Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 10 (2015) 134 – 144

Transportation Research Procedia

www.elsevier.com/locate/procedia

18th Euro Working Group on Transportation, EWGT 2015, 14-16 July 2015, Delft, The Netherlands

A robust framework for the estimation of dynamic OD trip matrices for reliable traffic management

Jaume Barceló*, Lídia Montero

Barcelona Tech – UPC, Campus Nord, Carrer Jordi Girona 1– 3, 08034 Barcelona, Spain

Abstract

Origin-Destination (OD) trip matrices describe the patterns of traffic behavior across the network and play a key role as primary data input to many traffic models. OD matrices are a critical requirement, either in static or dynamic models for traffic assignment. However, OD matrices are not yet directly observable; thus, the current practice consists of adjusting an initial or a priori matrix from link flow counts, speeds, travel times and other aggregate demand data. This information is provided by an existing layout of traffic counting stations, as the traditional loop detectors. The availability of new traffic measurements provided by ICT applications offers the possibility to formulate and develop more efficient algorithms, especially suited for real-time applications. However, the efficiently strongly depends, among other factor, on the quality of the seed matrix. This paper proposes an integrated computational framework in which an off-line procedure generates the time-sliced OD matrices, which are the input to an on-line estimator. The paper also analyzes the sensitivity of the on-line estimator with respect to the available traffic measurements.

© 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of Delft University of Technology

Keywords: Dynamic OD Matrices, Matrix Estimation, Kalman filtering, ICT data.

1. Motivation

In the context of estimating passenger-car transport demand, Origin-to-Destination (OD) trip matrices describe the number of trips between each origin-destination pair of transportation zones in a study area. For private vehicles,

* Corresponding author. Tel.: +34 93-401-69-41.
E-mail address: jaume.barcelo@upc.edu

2352-1465 © 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of Delft University of Technology
doi:10.1016/j.trpro.2015.09.063

route choice models describe how drivers select the available paths between origins and destinations and, as a consequence, the number of trips using a given path (or path flow proportions). The route choice proportions can vary depending on the time-interval in dynamic models, since they depend on traffic states changing over time.

All formulations of static traffic or transit assignment models (Florian and Hearn (1995)) as well as dynamic models involved in ATMS (Advanced Transport Management Systems, Ashok et al. (2000) assume usually that a reliable estimate of an OD matrix is available, and constitutes an essential input for describing the demand to estimate network traffic states and short term predict their evolution. Since OD trips are not yet directly observable, indirect estimation methods have then been proposed. These are the so-called matrix adjustment methods, whose main modeling hypothesis, in the static case, can be stated Cascetta (2001) as follows: if the assignment of an OD matrix to a network defines the number of trips in all network links, then the same OD matrix could be estimated, as the inverse of the assignment problem, as a function of the flows observed on the links of the network. However, since the resulting problem is highly undetermined, additional information is necessary to find suitable solutions and consequently this has been a fertile domain of research (see Lundgren and Peterson (2008) or Bullejos et al (2014).

The estimation of time-dependent OD matrices has been usually based on space-state formulations using Kalman Filtering approaches, Ashok et al (2000), as the most suitable to model dynamic phenomena. What these approaches share with the static ones is that they still require an assignment matrix, whose entries determine the proportion of trips between an OD pair, using a link at a given time interval, with a relevant role for those links where traffic detection stations were located. Chang and Wu (1994) proposed an Extended Kalman-filter approach to deal with the nonlinear relationship between the state variables and the observations. Variants of this approach have been explored by other researchers, using variants of Extended Kalman Filters to deal with the time dependencies of model parameters, which are usually included as state variables in the model formulation. Just to mention a few, Hu et al. (2001) which explicitly take into account temporal issues of traffic dispersion, or Lin and Chang (2007) which also assume that travel time information is available in order to deal with traffic dynamics. Other researchers, Dixon and Rilett (2002), Antoniou et al. (2004) or Work et al. (2008) also included other measurements as for example those supplied by GPS tracking of equipped vehicles or Automatic Vehicle Identification in the model formulation.

However, when real-time measurements from Information and Communications Technologies (ICT) are available, e.g. those supplied by Bluetooth/Wi-Fi devices, hypothesis on non-linear traffic flow propagation to estimate travel-times between pairs of points in the network are no longer necessary, since they can be measured by these technologies, and then state variables can be replaced by measurements and the Extended Kalman formulation can be successfully replaced by an ad hoc linear formulation, Barceló et al. (2013b). which uses deviations of OD path flows as state variables, as suggested by Ashok and Ben-Akiva (2000), and does not require an assignment matrix but instead a subset of the most likely OD path flows identified from a Dynamic User Equilibrium (DUE) assignment.

A relevant finding of this approach to estimate dynamic OD matrices exploiting ICT data, Barceló et al. (2013a), is that the three key design factors that determine the quality of the estimate are, respectively, the quality of the detection layout, the quality of the historic dynamic OD matrix used as a priori initial estimate, and the percentage of penetration of the technology. This influence can be clearly seen in the graphics in Fig. 1, Barceló et al. (2013a), where the results from a series of computational experiments are represented in terms o response surfaces where each one corresponds to a level of quality of the OD seed used. Clearly the two first factors are controllable design factors while the third one cannot be controlled by the analyst. Assuming that some of the ICT applications (e.g. Bluetooth/Wi-Fi antennas, License Plate Recognition or Electronic ID identifiers, etc.) require fixed locations in the network, the quality of the sensor layout can be guaranteed using an optimized detection layout, purposely

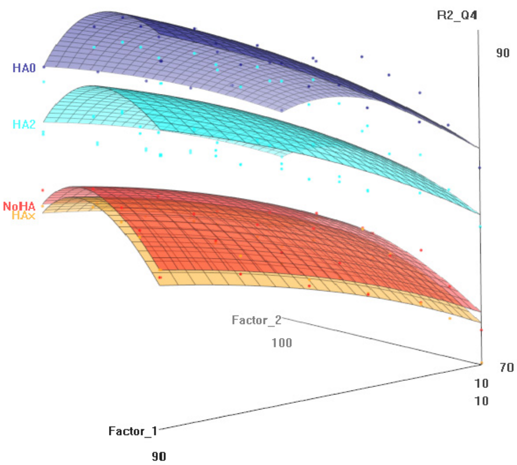


Fig. 1. R^2 Fitted vs Target OD flows (1h 15min) for OD pairs in the 4[th] quartile (**R2_Q4**) according to #BT Inner Sensors (Factor 1, Detection Layout), %BT Equipped Vehicles (Factor 2, Technology Penetration) and Quality of the OD seed (Factor 3).

suited to the OD estimation objectives (see Barceló et al. (2012)). Also the quality of priory initial OD matrices can be ensured, by specific off-line bilevel optimization procedures, Barceló et al. (2014) and Bullejos et al. (2014). The approach has proved to be very efficient when, further that the usual link flow counts, the model also includes travel time measurements between pairs of sensors (e.g. Bluetooth/Wi-Fi detection antennas when Bluetooth/Wi.Fi devices are set to discoverable mode) and the identification of the most likely used partial paths between them (e.g. as resulting from a DUE). The research reported in this paper is a direct consequence of the previous results, if the quality of the time-dependent OD estimates strongly depends on the controllable design factors and, if given a purposely designed detection layout and an associated traffic data collection procedure, the determinant factor is the quality of the input OD seed, which can be acceptably estimated off-line then, on one hand it is natural to integrate both procedures in a unified computational framework and on the other hand determine the robustness of the real-time estimation procedure when, in the given conditions the integrity of the detection is affected by detector malfunctions which perturb the quality of the expected input.

Consequently the paper is structured as follows, in Section 2, the architecture of the integrated computational framework will be introduced and discussed, and its components summarily described. Section 3 will present the modification required by the Kalman Filter approach when the available data inputs are a subset of the theoretically expected, this variant of the Kalman Filter will also analyze what else could be done when other data inputs are also available. Section 4 analyzes a simple case. Section 5 will draw the conclusions and identify the future work.

2. An integrated framework

The experience gained in Barceló et al. (2013a, 2013b, 2014) and Bullejos et al. (2014) lead to propose an integrated architecture, which combines off-line time sliced OD estimation procedures, with on-line time dependent OD estimation procedures that use the off-line as OD seeds, in the real-time estimation process. The resulting architecture for the integrated computational framework is depicted in Fig. 2, and can be described as follows:

- Off-line time-sliced OD estimation, corresponds to the upper box in Fig. 2, it assumes that:
  1. Traffic data from the available data sources (e.g. inductive loop detectors, magnetometers, License Plate Recognition (LPR) devices, Bluetooth/Wi-FI antennas, etc.) have been collect for a long period of time, along with other data (e.g. weather data, calendar events, etc.) which can influence traffic behavior and determine different behavioral patterns. An appropriate data analysis, filtering, fusing and clustering the available traffic data, can identify the traffic profiles corresponding to each relevant behavioral pattern. The available data and associated profiles are stored in an ad hoc Historical Traffic data Base.
  2. Profiles and associated data can be selected to generate primary estimates of time-sliced OD matrices that will be used as target OD matrices by the off-line matrix adjustment procedure. Heuristic procedures can be to generate these initial estimates from the available traffic data can be found in Spiess and Sutter (1990), Barceló et al. (2014) and Bullejos et al. (2014).
  3. The process assumes that a suitable network model is available from which a Dynamic User Equilibrium (DUE) can be conducted. The proposed DUE approach is based on a mesoscopic traffic simulation to specifically account for congestion propagation at each time slice and its influence in path uses, a phenomenon not properly captured by static assignment approaches. The results of the DUE allow the identification of the Most Likely Used Paths (MLU) between origins and destinations in the network, and a primary estimate of the expected paths flow proportions and paths travel times in equilibrium. The robustness and reliability of these paths is determined by the proper calibration of the mesoscopic model and the convergence of the DUE in terms of a Gap function.
  4. The traffic data associated to the currently selected time-slice, let's say the k-th, its corresponding initial OD matrix, and the MLU and associated path travel times, and path flow proportions, are the main input to the "Static Bilevel OD Adjustment Process", based on the approach described in Barceló et al. (2014) and Bullejos et al. (2014).
  5. The iterative application of the procedure to the time-slices in which has been split a time horizon and its profiles allows to generate a Database of adjusted time-sliced which will be used as OD seeds in the on-line adjustment process.
- On-line real-time OD estimation corresponds to the lower box in Fig. 2, it assumes that:

1. The detection layout, which is fixed, and the active detectors at the current time slice k, which can be variable, depending on the incidences, along with the structure of the most likely used paths (MLU) from a DUE, and the location of fixed detectors in this structure, are the main inputs to a procedure that generates the data structures of the Kalman model for the estimation and short term prediction of the OD matrix for time slice k+1.
2. Two other key inputs to the Kalman Filter are the OD seed for the current time slice k, selected from the off-line Database of time-sliced OD matrices, which depends on the current traffic data profile, and the traffic data measurements supplied in real-time by the available traffic detectors during that time period.
3. The Kalman Filter proposed in this paper that will be described next, is a variant of the versions in Barceló et al. (2013a, 2013b) purposely adapted to deal with variable configurations of the detection layout.

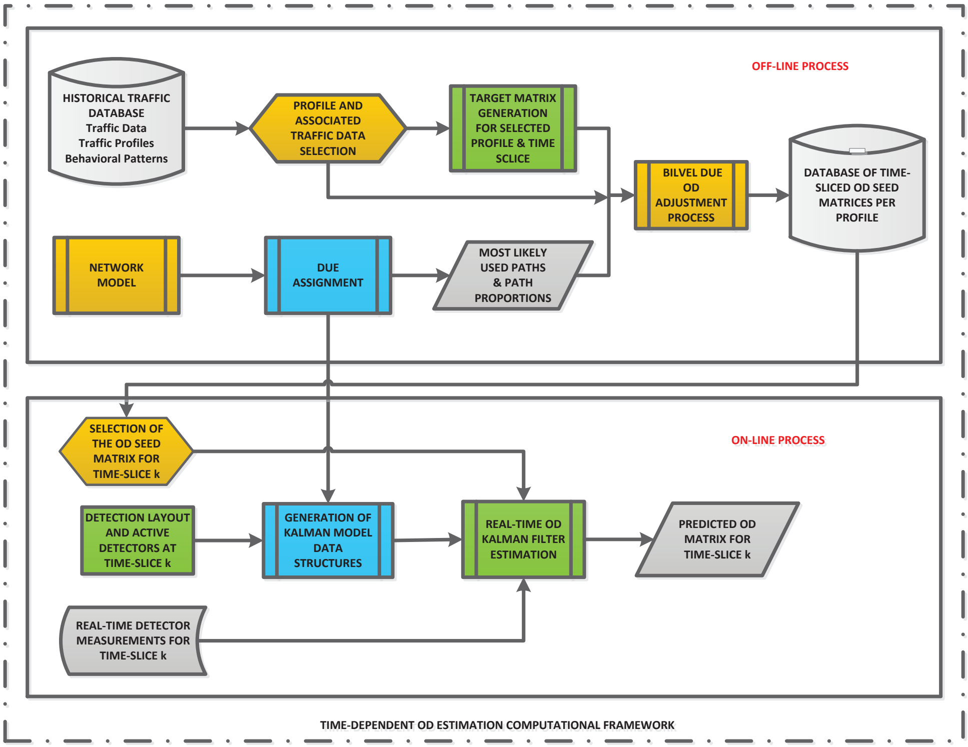


Fig. 2. Conceptual architecture of the computational framework for the estimation of time-dependent OD matrices

3. Space-state formulation revisited

The space-state formulation used in the on-line component is the recursive linear Kalman-Filter for state variable estimation discussed in Barceló et al. (2013a, 2013b), adapted to exploit traffic counts collected respectively by ICT sensors and conventional detection technologies, and travel times observed between any pair of ICT sensors. MLU Paths are the state variables in the state-space formulation used in the on-line procedure.The proposed approach initially assumes flow counting detectors and ICT sensors located in a cordon at each possible point for flow entry (*centroids* of the study area) and ICT sensors located at intersections in urban networks covering links to/from the intersection. Flows and travel times are available from ICT sensors for any 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 make 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 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.

The proposed linear formulation of the Kalman Filtering approach uses deviations of OD path flows as state variables, calculated in respect to DUE-based Historic OD path flows for equipped vehicles. A subset of the most likely OD path flows identified from a DUE assignment is used. The DUE is conducted with the historic OD flows, and the number of paths to take into account is a design parameter. 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. The time-varying dependencies between measurements (sensor counts of equipped vehicles) and state variables (deviates of equipped OD path flows), are used for estimating discrete approximations to travel time distributions. Since the approach uses the travel ICT time measurements from equipped vehicles, the nonlinear approximations can be replaced by estimates from a sample of vehicles, and then no extra state variables for modeling travel times and traffic dynamics are needed, since sampled travel times are used to estimate discrete travel time distribution, see Barcelo et al. (2013a) for details, splitting the time horizon into H subintervals. 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.

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 KFX2).

Let $Q_i(k)$ and $q_i(k)$ be respectively the number of vehicles and equipped vehicles entering from centroid i at time interval k. Conservation equations from entry points (centroids) are explicitly considered. Without $Q_i(k)$, a generic expansion factor has to be applied. The state equations are formulated as follows.

Let $\mathbf{\Delta g}(\mathrm{k})$ be the column vector of 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:

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

Where $\mathbf{w}(k)$ are zero mean with diagonal covariance matrix $\mathbf{W}_k$, and $\mathbf{D(w)}$ are 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 the implementation tested 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 $\mathbf{D}(\mathrm{w})$ is 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 vehicles) in a linear transformation that considers:

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

Where $u_{iq}^{h}(k)$ are the 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 $[(k-h-1)\Delta t \quad , \quad (k-h)\Delta t]$); and $u_{ijcq}^{h}(k)$ represent the 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*, where *i*=1,…,I , *j*=1,…, J, $h = 1\ldots M$, $q = 1\ldots Q$, and I, J and Q are, respectively, the number of origin centroids, the number of destination centroids and the number of ICT sensors. Direct samples of travel times from ICT sensors allow the updating of the H adaptive fractions u_{iq}^{h} and u_{ijcq}^{h}, making unnecessary to incorporate models for traffic dynamics. Time-varying model parameters u_{iq}^{h} and u_{ijcq}^{h} account for temporal traffic dispersion in affected paths and have to satisfy structural constraints, where H<M:

$$u_{ijcq}^{h}(k) \geq 0 \quad i = 1....I, \quad j = 1....J, \quad c = 1....K_{ij}^{max}, \quad q = 1,...Q, \quad h = 1...H$$
$$\sum_{h=1}^{H} u_{ijcq}^{h}(k) = 1 \quad i = 1....I, \quad j = 1....J, \quad c = 1....K_{ij}^{max}, \quad q = 1,...Q \tag{2}$$

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*.

$$\Delta z(k) = \begin{pmatrix} \mathbf{AU(k)^T} \\ \mathbf{E(k)} \end{pmatrix} \mathbf{\Delta g(k)} + \begin{pmatrix} v_1(k) \\ v_2(k) \end{pmatrix} = \mathbf{F(k)\Delta g(k) + v(k)} \tag{3}$$

Where $\mathbf{v}(k)$ are white Gaussian noises with covariance matrices $\mathbf{R_k}$. **U(k)** consists of diagonal matrices $U(k),\ldots,U(k-M)$ containing $u_{ijcq}^{h}(k)$. For $U(k-h)$ is a matrix with the estimated proportion of equipped vehicles whose travel time from the access point to the network takes *h* intervals and goes through the *q* sensor at interval *k*. **E(k)** is a matrix with I rows and non-zero columns only for the current time-interval *k* and defining conservation of flows for each entry at *k*. And **A** is a matrix that adds up sensor traffic flows from any possible entry, given time-varying model parameters at interval *k*. $\mathbf{F(k)}$ maps the state vector $\mathbf{\Delta g(k)}$ onto the current blocks of measurements at *k*: deviate counts of equipped vehicles by sensors and entries at centroids, accounting for time lags and congestion effects according to *time-varying model parameters*.

This paper reports on the revisited implementation KFX2, called KFX3, which extends the former in the following way:

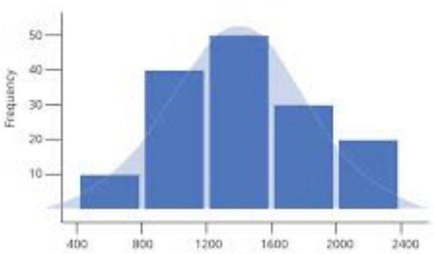
- Classes of vehicles are considered according to the available ICT technology, e.g. equipped vehicles with Bluetooth/Wi-Fi devices constitute a class.
- Several types of fixed location sensors for traffic data collection are considered, either located at links (e.g. loop detectors, magnetometers…) or at intersections (e.g. Bluetooth/Wi-Fi antennas) , counting vehicles or identifying the electronic footprint.
- In the current version the OD estimation procedure doesn't use vehicle trajectories but samples of travel times *between any pair of ICT sensors* to account for traffic congestion.
- From a design point of view, a critical attribute of a sensor is its state: active/non active. Non active sensors do not capture any data, but the on-line procedure is designed in such a way that dimensionality and complexity is not affected by non-active sensors.

DUE OD path proportions for MLU paths and interval are not an input to the Kalman Filter (KFX3) approach, only the description of the most likely OD paths. From our current version 3, KFX3, the use of travel times between any pair of regular ICT sensors has been incorporated to the formulation, as time-varying model parameters playing the role of discretization of travel time distributions. A data model dealing either with fixed link-based sensors or node-based sensors has been considered. New ICT sensors considered as mobile sources (GPS data) have also been included in the basic prototype through the definition of virtual ICT sensors link-level based. A basic hypothesis is that equipped and non-equipped vehicles follow common OD patterns. We assume that this holds true in what follows, but it requires a statistical contrast for practical applications.

In our experience the extended state vector has to be able to model trip travel times in the network in the 90% percentile order. Currently, the length of each interval (time-slice) has to be defined as a multiple of the subinterval length Δt, that it is assumed constant during the horizon of study.

The time-varying dependencies between measurements (sensor counts of equipped vehicles) and state variables make used of the estimation of discrete approximations to travel time distributions between pairs of *(r,s)* ICT sensors in KFX3. Sampled travel times from ICT equipped vehicles are used in KFX3 to approximate travel time distributions between pairs of *(r,s)* ICT sensors by discrete travel time distributions adapting the process described in Barceló et al (2013a), according to H non null bins. For example, let us assume that travel times between a pair of ICT sensors lie between 400 and 2400 seconds (see Fig. 3a) and a subinterval length of 400 seconds to simplify the figures and H=5 bins, a sample of 150 equipped vehicles provides travel time data that allows to approximate the distribution using H= 5 bins considering non null probability bins from one to five Δt (h 1 to 5), each with a corresponding probability as described in Fig. 3a.

For another OD pair and travel time distribution between 1200 and 3200 seconds, the discrete approximation would follow the Case (b) described in Fig. 3b. And finally, let us assume OD pair travel time distribution showing less dispersion, since the range between 1600 and 2800 sec coverts an observed probability of 1, and thus, the discrete approximation would follow the Case (c) described in Fig. 3 for h equal 4 covering 4 to 5 subintervals (1600 to 2000 sec) the observed probability would be 1/3, for h equal to 5 covering 5 to 6 Δts (2000 to 2400 sec) observed probability would be 5/12 and finally for the 2400 to 2800 sec bin, the observed probability is 3/12.



Case (a)

h	1	2	3	4	5
p	1/15	4/15	5/15	3/15	2/15

Case (b)

h	3	4	5	6	7
p	1/15	4/15	5/15	3/15	2/15

Case (c)

h	3	4	5	6	7
p	0	4/12	5/12	3/12	0

Fig. 3 Discrete travel time distribution approximation – Histogram representing Case (a)

Discrete distributions are updated every time subinterval, Δt. Clearly, the number of bins H depends on the dispersion of travel time distributions, in our experience H between 3 and 5 is enough to cope with travel time variability between pairs of points in medium size networks for a proper Δt. The same procedure has been applied in KFX3 for building and updating travel time distributions between OD pairs. There is a relationship between the subinterval length and the suitability of the scheme for approximating travel time distributions: a large Δt diminishes the ability to capture the variability of travel time distributions and thus large H are non-effective and rough approximations are provided. However a short Δt diminishes the ability to capture the central trend in travel time distributions providing inaccurate discrete travel time approximations unless the number of bins H is increased and thus increasing the computational burden. A trade-off has to be determined in a tuning process according to network characteristics and congestion issues.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. State equations are defined as in Eq. (4). The formulation is modified with respect to the time-varying model parameters (in form of fraction matrices, $\lfloor u_{ijcq}^{h}(k) \rfloor$ and $u_{ijc,rs}^{h}(k)$ for H<M) and the observation equations affecting the relationship between the state variables and the observations in a linear transformation that considers:

1. The number of equipped vehicles entering from each entry centroid during time intervals k,.., k-M, $q_i(k)$ if provided on-line by the active sensor layout or selected according to a priori historic data if not available. Total number of entry centroids is I.
2. The number of equipped vehicles tracked at each ICT sensor during time interval k if provided on-line by the active sensor layout. Total number of ICT sensors is Q.
3. The number of vehicles counted at each traditional sensor during time interval k if provided on-line by the active sensor layout. Total number of ICT sensors is P.
4. H<M time-varying model parameters in form of fraction matrices, $\lfloor u_{ijcq}^{h}(k) \rfloor$ and $u_{ijc,rs}^{h}(k)$.

Since travel times between any pair of *(r,s)* ICT are available, a more efficient use of the data makes allows to define $u_{rs}^{h}(k)$ the fraction vehicles that require *h* time intervals to reach sensor *s* at time interval k from sensor *r* (tracked at sensor *r* during time interval $[(k-h-1)\Delta t \quad , \quad (k-h)\Delta t]$); and $u_{ijc,rs}^{h}(k)$ that represent the fraction of equipped vehicles detected at interval k whose trip from sensor r to sensor s might use OD path *(i,j,c)* lasting h time intervals of length Δt to arrive from sensor r to sensor s, where i=1,…,I, j=1,…, J, $h = 1\ldots M$, $r,s = 1\ldots Q$, and I, J and Q are, respectively, the number of origin centroids, the number of destination centroids and the number of ICT sensors, and thus *(i,j)* identifies an OD pair. Previous model parameters, $u_{iq}^{h}(k)$ and $u_{ijcq}^{h}(k)$, are kept if no redundancy appears in the formulation.

Direct samples of travel times from ICT sensors allow the updating of the H adaptive fractions u_{iq}^{h} and u_{ijcq}^{h}, in particular, and also the general $u_{rs}^{h}(k)$ and $u_{ijc,rs}^{h}(k)$. Time-varying model parameters u_{iq}^{h}, u_{ijcq}^{h}, $u_{rs}^{h}(k)$ and $u_{ijc,rs}^{h}(k)$ are forced to satisfy nonnegativity structural constraints, where H<M:

$$\begin{aligned} &\sum_{h=1}^{H} u_{ijcq}^{h}(k) = 1 \quad i = 1....I, \quad j = 1....J, \quad c = 1....K_{ij}^{\max}, \quad q = 1,..Q \\ &\sum_{h=1}^{H} u_{ijc,rs}^{h}(k) = 1 \quad i = 1....I, \quad j = 1....J, \quad c = 1....K_{ij}^{\max}, \quad r = 1,..Q, \quad s = 1,..Q \end{aligned} \tag{4}$$

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.

$$\Delta z(k) = \begin{pmatrix} \mathbf{A}_1\mathbf{U}_1\mathbf{(k)}^{\mathbf{T}} \\ \mathbf{A}_2\mathbf{U}_2\mathbf{(k)}^{\mathbf{T}} \\ \mathbf{E(k)} \end{pmatrix} \Delta\mathbf{g(k)} + \begin{pmatrix} v_1(k) \\ v_2(k) \\ v_3(k) \end{pmatrix} = \mathbf{F(k)}\Delta\mathbf{g(k)} + \mathbf{v(k)} \tag{5}$$

Where,

- **v***(k)* are white Gaussian noises with covariance matrices $\mathbf{R_k}$.
- **U₁(k)** consists of diagonal matrices $U_1(k),\ldots,U_1(k-M)$ containing $u_{ijcq}^{h}(k)$. $U_1(k-h)$ is a matrix with the estimated proportion of equipped vehicles whose travel time from the access point to the network takes h intervals and goes through the q sensor at interval k.
- **U₂(k)** consists of diagonal matrices $U_2(k),\ldots,U_2(k-M)$ containing $u_{ijc,rs}^{h}(k)$. $U_2(k-h)$ is a matrix with the estimated proportion of equipped vehicles whose travel time from sensor r to s takes h intervals and goes through the s sensor at interval k.
- **E(k)** is a matrix with I rows and non-zero columns only for the current time-interval k and defining conservation of flows for each entry (I) at k.
- And A_1 is a matrix that adds up sensor traffic flows from any possible entry, given time-varying model parameters at interval k and A_2 is a matrix that adds up traffic flows from a pair (r,s) of ICT sensors, given time-varying model parameters at interval k.
- $\mathbf{F}(k)$ maps the state vector $\Delta\mathbf{g}(k)$ onto the current blocks of measurements at k: deviate counts of equipped vehicles by sensors and entries at centroids, accounting for time lags and congestion effects according to time-varying model parameters.

4. Computational Results

The experimental approach used in the computational experiments to test the proposed Kalman-Filter OD estimator has been based on the use of synthetic data to perform controlled experiments. This is an approach well suited to verify inverse problems where, typically the inputs to the forward problem are not measurable, as in the case of the estimation of OD matrices. In this case a synthetic demand is generated and used as input to the forward problem to produce synthetic measurements from which the inverse problem estimates the synthetic input. The values of the selected performance indicators measure the distance between the known synthetic input and the unknown estimated output, to determine the quality of the estimation procedure.

OD (h)	8	9
1	180	120
2	60	180

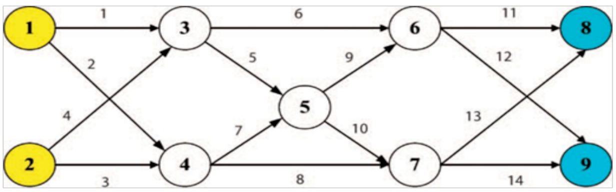


Fig 4. Parallel highway network of Hu, Peeta and Chu (2009): link, ICT sensors and node identifiers

The parallel highway network of Hu et al (2009) in Fig 4 has been used primarily for debugging and verification purposes. The OD matrix consists of four non-zero OD flows (1,8), (1,9), (2,8) and (2,9) (identified as 1 to 4). The set of MLU OD paths is composed of routes described in Table 1. ICT sensors are initially assumed available at entry nodes 1 and 2 and nodes 3 to 7.Thus, the number of OD pairs is 4, the number of MLU OD paths is 10 and there are 18 feasible pairs of ICT sensors *(r,s)*. We assume a subinterval of $\Delta t=5$ min. Since the time-horizon is one hour, 12 subintervals are defined and 15 min time slices account for (10-20-40-30) percentage of the total demand. No GPS virtual sensors have been initially considered and all vehicles are assumed ICT equipped for checking purposes. Clearly, the sensor layout is initially oversized, however the availability of sensors will be progressively reduced and the corresponding data model for the Kalman filtering formulation adapted automatically.

Table 1. Description of MLU OD paths related to DUE assignment

MLU OD Path id	OD path links	OD pair	ICT in node id	Entry id	MLU OD Path id	OD path links	OD pair	ICT in node id	Entry id
1	1-6-11	1=(1,8)	3,6	1	**6**	3-7-9-11	3=(2,8)	4,5,6	2
2	2-7-9-11	1=(1,8)	4,5,6	1	**7**	3-8-13	3=(2,8)	4,7	2
3	2-8-13	1=(1,8)	4,7	1	**8**	4-6-11	3=(2,8)	3,6	2
4	1-5-10-14	2=(1,9)	3,5,7	1	**9**	3-8-14	4=(2,9)	4,7	2
5	2-8-14	2=(1,9)	4,7	1	**10**	4-5-10-14	4=(2,9)	3,5,7	2

The integration and validation of the proposed framework for OD estimation with the new on-line estimation tool KFX3 is focused on the testing about the availability of ICT data. Two situations have been considered: Mx-All (origins and internal nodes 3 to 7) and Sx-All (Internal) Nodes. The quality of the computational results is measured in term of three performance indicators, Theil's coefficient, NRMSE (Normalized Root Mean Squared error) and R^2 (coefficient of determination in linear regression) goodness of fit statistics (sum of squared differences between *target* and estimated OD flows per interval, relative to total *target* flows). Robustness of the estimated OD flows is calculated by combining all 4 OD pairs in a Global NRMSE (GNRMSE), defined by:

$$GNRMSE = \frac{\sqrt{4 \cdot 12 \sum_{k=1:12} \sum_{od=1:4} \left(y_{od,k} - \hat{y}_{od,k}\right)^2}}{\sum_{k=1:12} \sum_{od=1:4} y_{od,k}} \tag{6}$$

For the given network and 'Target Matrix (h)' for the one hour horizon (170 30 54 186) (X axis in Fig 5), the aggregated estimate of the OD matrix for the whole interval is represented in Y-axis considering two different situations. In Fig 5 left, a complete sensor layout Mx-All (origins and internal nodes 3 to 7) . In the right of Fig 5, the estimated OD matrix when Sx, internal layout is assumed, is shown. In both situations, the coefficient of determination of the regression line is higher than 95%. Although, detailed Theil and NRMSE are not included for each OD pair in |Fig 5, when the sensor layout is Sx, the correspondent Theil coefficients are (0.19 0. 14 0.34 0. 15) and NRMSE are (0.44 0.33 0.72 0.32). OD pair number 3 shows a lack of fit, but the rest of OD pairs have a fine behavior.

5. Conclusions and future work

The proposed computational framework for the estimation of time-dependent OD integrates two procedures, an off-line, to estimate time-sliced OD seeds, which are the input to an on-line procedure, suitable for the real-time estimation necessary for real-time traffic management. The on-line procedure exploits the availability of ICT trarric measurements. This paper presents a revisited version of Kalman Filtering designed to deal with variable configurations of detection layouts. The approach has been successfully numerically tested with a small artificial network to verify and check the numerical correctness of the procedure. The future work will apply this modified variant of Kalman Filter to two real urban networks, the City of Vitoria and the CBD of Barcelonaal ready used in previous research.

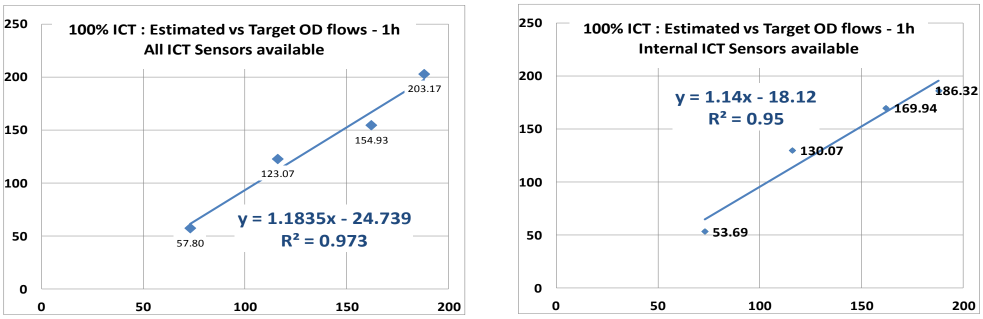


Fig 5. KFX3: Validation of Sensor availability implementation

Acknowledgments

This research has been partially funded by TRA2011-27791-C03-02 of the Spanish R+D National Programs and by the *Secretaria d'Universitats i Recerca de la Generalitat de Catalunya*.

References

Antoniou C., Ben-Akiva, M., Koutsopoulos, H.N., Incorporating Automated Vehicle Identification Data into Origin-Destination Estimation, TRR Vol 1882, 2004, pp 37-44.

Ashok, K. Ben-Akiva, M.: Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin-Destination Flows. In: Transportation Science vol.34 no.1, pp. 21-36 (2000).

Barceló, J., Montero, L., Bullejos, M., Linares, M.P., Serch, O. (2013a), Robustness And Computational Efficiency Of A Kalman Filter Estimator of Time Dependent OD Matrices. Transportation Research Record 2344-04 pp.31-39.

Barceló, J., Montero, L., Bullejos, M., Serch, O., Carmona, C. (2013b). A Kalman Filter Approach for the Estimation of Time Dependent OD Matrices Exploiting Bluetooth Traffic Data Collection. In: JITS Journal of Intelligent Transportation Systems, vol. 17(2) pp.123-141.

Barceló J.,. Montero L, Bullejos M., Linares M.P. (2014). A practical proposal for using origin-destination matrices in the analysis, modeling and simulation for traffic management, Paper 14-3793, presented at the 93rd TRB Annual Meeting

Barceló J., Gilliéron F., Linares M.P., Serch O., Montero L. (2012), Exploring Link Covering and Node Covering Formulations of Detection Layout Problem. Transportation Research Records: Journal of the Transportation Research Board, No. 2308, pp.17-26.

Bullejos M., Barceló J., Montero L. (2014), A DUE based bilevel optimization approach for the estimation of time sliced OD matrices, International Symposium of Transport Simulation 2014, to appear in Procedia Social and Behavioral Sciences.

Cascetta, E. Transportation Systems Engineering: Theory and Methods. Kluwer Academic Publishers, Norwell, Mass., 2001

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

Dixon, M.P, Rilett, L.R. , Real-Time OD Estimation using Automatic Vehicle Identification and Traffic Count data, Computer-Aided Civil and Infrastructure Engineering, 2002, Vol 17 pp 7-21.

Djukic, T., Barcelo, J., Bullejos, M., Montero, L.,Cipriani, E., van Lint, H., Hoogendorn, S. (2015) Advanced traffic data for dynamic OD demand estimation: The state of the art and benchmark study. Proceedings of the 94th Annual Meeting of the TRB Washington D.C

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

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.

Hu, S. R., S. Peeta, and C. H. Chu. 2009. "Identification of Vehicle Sensor Locations for Link-based Network Traffic Applications." Transportation Research Part B 43 (8–9): 873–894.

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

Lundgren J. T. and Peterson A. (2008), A heuristic for the bilevel origin–destination-matrix estimation problem, Transportation Research Part B 42 (2008) 339–354.

Spiess H. and Sutter D. (1990), Modeling the daily traffic flows on an hourly basis, Proceedings of the 18th Summer Annual Meeting of PTRC.

Work, D., Toossavainen, O-P. Blandin, S., Bayen, A, Iwuchukwu and Tracton, T., An Ensemble Kalman Filtering Approach to Highway Traffic Estimation Using GPS Enabled Mobile Devices, Proceedings of the 47th IEEE Conference on Decision and Control, pp 5062-5068