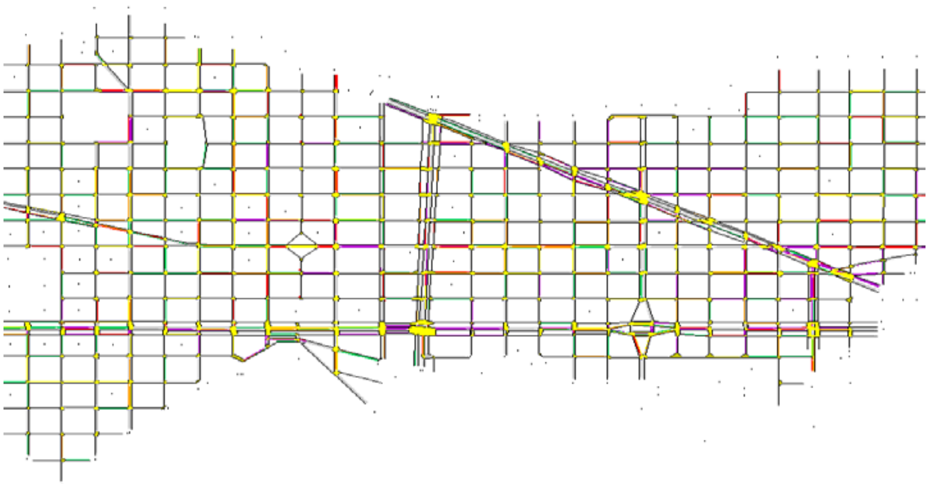


A Dynamic Estimation of Passenger OD Matrices based on space-state models

Authors: **L. Montero** and E. Codina
BarcelonaTech (UPC)

EURO 2013

(EURO-INFORMS Joint International Meeting)
Rome (Italy) July 1st-4th, 2013

Contents (I)

1. The aim of this work is …(I)

To explore the **design and implementation of efficient methods** to support the short-term and **real-time estimation of time dependent Origin-Destination Trip/Passenger matrices as long as** new detection technologies complement the traditional ones.

- This is the case of the new sensors detecting vehicles equipped with Bluetooth mobile devices, i.e. hands free phones, Tom-Tom, Parrot and similar devices. **AVI (Authomatic vehicle identification) technologies.**

- Since Real-time application in ATMS require efficient and robust estimation of dynamic OD matrices (time-sliced OD matrices).

L.Montero and E. Codina

1. The aim of this work is …(II)

To explore the **design and implementation of efficient methods** for **real-time estimation of time dependent Origin-Destination Trip/Passenger matrices as long as** new detection technologies complement the traditional ones.

- Therefore, from a research stand point this means starting to explore **the potential of new technologies in simplifying traffic/transit models involved in matrix estimation.**

- **Additionally for practitioners, we provide sound** applications, easy to implement, exploiting AVI technologies, as Bluetooth, **given that** penetration rates are increasing.

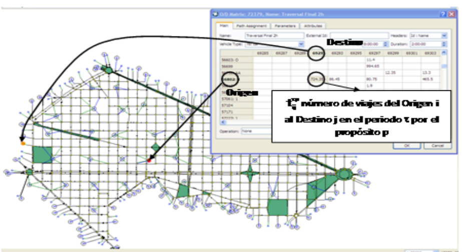
L.Montero and E. Codina

1-iii. Traffic state estimation at time τ from data fusion and traffic models accounting for traffic dinamics

TRAFFIC DATA FROM SENSORS → DATA FILTERING AND FUSION → DINAMIC TRAFFIC MODELS → URBAN NETWORK: TRAFFIC STATE

Author: Jaume Barceló

Data profile for time-period τ



HISTORICAL OD MATRIX ($t_{ij}^{p\tau}$) TIME PERIOD τ

Kalman filter model for dinamic OD matrix estimation for time period τ

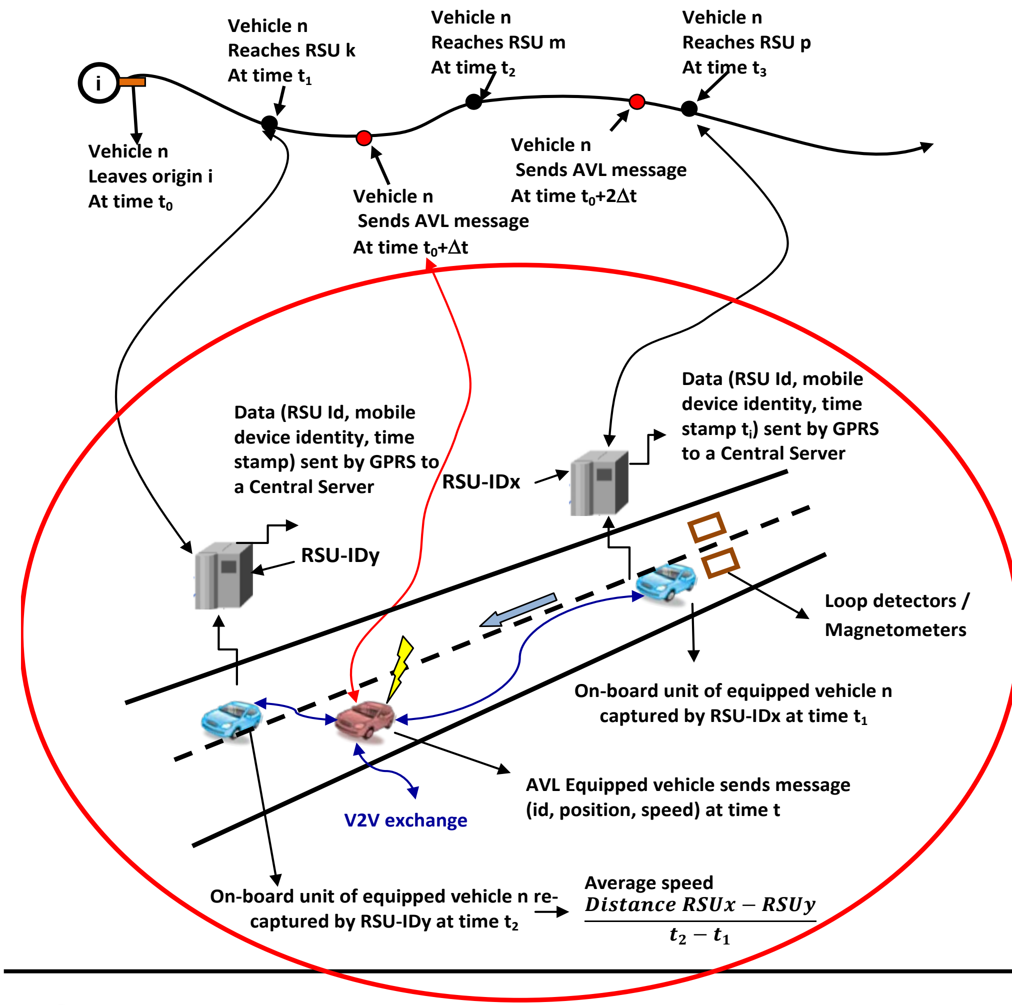
OD MATRIX FOR TIME PERIOD τ → MESOSCOPIC TRAFFIC SIMULATOR

STATE COMPLETION AND FORECASTING

Current TIS in BCN website (only measures) : non complete
Advance ATIS in(measures + fusión + models): complete description

L.Montero and E. Codina

1.iv-EXAMPLES OF DATA COLLECTION: Travel time between RSUs



FORTHCOMING TECHNOLOGICAL PLATFORM

DATA COLLECTION FROM:

- ETD (loops, magnetometers)
- EQUIPPED VEHICLES (FCD) GPS/GPRS, AVL, TAG…
- CCTV/LPR
- MOBILE DEVICES (Bluetooth)
- V2I TECHNOLOGIES
- V2V TECHNOLOGIES

GENERATING CONSISTENT AND HOMOGENEOUS DATA

- IDENTIFYING AND FILTERING OUTLIERS
- MISSING DATA MODELS
- DATA FUSION FROM HETEROGENOUS SOURCES

L.Montero and E. Codina

Contents

L.Montero and E. Codina

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
Departament d'Estadística i Investigació Operativa

2.i- A State Space Model approach

Traditional Statistical Model

- The parameter vector θ determines the distribution of the observed vector Y.
- A sample $y_1,\ldots,y_P$ is used to estimate θ.

$$Y \approx F\left(y \,|\theta\right)$$

$$\hat{\theta} = S\left(y_1,\ldots,y_n\right)$$

State Space Model

- The hyperparameter vector θ determines the distribution of the unobserved **state vector *X*.**
- The state vector ***X*** and the hyperparameter vector θ determine the distribution of the **observed vector *Y*.**
- A sample $y_1,\ldots,y_P$ is used to estimate θ.
- The estimated value of θ and the sample $y_1,\ldots,y_P$ are used to estimate $\hat{X}$.

$$X_i \approx g\left(x_i|\theta\right)$$

$$\hat{\theta} = S\left(y_1,\ldots,y_P\right)$$

$$Y_p \approx F\left(y_p\middle|\mathrm{X},\theta\right)$$

$$\hat{X}_i = T\left(y_1,\ldots,y_P,\hat{\theta}\right)$$

L.Montero and E. Codina

2.i- Why A State Space Model ?

State Space Model: hyperparameters	
Equipped passengers	All passengers

Unobserved States	
OD Pattern/OD Flows of passengers	OD Pattern/OD Flows of passengers

Observations	
counts, travel times for equipped passengers	counts on entries for all passengers

L.Montero and E. Codina

2.i- Kalman Filtering: State variables and Observations

- **State variables g(k)** constitude an stochastic non-white noise process (i.e., AR(r)) where time evolution is affected by a white Gaussian noise (w(k), assumed with zero mean).

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

- **State variables can not be measured**, however they are **related to measurements z(k)**, again affected by a white Gaussian noise (assumed with zero mean).

- **Measurements z(k)** constitude an stochastic non-white sequence.

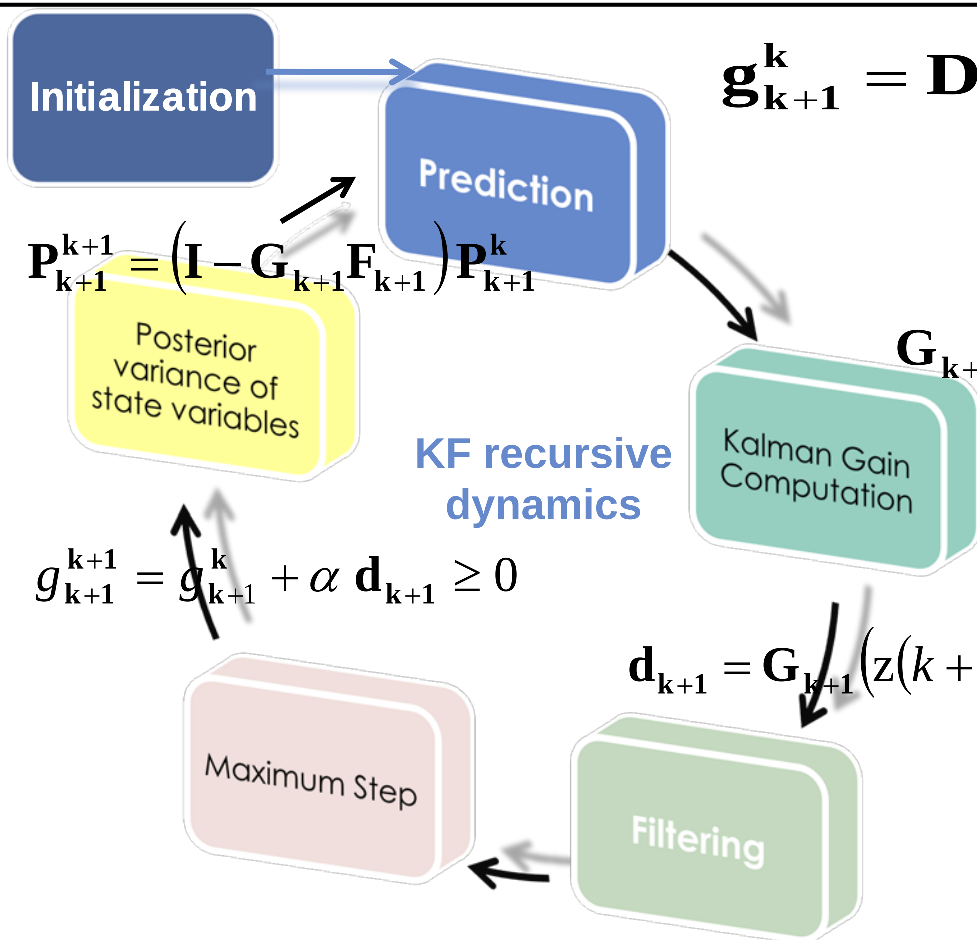
$$z(k) = \mathbf{F(k)\,g(k)} + v(k)$$

- White noise in state w(k) and observation v(k') equations are statistically independent for any k, k'.

2.i- Linear Kalman Algorithm

$$\mathbf{g}^{\mathbf{k}}_{\mathbf{k+1}} = \mathbf{D}\mathbf{g}^{\mathbf{k}}_{\mathbf{k}} \quad \mathbf{P}^{\mathbf{k}}_{\mathbf{k+1}} = \mathbf{D}\mathbf{P}^{\mathbf{k}}_{\mathbf{k}}\mathbf{D}^{\mathbf{T}} + \mathbf{W}_{\mathbf{k}}$$

What can be said before the observations are considered?



$$\mathbf{P}^{\mathbf{k+1}}_{\mathbf{k+1}} = \left(\mathbf{I} - \mathbf{G}_{\mathbf{k+1}}\mathbf{F}_{\mathbf{k+1}}\right)\mathbf{P}^{\mathbf{k}}_{\mathbf{k+1}}$$

$$\mathbf{G}_{\mathbf{k+1}} = \mathbf{P}^{\mathbf{k}}_{\mathbf{k+1}}\mathbf{F}^{\mathbf{T}}_{\mathbf{k+1}}\left(\mathbf{F}_{\mathbf{k+1}}\mathbf{P}^{\mathbf{k}}_{\mathbf{k+1}}\mathbf{F}^{\mathbf{T}}_{\mathbf{k+1}} + \mathbf{R}_{\mathbf{k}}\right)^{-}$$

What can be improved after the observations are considered?

$$g^{\mathbf{k+1}}_{\mathbf{k+1}} = g^{\mathbf{k}}_{\mathbf{k+1}} + \alpha\, \mathbf{d}_{\mathbf{k+1}} \geq 0$$

$$\mathbf{d}_{\mathbf{k+1}} = \mathbf{G}_{\mathbf{k+1}}\left(z(k+1) - \mathbf{F}_{\mathbf{k+1}}\mathbf{g}^{\mathbf{k}}_{\mathbf{k+1}}\right)$$

Innovations:

$$\boldsymbol{\varepsilon}_{\mathbf{k+1}} = \left(z(k+1) - \mathbf{F}_{\mathbf{k+1}}\mathbf{g}^{\mathbf{k}}_{\mathbf{k+1}}\right)$$

L.Montero and E. Codina

2.ii- Proposed state and measurement variables for OD passenger matrix estimation

- **State variables g(k)** are Origin-Destination passenger flows in a subset of Most Likely OD paths in the transit network. Usually those represented by path-finder transit assignment solutions.

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

- **State variables can not be measured**, however they are **related to measurements z(k)**.

- **Measurements z(k)** are counts of equipped passengers at Wifi Bus-Stops

$$z(k) = \mathbf{F(k)}\,\mathbf{g(k)} + v(k)$$

L.Montero and E. Codina

Contents

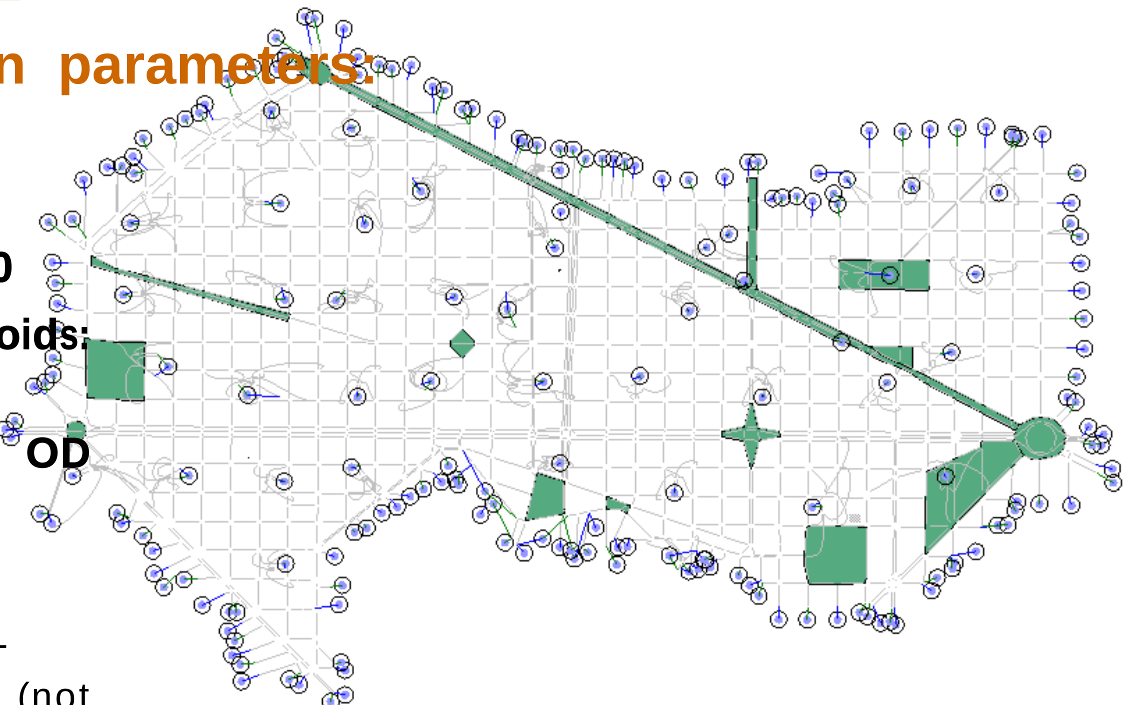
L.Montero and E. Codina

EXPERIMENT 3: BARCELONA'S CBD AIMSUN(©TSS) MODEL - Eixample District

Aimsun is a transport modelling software for real-time traffic simulation

Model dimension parameters:

- **OD PAIRS: 877**
- **Links: 2108**
- **Origin centroids: 120**
- **Destination centroids: 130**
- **2954 Most likely OD paths**
- **Sensors: 281**
  - Interior sensors: 151
  - Exit Gates: 130 (not used)



L.Montero and E. Codina

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística i Investigació Operativa

Contents

L.Montero and E. Codina

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
Departament d'Estadística i Investigació Operativa

4. i- A Linear State Space Model approach for dynamic passenger matrices (I)

- "**Public transport agencies have traditionally been hampered in planning, managing and evaluating their services by having to rely heavily on costly and unreliable manual data collection systems**" … (Zhao, Rahbee and Wilson, *Computer-Aided Civil and Infrastructure Engineering 22 (2007) 376–387).*
- **The general types and characteristics of ADC (Authomatic Data Collection) systems are classified as:**
  - AVL (Authomatic Vehicle Location) and tracking systems,
  - AFC (Authomatic Fare Collection) systems,
  - and APC (Authomatic Passenger Counting) systems
- **AFC systems are not always suitable for OD matrix estimation (usually passengers' exit points are not registered).**
- **APC does not differentiate between passengers, *just* counts, and it is expensive in terms of equipment when cameras are used.**

L.Montero and E. Codina

4. i- A Linear State Space Model approach for dynamic passenger matrices (II)

- **Recent static and non-real time applications are:**
  - An OD matrix inference from origin-only data has been addressed in the analysis of the Chicago Rail System using Authomatic Payment Systems by Zhao et al. (2007), amongst other authors as Cui (2006), Trepanier et al. (2007), and Barry et al. (2008), all them use trip-chaining methods with similar assumptions :
    - There is no private transportation mode trip segment.
    - Passengers will not walk a long distance
    - Passengers end their last trip of the day where they began their first trip of the day.
  - Wang et al (Journal of Public Transportation Vol 14, no 4, 2011) used Oyster smart card transactions and Automatic Vehicle Location (AVL) to *Bus Passenger Origin-Destination Estimation and Related Analyses* in London.
  - Jang (2010) further examined the possibilities of using the ADCS archived data for public transport planning in Seoul, South Korea. One feature that distinguishes the Seoul ADCS from many other cities is that it records each trip’s entry and exit times and locations.

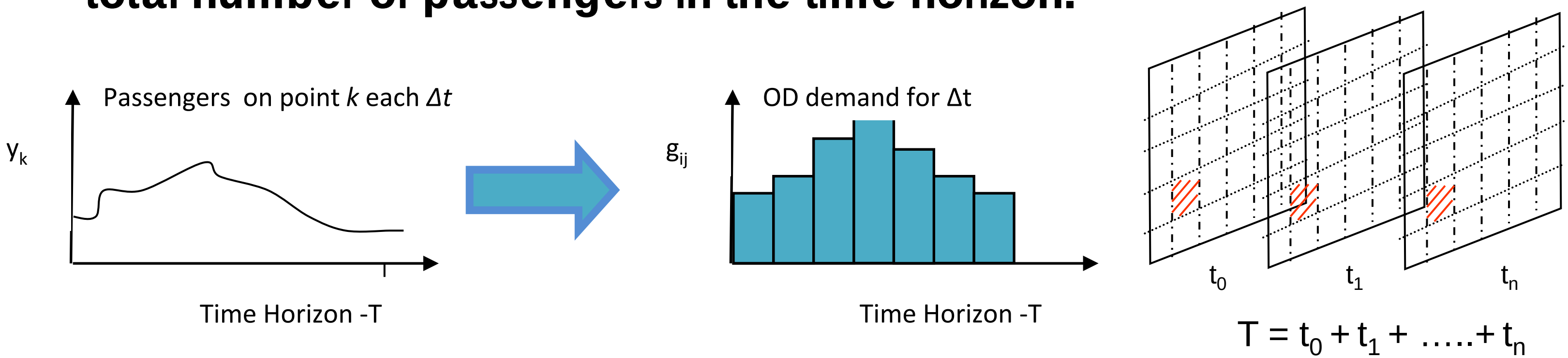
L.Montero and E. Codina

4. i- A Linear State Space Model approach for dynamic passenger matrices (III)

- **(Cont.) Recent static and non-real time applications are:**
  - Kostakos (2006, 2008) proposed a novel use of passengers' smartphones and BT antennas in bus units as a mean of capturing OD matrices in a field test project in Madeira.
  - To capture passenger trips on buses onboard cameras have also been used and automated head detection applied. But it is expensive.
- **We propose to use passengers' smartphones and BT antennas in bus-stops to estimate dynamic passenger matrices from origin to destination transportation zones:**
  - Expansion of BT data is not addressed in this work and a common expansion factor according to historical BT penetration rate in the population is considered for simplicity.
  - We are concerned with modeling passenger behavior based on the concept of strategy. At each possibly reached transit stop, the attractive line/s can be selected based on minimizing travel cost (**path-finder strategies).**

4. i- Formulation proposal for dynamic OD passenger matrices

- **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).**

- **The solution should provide estimations of the OD passenger matrices between network zones for each time interval up to the *k*-th interval once observations of BT equipped passengers upon to the k-th interval are available.**

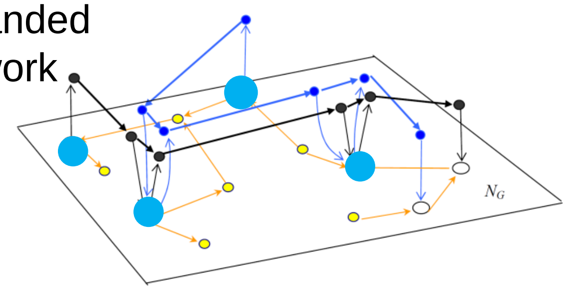
L.Montero and E. Codina

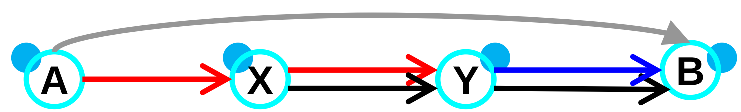
4. i- Formulation proposal for dynamic OD passenger matrices

State variable definition:

- **We propose to use deviates of state variables to include *a priori* structural information, and simplify properties of space-state-models based on Kalman filtering:** $$\Delta g_{ije} = g_{ije} - g_{ije}^{H}$$
- **Wifi antennas are proposed to be located on bus-stops:**
  - Only origin-destination trips for passengers in transit lines whose equipped stops, those covered with ICT sensors are observable.
  - Interferences with not covered transit-lines in some stops are not considered in this first approximation.

Expanded Network





L.Montero and E. Codina

4. i- Formulation proposal for dynamic OD passenger matrices

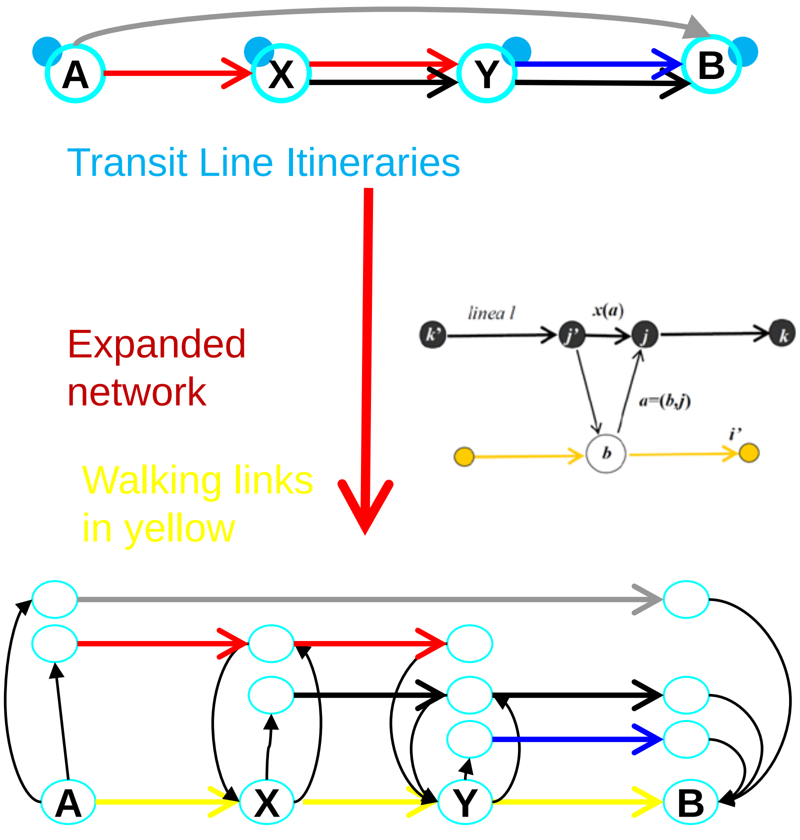
- **BT data is non-biased according to some studies as Rescott 2011** ***Feasibility of Bluetooth Data as a Surrogate Analysis Measure of Traffic*** **Degree thesis on Civil Engineering, Univ of Kansas.**
- **Either Historical profiles (for day-type and time-period) have to be used to expand BT samples of equipped passengers to the total number of passengers.**
- **Or to expand the BT sample a passenger counting system that records the comings and goings of passengers automatically, and precisely (less than 5% error according to Barcelona's tests, FGC) has to be considered in transit-units :**
  - The sensors are installed unobtrusively over the vehicle doors and deliver reliable passenger count data such as boarding/alighting passengers. One PCU (people counting unit) can have up to 16 connected elements and control up to 6 doors by means of a serial sensor link (cable).
  - Passenger count data are transmitted by cable or wirelessly to the data management system, via Ethernet, data bus, WLAN, GSM, GPRS or UMTS.

L.Montero and E. Codina

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
Departament d'Estadística i Investigació Operativa

4. i- Expanded Network Model (Graph)

State variables are noted as Δg_{ije} and are defined as deviations of OD passenger flows on most-likely paths *e* **from origin *i* to destination *j*** ***for smartphone equipped passengers*.**



An optimal strategy is represented as an hyperpath (set of paths) in the expanded transit network for the transit line itineraries in a network.

A simplified subset of **K-Shortest paths in the expanded network are stated as *state variables* (MLPaths,** most-likely paths)**.**

L.Montero and E. Codina

4. ii- KF Approach : simple path flow deviations on optimal strategy for OD as state variables (I)

$\tilde{g}_{ije}(k)$: Historical flow of equipped passengers on most-likely paths for i-j OD; i.e., first boarding on a transit-stop belonging to centroid *i* at time interval *k* and headed towards destination zone *j (the trip might have several transfers). Equal share among simple paths.*

***Most likely paths* are assumed to approximate optimal strategy transit assignment.**

Proportions assigned to most-likely paths constitute an output of the linear KF proposal.

$\Delta g_{ije}(k)$

$\Delta z_q(k)$

Deviates of OD flows on most-likely paths as state variables and deviates of counts on ICT sensors as observation equations

$\tilde{z}(k)$: The historic observation variables during interval k on **passenger counts on transit-stops** for time interval *length* Δt; **according to current time-varying model parameters**, i.e. a column vector of dimension Q+I, whose structure is

$$\mathbf{\Delta z(k) = z(k) - \tilde{z}(k) = F(k)}\left(\underbrace{\mathbf{g(k) - \tilde{g}(k)}}_{\mathbf{\Delta g(k)}}\right) + \mathrm{v(k)}$$

L.Montero and E. Codina

4. ii- KF Approach : simple path flow deviations on optimal strategy for OD as state variables(II)

Networks

- Most-likely paths are considered to approximate the optimal strategy for each selected OD pair and OD path flows are assumed to change smoothly.

Depends On experienced traffic state

- H(t) BT Travel Times
- Time-varying model parameters

Travel times to be considered

- **Several Δts (between 5 and 15 min) are to be considered: time lag effect is explicitly included in the formulation.**

State vector equations AR(r) on deviates:

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

Observation vector equations:

$$\Delta z(k) = \begin{pmatrix} \boxed{\mathbf{H(k)}} \\ \mathbf{E} \end{pmatrix} \Delta \mathbf{g(k)} + \begin{pmatrix} v_1(k) \\ v_2(k) \end{pmatrix} = \mathbf{F(k)}\Delta\mathbf{g(k)} + v(k)$$

<u>First block</u>: deviates of counts on sensor points

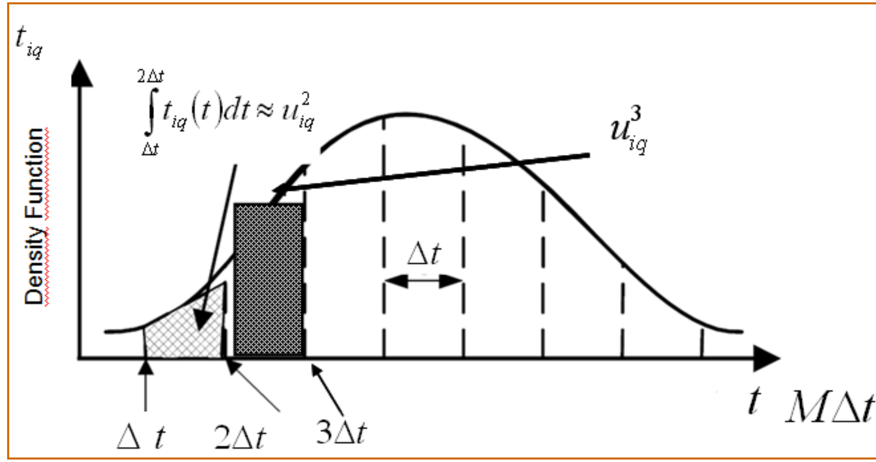
$$\mathbf{H(k)}\,\Delta\mathbf{g(k)} = \mathbf{AU(k)}^{T}\left(\mathbf{g(k)} - \tilde{\mathbf{g}}\mathbf{(k)}\right) \approx \left(y(k) - \tilde{y}(k)\right)$$

<u>Second block</u>: conservation flows for each entry and time interval k

L.Montero and E. Codina

4. iii- KF Approach : Time varying model parameters from ICT data – Travel times from equipped passengers (III)

- Travel time distribution of T_{iq} of passengers first boarding at any transit-stop at origin zone i in time interval k are headed during their journey through *sensor q*.
- **Approximate travel times distribution by discrete distribution with interval proportions updated according to on-line ICT data.**
- **Interval** proportions are considered time-varying model parameters: structural constraints to sum 1 are required.
  - Fraction of passengers that require *h* time intervals to reach sensor *q* at time interval k first boarding at centroid (transport zone) *i*.



$$u_{iq}^h(k) \geq 0 \quad i = 1\ldots I, \quad q = 1\ldots Q, h = 1\ldots H$$
$$\sum_{h=1}^{H} u_{iq}^h(k) = 1 \quad i = 1\ldots I, \quad ,q = 1\ldots Q,$$

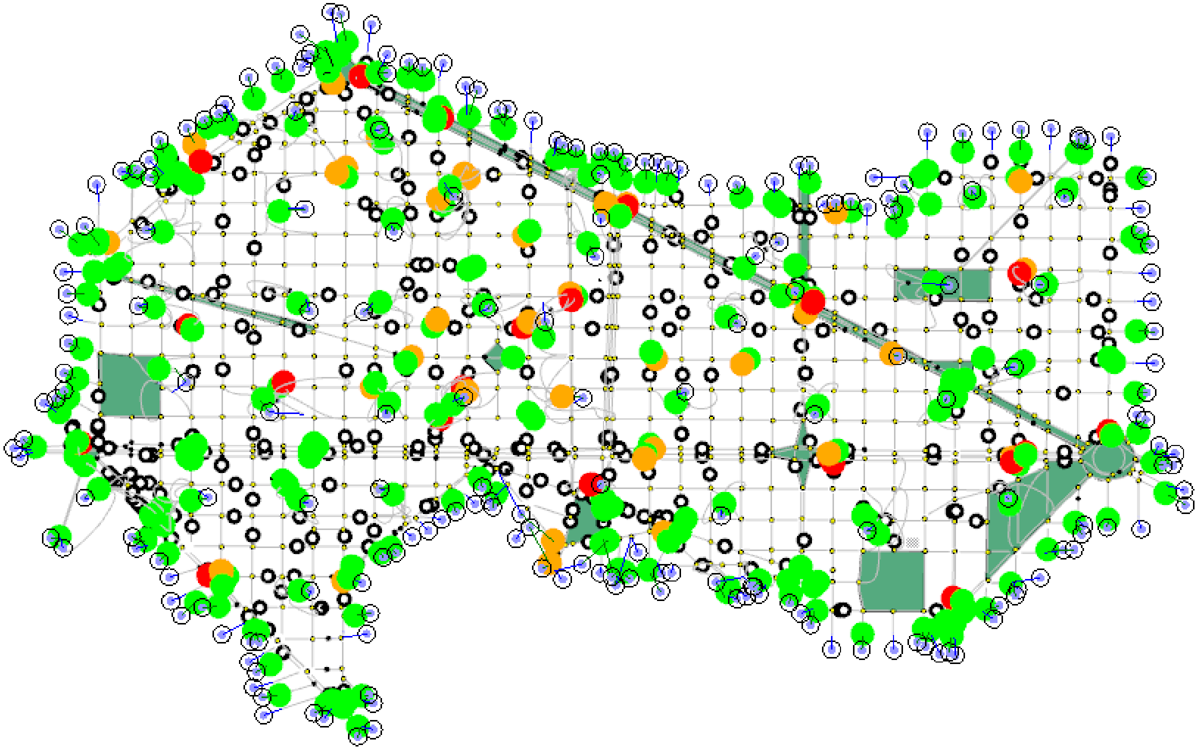
L.Montero and E. Codina

5. Conclusions and further research (I)

- A **linear KF formulation** for the dynamic estimation of OD passenger matrices for urban transit networks is being implemented in MatLab and tested by simulation.
- Further research **points to** decision methods to define detection layout **to be suitable to dynamic OD estimation in multimodal urban transit networks**.
- For railway networks a strategy-based transit assignment is not realistic since passengers do not arrive randomly at time to the stations. The formulation should be revised to deal with fix scheduled transit services.
- Formulation takes into account deviates of passenger flows on most likely paths as state variables and it is suitable for application to urban and regional transit networks.
- The **availability of ICT based measurements** allows a problem's formulation as a linear Kalman Filter that incorporate travels times and thus general traffic state and dinamics.

L.Montero and E. Codina

Thank you very much for your attention!



This research is funded by project TRA2011-27791-C03-02 of the Spanish R+D National Program.

L.Montero and E. Codina