Adapting a Dynamic OD Matrix Estimation Approach for Private Traffic based on Bluetooth data to Passenger OD Matrices

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 principal traffic models comes from Origin-Destination (OD) trip matrices, which describe the patterns of commuters across the network. In this way, OD matrices become a critical requirement in Advanced Transport Control and Management and/or Information Systems that are supported by Dynamic Traffic Assignment models (DTA models). Dynamic Transit Assignment models are a research topic, but once a dynamic transit assignment be available to practitioners, the problem of estimating the time-dependent number of trips between transportation zones shall be a critical aspect for real applications. However, OD matrices are not directly observable, neither for private nor public transport, and the current practice consists on adjusting an initial or seed matrix from link/segment counts which are provided by counting stations or data gathering in the field (detection layout). The emerging Information and Communication Technologies, 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 for dynamic matrix estimation. We present a linear Kalman filter approach that makes use of counts of passengers and travel times provided by Bluetooth devices to simplify an underlying space-state model. The formulation for dynamic passenger OD matrix estimation proposed was originally developed for auto trip matrices, but in this paper, we explore the possibility of adapting the approach to the estimation of OD matrices in public transport networks.

Keyword: Applied Science, Information Systems, Advanced Traffic Management, Kalman Filtering

1. Introduction

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. For private vehicles, route choice models describe how trips select the available paths between origins and destinations and, as a consequence, the number of trips using a given path in private transportation modes. 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. When a public transportation network is the object of study, Origin-to-Destination trip 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. Loop detectors are traditionally the means for measuring fundamental traffic variables (i.e. flows, speeds and occupancies), whose values determine the state of the traffic system; but, now ICT sensors (Information and Communication Technologies) provide a new source of data for a sample of commuters.

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 trips 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' mobile devices 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 to explore the possibility of making use of this new data to estimate real-time dynamic Origin/Destination matrices of passengers in the transit network.

2. Previous research

We have been working with counts and travel times provided by Bluetooth devices [5,6,7] for the estimation of dynamic OD matrices for commuters driving their own car on freeways, corridors and finally in urban networks. The space-state formulations based on Kalman Filtering (KF) have always been an appealing approach to the estimation of time dependent OD trips matrices and we developed several KF formulations during the last 3 years. We started with a KF formulation suitable for linear freeways where flow counts at entry ramps were required and Bluetooth equipped vehicles were identified by antennas (in the following ICT sensors) located on some freeway sections and exit ramps, equipped vehicles provide accurate travel time data from their entry point up to the freeway points they go through and are detected by ICT sensors [6]. State variables in the first KF formulation were stated as OD cars between the entry and the exit ramps for time-interval and observations were considered as the number of vehicles detected by ICT sensors for each time-interval and the total number of vehicles entering by each ramp per interval. A common dynamic horizon is defined to be 1.5 to 3 hours and it is divided into time-intervals of length between 1 to 5 minutes. Only one OD path for each OD pair is possible in linear structures, as freeways. Travel times from equipped vehicles from the origin entry ramp to ICT sensors **provide measures that allow to simplify the underlying KF formulation** that relates state variables and observations when congestion effects of traffic dispersion are taking into consideration, since traffic congestion variables do not need to be longer considered as extra state variables, but time-varying model parameters provided by measures. Linear KF formulations dealing with congestion are possible in freeways.

The estimation of time-dependent OD trip matrices in urban networks is a more complex problem given the existence of alternative paths between each OD pair, making that route choice becomes relevant. In our current experience [6,7], we propose an extended formulation suitable for urban networks, and for freeways as a particular case of a network, where state variables are defined as (deviates with respect to a priori historical values as suggested by Ashok and Ben-Akiva [2]) the number of OD equipped vehicles for interval along the most-likely used paths for each OD pair. Origin and destinations are no longer entry/exit ramps, but centroids in the study area and gates in a cordon. The measurements are the link flow counts on traditional detectors (usually at gates in a cordon and in a subset of network links), link flow counts of equipped vehicles in a subset of links where ICT sensors are located and travel times of equipped vehicles between ICT sensors. A basic hypothesis is that equipped and non-equipped vehicles follow common OD patterns. Formulations using deviates provide benefits in respect to those using OD path flows as state variables, because they allow to incorporate more historical data as *a priori* structural information in the model and the convergence of the algorithmic proposal is improved.

As a consequence of the experience gained, we implemented in MatLab, as described in [7], a linear Kalman filtering prototype that considers deviations of OD path flows as state variables, being calculated the most-likely used paths according to DUE-based Historic OD path flows. Our approach differs to other published proposals [2,8] in the fact that we do **not require an assignment matrix indicating the fraction of each OD pair traveling in a link for a time-interval**. DUE (Dynamic User Equilibrium) is conducted with the historic OD flows, and the number of paths that are taken into account is a design parameter (currently, only paths accounting for more than 5% of the total OD flow are considered, those being the *most important* according to the implications of DUE behavior). DUE OD path proportions for the most-likely OD paths and time-slices are not an input to the KF prototype, only the description of the most likely OD paths (and thus, a limited and usually small number of paths) in the form of a sequence of links. A list of paths going through each sensor is automatically built for each ICT sensor from the OD path description, ICT sensor location and the network topology. In this way, once an equipped car is detected by ICT sensor *j*, the travel time from its entry point to sensor *j* is available and it is used for updating *time varying model parameters* that affect travel times on OD paths (state variables) included in the list and thus, congestion effects are incorporated in the formulation (as measures, not as state variables).

We model the *time-varying dependencies* between measurements (counts of equipped vehicles) and state variables (deviates of equipped OD path flows in the subset of the most-likely paths), adapting an idea of Lin and Chang [9], for estimating discrete approximations to travel time distributions. Since our approach exploits the ICT travel time measurements from equipped vehicles, this has advantages that constitute a major contribution since sampled travel times are used to estimate discrete travel time distribution (H bins are used for adaptive approximations). Travel times collected from ICT sensors are incorporated into the proposed model [7] and it is not necessary that vehicles reach their destination, since at any intermediate sensor that they pass through, the measured travel time from the entry point to the network to the ICT sensor updates the discrete travel time approximations. Therefore, completed trips are not the only source for updating time-varying model parameters, but information about the exact trajectories of equipped vehicles is not used in our prototype. Expansion factors from OD path flows of equipped vehicles to total vehicles, in a given interval, can be estimated by using the inverse of the proportion of ICT counts to total counts at the corresponding origin-centroid; expansion factors are assumed to be shared by all OD paths and pairs with a common origin-centroid and initial interval. If total counts are not available than a generic expansion factor that depends on the penetration of the technology in the area of study must be used.

The Matlab code prototype is called KFX and it is able to handle large sites according to our recent computational experiments. We tested the prototype for estimation of dynamic OD matrices of trips by simulation (AIMSUN models [10] were built and API code was developed to emulate measures for BT equipped vehicles). Several scenarios according to technology penetration and parameter setting for the prototype have been considered and reported [8] in the following network tests:

- A urban freeway in Barcelona consisting on a 11.551-km-long section of the Ronda de Dalt, between the Trinitat and the Diagonal Exchange Nodes. The site has 11 entry ramps and 12 exit ramps (including main section flows) on the section being studied, which flows in the direction of Llobregat (to the south of the city).
- Amara District is a urban network with 232 links and 85 OD pairs, with a rich structure of alternative paths between OD pairs, totaling 358 paths according to the DUE for the selected demand matrix. The detection layout consists of 48 ICT sensors. Simulation horizon was set to 1h and time-intervals to 90 sec.

- Eixample's network is shown in Figure 1. It is Barcelona's CBD district consisting of 2108 links and 877 OD pairs, and according to DUE *2366 most likely OD paths* in the morning period. The detection layout has 281 ICT link detectors. The horizon of simulation is 1h15min and subintervals are set to 180 sec. KFX prototype needs less than 15 sec of CPU per iteration (time-interval) under Windows 7 – 64 bits – 8 GB RAM -Intel Core i7-2600 (8M, ,3.40 GHz) 4C/8T, once KFX parameters are properly set.

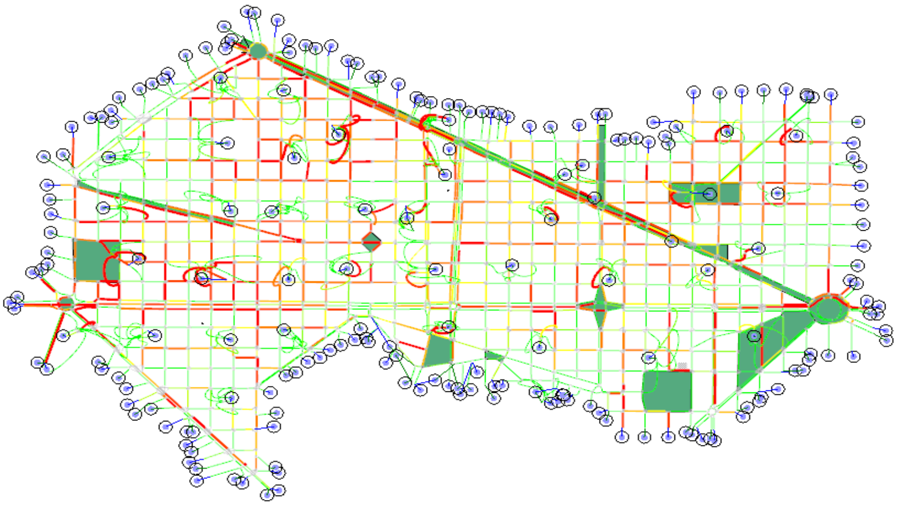

Figure 1. Eixample Network: peak morning congestion

The proposed linear KF approach to dynamic OD matrix estimation provides good estimates of target values in the synthetic simulation tests in network and freeway sites. ICT data simplifies the dynamic estimation of OD matrices by a Kalman-Filtering approach because the problem is formulated as a linear filter and reduces the computational burden when compared to well-known formulations in the literature that use Extended Kalman Filter [8]. The formulation takes into account network topology, multiple paths between OD pairs and the *most likely used paths* according to Dynamic User Equilibrium models (see [8] for details). It will be relevant to field test the approach in the forthcoming pilot project and define the product and interfaces to make KFX user friendly for practitioners.

3. 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 trips for equipped passengers 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 notation for the proposed formulation is the following:

$\tilde{Q}_i(k)$, $\tilde{q}_i(k)$	:	Historic total number of passengers and BT equipped passengers accessing a transit unit at any stop inside the transportation area modeled by centroid *i* at time interval k.
$Q_i(k)$, $q_i(k)$	:	Total number of passengers and BT equipped passengers accessing a transit unit at any stop inside the transportation area modeled by centroid *i* at time interval k.

$\tilde{y}_q(k)$, $y_q(k)$: Historic and actual number of equipped passengers crossing sensor q at time interval k

$G_{ije}(k)$, $\tilde{G}_{ije}(k)$, $g_{ije}(k)$, $\tilde{g}_{ije}(k)$: Total number of current $G_{ije}(k)$ and historic $\tilde{G}_{ije}(k)$ passengers as well as current $g_{ije}(k)$ and historic $\tilde{g}_{ije}(k)$ equipped passengers accessing at centroid i at time interval k headed towards j using strategy e.

$\Delta g_{ije}(k)$: ***State variables*** are deviates of equipped passengers accessing at centroid i during interval k headed towards centroid j using strategy e with respect to average historic data $\Delta g_{ije}(k) = g_{ije}(k) - \tilde{g}_{ije}(k)$.

$\mathrm{z}(\mathrm{k})$, $\tilde{\mathrm{z}}(\mathrm{k})$: The *current and historic measurements* of equipped passengers during interval k, a column vector of dimension P+I, whose structure is $z(k)^T = (y(k) \quad q(k))^T$

$u_{iq}^h(k)$: Fraction of equipped passengers that require h time intervals to reach sensor q at time interval k that accessed the network at centroid i (during time interval $[(k-h-1)\Delta t \quad , \quad (k-h)\Delta t]$).

$u_{ijeq}^h(k)$: Fraction of equipped passengers detected at interval k whose trip from centroid i to sensor q might use the OD strategy e and lasts h time intervals of length Δt to arrive from centroid i to sensor q, where i = 1,…,I , j = 1,…, J , $h = 1 \ldots M$, $q = 1 \ldots Q$

$\bar{t}_{iq}(k)$: Average measured travel time for equipped passengers accessing at centroid i and crossing sensor q during interval k

$u_{iq}^h(k)$ and $u_{ijeq}^h(k)$ are the time-varying model parameters. The values of $u_{iq}^h(k)$ are provided by the measurements of the ICT sensors.

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:

$$\Delta \mathrm{g}(\mathrm{k}+1) = \sum_{w=1}^{r} \mathrm{D}(\mathrm{w}) \Delta \mathrm{g}(\mathrm{k}-\mathrm{w}+1) + \mathrm{w}(\mathrm{k}) \quad (1)$$

Where w(k) are zero mean with diagonal covariance matrix $\mathbf{W_k}$, and $\mathbf{D(w)}$ are IJKxIJK transition matrices which describe the effects of previous OD deviates $\Delta g_{ije}(k-w+1)$ on current flows $\Delta g_{ije}(k+1)$ for w = 1,…,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 $\mathbf{D}(\mathbf{w})$ 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,..., k-M, $q_i(k)$.
- H<M *time*-varying *model parameters* in form of *fraction matrices,* $\left[u_{ijeq}^{h}(k)\right]$.

The H adaptive fractions that approximate u_{iq}^{h} and u_{ijeq}^{h} are updated from measures provided by ICT sensors. Direct samples of travel times 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, …k-M.*

$$\mathbf{\Delta z}(\mathbf{k}) = \mathbf{F}(\mathbf{k})\Delta\mathbf{g}(\mathbf{k}) + \mathbf{v}(\mathbf{k}) \quad (2)$$

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

4. Conclusions

Adapting a previous research on space-state formulation to the estimation of dynamic transit matrices between stops seems feasible and consistent. We make use of ICT data and take benefit of the previous experience of the authors in Kalman filters, by using OD historic matrices as a priori information, and adjusting it from available dynamic passenger counts and travel times from ICT data.

The formulation approach is going to be programmed as a MatLab prototype in the near future. Simulation data to test the approach seems difficult to emulate since microscopic simulators do not consider transit networks from the transit assignment point of view and no optimal behavioral rules for transit trips are included in available simulators to model the choice of current strategies by passengers.

5. Acknowledgments

This research is funded by projects MITRA (TRA2009-14270) and 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.

6. References

[1] Florian M. and Hearn D., Network Equilibrium Models and Algorithms, Chapter 6 in: M.O. Ball et al., Eds. (1995), Handbooks in OR and MS, Vol.8, Elsevier Science B.V.

[2] Ashok, K. and Ben-Akiva, M., *Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin-Destination Flows*, Transportation Science 34(1), 2000, pp. 21-36.

[3] Kostakos, V., Camacho, T. and Mantero, C. Wireless detection of end-to-end passenger trips on public transport buses. In proceedings of the IEEE Conference on Intelligent Transportation Systems, Funchal, Portugal, (2010). pp. 1795-1800.

[4] Wang, W., Attanucci, J. P. and Wilson, N. H.M., Bus *Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems*. Journal of Public Transportation, Vol. 14, No. 4, 2011, pp 131-150.

[5] Lídia Montero, Jaume Barceló, Laura Marquès, Carlos Carmona. *Dynamic Origin-Destination Trip Matrix Estimation In Urban Networks Based On A Kalman Filtering Approach And New ICT Sensor Data*. Presented at 19th IFORS 2011 - Melbourne – Australia 10-15 July 2011 .

[6] Barceló J.,.Montero L, Marqués L. and Carmona C., *Travel time forecasting and dynamic of estimation in freeways based on Bluetooth traffic monitoring*, Transportation Research Records: Journal of the Transportation Research Board, 2010, Vol. 2175, pp. 19-27.

[7] Barceló, J., Montero, L., Bullejos, M., Serch, O. and Carmona C., *A Kalman Filter Approach for the Estimation of Time Dependent OD Matrices Exploiting Bluetooth Traffic Data Collection*, (2011), submitted to the Journal of Intelligent Transport Systems (J-ITS) (on revision process).

[8] Antoniou, C., Ben-Akiva, M. and Koutsopoulos H.N., Nonlinear Kalman Filtering Algorithms for On-Line Calibration of Dynamic Traffic Assignment Models, (2007), IEEE Transactions on Intelligent Transportation Systems Vol. 8 No.4.

[9] Lin, P., Chang, G., *A generalized model and solution algorithm for estimation of the dynamic freeway origin-destination matrix*, (2007), *Transportation Research B 41 554-572*.

[10] AIMSUN 7.0 Traffic Simulator, distributed by Transport Simulation Systems, 2011.