Dynamic OD passenger matrix estimation: formulation and model-building environment

Authors: L. Montero, E. Codina and J. Barceló
Barcelona TECH - UPC (Technical University of Catalonia)

IC SEng 2014
(23rd INTERNATIONAL CONFERENCE ON SYSTEMS ENGINEERING)
LAS VEGAS, Nevada, August 19-21, 2014
Contents

1. Aim of this work

2. A Linear State Space Model approach

3. A Linear Kalman filter formulation for dynamic passenger matrices

4. KFX3T Prototype Testing by simulation

5. Automatization of KFX3T Model Building

6. Conclusions and further research
1. Aim of this work (I)

To explore the **design and implementation of efficient methods** to support the **short-term and real-time estimation of time dependent Origin-Destination matrices for transit trips** when, new detection technologies (**ICT**) complete the traditional ones:

- This is the case of the new sensors detecting **passengers equipped with Bluetooth mobile devices**,
  - Or for auto trip matrices, sensors detecting vehicles equipped with Tom-Tom, Parrot and similar devices allowing **AVI (Automatic vehicle identification) technologies**.

- From a research standpoint this means starting to explore the **potential of ICT new technologies in simplifying transportation models** and develop new applications for practitioners.
1. Aim of this work: Macro, meso and micro approaches to traffic modeling for private trips...

- Different types of network representations.
- All them share in common the demand model in terms of Origin-Destination Trip Matrices:
  - the number of trips from the selected origin to the chosen destination,
  - for a given time period
  - for a given trip purpose/mode
- Origins and destinations are represented in models in terms of artificial nodes, or centroids, where traffic flows are generated and sunk.

Source: J.Barceló – Sedona Workshop 2005
1. Aim… A main input to ATIS/ATMS models: real-time dynamic Origin to Destination (OD) matrices

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

Data Profile for period $\tau$

HISTORICAL OD MATRIX ($t_i^{\text{DT}}$)
TIME PERIOD $\tau$

Kalman filter model for dinamic OD matrix estimation for time period $\tau$

OD MATRIX FOR TIME PERIOD $\tau$

MESOSCOPIC TRAFFIC SIMULATOR

Current TIS in BCN website (only measures): non complete

Advance ATIS in (measures + fusión + models): complete description
Contents

1. Aim of this work

2. A Linear State Space Model approach

3. A Linear Kalman filter formulation for dynamic passenger matrices

4. KFX3T Prototype Testing by simulation

5. Automatization of KFX3T Model Building

6. Conclusions and further research
2. A State Space Model approach

Traditional Statistical Model

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

\[
Y \approx F(y \mid \theta) \\
\hat{\theta} = S(y_1, \ldots, y_n)
\]

State Space Model

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

\[
X_i \approx g(x_i \mid \theta) \\
\hat{\theta} = S(y_1, \ldots, y_P) \\
Y_p \approx F(y_p \mid X, \theta) \\
\hat{X}_i = T(y_1, \ldots, y_p, \hat{\theta})
\]
2. Why A State Space Model?

- **State Space Model: hyperparameters**
  - Equipped vehicles/passengers
  - All vehicles/passengers

- **Unobserved States**
  - OD Pattern/OD flows
  - OD Pattern/OD flows

- **Observations**
  - Counts, travel times for equipped vehicles/passengers
  - Counts at traditional sensors for all vehicles
  - Passenger counts at line-stops (off-line)
  - Passenger counts at some line-stops (on-line)
Contents

1. Objective of this work

2. A Linear State Space Model approach

3. A Linear Kalman filter formulation for dynamic passenger matrices
   i. Definition of state variables and observation variables.
   ii. Statement of the equations of the KF linear formulation.
   iii. Time varying model parameters: travel time discrete distributions from BT data

4. KFX3T Prototype Testing by simulation

5. Automatization of KFX3T Model Building

6. Conclusions and further research
3. A Linear Kalman filter 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 (Automatic Data Collection) systems are classified as:
  - AVL (Automatic Vehicle Location) and tracking systems,
  - AFC (Automatic Fare Collection) systems,
  - and APC (Automatic 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.
Recent static and non-real time applications are:

- An OD matrix inference from origin-only data has been addressed by several authors, all of them use trip-chaining methods with similar assumptions:
  - Zhao et al. (2007), Cui (2006), Trepanier et al. (2007), and Barry et al. (2008)


- Jang (2010) further examined the possibilities of using the ADCS archived data for public transport planning in Seoul, South Korea. Seoul ADCS records each trip’s entry and exit times and locations.

- Kostakos (2006, 2008) proposed a novel use of passengers’ smartphones and BT antennas in bus units as a mean of capturing off-line OD matrices in a field test project in Madeira.
3. A Linear Kalman filter formulation for dynamic passenger matrices (IV)

- We propose to use passengers’ smartphones and Wifi/BT antennas in bus-stops to estimate dynamic passenger matrices from origin to destination transportation zones:

- Expansion of Wifi/BT data is not addressed in this work and a common expansion factor according to historical 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 set of attractive lines can be selected in such a way that the strategy’s expected travel cost is minimized (optimal or shortest strategies).
3. A Linear Kalman filter formulation for dynamic passenger matrices (V)

- 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 $\Delta t$ (between 5 and 10 minutes for passenger matrices).

- The solution should provide estimations of the OD passenger matrices between network zones for each time interval up to the k-th interval and forecasting of K intervals ahead.
3. A Linear Kalman filter formulation for dynamic passenger matrices (VI)

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: 
  \[ \Delta g_{ije} = g_{ije} - g^H_{ije} \]

- Wifi/BT antennas are proposed to be located at some 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.
Formulation proposal for dynamic OD passenger matrices: use of BT data and APC data

- BT data provide anonymous identification of a sample of passengers and can offer itinerary follow-up of transit stops and travel times.

- APC data on equipped transit line units might provide accurate counts of all passengers at equipped stops (from http://www.dilax.com): less than 5% error according to Barcelona’s tests, done by FGC.

![Autonomous PCU based counting system in a vehicle: DILAX SYSTEM](image-url)
3. A Linear Kalman filter formulation for dynamic passenger matrices

- Optimal Strategies for transit trips can be computed by any transportation planning software including equilibrium transit assignment.

- Transit networks have been modeled with EMME4.

- Transit strategies can be decomposed into simple paths in the expanded transit network. Those simple path flows constitute the state variables in the space-state model proposed.

- Under uncongested transit assignment models optimal strategies do not depend on historic demand for the period of study computed.
Strategies and Hyperpaths

An Hyperpath for any (optimal) strategy can be decomposed into a finite number of acyclic simple paths and the expected cost of an hyperpath is *roughly speaking* the weighted average of its simple path costs.

Flows on physical transit segments
Historical flow of equipped passengers on simple paths belonging to an optimal strategy for i-j OD;

Most likely simple used paths are assumed to be the output of an optimal strategy transit assignment.

Proportions assigned to active paths constitute an output of the linear KF proposal.

Deviates of OD simple path flows on optimal strategy as state variables and deviates of counts on ICT sensors as observation equations

The historic observation variables during interval \(k\) on passenger counts on transit stops for time interval length \(\Delta t\); according to current time-varying model parameters, i.e. a column vector of dimension \(Q+1\), whose structure is

\[
\tilde{z}(k)^T = (\tilde{y}(k) \quad \tilde{q}(k))^T
\]

\[
\Delta z(k) = z(k) - \tilde{z}(k) = F(k) \left( g(k) - \tilde{g}(k) \right) + v(k)
\]
3. KF Approach: simple path flow deviations on optimal strategy for OD as state variables (II)

**State vector equations AR(r) on deviates:**

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

**Observation vector equations:**

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

**First block:** deviates of counts on sensor points

\[ H(k) \Delta g(k) = A U(k)^T (g(k) - \tilde{g}(k)) \approx (y(k) - \tilde{y}(k)) \]

**Second block:** conservation flows for each entry and time interval \( k \)

\[ \Delta g_{ijc}(k) \geq -\tilde{g}_{ijc}(k) \]

\[ q_i(k) = \sum_{j=1}^{J_i} g_{ij}(k) - \sum_{j=1}^{J_i} \sum_{c=1}^{K_{ij}} \Delta g_{ijc}(k) = q_i(k) - \tilde{q}_i(k) \]

**Networks**

- Several simple paths are considered to describe the optimal strategy for each selected OD pair and OD simple path flows change smoothly.
- Several \( \Delta t \)s (between 5 and 15 min) are to be considered: time lag effect is explicitly included in the formulation.

**Depends On experienced traffic state**

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

**Travel times are considered**

- Several \( \Delta t \)s are considered to describe the optimal strategy for each selected OD pair and OD simple path flows change smoothly.

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

- Several \( \Delta t \)s are considered to describe the optimal strategy for each selected OD pair and OD simple path flows change smoothly.

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

- Several \( \Delta t \)s are considered to describe the optimal strategy for each selected OD pair and OD simple path flows change smoothly.

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

- Several \( \Delta t \)s are considered to describe the optimal strategy for each selected OD pair and OD simple path flows change smoothly.

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

- Several \( \Delta t \)s are considered to describe the optimal strategy for each selected OD pair and OD simple path flows change smoothly.

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

- Several \( \Delta t \)s are considered to describe the optimal strategy for each selected OD pair and OD simple path flows change smoothly.

- Several \( \Delta t \)s (between 5 and 15 min) are to be considered: time lag effect is explicitly included in the formulation.
3. KF Approach: Time varying model parameters obtained from ICT data - Travel times from equipped passengers (III)

- Travel times $T_{rs}$ of passengers first boarding at any transit-stop with ICT sensor $r$ headed during their journey through ICT sensor $s$ at time interval $k$ follow a certain distribution, no matter the selected simple path.

- Approximate travel time distribution by discrete distribution with bin proportions updated according to on-line ICT data.

- Bin proportions from each first boarding transit-stop to any measurement point are considered time-varying model parameters: structural constraints to sum 1 are required.

  - Fraction of passengers that require $h$ time intervals to reach ICT sensor $s$ at time interval $k$ firstly captured at ICT sensor $r$ (related to transport zone) $i$.

$$\sum_{h=1}^{H} u_{rs}^h(k) = 1 \quad (r, s) \quad \text{pair of ICT sensors}$$

$$u_{rs}^h(k) \geq 0 \quad h = 1 \ldots H$$
3. A Linear Kalman filter formulation for dynamic passenger matrices: estimation process

\[
g_{k+1}^k = D g_k^k \quad P_{k+1}^k = D P_k^k D^T + W_k
\]

What can be said before the observations are considered?

\[
G_{k+1} = P_{k+1}^k F_{k+1}^T (F_{k+1} P_{k+1}^k F_{k+1}^T + R_k)^{-1}
\]

What can be improved after the observations are considered?

\[
d_{k+1} = G_{k+1} (z(k+1) - F_{k+1} g_{k+1}^k)
\]

Innovations:

\[
\epsilon_{k+1} = (z(k+1) - F_{k+1} g_{k+1}^k)
\]
Contents

1. Objective of this work
2. A Linear State Space Model approach
3. A Linear Kalman filter formulation for dynamic passenger matrices
4. KFX3T Prototype Testing by Simulation
5. Automatization of KFX3T Model Building
6. Conclusions and further research
4. KFX3T Prototype Testing by simulation

Test Network (EMME4)

• Headway L1 and L4 is 15 min and otherwise set to 10 min
• ICT sensors at all-stops (Sx – all nodes and S4 nodes 3, 4, 5 and 7)
• Travel speed is 20km/h
• Boarding time 0
• Horizon 60 min
• 4 Time-slices

<table>
<thead>
<tr>
<th>OD</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>180</td>
<td>120</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>180</td>
</tr>
</tbody>
</table>
4. KFX3T Prototype Testing by Simulation

**Test Network (EMME4)**

- Optimal Strategies Assignment for the global period (static)
- Flow is distributed at origins using a logit model with scale parameter 0.2.

- L1, L3, L4 for destination centroid 8.
- L2, L4 for destination centroid 9.
- Interarrival times of bus units assumed exponentially distributed.
4. KFX3T Prototype Testing by simulation

Test Network (EMME4)

- Optimal Strategies Assignment for the global period (static)
- Flow is distributed at origins using a logit model with scale parameter 0.2.

- L1, L3, L4 for destination centroid 8.
- L2, L4 for destination centroid 9.
- Common OD Pattern for each time-slice (15 min) accounting for 10-30-40-20 of total demand.
- Interval 5 min.
4. KFX3T Prototype Testing by Simulation

1. MatLab prototype implementation: KFX3T.

2. Test by simulation: Robustness to A priori OD and ICT Sensor Layout. Developed a discrete event simulator to emulate passenger counts and travel times in the test network.
4. KFX3T Prototype Testing by simulation: Design factors in computational experiments

1. **Factor 1**: Total number of passengers in the period. **Parameter**: 0.5, 1, 1.25, 1.5. **Split into time-slices fixed (10-30-40-20)**.

2. **Factor 2**: Quality of a priori Time-Sliced Passenger OD matrix. Strategies based on static EMME transit assignment.
   - **a) Level NoHA**: No deviates no a priori historical OD
   - **b) Level HA0**: The A priori OD is the true historical, equal fraction of use of all OD simple paths
   - **c) Level Ha1**: ‘Destroy’ OD pattern at entry 1
   - **a) Level Hax**: ‘Destroy’ pattern at all entries

3. **Factor 3**: Available ICT data. Levels: **Mx-All, Sx-All Bus Stops, S4-Sensors at Bus Stops 3,4,5,7**.
4. KFX3T Prototype Testing by simulation: KPIs - Example OD 1

1. **Theil’s Coefficient** is a measure on how close two time series are. 
   Ex: 0.143719. A threshold of 0.2 is admissible

2. **RMSEN- Normalized Root Mean Squared Error** (sum of squared differences between true and estimated OD flows per interval, relative to total true flows). Ex: 0.305397

\[
\text{RMSEN} = \sqrt{\frac{12 \sum (y_k - \hat{y}_k)^2}{\sum y_k}}
\]

Ex KFX3T: OD pair 1
- Factor 1: 1.5 Total volume increased by 50%
- Factor 2: Hx – No informative a priori matrix
- Factor 3: All Sensors
4. KFX3T Prototype Testing by simulation: KPIs - according to design factors

- Factors 1 and 2 - Robustness of the estimated OD flows (global KPIs)
  - 4 OD pairs
  - 12 five minute intervals: 60 min

\[
GRMSEN = \sqrt{4 \cdot 12 \sum_{k=1:12} \sum_{od=1:4} (y_{od,k} - \hat{y}_{od,k})^2} \div \left( \sum_{k=1:12} \sum_{od=1:4} y_{od,k} \right)
\]
4. KFX3T Prototype Testing by simulation: KPIs - according to design factors

- Factors 1 and 2 - Robustness of the estimated OD flows (global KPIs), while all ICT Sensors are available (Factor 3)

<table>
<thead>
<tr>
<th>Factor 2 – Perturbation to historic OD pattern</th>
<th>Factor 1 – A priori total OD flows</th>
<th>50%</th>
<th>100%</th>
<th>125%</th>
<th>150%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Theil Coefficient (GU) and RMSEN and R² according to nu parameter. In parenthesis KPIs for OD pair 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>GU</th>
<th>GRMSEN</th>
<th>R²</th>
<th>GU</th>
<th>GRMSEN</th>
<th>R²</th>
<th>GU</th>
<th>GRMSEN</th>
<th>R²</th>
<th>GU</th>
<th>GRMSEN</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA0-none</td>
<td>0.15</td>
<td>0.34</td>
<td>99.1%</td>
<td>0.15</td>
<td>0.34</td>
<td>99.1%</td>
<td>0.16</td>
<td>0.36</td>
<td>97.5%</td>
<td>0.17</td>
<td>0.38</td>
<td>94.7%</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.26)</td>
<td></td>
<td>(0.12)</td>
<td>(0.26)</td>
<td></td>
<td>(0.13)</td>
<td>(0.28)</td>
<td></td>
<td>(0.14)</td>
<td>(0.29)</td>
<td></td>
</tr>
<tr>
<td>HA1</td>
<td>0.14</td>
<td>0.32</td>
<td>98.8%</td>
<td>0.14</td>
<td>0.32</td>
<td>100%</td>
<td>0.14</td>
<td>0.33</td>
<td>99.8%</td>
<td>0.15</td>
<td>0.33</td>
<td>99.2%</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.26)</td>
<td></td>
<td>(0.11)</td>
<td>(0.25)</td>
<td></td>
<td>(0.11)</td>
<td>(0.25)</td>
<td></td>
<td>(0.11)</td>
<td>(0.25)</td>
<td></td>
</tr>
<tr>
<td>HAx-all</td>
<td>0.15</td>
<td>0.38</td>
<td>77.4%</td>
<td>0.20</td>
<td>0.54</td>
<td>3.8%</td>
<td>0.24</td>
<td>0.63</td>
<td>49.3%</td>
<td>0.27</td>
<td>0.73</td>
<td>73.1%</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.26)</td>
<td></td>
<td>(0.12)</td>
<td>(0.27)</td>
<td></td>
<td>(0.13)</td>
<td>(0.28)</td>
<td></td>
<td>(0.14)</td>
<td>(0.31)</td>
<td></td>
</tr>
<tr>
<td>No Deviates</td>
<td>0.14</td>
<td>0.33</td>
<td>95.7%</td>
<td>0.14</td>
<td>0.33</td>
<td>95.7%</td>
<td>0.14</td>
<td>0.33</td>
<td>95.7%</td>
<td>0.14</td>
<td>0.33</td>
<td>95.7%</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.27)</td>
<td></td>
<td>(0.12)</td>
<td>(0.27)</td>
<td></td>
<td>(0.12)</td>
<td>(0.27)</td>
<td></td>
<td>(0.12)</td>
<td>(0.27)</td>
<td></td>
</tr>
</tbody>
</table>
### 4. KFX3T Prototype Testing by simulation: KPIs - according to design factors

- **Factors 1 and 2** - Robustness of the estimated OD flows (global KPIs), while NO Entry Sensors are available (Factor 3)

<table>
<thead>
<tr>
<th>Factor 2 – Perturbation to historic OD pattern</th>
<th>Factor 1 – A priori total OD flows</th>
<th>GU</th>
<th>GRMSEN</th>
<th>R²</th>
<th>GU</th>
<th>GRMSEN</th>
<th>R²</th>
<th>GU</th>
<th>GRMSEN</th>
<th>R²</th>
<th>GU</th>
<th>GRMSEN</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>100%</td>
<td>125%</td>
<td></td>
<td></td>
<td>150%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HA0-none</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.19 (0.18)</td>
<td>0.42 (0.42)</td>
<td>94.4%</td>
<td>0.20 (0.19)</td>
<td>0.44 (0.43)</td>
<td>95.8%</td>
<td>0.20 (0.19)</td>
<td>0.45 (0.44)</td>
<td>95.3%</td>
<td>0.21 (0.20)</td>
<td>0.47 (0.45)</td>
<td>94.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.19 (0.18)</td>
<td>0.45 (0.42)</td>
<td>93.5%</td>
<td>0.18 (0.17)</td>
<td>0.41 (0.40)</td>
<td>95.5%</td>
<td>0.19 (0.17)</td>
<td>0.42 (0.40)</td>
<td>95.8%</td>
<td>0.19 (0.17)</td>
<td>0.42 (0.40)</td>
<td>95.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAx-all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.19 (0.18)</td>
<td>0.45 (0.42)</td>
<td>53.3%</td>
<td>0.24 (0.19)</td>
<td>0.60 (0.42)</td>
<td>4.2%</td>
<td>0.28 (0.19)</td>
<td>0.70 (0.43)</td>
<td>33.0%</td>
<td>0.32 (0.20)</td>
<td>0.79 (0.45)</td>
<td>56.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Deviates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.18 (0.18)</td>
<td>0.42 (0.41)</td>
<td>89.8%</td>
<td>0.18 (0.18)</td>
<td>0.42 (0.41)</td>
<td>89.8%</td>
<td>0.18 (0.18)</td>
<td>0.42 (0.41)</td>
<td>89.8%</td>
<td>0.18 (0.18)</td>
<td>0.42 (0.41)</td>
<td>89.8%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
... Good convergence instance: Factor 1 - 150%OD flows- Factor 2-H1- Factor 3-Only at Bus-Stops

- 'True Matrix' and Initial Total matrix

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>180</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>270</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>270</td>
</tr>
</tbody>
</table>

- Estimated Total matrix

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>172.5</td>
<td>127.5</td>
</tr>
<tr>
<td></td>
<td>60.9</td>
<td>200.0</td>
</tr>
</tbody>
</table>

- Theil's Coefficient

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
</tr>
</tbody>
</table>

- RMSE

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>0.21</td>
</tr>
</tbody>
</table>
Non-convergent instance: Factor 1 - 100%OD flows- Factor 2-Hx-all

- **Target OD matrix** (depends on simulation instance)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>162</td>
<td>116</td>
</tr>
<tr>
<td>73</td>
<td>188</td>
</tr>
</tbody>
</table>

- **‘True Matrix’ and Initial matrix**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>180</td>
<td>120</td>
<td>150</td>
</tr>
<tr>
<td>60</td>
<td>180</td>
<td>120</td>
</tr>
</tbody>
</table>

- **Estimated Total matrix**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>170</td>
<td>130</td>
</tr>
<tr>
<td>138</td>
<td>102</td>
</tr>
</tbody>
</table>

- **Theil’s Coefficient**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.19</td>
<td>0.13</td>
</tr>
<tr>
<td>0.33</td>
<td>0.33</td>
</tr>
</tbody>
</table>

- **RMSE**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.42</td>
<td>0.32</td>
</tr>
<tr>
<td>1.10</td>
<td>0.57</td>
</tr>
</tbody>
</table>
... Good convergence instance: Factor 1 - 150% - Factor 2-H1 - Factor 3- Entries + Bus-Stops (left) and (right) Bus-Stops

100% Equipped Passengers, Entries and Bus-Stops: Estimated vs Historic OD Passenger flows - 1h

\[ y = 1.17 \times - 22.45 \]

\[ R^2 = 0.9921 \]

100% Equipped Passengers and Bus-Stops: Estimated vs Historic OD Passenger flows - 1h

\[ y = 1.13 \times - 16.92 \]

\[ R^2 = 0.9572 \]
4. KFX3T Prototype Testing by simulation: KPIs according to design factors

- **Factors 3 and 1 - Robustness of the estimated OD flows (global KPIs)**

- **Factor 2: Perfect OD Pattern**

<table>
<thead>
<tr>
<th>Factor 3 – Availability ICT Sensors</th>
<th>Factor 1 – A priori total OD flows</th>
<th>50%</th>
<th>100%</th>
<th>125%</th>
<th>150%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Theil Coefficient (GU) and RMSEN and $R^2$ according to nu parameter. In parenthesis KPIs for OD pair 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GU</td>
<td>GRMSEN</td>
<td>$R^2$</td>
<td>GU</td>
<td>GRMSEN</td>
<td>$R^2$</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Mx-All Entries &amp; BusStops</td>
<td>0.15 (0.12)</td>
<td>0.34 (0.26)</td>
<td>99.1%</td>
<td>0.15 (0.12)</td>
<td>0.34 (0.26)</td>
</tr>
<tr>
<td>Sx-All Bus Stops</td>
<td>0.19 (0.18)</td>
<td>0.42 (0.42)</td>
<td>94.4%</td>
<td>0.20 (0.19)</td>
<td>0.44 (0.43)</td>
</tr>
<tr>
<td>S4-Bus Stops 3,4,5,7</td>
<td>0.36 (0.23)</td>
<td>1.47 (0.51)</td>
<td>87.2%</td>
<td>0.34 (0.19)</td>
<td>1.31 (0.42)</td>
</tr>
</tbody>
</table>
Contents

1. Objective of this work
2. A Linear State Space Model approach
3. A Linear Kalman filter formulation for dynamic passenger matrices
4. KFX3T Prototype Testing by simulation
5. Automatization of KFX3T Model Building
6. Conclusions and further research
5. Automatization of KFX3T Model Building

To deal with medium size networks: a platform integrating EMME in order to provide:

- A **graphic description** (centroids, nodes, links and connectors) and transit lines (stops, headways and itineraries).

- A **Transport Zoning System** and historic demand matrix for the period of study.

- The **location of ICT sensors at transit-stops** (identified by an additional segment attribute @nBT, in EMME terminology, and an additional line attribute, @sline, to identify lines affected by BT-equipped stops, either totally or partially).

- OD paths involved in **optimal strategies** for the historic transit demand according to an Extended Transit Assignment, where OD flows are split between connectors in the origin zone using a logit model (scale parameter 0.5) and are subject to an OD path proportion greater than 5%.
5. Automatization of KFX3T Model Building

• The total number of public transit trips was 47816.3 for a 2003 working day within 340 non-null cells, and an average value of 140 trips (max 1186 trips).

• We fixed ICT equipped transit-stops for three of the most important bus lines: L7, L8 and L13.

• After selecting OD pairs with captured flows greater than 10, these lines have a total load of 20765 transit trips in 151 OD pairs.
5. Automatization of KFX3T Model Building

- A strategy file is saved by EMME and the Extended Analysis tool reports path-based details on paths extracted from strategies in a text file.
  - This text file contains the path description included in the optimal strategies restricted to OD pairs and whose OD flow could be captured by some ICT sensors in accordance with historic demand and transit assignment.
  - The path file has a complex format and is post-processed using python script to generate the desired information in the data model of KFX3T.
  - For each ICT-equipped transit-stop, the captured optimal transit OD paths are identified. For each pair of ICT transit-stops, the paths and subpaths connecting both sensors are identified.

- Report files related to OD pairs, nodes, links, etc. were converted to worksheet and .csv format files, in order to be read by the KFX3T model building procedures. This led to:
  - A network spreadsheet containing worksheets for: centroids, nodes and links.
  - A demand spreadsheet containing worksheets for: OD pairs, Entrances, Exits and OD paths involved in optimal strategies.
  - A measurement spreadsheet containing worksheets: Measures, SODMeasures, ActiveMeas, CapODPaths and Global (parameters).
5. Automatization of KFX3T Model Building

Flowchart with the main elements of KFX3T in the MATLAB application:

- **Emme.m**
- **KFX3T**
- **Dynamic Transit OD matrix**
- **MAT files**
- **Data model workbooks**
- **Python scripts**
- **Itineraries**
- **OD flows**
- **Links**
- **Nodes**
- **Path details**
- **ICT equipped stops**
- **Emulation: passenger counts, travel times**

**EMME model**

- **Batch file**
- **Python script**
- **Historic Transit OD matrix**
5. Automatization of KFX3T Model Building

- The KFX3T internal data model is divided into MAT files that are loaded as needed into the program: Global.mat, Tuning.mat, Graph.mat, Demand.mat and Measures.mat.

- Additionally, two additional MAT files have been included for internal use in order to simplify the access to some critical structures.

- These two files are: AccDem.mat (related to accessing OD pairs and paths) and AccMes.mat (related to OD paths captured by each defined sensor and pairs of ICT sensors).
Contents

1. Objective of this work
2. A Linear State Space Model approach
3. A Linear Kalman filter formulation for dynamic passenger matrices
4. KFX3T Prototype Testing by simulation
5. Authomatization of KFX3T Model Building
6. Conclusions and further research
6. Conclusions and further research (I)

• A linear KF formulation for the dynamic estimation of OD passenger matrices for urban transit networks has been implemented in MatLab and tested by simulation.

• Formulation takes into account deviates of passenger flows on most likely simple paths for OD optimal strategies as state variables and it is suitable for application to urban and regional transit networks.

• The availability of these ICT based measurements allow a formulation of linear Kalman Filter approach that incorporate travels times and thus general traffic state as time-varying model parameters.

• Robustness of estimated OD flows per interval (5’) has been tested for:
  - A priori OD pattern and total flows per interval (Factors 1 and 2).
  - Availability of ICT Sensors (Factor 3)
  - Covariance matrices for state variables and counts. Surprisingly, not a priority concern.
6. Conclusions and further research (II)

- In the near future, to **complete the development of an automatic** Model Building framework to support the codification of data structures for KFX3T from the Transportation Planning Model in EMME.
  - Data structure files read for KFX3T are OD pairs, nodes, links, etc. structured in several (MS-Excel) spreadsheets and .csv format files,

- **Further research points to revise the** formulation to incorporate scheduled-based transit services.

- The quality of the historic time-sliced OD matrix used to initialize the KF algorithm is critical for convergence, hence further research **points to**:
  - An off-line procedure to keep updated Historical OD Matrices for the traffic profiles can not be avoided. We are working on this for auto networks.
6. Conclusions and further research (III)

Framework for Estimating and Forecasting OD passenger matrices in ATIS

Development of Dynamic Transit Assignment procedure:
• Near future strategy – based.
• Further: combines strategy-based and time-table behaviour
Thank you very much for your attention!

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