An Approach to use Cooperative Car Data in Dynamic OD Matrix Estimation

L. Montero and J. Barceló
Department of Statistics and Operations Research
Universitat Politècnica de Catalunya, UPC-Barcelona Tech

Abstract. Traffic management applications are supported by dynamic models whose input should be realistic real-time OD demand matrices in order to find efficient network state estimates and forecast their short-term evolution. OD matrices have been so far usually estimated from historic and/or real-time data collection and prior matrices. Some of those methods developed by the authors in previous works provide realistic matrices to cope with day-to-day demand variability and real-time traffic conditions. Off-line time-sliced OD matrices estimation based on a simulation-optimization Bilevel-DUE approach has proved to provide appropriate initializations for on-line and real-time dynamic OD estimation methods based on specific versions of Kalman filtering whose input data requirements are traffic counts collected from traffic detection stations and other data supplied by ICT (Information and Communication Technologies) sensors, as for instance travel times between pairs of fixed ICT sensors . In this work, we present a review of the contributions to the on-line/off-line estimation of Dynamic OD matrices and we examine how new data provided by cooperative vehicles as the tracking along trajectories giving travel times between intermediate points of OD trips can be incorporated into the Kalman filtering equations for on-line dynamic OD estimation. Cooperative vehicles can be considered as mobile sensors, generating data from any point of the network, the computational burden of the adapted Kalman approach to cope with tracking data is examined in depth in order to guarantee the on-line applicability of the proposed approach for a mid-sized urban network.

Keywords: Demand Estimation, Information Systems, Advanced Transport Information Systems, Kalman Filtering, Cooperative car data

1. Introduction: Background, motivation and objectives

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, 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.

While an average OD table for a whole period of interest is acceptable for an urban transportation planning study, OD matrices for consecutive time intervals are required for appropriately modeling and/or optimizing dynamic system operations. For all formulations of static traffic or transit assignment models (Florian and Hearn, 1995), as well as dynamic models involved in ATIS (Advanced Transport Information Systems) (see Ashok et al., 2000), the usual assumption is that a reliable estimate of an OD matrix is available and constitutes an essential input for describing the demand in predicting traffic state evolution over the network. Since OD trips are not directly observable, indirect estimation methods have then been proposed. These are the so-called matrix adjustment methods, whose main modeling hypothesis can be stated as follows (Cascetta, 2001): if the assignment of an OD matrix to a network defines the number of trips in all network links, then the same OD matrix can be estimated as the inverse of the assignment problem as a function of the flows observed on the links of the network.

We focus on dynamic OD matrix estimation. In a previous research, (Barceló et al., 2013), we have found that the quality of the detection layout and the quality of the historic dynamic OD matrix are key controllable design factors. Another relevant design factor the penetration rate of the selected ICT technology, which unfortunately is not controllable. In our study we dealt in particular with Bluetooth antennas to identify Bluetooth , devices set to *discoverable mode*. So, given a suitable Bluetooth sensor layout and the most likely used paths between BT sensors, the historical measurements of the associated travel times between pairs of BT sensors for predetermined time-slices were included in an off-line (simulation-optimization Bilevel-DUE approach) to generate a set of initial OD matrices. These estimated time-sliced OD matrices and the real-time measurements are the inputs to on-line procedures (linear Kalman filtering approach) to estimate dynamic OD matrices. The computational testing of the dynamic OD matrix estimation was conducted by simulation, and the most likely used paths between OD pairs and between all BT antennas were assumed to be those computed by a Dynamic User Equilibrium, achieved in our case by a mesoscopic traffic simulation conducted with the simulator Aimsun Meso. Aimsun Meso generates not only the simulated flows and speeds at traditional detection stations and BT antennas, but also the simulated travel time estimates from Bluetooth antennas along the corresponding paths (Barceló et al., 2014),(Bullejos et al.2014).

Almost 20 years ago, memory space was expensive and unavailable on ordinary laptop computers. Because of this, linear Kalman filtering (KF) approaches for estimating dynamic trip matrices were considered inefficient, and unable to satisfy the requirements of on-line applicability. A linear KF prototype suited to the specific characteristics of the problem, coded in MATLAB (Barceló et al, 2013a and b), has proven to meet the requirements for on-line applications using traditional data collection and new ICT data.

The off-line dynamic OD matrix estimation procedure provides updated and high quality time-sliced OD matrices to be used as initial or seed matrices by the on-line estimation procedure. As reported in (Barceló et al. 2013a), the quality of this input is a key factor to achieve the fast convergence required for real-time applications requiring a response fast enough to support them. A relevant example are the Real-Time Traffic Management Systems whose decisions have to be supported and evaluated by traffic models based on the estimated dynamic OD matrices at short time-intervals, typically shorter that 15 minutes, and forecasting is performed for time horizons of about 30 minutes.

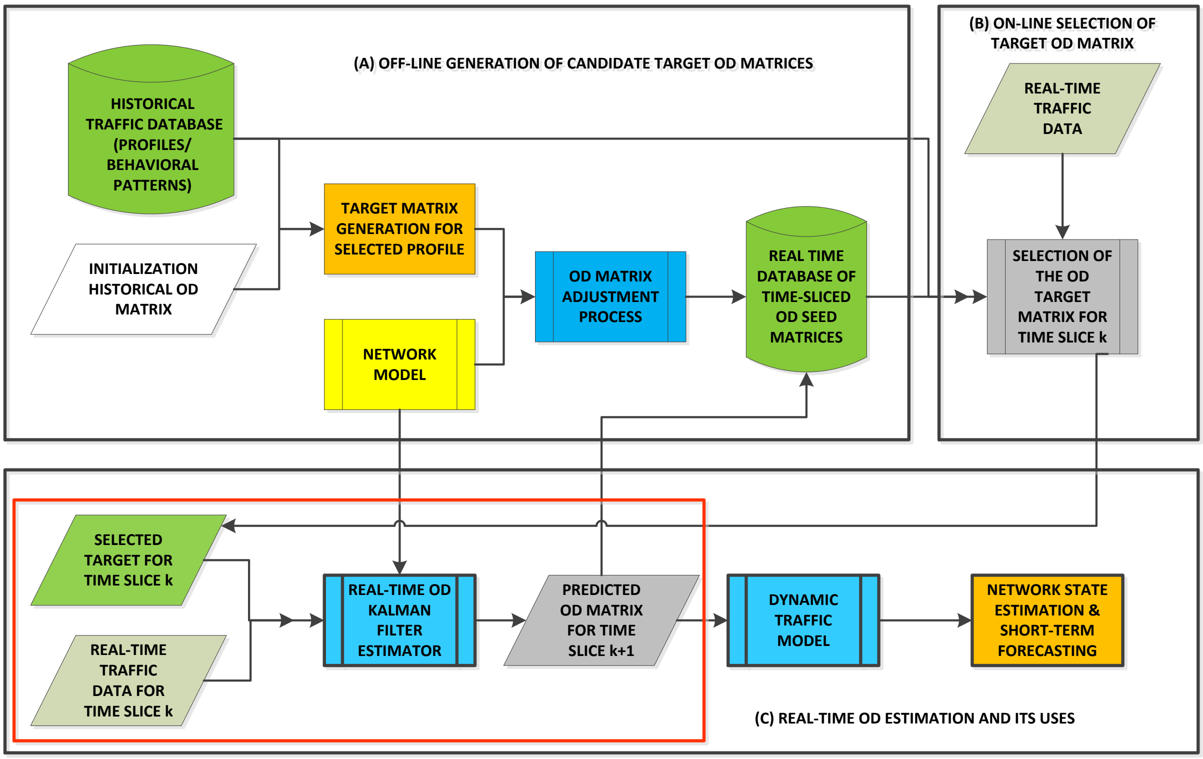

Figure 1: Methodological process for the estimation of time dependent OD matrices for real-time applications

The reported research corresponds to a methodological framework whose conceptual logic is depicted in Figure 1. The off line generation of the seed OD matrices exploiting all sources of traffic data is described in Box-A and the on-line process in Box-C, with an intermediate selection process of the suitable inputs in term of the observed evidence, described in Box-B.

The process in Box-A has been reported in (Barceló et al., 2014), (Bullejos et al.2014), and process in Box-C in ([Barceló et al. 2013a and 2013b), the objective of this paper is to explore the way, and expected improvements, of expanding the off-line and real-time estimation processes when new data are available. Our interest addresses two main mobile data sources that are either currently available or will become available in the near future: GPS data and data from Cooperative Cars. The advantage that they are, and presumably will become, pervasive, providing in this way a wide coverage of the traffic network, while the challenge is that they are mobile and the data captures are randomly distributed in the space, while in time randomness can be reduced by sampling policies (e.g. reporting every Δt time units, as in GPS data collection, for example), or can be a continuous tracking. The research questions addressed in this paper are then, find how the Kalman Filtering framework developed to deal with ICT data can be adapted to tackle with the new data and whether the computational burden doesn't hinders the possibility of the real-time applications.

2. Proposed approach

The scenario addressed in this research is a traffic network whose potential data sources are illustrated schematically in Figure 2, and will be:

- Inductive loop detectors supplying: Traffic volumes (in vehicles/hour), Occupancies (in time percent) and Spot speeds (in Km/h). With a given time resolution (e.g. 1 minute, 5 minutes...).
- Magnetometers, supplying: Individual vehicle detection, that can be aggregated to estimate traffic volumes at the desired time resolution, Individual vehicle's time on the detector, that can be aggregated to estimate percent time occupancy at the desired time resolution, Time headways between successive vehicles passing over the detector.
- Bluetooth/Wi-Fi antennas, supplying: Time continuous sampling of Bluetooth/Wi-Fi devices passing through the detection lobule of the antenna, Travel times of individual devices between two antennas (tt_{AB} = travel time between antenna BT_A and antenna BT_B)

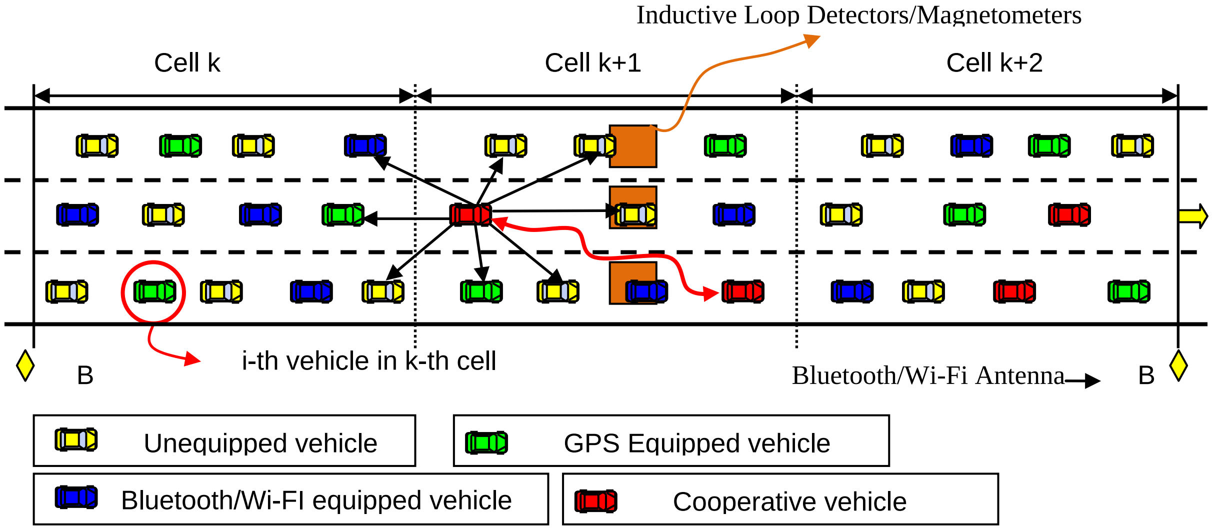

Figure 2: Potential traffic data sources in the research scenario

- Additionally there will be available at the site data from GPS equipped vehicles providing: Vehicle's identity, Time tag (data collected every Δt time units), Map-matched vehicle's position, Speed (in Km/h), Heading (in º from due north; i.e. north = "0").

Physical cooperative vehicles are not yet available in the envisaged scenario, the research plan consists of the microscopic simulation model of an urban scenario in which data collection is conducted emulating the technologies by means of specific ad hoc APIs and scripting. Concerning the Cooperative vehicles they are based on the specifications about the local data related to traffic conditions that cooperative vehicles will collect by the sensors equipping them, making them aware of local traffic conditions surrounding a cooperative vehicle, that is, the approach assumes that cooperative vehicles continuously measure:

- Its position, speed and acceleration (longitudinal and transversal)
- The distances and relative speeds all surrounding vehicles, as depicted in the Figure 3.

Additionally equipped vehicles (red vehicles in Figure 3) exchange this information between them.

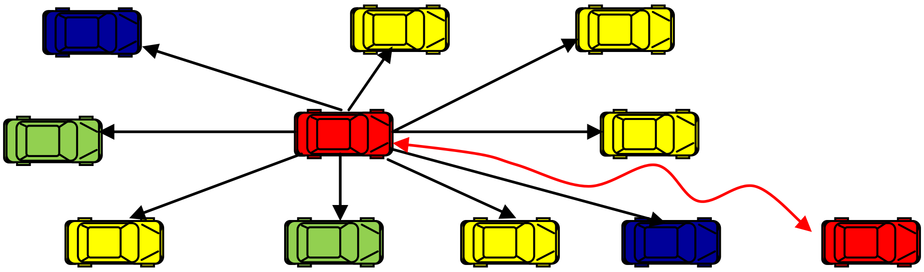

Figure 3: Surrounding "awareness" of an equipped vehicle

Adopting the usual terminology in traffic flow theory let's assume a couple of vehicles, a leader, denoted by l, and a follower, denoted by f. From the cooperative data gathered at time t, assuming that the cooperative vehicle is the red one in Figure 3, the system will get (see also Figure 4):

- The relative distance d_{fl} between the follower and the leader
- The relative speed v_{fl} between the follower and the leader
- The follower's map matched position (x_f,y_f)

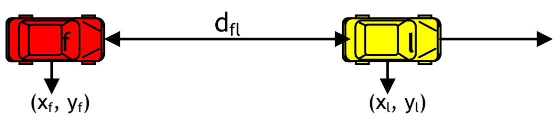

Figure 4: Local traffic related data collected by an equipped vehicle

These are the data that, assuming a hypothetical fleet of cooperative vehicles is emulated, have been used to estimate by simulation, local densities and speeds, and their changes over time, to estimate local traffic state and identify congestion generation (if necessary I can add here a list of related publications), assuming that the penetration of the cooperative technology is large enough.

Data from Cooperative vehicles provide local speed and position at the corresponding time tag allow tracking the trajectory of each cooperative vehicle. Time tags of recorded data allow estimating travel times between pairs of points along the trajectory of the vehicle, this is the characteristic used to adapt the specific Kalman Filter developed in (Barceló et al. 2013a) to exploit the travel times between pairs of BT Antennas, without knowledge of the paths used between antennas. The information about these paths provided by tracking the Cooperative vehicles adds a new feature to the supplied data. How to deal with the subpath information in the framework and the benefits of this new source has been the aim of this research.

3. Model Formulation

Cooperative cars are mobile sources and the network framework will be considered as covered by Virtual Cooperative Car sensors (VCC sensors). Each VCC sensor is referred to a junction or to a link section in dense urban areas. For test site with long links several VCC sensors have been proposed providing a complete and non-redundant coverage of the network links. A threshold parameter has been defined as the maximum length in a section to be covered by a virtual cooperative sensor (for the current tests it is set to 130 m). VCC sensor might have an active or non-active state, VCC data is considered for active states and the active state occurs when a Cooperative vehicle in the coverage area of the virtual sensor. We need to rely on this activity state since only active sensors are considered in the KF approach that we have developed, this allows us to skip the computational burden related to sensor definition. Of course VCC on/off activity is a dynamic feature that can vary each interval (an interval is usually considered between 0.5sec to 3 min). Cooperative vehicle OD data are assumed to be unknown.

The total number of origin and/or destination centroids (related to transportation zones) is I; the total number of ICT sensors is P, located at in the inner network; the total number of VCC sensors is V; and the total number of paths corresponding to the most likely used paths obtained by a DUE (Dynamic User Equilibrium) from the historic OD matrix in the horizon of study is K. QR is the number of pairs of ICT sensors *(r,s) plus* individual sensor counts *qr* to be **considered.** QV is the total number of pairs of VCC sensors *(r,s) plus* individual VCC sensor counts *qv*, only those active will be considered in a given interval *k* in the KF equations. Q is the number sensors (ICT real sensors plus VCC virtual sensors). We finally estimate OD trips between OD pairs for all cars, but OD path trips for ICT equipped cars and Cooperative car has to be internally considered and the percentage of ICT equipped vehicles for each sensor technologyas , not between transit-stops. 5

Table 1. Definition of model variables

$\tilde{Q}_l(k)$, $\tilde{q}r_l(k)$, $\tilde{q}v_l(k)$:	Historic total number of cars, Cooperative cars and ICT equipped cars first detected at *r*, related to centroid *i* in flow conservation *l* at time interval *k*.
$Q_l(k)$, $qr_l(k)$, $qv_l(k)$:	Total number of cars, Cooperative cars and ICT equipped cars first detected at *r*, related to centroid *i* in flow conservation *l* at time interval *k*.
$\tilde{y}_q(k)$, $y_q(k)$:	Historic and actual number of cars crossing at time interval *k* the *q* sensor or destination sensor *s, for a pair of either ICT o VCC sensors q=(r,s)*
$G_{ije}(k)$, $\tilde{G}_{ije}(k)$ $g_{ije}(k)$, $\tilde{g}_{ije}(k)$:	Total number of current $G_{ije}(k)$ and historic $\tilde{G}_{ije}(k)$ cars, as well as the current $g_{ije}(k)$ and historic $\tilde{g}_{ije}(k)$ ICT equipped cars accessing centroid *i* at time interval *k* and heading towards *j* using path *e*.
$\Delta G_{ije}(k)$:	State variables are deviates of cars from centroid *i* during interval *k* headed towards centroid *j* using path *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 / Cooperative cars during interval *k*, a column vector of dimension Q (counts) plus L (flow conservation equations). Expansion equations.
$u_{rs}^{h}(k)$:	Fraction of cars that require *h* time intervals to reach sensor *s* at time interval *k from sensor r* during time interval $[(k-h-1)\Delta t \quad , \quad (k-h)\Delta t]$. Time-varying model parameters. Obtained from ICT equipped and/or Cooperative cars
$\bar{t}_{rs}(k)$:	Average measured travel time for ICT equipped and/or Cooperative that reach sensor *s* at time interval *k from sensor r*

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

$$\mathbf{\Delta G}(k+1) = \sum_{w=1}^{r} \mathbf{D}(w)\mathbf{\Delta G}(k-w+1) + \mathbf{w}(k), \qquad (1)$$

where $\mathbf{w}(k)$ is zero mean with diagonal covariance matrix $\mathbf{W_k}$, and $\mathbf{D}(w)$ are transition matrices which describe how previous OD deviates $\mathbf{\Delta G_{ije}}(k-w+1)$ affect current flows $\mathbf{\Delta G_{ije}}(k+1)$ for $w = 1,\ldots,r$. In our tests, we assume simple random walks to provide the most flexible framework for state variables, if no convergence problems are detected. Thus, our first trial is *r=1*, and $\mathbf{D}(\mathbf{w})$ becomes the identity matrix.

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

- The number of ICT equipped cars first detected in the system at real sensor *r*, related to origin zones through explicit flow conservation equations *l* during time intervals in *k* ,..., *k-M*, $qr_l(k)$.
- The number of Cooperative cars first detected in the system at virtual sensor *r*, related to origin zones through explicit flow conservation equations *l* during time intervals in *k* ,..., *k-M*, $qv_l(k)$.
- The number of cars related to origin zones *i* during time interval *k*, $q_l(k)$.
- H<M *time*-varying *model parameters* in form of *fraction matrices*, $\left[u_{rs}^{h}(k)\right]$, where the H adaptive fractions are updated from measures provided by ICT sensor and Cooperative car data. 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*.

$$\Delta\mathbf{z}(k) = \mathbf{F}(k)\Delta\mathbf{G}(k) + \mathbf{v}(k), \qquad (2)$$

where $\mathbf{v}(k)$ is white Gaussian noise with covariance matrix $\mathbf{R_k}$. $\mathbf{F}(k)$ maps the state vector $\Delta\mathbf{G}(k)$ onto the current blocks of measurements at time interval *k*: counts of cars at sensors, either ICT or VCC active sensors, accounting for time lags and congestion effects.

The solution should provide estimations of the OD matrix between OD pairs for each time interval up to the *k-th* interval, once observations of equipped and cooperative cars up to the *k-th* interval are available. KF prediction of OD trips up to some intervals ahead must be considered, in order to feed a dynamic transit assignment tool that will provide the forecasted travel times and traffic variables in the short-future.

The key aspect relies on how to combine Cooperative car data for subpaths (counts) to OD path flows affected by time-varying parameters. Time-varying parameters related to travel times from ICT equipped cars have to be updated with considering all the possible paths between a pair of sensors, but time-varying parameters updated from Cooperative cars data are updated knowing the exact used route that can affect to several OD path travel times.

Figure 5 shows the flowcharts between the dynamic estimation KFX4 tool and the emulation of data. We have completed development of API emulating the cooperative data flow, the KFX4 filtering tool implemented in MATLAB is in process of programming. Once available, the estimation and forecasting tool (KFX4) could be validated using medium size network, the testing of the filtering tool is being performed using Vitoria’s network.

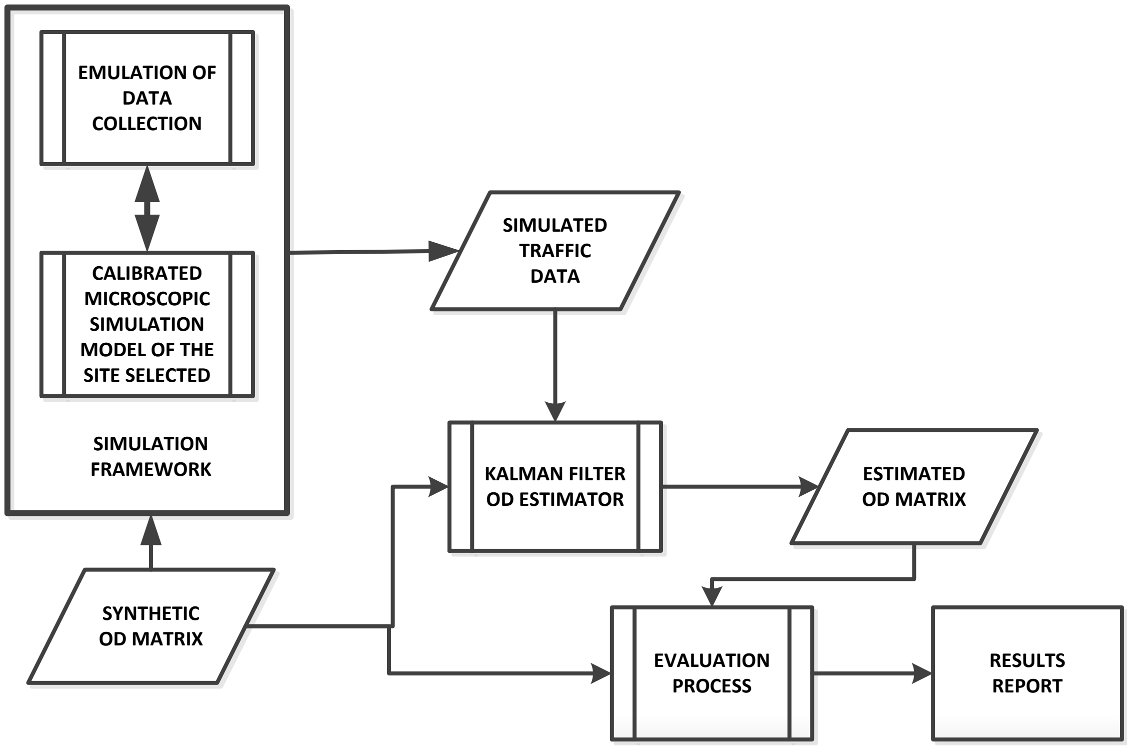

Figure 5. Conceptual Framework for the Computational Experiments

4. Acknowledgments

The research is funded by TRA2011-27791-C03-02 of the Spanish R+D National Programs.

5. References:

1. 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).
2. 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).
3. Barceló, J., Montero, L., Bullejos, M., Linares, M.P., Serch, O. : Robustness And Computational Efficiency Of A Kalman Filter Estimator of Time Dependent OD Matrices. Transportation Research Record 2344-04 pp.31-39 (2013a).
4. Barceló, J., Montero, L., Bullejos, M., Serch, O., Carmona, C.: A Kalman Filter Approach for the Estimation of Time Dependent OD Matrices Exploiting Bluetooth Traffic Data Collection. In: GITS Journal of Intelligent Transportation Systems, vol. 17(2) pp.123-141 (2013b).
5. MatLab (R2012a) by MathWorks, Inc. (www.mathworks.com) (2012).
6. J. Barceló, L. Montero, M. Bullejos, M.P. Linares, (2014b). 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, included in the CD-Compendium of Papers
7. M.Bullejos, J.Barceló, L.Montero (2014), A DUE based bilevel optimization approach for the estimation of time sliced OD matrices, International Symposium of Transport Simulation 2014, Procedia Social and Behavioral Sciences.
8. Cascetta, E. *Transportation Systems Engineering: Theory and Methods.* Kluwer Academic Publishers, Norwell, Mass., 2001