A Dynamic Estimation of Passenger OD Matrices based on space-state models

Lídia Montero, Jaume Barceló, Esteve Codina
TECHNICAL UNIVERSITY OF CATALONIA- UPC
Statistics and Operations Research Department-PROMALS Group
Carrer Jordi Girona 1-3 08034 BARCELONA-SPAIN
{lidia.montero, jaume.barcelo, esteve.codina}@upc.edu

The corresponding author: Lídia Montero

Abstract
The primary data input used in transit models comes from Origin-Destination (OD) passenger matrices, which describe the patterns of commuters across the network. In this way, OD matrices become a critical requirement in Transport Management and/or Information. Dynamic Transit Assignment models are a research topic, but dynamic OD passenger matrices are a critical aspect in applications. However, OD matrices, either of trips or passengers, are not directly observable and the current practice consists for static estimation consists on adjusting an initial or seed matrix from boarding/link/segment passenger counts which are provided by surveys or by automatic fare systems in a daily-basis or peak-period point of view. A space-state model for dynamic passenger matrices estimation has been formulated. The Kalman filter approach incorporates the emerging Information and Communication Technologies (ICT data), especially those based on the detection of the electronic signature of on-board devices provide a rich source of data that can be used in space-state models to simplify the underlying formulation, by reducing the dimensionality of the state vector and allowing a linear mapping between state variables and measurements. The formulation for dynamic passenger OD matrix estimation proposed has been derived from an originally one developed for auto trip matrices, but in this paper, we present the adapted approach to the estimation of dynamic OD matrices in public transport networks.

Keyword: Dynamic OD matrix estimation, Information Systems, Transit Management, Kalman Filtering

1. Introduction
When a public transportation network is the object of study, Origin-to-Destination passenger matrices describe the number of passengers between OD pairs or origin-to-destination stations. Either under the static or the dynamic scope, Origin-Destination (OD) matrices are a major data input for describing the demand. All formulations of static traffic or transit assignment models (Florian and Hearn [1]), as well as dynamic models involved in Advanced Traffic Control and Management (see Ashok et al. [2]) assume that a reliable estimate of an OD is available. In the private transport context for demand estimation, Origin-to-Destination trip matrices describe the number of trips between any origin-destination pair of transportation zones in a study area. In other words, they describe the path flows or the path flow
proportions, depending on whether we refer to the number of trips using a path or to the fraction of trips using a path, with respect to the total number of trips between the corresponding origin and destination. Route choice proportion can vary depending on the time-interval in dynamic models, since the traffic state and the temporal dimension are considered.

Applications of Automated Data Collection Systems (ADCS) to transport planning with a focus on transit Origin-Destination (OD) inference have been proposed by several authors ([3,4]). However, OD matrices are not yet directly observable, even less so in the case of the time-dependent OD matrices; consequently, it has been natural to resort to indirect estimation methods. These indirect estimation methods are the so-called matrix adjustment methods, whose main modeling hypothesis can be stated for transit networks as follows: if the number of passengers in the segments of the transit line itineraries defined in a network are the consequence of the assignment of an OD matrix onto a network, then, if we are capable of measuring passengers by segment (or for the total number of passengers in all the line-segments sharing a network link), the problem of estimating the OD matrix that generates such line loading can be considered as the inverse of the assignment problem.

Kostakos et al [3] proposed the use of passengers’ Bluetooth mobile devices to derive passenger OD matrices in a simplified context. A Bluetooth device set to discoverable mode must respond to a discovery request by transmitting its unique Bluetooth identifier (12 hex digits) and device class (6 hex digits), since a Bluetooth scanner located on vehicle-units constantly scans the presence of the various devices it encounters (along with the date and time). Extending this idea, passengers’ (Bluetooth) smartphones might be captured by Bluetooth of Wi-Fi detection located at Road Side Units (RSU) (intersections or bus-stops) and be transmitted by a Wifi protocol in real-time to a Central Process Center. This ICT sensor for data collection provide for the authors with two classes of data: primary data, the identity of the devices, the position at which the device is detected and the detection time, but our aim is making use of this new data to estimate real-time dynamic Origin/Destination matrices of passengers in the transit network.

2. Formulation proposal for passenger trips

Some previous considerations to the model statement are:

- The demand matrix for the period of study is assumed to be divided into several time-slices, accounting for different proportions of the total number of passengers in the time horizon.
- The approach assumes an extended space state variable for M+1 sequential time intervals of equal length Δt (between 5 and 10 minutes for passenger’s matrices), in order to consider non-instantaneous travel times as in the private case. M should guarantee to traverse the network (at least in 90-95% of the trips according to our experience).
We propose to use deviates of state variables since such model formulations indirectly take into account all the available a priori structural information, and additionally making the transformed variables easier to follow a normal distribution as space-state-models based on Kalman filtering require.

Only origin-destination passenger trips for equipped passenger in transit lines whose stops are covered with ICT sensors are observable. Interferences with not covered transit-lines in some stops are not considered in this first approximation.

The solution should provide estimations of the OD passenger matrices between transit stops for each time interval up to the \( k \)-th interval once observations of BT equipped passengers at the bus-stops equipped with wifi antennas upon to the \( k \)-th interval are available.

Historic profiles (for day-type and time-period) have to be used to expand BT samples of equipped passengers to the total number of passengers.

Strategies for transit trips can be computed by any transportation planning software that includes equilibrium transit assignment. The transit network has to be modeled and the most-likely transit strategies according to optimal transit assignment for a given historic demand for the period of study computed. The mapping from the most-likely transit strategies for transit trips from centroid \( i \) to \( j \) going through ICT sensors can be systematically programmed in any language, including MatLab.

State variables are noted as \( \Delta g_{ije}(k) \) and are defined as deviations of OD passenger flows on strategy \( e \) from origin \( I \) to destination \( j \) \( g_{ije}(k) \) relative to historic OD passenger flows on strategy \( e \) \( \tilde{g}_{ije}(k) \) for equipped passengers. If dynamic historical OD matrices are not available, then the formulation reduces to the case where state variables are directly the equipped passenger OD flows on strategies. The total number of OD passenger flow on strategies \( G_{ije}(k) \) for all passengers (equipped or not) are computed according to expansion factors \( Q_i(k)/q_i(k) \) provided by historic profiles.

The total number of origin and/or destination centroids is \( I \) (bus-stops), identified by index \( i, i = 1,..,I \); the total number of ICT sensors is \( Q \), identified by index \( q, q = 1,..,Q \), where \( Q = I+P \), I ICT sensors located at bus-stops and \( P \), ICT sensors located at segments in the inner network; and the total number of most likely used strategies between origins and destinations is \( K \). Each equipped transit stop might be considered either as an origin or as a destination, and models a transit-stop that might be shared by several transit lines.

The state variables \( \Delta g_{ije}(k) \) are assumed to be stochastic in nature, and OD strategy flow deviates at the current time \( k \) are related to the OD strategy flow deviates of previous time intervals by an autoregressive model of order \( r << M \); the state equations are:

\[
\Delta g(k+1) = \sum_{w=1}^{r} D(w) \Delta g(k-w+1) + w(k) \quad (1)
\]
Where \( w(k) \) are zero mean with diagonal covariance matrix \( W_k \), and \( D(w) \) are \( IJK \times IJK \) transition matrices which describe the effects of previous OD deviates \( \Delta g_{ij}(k-w+1) \) on current flows \( \Delta g_{ij}(k+1) \) for \( w = 1, \ldots, r \). In our research, we assume simple random walks to provide the most flexible framework for state variables, if no convergence problems are detected. Thus, our first trial will be \( r=1 \) and \( D(w) \) matrix becomes the identity matrix.

The relationship between the state variables and the observations involves *time-varying model parameters* (congestion–dependent, since they are updated from sample travel times provided by equipped passengers) in a linear transformation that considers:

- The number of equipped passengers accessing transit-stops during time intervals \( k, \ldots, k-M, q(k) \).
- \( H < M \) *time-varying model parameters* in form of fraction matrices, \( u_{ijeq}^h(k) \).

The \( H \) adaptive fractions that approximate \( u_{ijq}^h \) and \( u_{ijeq}^h \) are updated from measures provided by ICT sensors or historic distributions if real-time data is not available. Direct on-lines samples of travel times would allow the updating of discrete approximations of travel time distributions. At time interval \( k \), the values of the observations are determined by those of the state variables at time intervals \( k, k-1, \ldots, k-M \).

\[
\Delta z(k) = F(k)\Delta g(k) + v(k) \quad (2)
\]

Where \( v(k) \) are, respectively, white Gaussian noises with covariance matrices \( R_k \). \( F(k) \) maps the state vector \( \Delta g(k) \) onto the current blocks of measurements at time interval \( k \): counts of equipped passengers at ICT sensors and accesses at transit-stops (centroids), accounting for time lags and congestion effects. Deviate counts at \( k \) mean the observed counts minus the historical demand \( \tilde{g}_{ij}(k) \) counts, given the current traffic and transit conditions (according to *time-varying model parameters*).

The formulation approach is being programmed as a MatLab prototype in the near future. Simulation data to test the approach has to be provided to emulate a real system.

### 3. Acknowledgments

This research is funded by project TRA2011-27791-C03-02 of the Spanish R+D National Programs and has benefited from participation in EU COST Action TU0903 MULTITUDE – Methods and tools for supporting the Use caLibration and validaTIon of Traffic simUlation models.
4. References


