-linear OD estimation problem that is frequently found in congested networks. The upper-level is formulated as a generalized least square function of OD demand and traffic counts. The framework is modified by adding a recursive step to account explicitly the impact of OD demand variation on traffic observations, leading to improvement of the optimization function performance. The recursive step involves evaluation of the marginal effects for the subset of significant OD pairs whose variation leads to large changes in link flow proportions and traffic flows. Furthermore, to overcome the extra computational requirement, an heuristic method for obtaining a reduced set of OD pairs in the evaluation of marginal effects is proposed. In this way, the number of optimization function evaluations is reduced allowing the modeller to control the trade-off between simplicity of the model and the level of realism for large-scale, congested networks. A conventional bi-level optimization solutions approach is used in the performance assessment study. All the OD demand estimation approaches are implemented in a mesoscopic traffic simulation tool, Aimsun, to perform the traffic network loading on a large-scale network: the Vitoria urban network with 3249 OD pairs, 389 detectors, and 600km road network. The results demonstrate the applicability of the proposed solution approach to efficiently obtain dynamic OD demand estimates for large-scale, congested networks within computationally short periods.

INTRODUCTION

This paper focuses on the efficient estimation of OD matrices for congested networks, which are essential inputs for dynamic traffic management and Dynamic Traffic Assignment (DTA) applications. The absence of reliable OD matrices, especially in the peak hours, limits the potential for DTA deployment to analyse and alleviate traffic congestion as part of Intelligent Transportation System (ITS) measures. In order to remedy this problem a number of solutions are being examined and developed as part of the EU Horizon2020 SETA project which aims to examine the impact of multiple dynamic data sources, many of which, are sourced through telematics or mobile phones.

The OD estimation problem itself is computationally intensive due to the complexity of the demand estimation problem, the approaches, and the fact that dynamic OD matrices for real-life transport networks typically constitute high dimensional data structures. One of the problems with estimating OD demand is that, in many cases, development and availability of emerging technologies for their direct observation is still in the early stages. Thus, OD trips have to be inferred from alternative, available traffic observations (e.g. link traffic counts, speeds). From a modelling point of view, the key difference between the OD demand estimation approaches, is how the relationship between OD demand and any available traffic data is defined, calculated and re-calculated throughout the estimation process. Therefore, an exact description of this relationship leads to an accurate description of traffic state reality in the network, but to more complexity as well.

Initial efforts in research and practice have defined this relationship as linear with the assumption that variations in OD demand do not impose changes in all the traffic flow observations. However, the linear relationship may be invalid when congestion builds up in the network, resulting in a non-linear relationship between OD flows and link traffic observations. Consequently, the existence of non-linearity may lead to non-optimal solutions. In the past decades, researchers have attempted to develop new methods and techniques to capture the relationship and effects of OD demand variation on traffic observations. These methods can be categorized into three fundamental alternatives used to express this relationship: analytical derivation, simulation-based, and numerical-based approximation.

Analytical derivation: Dynamic link-flow proportions, expressed in the assignment matrix form, are typically used to express the weights between OD flows and link traffic counts. Theoretically, these link-flow proportions can be analytically derived using network topology, path choice set, current route choice model and equilibrium travel times (*1*). However, it is recognised that the complexity of the problem at hand can quickly lead to intractable situations (*1*).

Simulation-based approximation: The relationship between demand flows and traffic observations (e.g. link traffic counts, speeds, density) is uncovered by using traffic simulation without the direct derivation of the assignment matrix. The most studied method is the Simultaneous Perturbation Stochastic Approximation (SPSA) method (*2*), (*3*), (*4*), (*5*), (*6*), (*7*) which allows one to approximate a descent gradient direction with significantly lower computational resources than through the explicit derivation. However, due to the stochasticity of the simulation model, for each perturbation of the SPSA or W-SPSA where gradient needs to be determined, the DTA has to be replicated R times, leading to $2R$ runs (*7*).

Numerical approximation: Recent studies by ((*8*), (*9*), (*10*)) rely on linear approximation of the assignment matrix, which explicitly accounts for congestion effects. This definition requires the computation of the marginal effects of demand flow change on the link-flow proportions at the current solution of each iteration. It is possible to use the finite differences approach to numerically approximate the Jacobian matrix by using a traffic simulation, but it would be required in every iteration of the gradient solution to perturb each element in the OD demand vector, one at a time, leading to $2RDK$ runs, where D the number of OD pairs in the network and K the number of time intervals for the simulation period. In their studies relationship between demand flows and link traffic counts has been examined.

Given the high computational costs involved in evaluating marginal effects of OD demand changes on traffic observations, there is a need for heuristic approaches and algorithms that can identify solutions with a good performance at a low computational cost. To overcome the computational overhead, a number of authors have proposed **heuristic-based approaches**. Toledo and Kolechkina (2013) neglected the effect of changes in one OD pair over the other OD pairs in the assignment matrix. Frederix, Viti and Tampre (2013) implemented space decomposition of the network in the congested and non-congested sub-networks, where derivatives were computed only for the congested area. Shafiei et al. (2017) reduced computation costs through iterations progress and computing derivatives on OD pairs whose flows have a higher tendency to vary during the dynamic OD demand estimation process. However, their approach requires in the second iteration step evaluation of the derivatives for all the OD pairs that still leads to $3DK$ simulation runs and high computational costs. Note that three simulation evaluations per each OD pair within one iteration step are required to compute the numerical derivatives of the first Taylor approximation. Also, all these approaches rely on strong heuristic assumptions such as ignoring the effect of OD demand changes outside of congested area as shown in Djukic et.al (2017) ((*11*)) or have been tested on relatively small or medium sized networks. Further research is therefore necessary to develop approaches to solve nonlinear OD estimation problems that will guarantee reliability and computational efficiency in the large-scale networks.

Here we extend the previous work by proposing a modified bi-level optimization solution approach to estimate dynamic OD demand for large-scale, congested networks that accounts for relationship approximation between traffic counts and OD flows. This relationship has been computed for the subset of the OD pairs when performance of the objective function has been deteriorated. The subset of the most important OD pairs in the network has been identified based on the highest variation in the link flows obtained with demand derived in two consecutive iteration steps. Reducing the problem dimensionality through selection of the most significant OD pairs replaces the conventional approach of computing derivatives for all OD pairs, whether through all the iteration steps or in the initial step as proposed in (*10*). The importance of this approach lies in the possibility to capture the most important effects of congestion, and not only at congested links, by relaxing assumption of constant link-flow proportions without loss of accuracy and considerable decrease in model dimensionality and computational complexity. In addition, model formulation and solution is not limited only to link traffic counts. For example, other traffic observations can be added in traffic measurements vector as well as in function for the selection of the most important OD pairs.

The paper is organized as follows. In the first Section, we summarize the main challenges in defining the non-linear relationship between traffic observations and OD demand. In the second Section, we present the modified bi-level optimization framework with an additional recursive step to overcome the divergence of the optimization function performance. Next, we explore the properties of reducing the number of optimization function evaluations by defining the subset of significant OD pairs whose variation leads to large changes in link-flow proportions and traffic flows. Subsequently, we demonstrate the performance of the proposed OD estimation model on a large-scale network, Vitoria, Spain. The paper closes with a discussion on further research perspectives of the OD demand estimation model.

THE PROBLEM FORMULATION

This section describes the most critical issue in OD matrix estimation, whether static or dynamic, the relationship (mapping) of the observed traffic condition data with unobserved OD flows. This relationship is often accomplished by means of an assignment matrix. In the dynamic problem, the assignment matrix depends on link and path travel times and traveller route choice fractions - all of which are time-varying, and the result of dynamic network loading models and route choice models. These dynamics are reflected in travel times between each origin and destination trips on a network, influenced by traffic link flow. While a vast body of literature has been developed in this area over the past two decades, this section focuses on some of the efforts that highlight the basic dimensions of the problem.

The general OD estimation problem is to find an estimate of OD demand by effectively utilizing traffic and demand observations. Let $\Omega \subseteq N \times N$ be set of all n OD pairs in the network, and $L^{'} \subseteq L$ be the set of l links where traffic data observations are available. The time horizon under consideration is discretised into R time intervals of equal duration, indexed by $r = 1, 2, ..., R$. The OD matrix, $\mathbf{x} = \{x_{nr}\}$, defines the demand for each OD pair $n \in N$ with departure time interval $r \in R$. Prior information on the OD matrix is defined as vector, $\tilde{\mathbf{x}} = \{\tilde{x}_{nr}\}$. The vector $\tilde{\mathbf{y}} = \{\tilde{y}_{lt}\}$ defines traffic flow observations for time interval $t = 1, 2, ..., T$, for each link in $L^{'}$. It is also assumed that T and R describe the same length of time interval, but their decomposition to time intervals can be different.

The dynamic OD estimation problem can be formulated as a constraint optimization problem (*12*) as:

$$\underset{x \geq 0}{min} Z(\mathbf{x}) = \alpha f(\mathbf{x}, \tilde{\mathbf{x}}) + (1 - \alpha) f(\mathbf{y}, \tilde{\mathbf{y}}) \tag{1}$$

Regardless of the function f used, the purpose is to obtain an OD demand that yields OD flows and traffic data as closely as possible to their observed values. When solving the OD problem in Eq. (1) the relationship between traffic observations and OD demand has to be defined, implicitly or explicitly. Most dynamic OD demand estimation methods, define this relationship implicitly by the traffic assignment model that can be expressed as:

$$\hat{\mathbf{y}}_t = \sum_{h=r-k}^{r} A_t^h \mathbf{x}_h \tag{2}$$

There are two main drawbacks of relationship defined in Eq (2):

1. *Separability of traffic count observations*: it assumes that the traffic flow observed at the link l during time interval t can always and only be changed by changing one of the OD flows that passes link l when $\mathbf{x}_h$ is assigned in the network. This assumption of separability is incompatible with some typical phenomena in congested networks, such as congestion spillback between links and time lags due to the delay during congestion. In these cases, it is very likely that increasing an OD flow will cause delays to other flows that do not pass that time-space interval, hereby altering the amount of flow passing the link in the considered time interval. This issue has been addressed in past studies ((*13*), (*14*), (*15*)). Frederix, Viti and Tampre (*9*) suggested using a Taylor approximation to specify the linear approximation of Eq. (2) using a non-separable response function, given by

$$\hat{\mathbf{y}}_t = \sum_{h=r-k}^{r} A_t^h(\mathbf{x}_0)\mathbf{x}_r + \sum_{h=r-k'}^{r} (\mathbf{x}_h - \mathbf{x}_{0h}) \left[\sum_{h'=r-k'}^{r} \frac{\mathrm{d}(A_t^{h'}(\mathbf{x}_{h'}))}{\mathrm{d}\mathbf{x}_h} \mathbf{x}_{0h'} \right] \tag{3}$$

2. *Limited only to one data source*: formulation of relationship by assignment matrix in Eq. (2) and Eq. (3) restricts dynamic OD demand estimation problem to the use of traffic count data only, which can potentially over-fit to counts at the expense of traffic dynamics. Relationships between traffic condition data, such as speeds and densities, and OD flows are expected to be non-linear and approximations similar to the assignment matrix cannot be justified (*16*). This issue has been addressed in the past studies ((*16*), (*4*), (*6*), (*7*)) who proposed use of traffic simulation models to capture the nonlinear relationship between OD flows and traffic observations instead of the assignment matrix.

Although presented solutions significantly contributed to quality improvement of dynamic OD demand estimates, they still share a common challenge to overcome high computational costs. A complicating factor in utilizing these methods for estimation or prediction purposes, is that OD matrices are very large data structures, that grows rapidly in large networks. Even in case where high-dimensional OD flows can be reduced (see e.g. (*17*) and this is not entirely unlikely, there are serious methodological difficulties in finding optimal solutions (e.g. getting stuck in local minima, slow convergence, high number of simulation runs, etc.), aside from the computational and memory requirements for such a procedure on the basis of thousands (to millions) of traffic observations. For example, computing the exact Jacobian vector in the second term of Eq. (3) with respect to changes in OD flows for each OD pair remains intractable even when an efficient, well calibrated, DTA model is used.

METHODOLOGY

Algorithms proposed in literature to solve the problem given in Eq. (1) that incorporate computation of the marginal effects of demand changes on traffic observations, lead to high computational costs for medium- or large-scale networks. In this situation, dimensionality reduction of simulation runs required to capture these effects is necessary, leading to improve computational performance. In order to overcome problems related to the dimensionality of OD demand problem we propose the following heuristic approach. First of all, we propose the use of Eq. (3) to capture marginal effects with respect to changes in OD flows, rather than using a more conventional approach with linear assignment proportions given by Eq. (2). Secondly, we propose to use Eq. (3) on the subset of the OD pairs whose variation in demand creates the divergence of the cost function given by the objective function defined in Eq. (1). Lastly, we suggest using an initial OD matrix that produces similar congestion patterns as those observed in reality, i.e. that allows one to start with the correct traffic regime. It is convenient to start the presentation of the proposed solution approach with reference to the idea of OD demand estimation problem formulation as bi-level optimization framework. Then, we provide modified bi-level optimization framework with recursive step to account the marginal effects of OD demand on the link-flow proportions.

Conventional OD model formulation in bi-level optimization framework

Dynamic OD demand estimation problem can be defined as a bi-level optimization framework. The main advantage of using the bi-level formulation is the ability to capture network congestion effects in the dynamic OD demand estimation problem, as the traffic assignment model can be defined as an optimization problem in itself. The upper level is formulated as an ordinary least square (OLS) problem, which estimates the dynamic OD demand based on the given link-flow proportions. Assuming that errors are independently and identically normally distributed, the objective function aims to minimize the square distance between estimated and observed traffic flows, and the estimated and prior OD demand matrix, defined in Eq. (4) as follows:

$$\underset{x \geq 0}{min} Z(x) = \alpha \left\| x - \tilde{x} \right\|^2 + \left\| (A(x)x - \tilde{y} \right\|^2 \tag{4}$$

subject to

$$y = DTA(x) \tag{5}$$

Here we assume that the entire set of link traffic counts for the analysis period, $L^{'} \times T$, is used to simultaneously estimate the OD demand for all time intervals, $N \times R$. The link-flow proportions are, in turn, generated from the dynamic traffic network loading problem at the lower level, which can be solved through a simulation-based DTA procedure (in this case, Aimsun software (*18*)).

In general terms, all dynamic OD demand estimation methods defined as a bi-level optimization problem aim to find the most probable OD matrix by iteratively solving problems defined at upper and lower-level. The iterative solution algorithm is given as follows:

Algorithm 1 The conventional bi-level optimization algorithm

Step 1. Initialization. Initiate prior OD demand matrix, set $k = 0$.

Step 2. Assignment. Assign the demand to the network to obtain assignment matrix, A_k^h and estimated link traffic counts on the links with traffic observations, by Eq. (2) or Eq. (3).

Step 3. Convergence test. Check objective function value convergence. If objective function value has converged, stop and accept the current demand. Otherwise, proceed to step 4.

Step 4. Update OD demand. Estimate OD demand with link flows obtained from DTA, as given by Eq. (2). Go to step 2, $k = k + 1$.

End

When non-separability of traffic observations is considered, as was shown in the previous section, Eq.(3) has to be applied in Step 2 to capture the marginal effects of the demand variation on the changes in traffic flow observations. The traffic assignment relationship given in Eq.(3) can be solved by computing the numerical derivatives using a finite or central differences method of the traffic link flows with respect to changes in all OD pairs. This requires perturbing each OD pair in the OD demand two times, one at the time, resulting in $2NR$ traffic simulation runs and objective function evaluations per iteration step. It is obvious that such an approach will result in computationally expensive tasks, that have to be overcome.

Modified bi-level optimization framework

The bi-level optimization framework presented in the previous section is modified to meet the following requirements for congested, large-scale networks:

- Relax assumption on link-flow proportions derived from DTA by computing the marginal effects of the demand deviations on link flows given by Eq.(3);
- Reduce the number of OD variables in Eq.(3) through the inclusion of only those OD pairs whose change in demand values cause significant deviations in the link flows;
- Keep the computational costs lower.

These requirements are implemented through the following modified iterative solution algorithm with recursive step:

Algorithm 2 The modified bi-level optimization algorithm

Step 1. *Initialization.* Initiate prior OD demand matrix, set $k = 0$, $I^{'} = \emptyset$.
Step 2. *Assignment.* Assign the demand to the network to obtain assignment matrix, A_k^h and estimated link traffic counts on the links with traffic observations, by Eq.(2) or Eq.(3).
Step 3. *Convergence test.* Check objective function value convergence. If objective function value has converged, stop and accept the current demand. Otherwise, proceed to step 4.
Step 4. *OF performance test.* Check performance of the objective function value. If objective function decreases proceed to step 5. Otherwise, proceed to step 6, $k = k - 1$.
Step 5. *Update OD demand.* Estimate OD demand with link flows obtained from DTA, as given by Eq.(2). Otherwise, proceed to step 2, $k = k + 1$.
Step 6. *Select OD pairs.* Determine OD pairs whose variation has a considerable impact on link flow variation in the previous iteration and insert them in $I^{'}$.
Step 7. *Update assignment.* Update the link-flow proportions in the assignment matrix A_{k-1}, with values obtained from Equation (3) for the selected OD pairs in $I^{'}$.
Step 8. *Update OD demand.* Estimate OD demand with link flows obtained from Eq.(3). Go to step 2, $k = k + 1$.
End

The common and modified bi-level optimization framework with inputs and outputs for OD demand estimation is illustrated in Figure 1.

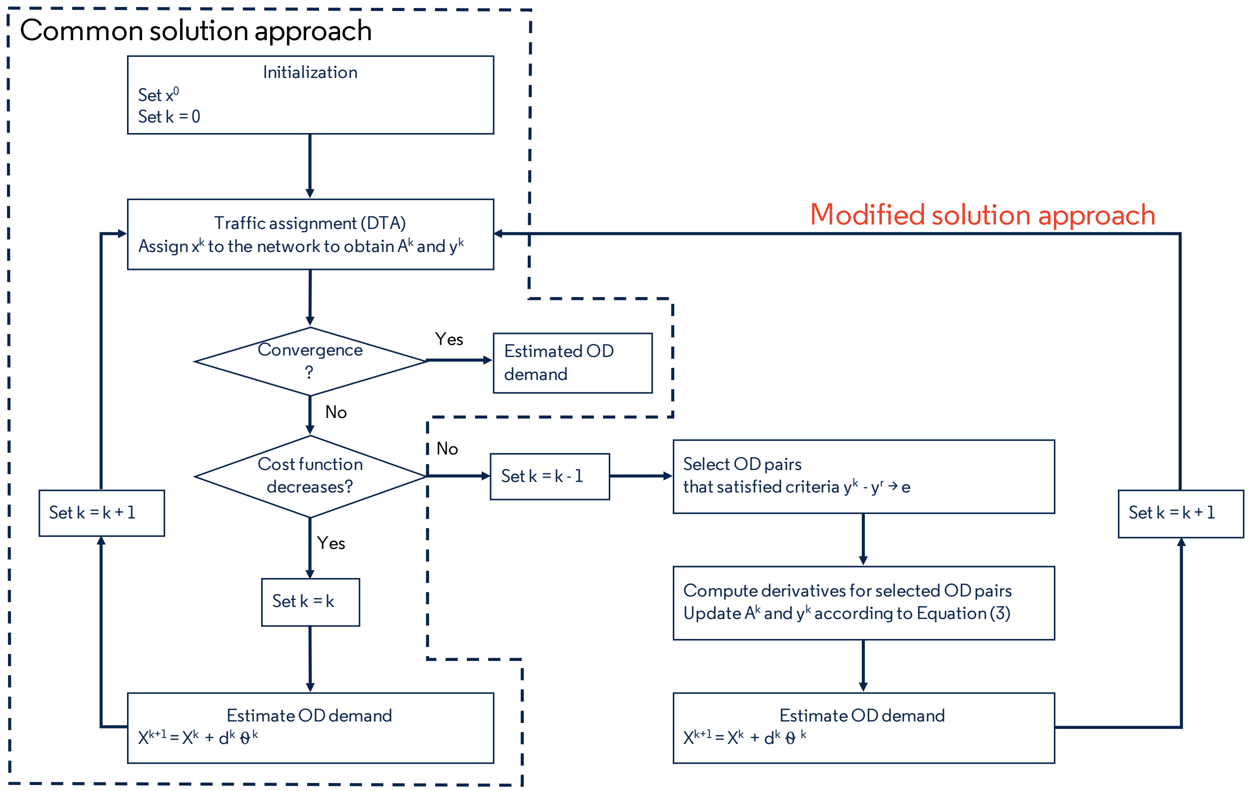


FIGURE 1 Generic algorithm based on the proposed modified bi-level solution framework.

Note that the proposed solution algorithm in Steps 6 and 7 uses a demand with one iteration step latency because the updated demand at iteration step k caused the increase of the objective function value. Therefore, following the modified bi-level framework, Step 8 denotes correction of the state variable for iteration k, using the information from the link-flow proportions and link flows for iteration $k = k - 1$, obtained in Steps 6 and 7. In Step 6, we analyse the change in the link flows obtained with demand assigned in iteration step k and $k = k - 1$, and determine the link flows with the highest variation. Using the information from the link flow proportions matrix, we can identify which OD flows are crossing these links with the highest flow variation and set them in $I^{'}$. Then, in Step 7 the elements of the link-flow proportion matrix are corrected for these OD pairs using the Eq.(3). The upper-level problem in Steps 5 and 8 is solved using the gradient decent method. The OD demand estimation results are evaluated in Step 3 against termination criteria and the procedure would continue if termination criteria is not met. Finally, the Steps 6-8 reflect our corrected knowledge on the OD demand state at iteration $k = k - 1$ to improve the performance of the solution algorithm.

Method for solving the upper-level problem

Let $\mathbf{x}_k$ be the demand at iteration step k, and A_k and $\mathbf{y}_k$ the assignment matrix and simulated traffic counts given by this demand. As an approximation to the OD estimation objective function given in Eq. (4) that we want to minimize, we consider the following auxiliary objective function:

$$Z_k(x) = \|\tilde{y} - y_k - A_k(x - x_k)\|^2 + \alpha \|x - \tilde{x}\|^2 \tag{6}$$

There are different types of exact and heuristic methods proposed in the literature that can be employed to solve the optimization problem defined in Eq. (6) with non-negative variable constraints. At every outer iteration step, the gradient descent method is selected to minimize the objective function defined in Eq. (6), which uses the gradient as search direction:

$$d = -\nabla Z_k \tag{7}$$

where

$$\nabla Z_k(x) = 2\alpha(x - \tilde{x})^2 + 2(A_k^T A_k x - A_k^T \tilde{y} + A_k^T y_k - A_k^T A_k x_k) \tag{8}$$

To perform this gradient method, we start at $x = x_k$ and we perform M gradient steps, the direction being given by the latter Eg. (8). At internal step $m \leq M$, let us denote the estimated demand by $\mathbf{x}_k^m$. After determining the search direction, which is given by $\nabla Z(\mathbf{x}_k^m)$, the optimal step length, θ^m needs to be obtained in each internal iteration step. The following criterion is used to compute the step size:

$$\theta^m = \underset{\theta^m}{min} Z(\mathbf{x}_k^m - \theta^m \nabla Z(\mathbf{x}_k^m)) \tag{9}$$

The exact line search procedure proposed by Cauchy (1847) (*19*) is used to compute the step size. In the case where Z is a quadratic function, the optimal step can be computed analytically. In this case, the optimal step size is computed using the following expression:

$$\theta^m = \frac{\|\nabla Z(\mathbf{x}_k^m)\|^2}{\left\|\nabla Z(\mathbf{x}_k^m)\right\|^2 + \left\|A_k \nabla Z(\mathbf{x}_k^m)\right\|^2} \tag{10}$$

NUMERICAL EXPERIMENT DESIGN

In this section, we will first describe the input data used, e.g. historical OD demand generation and the DTA traffic assignment procedure. We consider three assessment scenarios in terms of link-flow proportions derivation (i.e. with and without computation of marginal effects). Numerical experiments are performed on a large-scale network, (Vitoria, Basque Country, Spain) with real data to evaluate the performance of the proposed approach.

DTA with mesoscopic simulation model

In the experiments, we use the mesoscopic event-based demand and supply models in Aimsun, each synthesizing microscopic and macroscopic modelling concept. The travel demand in Aimsun is represented by dynamic OD demand matrices. Vehicle generation is performed for each OD pair separately with arrival times that follow an exponential distribution. The iterative interaction between demand and supply models allows the system to update the set of routes and the travel times after each iteration leading to robust estimation and prediction of traffic conditions in the network. For this study, a route choice set will be pre-computed in Aimsun and used as fixed for all the simulation runs in performance analysis. In this way, dependence of re-routing effects on the changes in the OD demand is ignored. Here we focus on investigating the effects of travel time variation and congestion spill-back on traffic observations in the network.

Network and traffic data

The proposed OD estimation approach is evaluated for the large-scale network in Vitoria, consisting of 57 zones, 3249 OD pairs (57 x 57) with 2800 intersections and 389 detectors, presented as black dots in Figure 2. This network is available in the mesoscopic version of the Aimsun traffic simulation model for the reproduction of traffic propagation over the network. The true OD demand is available for this network, which allows analysts to assess the performance of the proposed method. The true assignment matrix and traffic counts on detectors are derived from the assignment of true OD matrix in Aimsun for the evening period from 19:00 to 20:00 reflecting a congested state of the network. The simulation period is divided into

15 minute time intervals with an additional warm-up time interval, $R = 5$. The link flows resulting from the assignment of the true OD demand are used to obtain the real traffic count data per observation time interval.

FIGURE 2 The Vitoria network, Basque Country, Spain

The historical OD demand flows are derived by adding a uniform normal component in the range of $\pm$ 40% to the real OD demand to produce uncertainty in the historical demand and congestion in the network.

Assessment scenario

Three assessment scenarios have been defined for the performance assessment of the proposed solution approach. The main goal of this task is to evaluate the expected improvements due to exact implementation of marginal effects of OD demand variations. Thus, dynamic OD estimation method that shares same performance measure and solution framework, i.e. least square (LS) error measure defined in bi-level solution framework is selected as a benchmark scenario. Further, two solution strategies in selecting the number of OD pairs have been defined to assess the performance of the proposed OD estimation method. Subsets of OD pairs involved in proposed modified solution approach are defined, such that in every iteration step we can identify potential set of OD pairs that impacts the 90% and 80% of variation in traffic counts. For example, the distance between observed and simulated traffic counts is computed, and arranged in decreasing order of deviation magnitude in each iteration step. Then, OD pairs that dominate changes in the traffic detectors with deviation higher then 90% are selected for their evaluation of the marginal effects in demand estimation process, here denoted as $I^{'} = 90\%$. In this evaluation task, three assessment scenarios are considered:

1. **Conventional bi-level approach**: LS solved by conventional bi-level solution approach without explicit non-linear relationship formulation;
2. **Modified bi-level approach with** $I^{'} = 90\%$
3. **Modified bi-level approach with** $I^{'} = 80\%$

A point of interest now is finding out to which extent the estimation accuracy and computational time are improved. To get a better grasp of the algorithms real world performance, results are presented in the following section.

RESULTS

The performance of the objective function for all three scenarios are presented in Figure 3. For the purpose of this study, convergence was defined as reaching an objective function value that is three times lower than the initial value obtained (by any of the algorithms) within 20 iterations. The performance of the proposed modified bi-level solution approach demonstrates satisfying results, since it is able to maintain the decrease of the objective function value through iteration steps. Modified bi-level approach with both OD pair subsets demonstrate convergence trend to a local minimum, in contrast to conventional bi-level optimization framework. Note that conventional bi-level approach did not reach convergence, which seems to indicate that search directions that algorithm produces are increasingly inefficient as the algorithm progress. For the purpose of visualisation, we have shown results for the conventional bi-level approach up to iteration step 13 and stored the results for further analysis from this step.

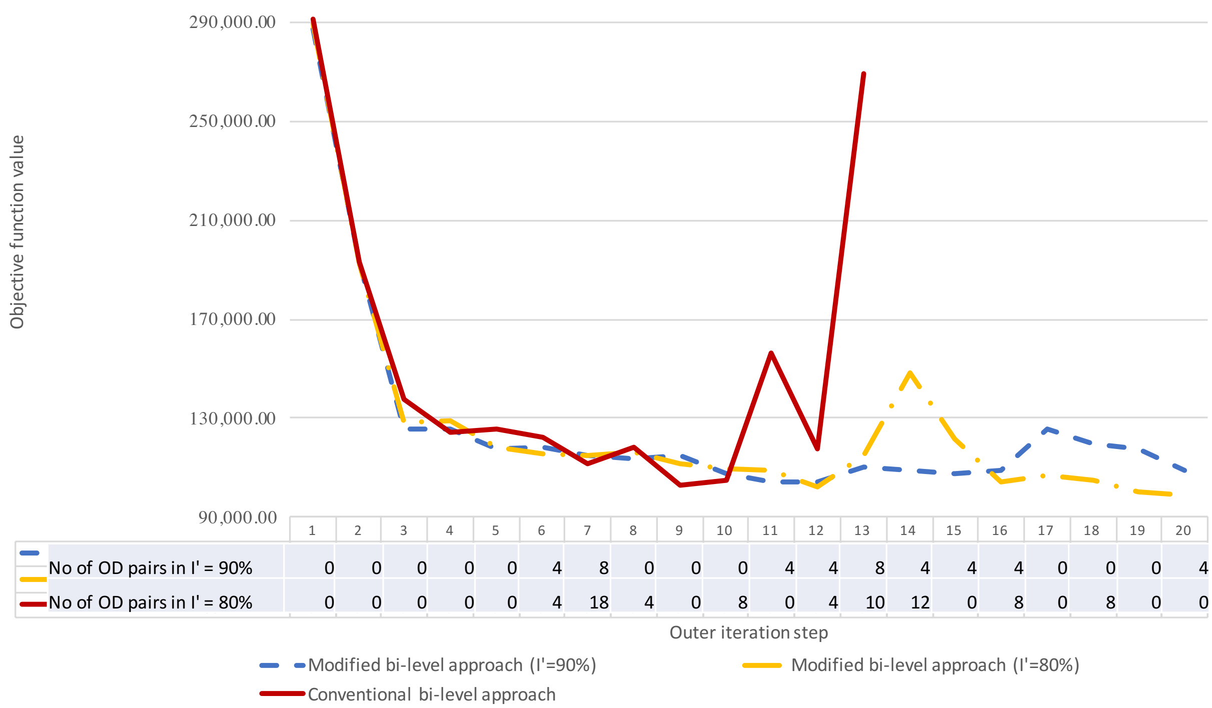


FIGURE 3 Comparison of the objective function performance

It can be observed from Figure 3 that the conventional bi-level approach runs into an unstable convergence trend from the iteration step 9. This effect is a consequence of initial OD matrix that reflects over fitted demand, very close to the congestion level. When demand obtained in iteration step 9 is assigned in the network it will result in the network blockage due to spill-back propagation. Thus, assignment matrix obtained out of this simulation step is not realistic and if used further by algorithm in iteration step 10 results in sudden divergence of cost function and intractable solution. Results shown in Figure's 3 table reveal that the proposed modified bi-level approach identified the cause of function deterioration in the iteration step 6, and by updating the elements of the assignment matrix in the recursive step guaranteed more stable objective function convergence. Also, results show that extending the number of OD pairs involved in updating the link flow proportions leads to objective function performance improvement (e.g., see modified bi-level approach with $I^{'} = 80\%$ in Figure 3) although with slight higher objective function deterioration as shown in iteration step 12 and 13. This effect can be explained by definition of the recursive step in modified approach, where derivatives have been computed for the larger set of OD pairs and whose effect has a better impact in finding more accurate demand solution. However, this effect can be further explored by extending the list of OD pairs whose variance dominates deviations in the traffic flows.

Figure 4 shows the estimated total OD demand per departure time interval for OD demand solution approaches. We have included the real OD demand in the figure as a point of reference. All three tested solutions demonstrate a tendency to slightly overestimate demand as a consequence of incorrect data interpretation from loop detectors when congestion level increase. However, performance of the proposed modified bi-level approach demonstrates a capacity to recognize an overestimation trend and improves demand estimation using the first order approximation given by Eq.(3) to update elements of the assignment matrix in time intervals when congestion occurs in the network. In addition, results indicate unstable performance of the conventional bi-level approach, where the misinterpreted impact of the congestion led to under- and overestimation of total demand.

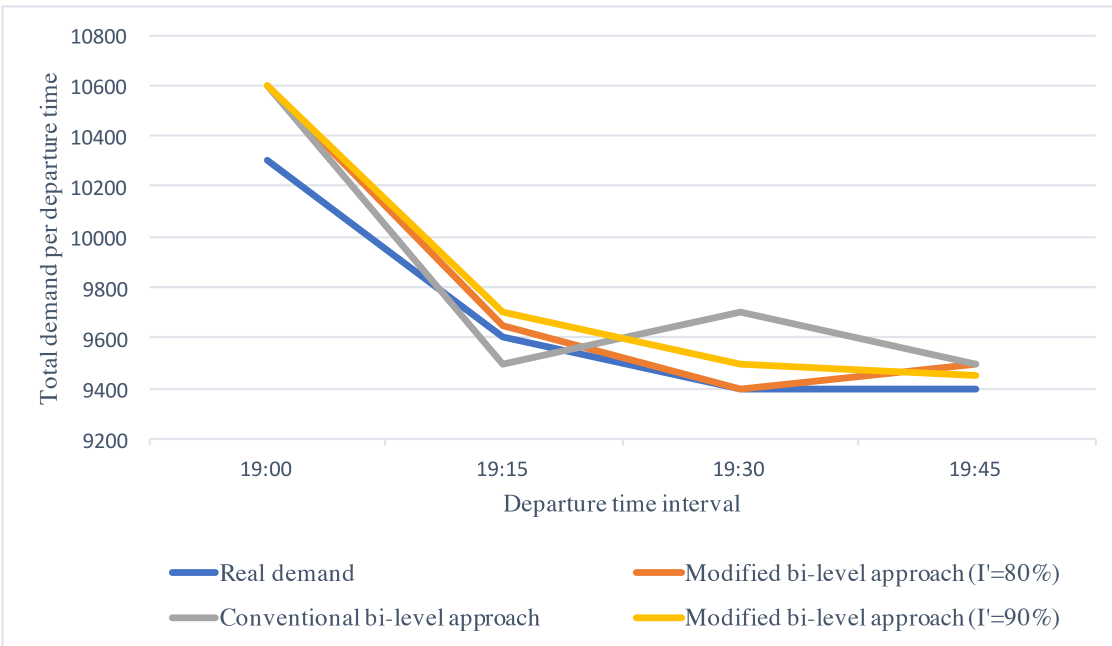
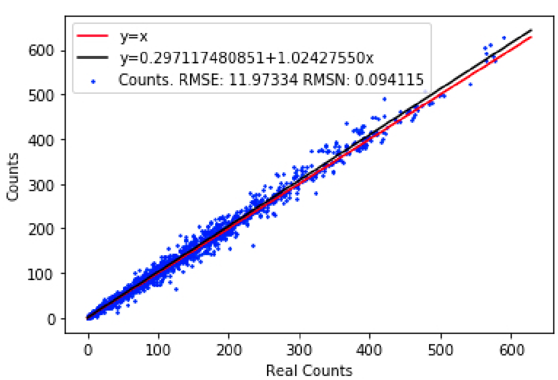
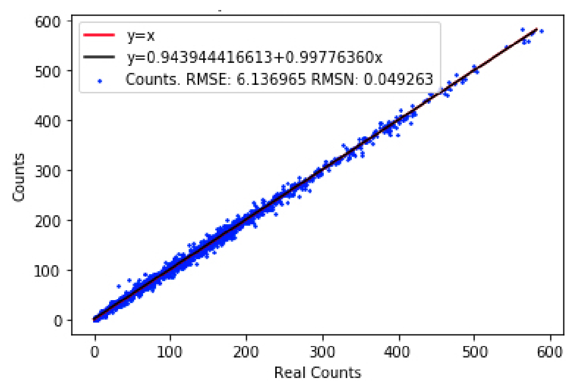


FIGURE 4 Estimated OD demand per departure time interval

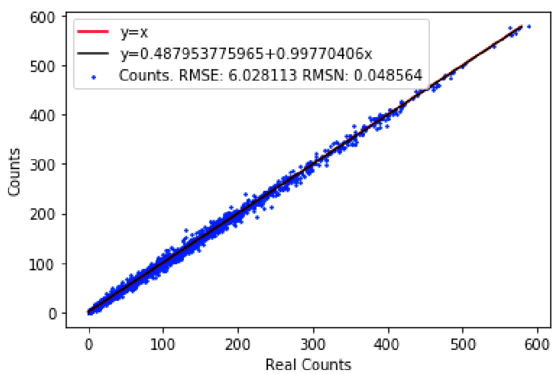
Next, it is important to investigate how these estimated OD demands once assigned to the network can produce traffic counts close to their real observations. Figure 5 provides a performance overview of approaches in terms of the relationship between simulated and observed traffic counts. It appears clearly from Figure 5 that all the approaches show correlation increase and significant reductions in RMSE and RMSN error. Results in Figure 5(d) indicate that proposed modified bi-level approach with higher sensitivity to demand changes has the best performance with improvement of the RMSE value for 52.29% compared to the initial value before OD demand estimation. It is interesting to observe from Figure 5(b) that a conventional bi-level approach demonstrates slightly lower reduction of the RMSE value for 45.19%. However, note that results presented here for the conventional bi-level approach are obtained from iteration step 13, since the method did not show a tendency to converge until iteration step 20 as discussed above. In turn, performance results of conventional bi-level approach should be considered with caution. A detailed examination of Figures 4, 5(d) and 5(c), shows that the more OD pairs are included in the evaluation of the marginal effects of demand variation on the traffic flows the better the demand estimation results obtained, in terms of total demand and traffic flows.



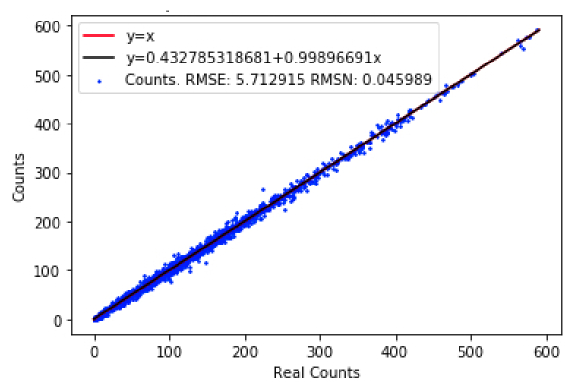
(a) Initial state from prior OD demand



(b) Conventional bi-level framework



(c) Modified bi-level approach with $I^{'} = 90\%$



(d) Modified bi-level approach with $I^{'} = 80\%$

FIGURE 5 Scatterplot of the observed and simulated traffic counts per solution approach

Note that initial idea was to solve the computational complexity of the OD demand estimation problem for real case applications while maintaining reliable estimation results in the congested networks. Therefore, Table 1 shows the run time and number of simulation runs for each of the tested solutions.

TABLE 1 CPU computation time and the number of the DTA simulations

CPU time	No. of assignment simulations	Aimsun simulation time	Demand estimation time (Python)
Conventional bi-level method	13	26min	31min
Modified bi-level method $I' = 90\%$	64	2h 2 min	51min
Modified bi-level method $I' = 80\%$	96	3h 4min	1h 16min
Benchmark method (Shafiei (2017))	366	12h 12min	14h 32min

Table 1 shows that the conventional bi-level method requires the least number of simulation runs. This can be explained as follows: the conventional algorithm requires one simulation run in each iteration, compared to the other two solutions that require three simulation runs for each OD pair within one iteration step to compute the numerical derivatives defined by Eq. (3). As a result, the gain in terms of run times obtained by the use of an assignment matrix without updates is large, but with a trade off on the lower quality of estimation results. Furthermore, this degradation in estimation accuracy is expected to increase for larger and more complex networks. We can observe significant CPU computation time reduction of the proposed modified bi-level solution approach compared to the benchmark method proposed by (*10*). This effect can be explained by definition of solution approaches. The proposed modified bi-level approach calculates derivatives for the subset of the OD pairs when deterioration of the objective function is observed in contrasted to the benchmark method that requires in the second iteration step evaluation of the derivatives for all the OD pairs. These times have been obtained by running Aimsun and Python on DELL Latitude E6430 with processor Intel Core i5-3320M, and2.6GHz memory.

CONCLUSIONS

The common approach usually adopted in dynamic OD demand estimation and prediction consists of solving an optimization problem in which the distance between observed and simulated traffic conditions is minimized by assuming the relationship between OD flows and traffic observations is independent of traffic conditions in the network. This approach has a severe shortcoming as it does not take into account the impact of demand flow variation on traffic observations in congested networks. Modelling of traffic observations dependency on variations in OD flows has been identified by many researchers as a key challenge in the estimation and prediction of high-quality OD matrices.

In this paper, we proposed a modified bi-level optimization framework to solve the high-dimensionality of non-linear OD estimation problem by computing the marginal effects of demand flow variation only for the most significant OD pairs with respect to traffic observations. This approach allows the modeller to control the trade-off between simplicity of the model and the level of realism. Several specific solution approaches that differ in the assumptions on the link-flow proportions derivation and solution algorithms were used in the performance evaluation study. From the results presented in this contribution, modified bi-level approach appears to outperform conventional bi-level solution with fixed linear relationship significantly and achieves great improvement over the reference case. Results show that proposed approach is able to capture the effect of congestion in the network and to reproduce the observed traffic conditions with high level of accuracy. Furthermore, we show that deriving a non-linear relationship between OD demand and traffic counts for the subset of the OD pairs will lead to computational efficiency with a guaranteed improvement in result's accuracy.

An improvement of the algorithm presented in this paper can been seen in two directions: 1) extension of the model as a multi-objective function with traffic condition data (i.e., speed, density, demand derived from floating car data) can be considered to overcome limitation of the method relying only on traffic count data; 2) explore alternative gradient solution approaches in solving the upper-level problem to avoid convergence in local minima, especially when initial OD matrix does not reflect congestion pattern in reality.

ACKNOWLEDGMENT

We acknowledge the support of the SETA project funded from the European Union Horizon 2020 research and innovation program under grant agreement No 688082.

REFERENCES

[1] Ashok, K. and M. E. Ben-Akiva. Estimation and Prediction of Time-Dependent Origin-Destination Flows with a Stochastic Mapping to Path Flows and Link Flows. *Transportation science*, Vol. 36, No. 2, 2002, pp. 184–198.

[2] Balakrishna, R., M. Ben-Akiva, and H. Koutsopoulos. Off-line Calibration of Dynamic Traffic Assignment: Simultaneous Demand and Supply Estimation. *Transportation Research Record*, 2007.

[3] Ben-Akiva, M., H. N. Koutsopoulos, and A. Mukundan. A dynamic traffic model system for ATMS/ATIS operations. *I V H S Journal*, Vol. 2, No. 1, 1994, pp. 1–19.

[4] Cipriani, E., M. Florian, M. Mahut, and M. Nigro. A gradient approximation approach for adjusting temporal origindestination matrices. *Transportation Research Part C: Emerging Technologies*, Vol. 19, No. 2, 2011, pp. 270–282.

[5] Cipriani, E., A. Gemma, and M. Nigro. A bi-level gradient approximation method for dynamic traffic demand estimation: sensitivity analysis and adaptive approach. *Proceedings of the IEEE conference on Inteligent TRansport Systems, 16th IEEE ITSC*, Vol. 1, No. 2, 2013.

[6] Cantelmo, G., E. Cipriani, A. Gemma, and M. Nigro. An Adaptive Bi-Level Gradient Procedure for the Estimation of Dynamic Traffic Demand. *Intelligent Transportation Systems, IEEE Transactions on*, Vol. 15, No. 3, 2014, pp. 1348–1361.

[7] Antoniou, C., C. L. Azevedo, L. Lu, F. Pereira, and M. Ben-Akiva. W-SPSA in practice: Approximation of weight matrices and calibration of traffic simulation models. *Transportation Research Part C: Emerging Technologies*, Vol. 59, 2015, pp. 129–146.

[8] Toledo, T. and T. Kolechkina. Estimation of Dynamic Origin-Destination Matrices Using Linear Assignment Matrix Approximations. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 2, 2013, pp. 618–626.

[9] Frederix, R., F. Viti, and C. M. J. Tampére. Dynamic origin-destination estimation in congested networks: theoretical findings and implications in practice. *Transportmetrica A: Transport Science*, Vol. 9, No. 6, 2013, pp. 494–513.

[10] Shafiei, S., M. Saberi, A. Zockaie, and M. Sarvi. A Sensitivity-Based Linear Approximation Method to Estimate Time-Dependent Origin-Destination Demand in Congested Networks. *Proceedings of Transportation Research Board - 96th Annual Meeting*, , No. Washington D.C., 2017, pp. 1–16.

[11] Djukic, T., M. Breen, J. , Perarnau, D. Masip, and J. Casas, Marginal effects evaluation with respect to changes in OD demand for dynamic OD demand estimation. In *Proceedings of the International Conference on Intelligent Transport Systems in Theory and Practice, mobil.TUM 2017*. Procedia, Munich, Germany, 2017, 1, pp. 1–11.

[12] Cascetta, E., D. Inaudi, and G. Marquis. Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts. *Transportation science*, Vol. 27, No. 4, 1993, pp. 363–373.

[13] Yang, H. and J. Zhou. Optimal traffic counting locations for origin-destination matrix estimation. *Transportation Research Part B: Methodological*, Vol. 32, No. 2, 1998, pp. 109–126.

[14] Tavana, H. and H. S. Mahmassani. Estimation of Dynamic Origin-Destination Flows from Sensor Data using Bi-level Optimization Method. *Presented at the 80th annual meeting of the Transportation Reasearch Board, Washington DC, USA*, 2001.

[15] Lundgren, J. T. and A. Peterson. A heuristic for the bilevel origin?destination matrix estimation problem. *Transportation Research Part B: Methodological*, Vol. 42, No. 4, 2008, pp. 339–354.

[16] Balakrishna, R. and H. Koutsopoulos. Incorporating Within-Day Transitions in the Simultaneous Off-line Estimation of Dynamic Origin-Destination Flows without Assignment Matrices. *Transportation Research Record*, 2008.

[17] Djukic, T., G. Flotterod, H. J. Van Lint, and S. P. Hoogendoorn, Efficient real time OD matrix estimation based on Principal Component Analysis. In *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE*, 2012, October, pp. 115–121.

[18] TSS-Transport Simulation Systems. *Aimsun Dynamic Simulators Users Manual v8*, 2015. Barcelona, Spain. Barcelona, Spain.

[19] Cauchy, A. M. Methode generale pour la resolution des systemes d'equations simultanees. *Comptes Rendus Hebd. Seances Acad. Sci.*, Vol. 25, 1847, pp. 536–538.