Origin-Destination Matrices Estimation for Dynamic and Urban Context

Emmanuel Bert\textsuperscript{1, 3}, Edward Chung\textsuperscript{2}, André-Gilles Dumont\textsuperscript{3}

emmanuel.bert@aimsun.com
Tel.: +34 933 171 693
Fax: +34 933 171 716

\textsuperscript{1} TSS-Transport Simulation Systems
Passeig de Gracia 12
08007 Barcelona
Spain

\textsuperscript{2} Queensland University of Technology
Gardens Point Campus
Brisbane 4001
Australia

\textsuperscript{3} Traffic Facilities Laboratory
Ecole Polytechnique Fédérale de Lausanne
CH-1015 Lausanne
Switzerland

Keywords:
Introduction

Nowadays, modelling traffic simulation is an increasingly widely used tool in transportation research. Traffic modelling can evaluate and quantify generated scenarios based on actual data; it can help transport managers in operational (e.g. routing or traffic management) and planning studies; and is also a helpful decision tool for short, medium and long-term studies. Nevertheless, the quality of inputs is not always best adapted for this kind of tool.

Traffic simulations are based on modelling of the supply (road infrastructure) and demand (traffic data). The most common technique for collecting traffic data is using sensors (loop detectors) for traffic counts. Even though this tool can give detailed time-dependent information about utilization at specific points, in most of the applications and particularly in the case of traffic simulations, this type of data does not produce a sufficiently accurate idea of the vehicle utilization of the network. Nevertheless, based on this input (available at small time intervals) and using other sources of information, dynamic Origin-Destination (OD) matrices, which represent the flow of vehicle between two points of the network, can be estimated to model dynamic and detailed demand.

Therefore, to carry out a transportation study using simulators, OD estimation is a crucial step as it represents the transport demand for the network. In this way, its quality has a great influence on the results of analyses based on this traffic representation. Quality and quantity must be as close as possible to the real situation. OD estimation (which consists of modification of an initial and approximate OD demand based on traffic count data) is solved as an optimization problem. This proposes an infinite number of solutions, because in almost all of the cases, there are more unknown parameters (OD pairs flows) to estimate than information (traffic counts data) available. The adopted methodology must find the optimal solution depending on the modelling constraints.

Dynamic OD estimation presents various challenging aspects. Demand and path evaluation must be carried out in time slices. From these time periods, OD estimation must be achieved taking into account link flows and the relation between them. Indeed, depending on the size of the network and its complexity (speed and distance from origin to destination), some of the vehicle might need more than one time period to reach their destination or counting sensor. This statement leads to the fact that counting values of one time interval could be influenced by the previous interval or intervals. As a consequence, demand generation (OD flows) for period n must take into account the action of the time period n, n+1, n+2… and N (Total study period is divided in N equal time periods). To achieve this, the different stages of the OD estimation process must be adapted to capture this evolution. Firstly, traffic assignment needs to be dynamic in order to propose route choice solutions on time depending on traffic conditions. In additions, the OD adjustment needs to take into account the evolution of trips in the
network. The algorithm must be able to make the distinction between the entrance time (in the network) and the time period at the traffic count point.

This paper focuses on dynamic OD estimation in urban networks. This kind of network presents particular characteristics which strongly influence traffic flows. Laminar flows are disturbed by traffic conflicts or signalisations (Stops, give ways, signalized intersections, etc.). Platoons of vehicles are interrupted and delayed by priorities between traffic streams. This discontinuity causes great variation in flow spreading and could lead to congestion (added to high demand) and high variation in travel time experienced within the network. Route choice possibilities in urban areas are usually greater than in other types of network. Traffic is then spread over a higher number of paths from an origin to the destination. Moreover, urban networks, due to the high density of traffic interfaces, present a higher number of OD pairs in most cases. This heterogeneity and distribution of the traffic in a large urban network make behaviour evaluation and modelling of the situation highly complex.

Given this situation, this paper focuses on ways to estimate relevant dynamic demand inputs based on traffic data for detailed traffic analysis and, in particular, for traffic simulation use in urban context.

**State of the Art**

State of the Art OD matrix estimation highlights different weaknesses of the current methodologies. Most commonly used methods deal with the OD-flow estimation problem using static approaches. They estimate a unique OD matrix for the complete study period. Disadvantages or weaknesses in the static method can lead to outputs inadequately adapted or incompatible for an exploitation of the data for detailed and dynamic analyses because it does not estimate time dependent traffic variation adapted for dynamic flow modifications. The propagation of flow is not considered. In this way, dynamic characteristics of the demand, particularly in an urban context, could not be modelled in an efficient manner.

Furthermore, in the literature, we find very little consideration for complex route choice possibilities in the evaluation of the assignment matrix (lower level problem). It could be carried out by observation, analytically or heuristically, but depending on the method, the quality of the assignment might be very different. To catch the complex evolution of traffic through time, heuristic approaches using traffic simulators are better adapted and cheaper (in terms of both time and cost) than other approaches.

Most traffic problems are observed in urban areas and present a greater challenge to traffic engineers due to the large effects of traffic perturbations (delays, pollution, costs, etc.). There are very few tools adapted for medium-to-large complex urban networks. For the rare cases dealing with urban typography, they are usually small networks with low route choice and signalized capabilities.
Papers from Balakrishna [2], Chang & Tao [10] and Bierlaire [6] are the most relevant for urban applications. The first uses a small and theoretical network and analytic approaches for the assignment matrix. The paper concludes that “much remains to be done to have a reliable dynamic OD system for efficient use in practice”. The second paper takes into account only freeways and main arterials. Bierlaire and Crittin apply Kalman Filtering algorithm in the Irvine network. This network is quite large in area and offers route choice capabilities (even if limited). However, in terms of link density (number of roads per unit area) and OD matrix size, it is not large. Moreover, the paper provides no information concerning traffic lights. In addition, the process of obtaining an assignment matrix (using DynaMIT) is not detailed. This research focuses more on computation time efficiency than OD estimation performances.

Finally, very few papers about OD matrix estimation consider detector layouts to be problematic (for instance [14]). The quantity and the position of these detectors are never discussed and there is no information concerning the rate of interception of flows or OD pairs by detectors (assuming that 100% are intercepted, but this hypothesis is seldom attainable in practical applications). Nevertheless, this issue is critical for OD estimation formulation. Indeed, quality of the estimation is directly linked to the number of OD pairs intercepted by the detection layout.

Therefore, by reviewing existing methods, from static to dynamic OD matrix evaluation, deficiencies in the approaches are identified: mainly, the level of detail in the traffic assignment for complex urban networks and the lack of dynamic approaches (see [4] for more detailed State of the Art).

**Methodology**

Based on the deficiencies identified, the proposed methodology focuses on several improvements of current solutions. The methodology for achieving dynamic OD estimation in an urban context is comprised of a heuristic bi-level approach. Assignment of the initial demand is performed by mesoscopic simulation, adapted for large and complex urban networks, based on the Dynamic User Equilibrium (DUE) to model detailed dynamic complex traffic patterns without numerous calibration parameters. For this step, the quality of the dynamic equilibrium and level of detail of the network are important features in providing an assignment representative of the actual one in whatever the traffic situation. Afterwards, OD flow adjustment is executed by an efficient least square solution which takes into account dynamic aspects of the flow propagation and traffic counts. For this task, an LSQR algorithm has been selected for its capacity to deal with large matrices and its ability to constrain outputs.

Figure 1 shows the details of the bi-level mechanism of the new approach:
Traffic assignment

The aim of the lower level is to assign the demand in the network and to know how it influences traffic sensors. Using an appropriate simulator in the lower level allows performing dynamic assignment on urban network and extracting all the necessary information useful for the process. It must achieve path estimation from origin to destination depending on urban constraints (signalisation, congestion, etc.). From the dynamic best paths evaluation, we need to extract useful information for the next step (upper level). Entrance time period, counted time period, proportion of the global flows of the OD pair concerned or counting location are established for the whole study period.

Based on these needs, the simulator "Aimsun 6.1" [3, 19], developed by the Polytechnic University of Catalunya in Spain and commercialised by TSS, has been chosen for this task. Indeed, it offers an API\(^1\) that allows the possibility of exporting/importing all the necessary information. Its high level of detail and low number of parameters make it a perfect tool for urban complex networks simulations.

The Aimsun mesoscopic model is based on the event scheduling approach. It means, instead of focusing on the trajectory of each vehicle, this model is interested in different events in the network (generation of a new vehicle, entrance into a link, transfer from one link to another, etc.).

\(^1\) Application Programming Interface.
Network loading is simplified and focuses on flow density, speed or queues. Moreover, lane changing, car following, and gap acceptance models are simplified to reduce the number of calibration parameters. Given that the network loading is based on a heuristic simulation approach, analytical proof of convergence to a user equilibrium cannot be provided but empirically, convergence with an equilibrium solution can be provided by the Rgap function, measuring the distance between the current solution and an ideal equilibrium solution [12, 15]. A small value of Rgap expresses equilibrium in the network close to the Dynamic User Optimal ([11-13]).

\[
Rgap(r) = \frac{\sum_{i \in I} \sum_{k \in K_i} h_k^n(r) \cdot (s_k^n(r) - u_i^n(r))}{\sum_{i \in I} g_i^n \cdot u_i^n(r)}
\]

Where \( u_i^n(r) \) are the travel times on the shortest paths for the \( i^{th} \) OD pair at time interval \( n \) for iteration \( r \), \( s_k^n(r) \) is the travel time on path \( k \) at time interval \( n \) for iteration \( r \), \( h_k^n(r) \) is the flow on path \( k \) at time \( n \) for iteration \( r \), \( g_i^n \) is the demand for the \( i^{th} \) OD pair at time interval \( n \), \( K_i \) is the set of paths for the \( i^{th} \) OD pair, and \( I \) is the set of all OD pairs. Rgap value for a simulation is the maximum Rgap observed during the whole time period \( n \) of the simulation. Practically speaking, an Rgap of around 10-15% is considered as satisfactory (as the difference in the travel time perception by drivers).

Concerning traffic assignment, the Aimsun mesoscopic simulator performs Dynamic User Equilibrium assignment by iteration ([18]). The approach taken in Aimsun meso to solve the dynamic equilibrium problem assumes that, according to Friesz et al. ([13]), it can be formulated in the space of path flows \( h_k(n) \), for all paths \( k \in K_i \), where \( K_i \) is the set of feasible paths for the \( i^{th} \) OD pair at time \( n \). The path flow rates in the feasible region \( \Omega \) satisfy at any time \( n \in (0,N) \) the flow conservation and non-negativity constraints ([11, 15]):

\[
\Omega = \{ h(n) \mid \sum_{k \in K_i} h_k(n) = g_i(n), i \in I ; h_k(n) \geq 0 \} \text{ for almost all } n \in (0,N)
\]  

(1)

Where \( I \) is the set of all OD pairs in the network, \( N \) is the time horizon, and \( g_i(n) \) is the fraction of the demand for the \( i^{th} \) OD pair during the time interval \( n \). The approach assumes that the optimal user equilibrium conditions can be defined as a temporal version of the static Wardrop user optimal equilibrium conditions, which can be formulated as:

\[
S_k(n) \left\{ \begin{array}{ll}
= u_i(n) & \text{if } h_k(n) > 0 \\
\geq u_i(n) & \text{Otherwise}
\end{array} \right.
\]

\[
u(n) = \min_{k \in K_i} \{ S_k(n) \}
\]

\[
\text{for } \forall k \in K_i, \forall i \in I, \text{ for almost all } n \in (0,T) h_k(n) \in \Omega
\]  

(2)

Where \( S_k(n) \) is the path travel time on path \( k \) determined by the dynamic network loading. Friesz et al., show that these conditions are equivalent to the variational inequality problem consisting on finding \( h^* \in \Omega \) such that:
This problem is usually solved numerically discretising the time horizon $N$ into discrete time periods $n = 1, 2, ..., \left\lfloor \frac{N}{\Delta n} \right\rfloor$ of length $\Delta n$, corresponding to equilibrium flows according to (1) and (2) where the feasible flows $h_k(n)$ are the solution of the discretisation of (3):

\[ \sum_{K \in K} S_k(n)[h_k(n) - h_k'(n)] \geq 0 \]  

Where $K = \bigcup_{i \in I} K_i$ is the set of all paths for all OD pairs. The approach taken in the Aimsun mesoscopic simulator analytically solves this problem at each time interval by the Method of Successive Averages (MSA, [11, 16]). Once the paths and the path flows for the current time interval have been calculated, the dynamics of the flows in the link are simulated according to the approach described above.

### OD adjustment problem

Least square approaches present an interesting compromise between the inputs utilization and the solving complexity for OD adjustment problems. The use of this approach to solve the dynamic OD estimation linear problem was originally proposed by Cascetta ([9]). The least square formulation has been carried out in a similar manner to the process described in [6]. In this paper, Bierlaire and Crittin build on their modelling framework by exploiting the Ashok and Ben-Akiva ([1]) proposal of using deviation as a state variable. The analysis period is divided into equal intervals $n = 1, ..., N$. The network is modelled by directed graph ($\varnothing, B$), where $\varnothing$ is the set of nodes and $B$ is the set of links. TOD is the number of OD pairs considered. $x_n$ is the actual OD table capturing all trips departing during time interval $n$ and $x^H_n$ is the associated historical OD table. The vector of deviations is denoted by $\partial x_n = x_n - x^H_n$. TC is the number of link $b$ from $B$ equipped with sensor $d$ able to count the number of vehicles during the time period. $y_{dn}$ is the number of vehicles crossing sensor $d$ during time interval $n$ and $y_n$ the vector gathering all such counts. In our case, apriori (or initial) OD matrices are considered historical data and the estimation works off-line, without prediction of the further states. The model is defined by two equations, which model the evolution of the OD flows. The Transition equation captures the dynamic of the system. It is based on an auto-regressive process on the OD flows deviation:

\[ \partial x_n = \sum_{p=1}^{n-1} f^{p}_{n} \cdot \partial x_p + w_n \]  

Where $f^{p}_n$ describes the effect of $x_p$ on $x_n$ and $w_n$ is a random error. $q'$ is the number of lagged OD flow assumed to affect the OD flow in interval $n + 1$. The following assumption is made on $w_n$, the vector of random variables capturing the error:
\[
E[w_n] = 0 \\
E[w_n w_n'] = Q_n \delta_{nt}
\]

Where \(Q_n\) is a \([TOD, TOD]\) variance-covariance matrix and \(\delta_{nt}\) is the Kronecker symbol.

The measurement equation maps the state variable onto data. It captures the relationship between the state variable (OD deviation) and the measurements (traffic counts):

\[
y_n = \sum_{p=n-p'}^n \alpha_n^p \cdot x_p + v_n
\]  
(6)

Where \(y_n \in \mathbb{R}^{TC}\) contains the traffic counts data from time period \(n\), \(\alpha_n^p\) \([TC, TOD]\), the assignment matrix is the fraction of the \(i^{th}\) OD flow that departed from its origin during interval \(p\) and is on sensor \(d\) during interval \(n\) (this matrix is spare because not all the OD flows are captured by all sensors on the network at each time interval). \(p'\) is the maximum number of time intervals taken to travel between any OD pair in the network. \(v_n\) is the vector of random variable capturing the error measurement on traffic count data during time period \(n\). The following assumption is made on \(v_n\):

\[
E[v_n] = 0 \\
E[v_n v_{nt}'] = R_n \delta_{nt}
\]

Where \(R_n\) is a \([TC, TC]\) variance-covariance matrix and \(\delta_{nt}\) is the Kronecker symbol.

The following equation presents the formulation based on deviation:

\[
\partial y_n = \sum_{p=n-p'}^n \alpha_n^p \cdot \partial x_p + v_n
\]  
(7)

Where \(\partial y_n = y_n - \sum_{p=n-p'}^n \alpha_n^p \cdot x_p^H\)

In these equations, at each time interval \(n\), \(f_n^p\), \(x_p\) and \(y_n\) could be extracted from inputs of the chosen simulator and \(\alpha_n^p\) could be calculated using traffic assignment of the DUE simulation.

The implementation of the least square formulation has been carried out in a similar manner to the process described in [7]. From this least square formulation, the approach could be solved using the LSQR approach (based on [5, 6] works) to achieve 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-diagonalisation procedures ([7, 17]). It generates a sequence of output which monotonically decreases the associated sequence of the residual's norms. It is a global approach dealing with all time periods at the same time. In this way, all OD tables for all time intervals within the considered horizon must be included in a state vector.

Full details of the algorithm implementation in the Aimsun plug-in are presented in [7].
The LSQR method involves two distinct algorithmic layers:

- The outer layer handles the status of the variables taking into account their bounds. This layer process called Major Iteration declares each variable "free" if it is strictly between its bounds or "fixed" if it is at one of its bounds (using the Inexact Line Search loop).
- The inner layer solves the problem by a conjugate gradient algorithm on "free" variables only, at every major iteration. This layer, called Minor Iteration, keeps the "fixed" variables constant.

### Network and data used

For a practical assessment of the method, the centre of the city of Lausanne (Switzerland) has been chosen. This is a 2.5 km x 2.5 km (6.25 Km²) perimeter area representing a dense network where all the roads and signals have been considered. The OD matrix size is 60 x 60 (3600 OD pairs). About 35 traffic counts will be used to adjust the demand (ensuring that the majority of OD paths are covered - 95% of the flows), using the approach developed in [4]).

To evaluate the performance of the proposed methodology, the true/reference OD matrix (corresponding to the traffic count values used) has been deformed (random multiplication of each cell by a number between 0.7 and 1.1) and used as the initial OD demand. The process then modifies this initial matrix based on traffic count, and the result is compared with the reference matrix.

### Results

The results obtained, using the proposed methodology based on the mesoscopic assignment and the LSQR algorithm for OD adjustment are satisfactory (see Figure 2). After only a few iterations of the global bi-level process (less than 10), stabilisation is observed close to the reference target demand. ME graphs present an error of 10% (0.1/0.9) on OD flow compared to reference matrix and 6% on traffic counts (10/170). These good results are similar for all time intervals of the study period (Full results could be found in [4]).

In these graphs, the distance between the reference matrix/traffic count and those estimated (for each time period) are evaluated using: Mean Square Error (MSE) indicator and Mean Error (ME) indicator.

\[ \text{MSE} = \frac{\sum_{E} (E_{\text{sim}} - E_{\text{ref}})^2}{\#E} \]

\[ \text{ME} = \frac{\sum_{E} (E_{\text{sim}} - E_{\text{ref}})}{\#E} \]

Where: \#E, the number of measurements, \( E_{\text{sim}} \) data from the simulation and \( E_{\text{ref}} \) reference data (initial or real data). The MSE indicator gives information about the
square of the error which penalizes large errors and ME give more information about the sign of the error.
Moreover, a sensitivity analysis is carried out on the number and the "quality" of traffic counts used for flow adjustment. Indeed, the number of detectors and vehicle intercepted can be reduced to assess the approach. The goal is to compare results obtained using different detection layout configurations. Added to the initial configuration of 35 traffic detectors layout, three other configurations are evaluated considering:

- Traffic detectors with 25 highest flow (called 25 TC)
- Traffic detectors with 15 highest flow (called 15 TC)
- Traffic detectors with 15 lowest flow (called 15 TC Low)
FIGURE 3 - Results summary X-Y OD flows plots, detection layout configuration

Figure 4 shows the coefficient of determination obtained from the clouds of dots presented in the previous figure. They are a global value for all OD flows for all time intervals.
Figure 3 and Figure 4 show that the methodology manages to satisfactorily fit the different traffic counts used by the configurations. The OD adjustment based on the 15 highest detector flows induces a satisfactory fitting of the OD flows. Using a greater number of traffic counts leads to observed "noise" on small OD pair flows. Using the 15 lowest traffic count (15 TC L) flows for adjustment leads to poorer results due to the un-interception of the main and most influential OD pair flows. These results help to conclude that a smaller number of traffic counts produces more accurate and fitting results. This observation is true for 25 TC and 15 TC but not for 15 TC Low. It demonstrates the limitation of the minimum quality required for the detection layout to perform an OD estimation.

Based on the results of this sensitivity analysis, this process has demonstrated its great sensitivity to the quality of the detection layout configuration. Therefore, particular attention must be paid to this traffic counts input.

**Conclusion**

The methodology developed is innovative in various aspects. It is comprised of a heuristic bi-level approach. Assignment of the initial demand is performed by a mesoscopic simulation based on the Dynamic User Equilibrium to model detailed dynamic traffic patterns, and OD flow adjustment is executed by an efficient least square solution, which takes into account dynamic aspects of the flow propagation and traffic counts. Based on the different outputs and indicators selected, this method demonstrates robustness and consistency. Moreover, this research has shown the importance of input data for the OD estimation process and mainly the detection layout configuration used for traffic count data. Indeed, OD flows must be intercepted by detectors to be estimated. Therefore, the choice of the vehicle intercepted is crucial to
adjust most influential flows. Sensitivity analysis has shown that adding detectors does not necessarily improve the OD estimation output and that flows of little importance could be ignored in order to increase the efficiency of the detection layout configuration and decrease the computation time.

References


