Dynamic OD Matrix Estimation Exploiting Bluetooth Data in Urban Networks

J.BARCELÓ, L.MONTERO, M.BULLEJOS, O. SERCH and C. CARMONA
Department of Statistics and Operations Research and CENIT (Center for Innovation in Transport)
Technical University of Catalonia,
Campus Nord, Jordi Girona, 29, 2A, Building Nexus II 08034 Barcelona, SPAIN
(jaume.barcelo, lidia.montero, manuel.bullejos, oriol.serch, carlos.carmona)@upc.edu

Abstract: - Time-Dependent Origin-Destination (OD) matrices are a key input to Dynamic Traffic Models. Microscopic and Mesoscopic traffic simulators are relevant examples of such models, traditionally used to assist in the design and evaluation of Traffic Management and Information Systems (ATMS/ATIS). Dynamic traffic models can also be used to support real-time traffic management decisions. The typical approaches to time-dependent OD estimation have been based either on Kalman-Filtering or on bi-level mathematical programming approaches that can be considered in most cases as ad hoc heuristics. The advent of the new Information and Communication Technologies (ICT) makes available new types of traffic data with higher quality and accuracy, allowing new modeling hypotheses which lead to more computationally efficient algorithms. This paper presents a Kalman Filtering approach, that explicitly exploit traffic data available from Bluetooth sensors, and reports computational experiments for networks and corridors.

Key-Words: - Dynamic OD Matrix, Kalman Filter, Advanced Traffic Management, Data Collection.

1 Introduction

We draw attention in this paper to a main requirement of ATIS/ATMS: the estimation of time-dependent Origin to Destination (OD) matrices from measurements of traffic variables. We assume that the usual traffic data collected by inductive loop detectors (i.e. volumes, occupancies and speeds) are complemented by accurate measurements of travel times and speeds between two consecutive sensors based on new technologies able to capture the electronic signature of specific on-board devices, such as a Bluetooth device on board a vehicle. The sensor captures the public parts of the Bluetooth or Wi-Fi signals within its coverage radius, the most relevant being the MAC address, whose uniqueness makes it possible to use a matching algorithm to log the device when it becomes visible to the sensor. A vehicle equipped with a Bluetooth device traveling along the freeway is logged and time-stamped at time t_1 by the sensor at location 1. After traveling a certain distance it is logged and time-stamped again at time t_2 by the sensor at location 2 downstream. The difference in time stamps $\tau = t_2 - t_1$ measures the travel time of the vehicle equipped with that mobile device. The speed is also measured, assuming that the distance between both locations is known.

Data captured by each sensor is sent to a central server by wireless telecommunications for processing. Raw measured data cannot be used without pre-processing aimed at filtering outliers that could bias the sample. The quality of the input data is crucial for the applications, therefore it is a subject of intense research. A variety of nonlinear filters have been proposed recently by Van Lint and Hoogendoorn [1] and Treiber et al, [2]. A refined version of the Kalman Filter (Kalman [3], proposed in Barceló et al. [4,5]) has been used in this paper.

2 Problem Formulation

The space-state formulations based on Kalman Filtering have always been an appealing approach to the estimation of time dependent OD matrices. In this paper we propose a recursive **linear** Kalman-Filter for state variable estimation that combines and modifies the earlier work of Chang and Wu [6], Hu *et al* [7], Choi *et al*, [8] and Van Der Zijpp and Hamerslag [9], adapting their models to take advantage of ***travel times and traffic counts*** collected by tracking Bluetooth equipped vehicles and conventional detection technologies.

Chang and Wu [6] proposed a model for freeways that, for each OD pair that estimates time-varying travel times, uses time dependent traffic measures and implicit traffic flow models to account for flow propagation. An Extended Kalman-filter approach is proposed to deal with the nonlinear relationship between the state variables and the observations.

Hu et al. [7] proposed an Extended Kalman Filtering algorithm for the estimation of dynamic

OD matrices in which time-varying model parameters are included as state variables in the model formulation. The approach takes into account temporal issues of traffic dispersion. Lin and Chang [10] proposed an extension of Chang and Wu [6] in order to deal with traffic dynamics, assuming that travel time information is available.

We assume flow counting detectors and ICT sensors located in a cordon and at each possible point for flow entry (*centroids* of the study area). ICT sensors are located at intersections in urban networks and cover access and links to/from the intersection. Flows and travel times are available from ICT sensors for any selected time interval length higher than 1 second. Trip travel times from origin entry points to sensor locations are measures provided by the detection layout. Therefore, they are no longer state variables but measurements, which simplify the model and makes it more reliable.

A basic hypothesis is that equipped and non-equipped vehicles follow common OD patterns. We assume that this holds true in what follows and that it requires a statistical contrast for practical applications. Expansion factors for everything from 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 centroids; expansion factors are assumed to be shared by all OD paths and pairs with a common origin centroid and initial interval.

We propose a linear formulation of the Kalman Filtering approach that uses deviations of OD path flows as state variables, as suggested by Ashok and Ben-Akiva[11], calculated in respect to DUE-based Historic OD path flows for equipped vehicles. But our approach differs in that we **do not require an assignment matrix**. We use instead the subset of the most likely OD path flows identified from a DUE assignment with Dynameq [12]. The DUE is conducted with the historic OD flows, and the number of paths to take into account is a design parameter, only the description of the most likely OD paths is needed. A list of paths going through the 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 OD paths (state variables) which are included in the list.

We model the time-varying dependencies between measurements (sensor counts of equipped vehicles) and state variables (deviates of equipped OD path flows), adapting an idea of Lin and Chang [10], for estimating discrete approximations to travel time distributions. The estimation of these distributions is made on the basis of flow models which induce nonlinear relationships that require extra state variables, leading to a non linear KF approach. Since our approach exploits the travel ICT time measurements from equipped vehicles, we can replace the nonlinear approximations by estimates from a sample of vehicles. This has advantages that constitute a major contribution of this paper because no extra state variables for modelling travel times and traffic dynamics are needed, since sampled travel times are used to estimate discrete travel time distribution (H bins are used for adaptive approximations). Additionally, travel times collected from ICT sensors are incorporated into the proposed model and it is not necessary that vehicles reach their destination, since at any intermediate sensor that they pass by the travel time measured from the entry point (centroid) to that sensor updates the discrete travel time approximation. No information about trajectories of equipped vehicles is used in this version.

The demand matrix for the period of study is divided into several time-slices, accounting for different proportions of the total number of trips in the time horizon.

The approach assumes an extended state variable for M+1 sequential time intervals of equal length Δt, M is the maximum number of time intervals required for vehicles to traverse the entire network in a congested scenario.

The solution provides estimations of the OD matrices for each time interval up to the k-*th* interval. State variables $\Delta g_{ijc}(k)$ are deviations of OD path flows $g_{ijc}(k)$ relative to historic OD path flows $\tilde{g}_{ijc}(k)$ for equipped vehicles. A MatLab prototype algorithm has been implemented to test the approach (named KFX).

2.1 Notation

The total number of origin centroids is I, identified by index i, i = 1,..,I; the total number of destination centroids J, identified by index j, j = 1,…,J; the total number of ICT sensors is Q, identified by q=1,…Q, where Q = I+J+P, I ICT sensors located at origins, J, ICT sensors at destinations and P, ICT sensors located in the inner network; and the total number of most likely used paths between origins and destinations is K. The notation used in this paper is defined in Table 1.

$\tilde{Q}_i(k)$ $\tilde{q}_i(k)$	:	Historic total number of vehicles and equipped vehicles entering from centroid *i* at time interval *k*
$Q_i(k)$, $q_i(k)$	:	Total number of vehicles and equipped vehicles entering from centroid *i* at time interval k.
$\tilde{y}_q(k)$, $y_q(k)$	:	Historic and actual number of equipped vehicles crossing sensor *q* at interval *k*
$G_{ijc}(k)$, $\tilde{G}_{ijc}(k)$, $g_{ijc}(k)$, $\tilde{g}_{ijc}(k)$	:	Total number of current $G_{ijc}(k)$ and historic $\tilde{G}_{ijc}(k)$ vehicles as well as current $g_{ijc}(k)$ and historic $\tilde{g}_{ijc}(k)$ equipped vehicles entering the network from centroid *i* at interval *k* headed towards destination *j* using path *c*.
$\Delta g_{ijc}(k)$	:	***State variables*** are deviates of equipped vehicles entering from centroid *i* during interval *k* headed towards centroid *j* using path *c* with respect to average historic flows $\Delta g_{ijc}(k) = g_{ijc}(k) - \tilde{g}_{ijc}(k)$
$\mathbf{z}(k)$, $\tilde{\mathbf{z}}(k)$	:	The *current and average historic measurements* during interval k, vector, $\mathbf{z}(k)^T = (\mathbf{y}(k) \quad \mathbf{q}(k))^T$
IJ, IJK	:	Number of feasible OD pairs and *most likely* OD paths depending on the zoning system defined in the network.
$u_{iq}^h(k)$	:	Fraction of vehicles that require *h* time intervals to reach sensor *q* at time interval k that entered the system from centroid *i* (during time interval $[(k-h-1)\Delta t \;,\; (k-h)\Delta t]$).
$u_{ijcq}^h(k)$	:	Fraction of equipped vehicles detected at interval k whose trip from centroid *i* to sensor *q* might use OD path (i,j,c) lasting *h* time intervals of length Δt to arrive from centroid *i* to sensor *q*.

Table 1. Notation

Conservation equations from entry points (centroids) are explicitly considered. Without $Q_i(k)$, a generic expansion factor has to be applied.

2.1.1 State Equations

Let $\Delta\mathbf{g}(k)$ be a column vector of dimension IJK containing the state variables $\Delta g_{ijc}(k)$ for each time interval k for all *most likely* OD paths (i,j,c). The state variables $\Delta g_{ijc}(k)$ are assumed to be stochastic in nature, and OD path flow deviates at current time k are related to the OD path flow deviates of previous time intervals by an autoregressive model of order r <<M; the state equations are:

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

Where $\mathbf{w}(k)$ 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 path flow deviates $\Delta g_{ijc}(k-w+1)$ on current flows $\Delta g_{ijc}(k+1)$ for w = 1,…,r . In this paper we assume simple random walks to provide the most flexible framework for state variables, since no convergence problems are detected. Thus r=1 and matrix $\mathbf{D}(w)$ is the identity matrix.

2.1.2 Observation Equations

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 vehicles) in a linear transformation that considers:

- The number of equipped vehicles entering from each entry centroid during time intervals k, k-1, k-M, $q_i(k)$.
- H<M *time*-varying *model parameters* in form of *fraction matrices,* $[u_{ijcq}^h(k)]$.

The H adaptive fractions that approximate u_{iq}^h and u_{ijcq}^h are updated from measures provided by ICT sensors. Direct samples of travel times allow the updating of discrete approximations of travel time distributions, making it unnecessary to incorporate models for traffic dynamics. This model simplification, due to the availability of the new ICT, is another major novelty in our proposed formulation.

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.*

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

Where $\mathbf{v}(k)$ are, respectively, white Gaussian noises with covariance matrices $\mathbf{R}_k$. $\mathbf{F}(k)$ maps the state vector $\Delta\mathbf{g}(k)$ onto the current blocks of measurements at time interval *k*: counts of equipped vehicles by sensors and entries at centroids, accounting for time lags and congestion effects. Deviate counts at *k* mean the observed counts minus the historical demand $\tilde{g}_{ijc}(k)$ counts, given the

ISBN: 978-1-61804-080-0

current traffic conditions according to *time-varying model parameters*).

3 Simulation Tests

3.1 Design of Computational Experiments

The simulation experiments for testing the proposed approach have been conducted using dynamic OD matrix sliced into four 15-minute slices --each one accounting for 15%, 25%, 35% and 25% of the total number of trips-- to emulate demand variability, corresponding to a rise in congested conditions. This OD, considered the true historical OD matrix, has been determined through simulation by building the Macro Fundamental Diagram (MFD) for the network (Daganzo and Geroliminis [13]).

Microscopic simulation with AIMSUN induces variability in the historical inputs in a realistic way and produces target OD flows per interval (related to the true historical OD matrix). Therefore, congestion is not a design factor in these experiments. This OD is sliced into four 15-minute slices --each one accounting for 15%, 25%, 35% and 25% of the total number of trips-- to emulate demand variability. The simulation experiments with Aimsun use the described OD matrix, emulating traffic detection, and provide the inputs to the KFX model in terms of flow counts for ordinary and BT sensors and travel times from entry cordon *i* to *q* BT sensor (in general, ICT sensor).

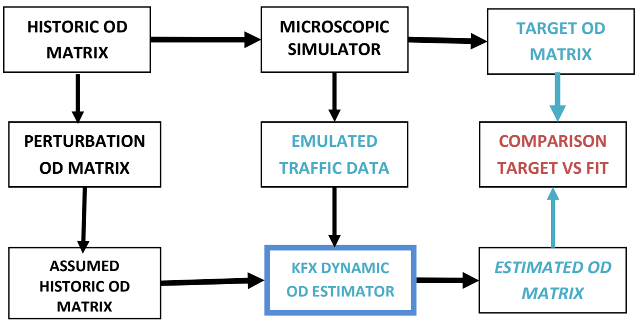

Figure 1. Methodological Design Of The Experiments By Simulation

Figure 1 depicts the methodological framework for the simulation experiments. The assumed OD matrix is the result of applying some perturbation to the true historical matrix.

Target OD flows per interval are compared with estimated OD flows (filtered OD flows) per interval at OD pair level by means of U Theil's coefficient and the normalized root mean square error (RMSEN). Weighted indicators for subsets of OD pairs (usually subset of OD pairs whose hourly flow is in 25% of higher flows) and a weighted global indicator for the whole set of OD pairs are also computed (GU, GRMSEN).

3.2 Results for Amara Test Site

Amara District is an urban network with 232 links, 142 nodes and 85 OD pairs, with a rich structure of alternative paths between OD pairs, totaling 358 most likely used paths according to the DUE with Dynameq [12].

The detection layout of 48 detectors raises some methodological concerns. As we have specified, it consists of two components: the cordon component encircling the network with sensors at input-output gates (currently available in most of the urban pricing systems), and the detection layout at the interior of the encircled area. However, when dealing with sensors capturing the electronic signature, such as the detectors of Bluetooth devices on board vehicles, the detector location requires a new approach based on a node covering formulation of the detection layout problem that has been developed and tested in Barceló *et al.* [14].

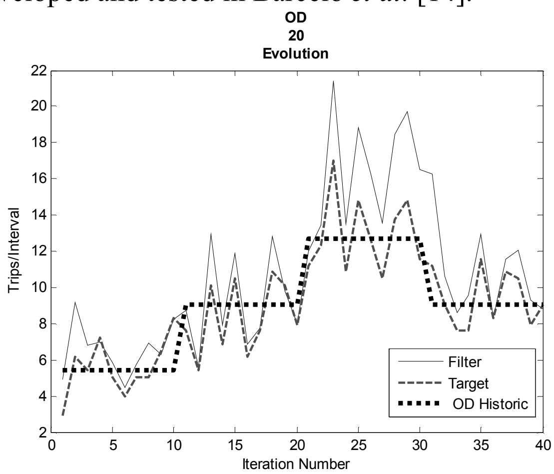

Figure 2 (100% Equipped Vehicles): Filtered OD Flows, Target OD Flows And Historic OD Flows For OD-Pair 20.765-777.

Figure 2 visualizes the results for the most relevant OD flow when the assumed historical OD matrix is reliable and the OD pattern is preserved The initialization of the state variables $\Delta g_{ijc}(0)$ is set to 0 and path proportions for each OD pair are taken as constant and equal. The concordance is numerically quantified in terms of RMSEN and U for OD flows from origin 765 to all destinations in the network in Table 2. U Theil's coefficients is 0.11 and RMSEN is 0.26. The low values of RMSEN and U prove good behaviour of KFX for OD pairs with large demand (those in the 4th quantile of OD flows).

Average Historic (veh/90s)		OD pairs from centroid 765 to all – 4 Time Slices					
		762	**768**	**777**	**783**	**789**	**806**
765	**Flow**	1.00	4.83	**8.98**	0.18	1.13	7.73
	RMSEN	0.56	0.52	**0.26**	0.47	0.20	0.22
	U	0.34	0.31	**0.11**	0.27	0.10	0.09

Table 2. OD Pairs From 765 To All – Reliable assumed historical OD - 100% BT Equipped

The aggregated hourly fit between estimated KFX *versus* Target OD flows shows a regression R squared above 95% when all 48 sensors are active, but the performance of the global fit decreases as the number of BT sensors is reduce. In Figure 3, an scatterplot shows the regression fit when BT sensor data is emulated for half of the sensors (the most important according to the number of collected vehicles), the R^2 is 90% (colors for dots are selected according to quantiles in OD flows).

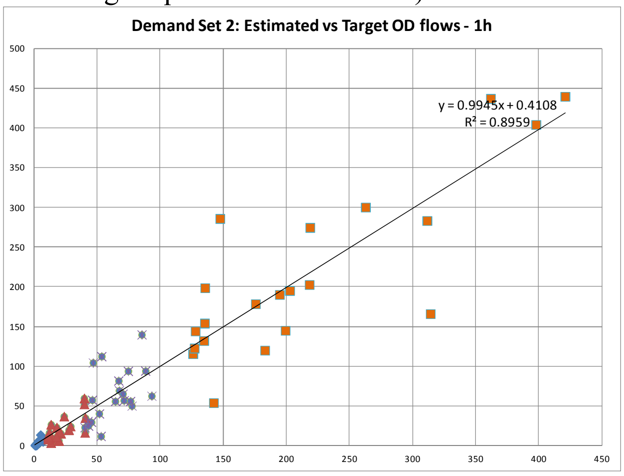

Figure 3. Fit estimated KFX versus Target OD flows. 100% BT equipped and reliable initial OD matrix. Detection layout reduced to 50%

OD pairs with large flows converge to correct values but, when sparse observations are available, no correction for *a priori* assumed flows is possible and, thus, no convergence is achieved (for OD pairs in the 1st and 2nd quantile of OD flows, GU ranges from 0.28 to 0.74, indicating a bad fit).

Although not shown in Table 2, we studied the effect of decreasing the penetration rates of BT technology in the selected OD pairs, showing an increase in the fit indicators (U, RMSEN) and, thus, a decrease in the quality of the results as equipped rate decreases. So, the decrease in BT penetration means fewer observations and larger Δt needed for practical purposes. Jointly, small OD flows show worse behaviour, due to few observations per interval and, thus, design parameter Δt should be increased to properly deal with the sparse reality.

For practical purposes, the effect of the reliability of the *assumed historical matrix* on the convergence of KFX estimates to the target OD matrix is critical. When fixing 100% equipped (not realistic), the historical matrix reliability does not seriously affect the global results (U and RMSEN global indicators): even under the most perturbed historical OD matrix tests, RMSEN stays around 0.3, when considering the 25% of the most important OD pairs, and it is reduced to 0.16 for pair 20 (the largest one). In terms of serial matching for the overall 4th quantile OD flows, the effect of an unreliable input matrix is minor: GU never increases to 0.20 but, for the most important pair 20, U values range from 0.07 to 0.17 (excellent coefficients). The behavior shows that OD pairs with large flows converge to correct values but, when sparse observations are available, no correction for *a priori* assumed flows is possible. The elapsed time for each experiment is around 1.5 min in Intel Core2 Duo T9550@2,66GHz with 4 GB RAM, under Windows XP.

3.3 Results for Ronda de Dalt Test Site

An urban freeway in Barcelona has been selected as a second test site --a 11.5-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). Traditional and BT sensors are modeled for entry and exit ramps and main sections.

Time horizon is defined as 100 min and a 150-sec subinterval length is considered. There are 74 OD pairs and OD path flows since a unique path is associated to each OD pair

Perturbation Level to Historic OD	Indicators for 4th Quantile OD flows in parenthesis					
	20%		**30%**		100%	
	U	**RMSEN**	**U**	**RMSEN**	**U**	**RMSEN**
0-none	0.22 (0.11)	0.39 (0.24)	0.20 (0.10)	0.33 (0.22)	0.17 (0.09)	0.25 (0.18)
1	0.24 (0.12)	0.43 (0.29)	0.22 (0.11)	0.37 (0.26)	0.19 (0.10)	0.29 (0.23)
2	0.25 (0.11)	0.43 (0.28)	0.23 (0.10)	0.36 (0.25)	0.20 (0.09)	0.28 (0.22)
3	0.27 (0.11)	0.44 (0.27)	0.25 (0.10)	0.37 (0.24)	0.22 (0.09)	0.30 (0.21)
all	0.40 (0.14)	1.40 (0.34)	0.39 (0.14)	1.29 (0.31)	0.38 (0.13)	1.17 (0.27)

Table 3. Ronda de Dalt simulation results according to BT rates and OD Reliability

ISBN: 978-1-61804-080-0

Results included in this paper affect the quality of the *assumed historic OD flows* used for computing deviates as state variables and the percentage of BT equipped vehicles) with levels 20%, 30%, 50% and 100% (see Table 3). Results are consistent with those obtained in the Amara Test site.

4 Conclusions

The computational experiments show that the proposed linear Kalman-Filtering approach provides good estimates of target values in the simulation tests in network and freeway sites. BT data simplifies the dynamic estimation of OD matrices by a KF approach because it is a linear filter and reduces the computational burden when compared to well-known formulations in the literature that use Extended Kalman Filter.

The strategy of collecting the DUE most likely used paths according to Dynamic User Equilibrium models --and, thus, defining KFX state variables-- seems promising, since time-dependent path proportion shares and assignment matrices are not employed in the formulation.

The horizon of study has to be divided in time intervals of length Δt, usually 1.5 to 5 minutes, depending on the network size (and travel times involved), OD flows and BT penetration rates.

OD pairs with large volumes are not seriously affected neither by the quality of the *assumed historical matrix* nor by *BT penetration rates* under 50% (around 30% in the real sites) and thus the approach exhibits good convergence properties that are needed for applications in urban networks, corridors and freeways.

A new site in Barcelona's CBD district is being prepared to field test the approach in a forthcoming pilot project.

References:

[1] Van Lint J.W.C. and Hoogendoorn S. P., A Robust and Efficient Method for Fusing Heterogeneous Data from Traffic Sensors on Freeways *Computer-Aided Civil and Infrastructure Engineering* 24**, 2009,** pp. 1–17.

[2] Treiber M., and Kesting A., Wilson R. E., Reconstructing the Traffic State by Fusion of Heterogeneous Data, *Computer-Aided Civil and Infrastructure Engineering*, No26, 2011, pp. 408-419.

[3] Kalman, R.E. A new approach to linear filtering and prediction problem. *Journal of Basic Engineering* **82** (1) , 1960, pp. 35–45.

[4] 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*, Vol. 2175, 2010, pp. 19-27.

[5] Barceló J., Montero L., Marqués L. and Carmona C., A Kalman-Filter Approach For Dynamic OD Estimation In Corridors Based On Bluetooth And Wifi Data Collection, *Proceedings 12th World Conf. on Transportation Research,* Lisbon, July 11-15, 2010.

[6] Chang, G.L. and Wu, J., Recursive estimation of time-varying Origin—Destination flows from traffic counts in freeway corridors. *Transportation Research B* Volume 28B No.2, 1994, pp. 141-160

[7] Hu, S., Madanat S.M., Krogmeier J., Peeta S. Estimation of Dynamic Assignment Matrices and OD Demands using Adaptative Kalman filtering, *Intelligent Transportation Systems Journal* vol. 6, 2001, pp 281-300.

[8] Choi, K., Jayakrishnan, R., Kim, H., Yang, I., Lee, J., Dynamic OD Estimation using Dynamic Traffic Simulation Model in an Urban Arterial Corridor, *Presented at the 88th TRB Meeting* Washington D.C., 2009.

[9] Van Der Zijpp, N.J. and Hamerslag, R., An improved Kalman filtering approach to estimate Origin—Destination matrices for freeway corridors. In: *Transportation Research Records 1443*, National Research Council, 1994, pp. 54–64.

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

[11] 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, 21-36

[12] Dynameq, Version 1.4.6, INRO, 2010.

[13] Daganzo, C. F. and N. Geroliminis An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Trans. Res. B* 42, 2008, pp. 771-781.

[14] Barceló J., Gilliéron F., Linares M.P., Serch O. and Montero L., The Detection Layout Problem, *Paper #12-2056, Presented at the 91st Annual TRB Meeting*, 2012.