Calibration and Validation I: Estimation of Origin to Destination Flows from Counts

Ennio Cascetta

Modeling and Simulation of Transportation Networks
July 31, 2015

OUTLINE

• PRELIMINARY CONSIDERATIONS

• PART I: STATIC O-D ESTIMATION

• PART II: DYNAMIC O-D ESTIMATION

• PART III: QUASI-DYNAMIC OD ESTIMATION
OUTLINE

• PRELIMINARY CONSIDERATIONS

• PART I: STATIC O-D ESTIMATION

• PART II: DYNAMIC O-D ESTIMATION

• PART III: QUASI-DYNAMIC OD ESTIMATION

PRELIMINARY CONSIDERATIONS

• O-D trip matrices are a fundamental input for most problems related to the management and planning of transportation systems.

• O-D trip matrices were typically estimated by using sample surveys of various types (expansive, time consuming, not easily repeatable, etc.) and/or travel demand models.

• Recently, considerable work has been devoted to improve the quality of O-D demand estimates by using cheap, easily and automatically collectable traffic counts.
PRELIMINARY CONSIDERATIONS

**Example**

Initial estimate of the O-D flows:

<table>
<thead>
<tr>
<th>O-D Flows</th>
<th>Distance Measures</th>
<th>$\sum (d_{nm} - \hat{d}_{nm})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>$d_{12}$</td>
<td>66</td>
</tr>
<tr>
<td>1-3</td>
<td>$d_{13}$</td>
<td>8.702</td>
</tr>
<tr>
<td>2-3</td>
<td>$d_{23}$</td>
<td>2.226</td>
</tr>
<tr>
<td>2-4</td>
<td>$d_{24}$</td>
<td>2.286</td>
</tr>
<tr>
<td>2-5</td>
<td>$d_{25}$</td>
<td></td>
</tr>
<tr>
<td>2-6</td>
<td>$d_{26}$</td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>$d_{34}$</td>
<td></td>
</tr>
<tr>
<td>3-5</td>
<td>$d_{35}$</td>
<td></td>
</tr>
<tr>
<td>3-6</td>
<td>$d_{36}$</td>
<td></td>
</tr>
<tr>
<td>4-5</td>
<td>$d_{45}$</td>
<td></td>
</tr>
<tr>
<td>4-6</td>
<td>$d_{46}$</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>$d_{56}$</td>
<td></td>
</tr>
</tbody>
</table>

**Distance Measures**

$\hat{d}_{nm} = \sum (d_{nm} - \hat{d}_{nm})$
PRELIMINARY
CONSIDERATIONS

Example

Initial estimate of the O-D flows:

\[ d = \begin{bmatrix} 3 & 4 \\ 2 & 7 \end{bmatrix} \]

The balance between unknown O-D flows \( d_{od} \) and “independent” information about counts \( n_i \)

\[ f_{1,5} = d_{13} + d_{14} \]

\[ f_{5,6} = d_{13} + d_{14} + d_{23} + d_{24} \]

<table>
<thead>
<tr>
<th>Application</th>
<th>Estimation</th>
<th>Prediction</th>
</tr>
</thead>
</table>
| **off-line**
  to design and evaluate traffic management schemes and systems | * | * |
| **on-line**
  to support operation of real-time control strategies (e.g. predictive control) | * | * |

- to update an estimate of OD matrix using observed link flows (counts)
- to predict OD matrices for future time slices combining historical OD matrices, OD matrices estimated for previous time slices, and observed link flows
OUTLINE

• PRELIMINARY CONSIDERATIONS

• PART I: STATIC O-D ESTIMATION

• PART II: DYNAMIC O-D ESTIMATION

• PART III: QUASI-DYNAMIC OD ESTIMATION

STATIC O-D ESTIMATION

Contents

• Notation and terminology

• Relationship between traffic counts and O-D demand flows

• Estimators of O-D trip matrices

• Solution methods
STATIC O-D ESTIMATION
Notation and terminology

• Transport Network \((N,L,C)\)
  \(N\): set of nodes \(N_c \subseteq N\) set of centroids
  \(L\): set of links
  \(C\): set of link cost functions \(c_l = c_l(v)\)

• O-D Trip vector \(d \equiv \{d_{od}\}\) \(o,d \in N_x\)
  \(d_{od}\): average number of trips going from origin \(o\) to destination \(d\) with a given mode within a given time period.

• Link flow vector \(f \equiv \{f_l\}\) \(l \in L\)
  \(f_l\): n. of trips using link \(l\) in a fixed time period.

• Link cost vector \(c \equiv \{c_l\}\) \(l \in L\)
  \(c_l\): average gen. cost on link \(l\).

• Path flow vector \(h \equiv \{h_k\}\)
  \(h_k\): n. of trips using path \(k\), connecting O-D pair \((o,d)\), in a fixed time period \((k \in I_{od})\).

• Path choice fraction matrix \(P \equiv \{p_{k,od}\}\)
  \(p_{k,od}\): \(p(k/od)\) fraction of trips using path \(k\) (connecting O-D pair \(o,d\)) in a fixed time period.

• Link-path incidence matrix \(\Delta \equiv \{\delta_{lk}\}\)
  \(\delta_{lk}\): 1 if link \(l\) belongs to path \(k\); 0 otherwise.

• Assignment matrix \(M = \Delta P \equiv \{m_{l,od}\}\)
  \(m_{l,od}\): fraction of O-D flow \(d_{od}\) contributing to \(f_l\).

Example
STATIC O-D ESTIMATION

Relationship between traffic counts and O-D demand flows

• EXAMPLE OF ASSIGNMENT MATRIX

\[
\begin{bmatrix}
1 & 4 & 6 & 7 & 9 & 3 \\
1 & 4 & 5 & 7 & 9 & 3 \\
1 & 4 & 5 & 7 & 9 & 8 & 3 \\
1 & 4 & 6 & 7 & 9 & 8 & 3 \\
2 & 5 & 7 & 9 & 3 \\
2 & 5 & 7 & 9 & 8 & 3
\end{bmatrix} =
\begin{bmatrix}
100 \\
80 \\
30 \\
20 \\
70
\end{bmatrix}
\]

\[N=2\text{ (link 9-3 e link 8-3)}\]

\[
\Delta \quad P = M
\]

\[
\begin{bmatrix}
1 & 0.3 & 0.3 \\
2 & 0.3 & 0.3 \\
3 & 0.2 & 0.2 \\
4 & 0.2 & 0.2 \\
5 & 0.7 & 0.7 \\
6 & 0.3 & 0.3
\end{bmatrix}
\]

\[
f_i = \sum_k \delta_{lk} h_k
\]

\[
f_i = \sum_k \delta_{lk} h_k = \sum_k \delta_{lk} \sum_{od} p_{k,od} d_{od}
\]

\[
f_i = \sum_{od} d_{od} \sum_k \delta_{lk} p_{k,od} = \sum_{od} m_{i,od} d_{od}
\]

\[
f_i = m_i^T d
\]

• ASSIGNMENT MAP

\[f = \Delta h = \Delta Pd = Md\]

Example
STATIC O-D ESTIMATION

Relationship between traffic counts and O-D demand flows

- \( f = Md \): NUMERICAL EXAMPLE

\[
\begin{bmatrix}
100 \\
80
\end{bmatrix} =
\begin{bmatrix}
1 & 3 & 2 & 3 \\
9 & 3 & 1 & 16 \\
8 & 3 & 6 & 4
\end{bmatrix}
\begin{bmatrix}
1 & 3 & 100 \\
2 & 3 & 80
\end{bmatrix}
\]

The balance between unknown O-D flows \( d_{od} \) and “independent” information about counts \( n_l \) for real networks:

\[
n_l = 1,000 \div 10,000
\]

\[
n_{od} = (200 \times 200) \div (1000 \times 1000)
\]
STATIC O-D ESTIMATION

Relationship between traffic counts and O-D demand flows

- $\hat{M} = \text{estimate of } M \text{ from the assignment model}$
- $g = \text{overall path cost vector}$

\[
\hat{p}_{k,od} = p[k/od](g)
g = \Delta^T c
\]

\[
\hat{M}(c) = \Delta \hat{P}(c)
v = \hat{M}d = v(d)
\hat{M} = M + E
f = (\hat{M} - E)d = \hat{M}d - Ed
Ed \Rightarrow \varepsilon^{SIM}
f = \hat{M}d + \varepsilon^{SIM}
\]

Observed link flow vector (traffic counts) $\hat{f} = \{\hat{f}_i\}$:

\[
\hat{f} = f + \varepsilon^{OBS}
\hat{f} = \hat{M}d + \varepsilon^{SIM} + \varepsilon^{OBS} = v(d) + \varepsilon
\]

$\varepsilon$: vector of random error terms with $E(\varepsilon)=0$
- measurement errors
- temporal fluctuation of demand
- assignment model (path choice and network) errors
- etc.
STATIC O-D ESTIMATION

Estimators of O-D trip demand

• PROBLEM STATEMENT
Estimate the O/D trip demand $d$ by “efficiently” combining traffic counts with all other available information

• ESTIMATION PROBLEMS
– Experimental information (sample surveys) + traffic counts
  \[ \text{"CLASSIC" INFERENCE} \]

– Non-experimental information (“a priori” information) + traffic counts
  \[ \text{BAYESIAN INFERENCE} \]

GENERAL ESTIMATION PROBLEM
Several models have been proposed to estimate O-D flows by combining counts with other information sources.

General Formulation
\[
d^* = \arg\min_{x \in S} [z_1(x, d) + z_2(v(x), f)]
\]

where:
- $z_1(x, d)$ is a measure of the “distance” between the unknown demand $x$ and the a priori demand $d$;
- $z_2(v(x), f)$ is a measure of the “distance” between the link flows resulting from the unknown demand $x$ and the observed link flows $f$. 
**STATIC O-D ESTIMATION**

Estimators of O-D trip demand

### POSSIBLE FUNCTIONAL FORMS for $z_1(.)$ and $z_2(.)$

<table>
<thead>
<tr>
<th>Distance from the initial estimate $z_1(x, d)$</th>
<th>Distance from flows counts $z_2(x, f)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized least squares (GLS) $(\hat{d} - x)^T Z^{-1} (\hat{d} - x)$</td>
<td>Generalized least squares (GLS) $(f - \hat{v}(x))^T W^{-1} (f - \hat{v}(x))$</td>
</tr>
<tr>
<td>$\sum_{od} (d_{od} - \hat{d}<em>{od})^2 / \text{Var}(\hat{d}</em>{od})$</td>
<td>$\sum_{od} (f_{od} - \hat{v}(x_{od}))^2 / \text{Var}(\hat{v}(x_{od}))$</td>
</tr>
<tr>
<td>Maximum Likelihood (ML) Poisson $\sum (n_{od} \ln(\hat{n}<em>{od}) - n</em>{od})$</td>
<td>Maximum Likelihood (ML) Poisson $\sum (f_{od} \ln(\hat{v}(x_{od})) - f_{od})$</td>
</tr>
<tr>
<td>MVN $(\hat{d} - x)^T Z^{-1} (\hat{d} - x)$</td>
<td>MVN $(f - \hat{v}(x))^T W^{-1} (f - \hat{v}(x))$</td>
</tr>
<tr>
<td>$\sum (d_{od} - \hat{d}<em>{od})^2 / \text{Var}(\hat{d}</em>{od})$</td>
<td>$\sum (f_{od} - \hat{v}(x_{od}))^2 / \text{Var}(\hat{v}(x_{od}))$</td>
</tr>
<tr>
<td>Bayes Poisson $\sum \hat{n}<em>{od} \ln(n</em>{od} / \hat{n}_{od})$</td>
<td>Bayes Poisson $\sum f_{od} \ln(\hat{v}(x_{od})) / \hat{v}(x_{od})$</td>
</tr>
<tr>
<td>MVN $(\hat{d} - x)^T Z^{-1} (\hat{d} - x)$</td>
<td>MVN $(f - \hat{v}(x))^T W^{-1} (f - \hat{v}(x))$</td>
</tr>
<tr>
<td>$\sum (d_{od} - \hat{d}<em>{od})^2 / \text{Var}(\hat{d}</em>{od})$</td>
<td>$\sum (f_{od} - \hat{v}(x_{od}))^2 / \text{Var}(\hat{v}(x_{od}))$</td>
</tr>
<tr>
<td>Multinomial $\sum \hat{n}<em>{od} \ln(n</em>{od} / \hat{n}<em>{od}) / \sum\hat{n}</em>{od} = \text{const.}$</td>
<td>$\sum f_{od} \ln(\hat{v}(x_{od})) / \hat{v}(x_{od}) = \text{const.}$</td>
</tr>
</tbody>
</table>

### GLS ESTIMATORS

Experimental and numerical tests show that Generalized Least Square method is robust.

$$\hat{d} = x + \eta$$
$$\hat{f} = \hat{M}x + \epsilon$$

$$d^{\text{GLS}} = \arg\min_{x \in S} [(\hat{d} - x)^T Z^{-1} (\hat{d} - x) + (\hat{f} - \hat{M}x)^T W^{-1} (\hat{f} - \hat{M}x)]$$

Simplified Version:

$$d^{\text{GLS}} = \arg\min_{x \geq 0} \left[ \sum_{od} \left( \frac{(d_{od} - x_{od})^2}{\text{var}(\hat{d}_{od})} \right) + \sum_{od} \left( \frac{(f_{od} - \hat{M}x_{od})^2}{\text{var}(\hat{v}(x_{od}))} \right) \right]$$
STATIC O-D ESTIMATION
Solution methods

• LINK COSTS KNOWN
( uncongested networks or measured congested travel times)

\[ d^* = \arg\min_{x \geq 0} \left[ z_1(x, \hat{d}) + z_2(\hat{M}(\hat{c})x, \hat{f}) \right] \]

standard linearly constrained optimization algorithms

projected gradient algorithm
STATIC O-D ESTIMATION

Solution methods

• LINK COSTS UNKNOWN
  (congested networks)

⇒ FIXED-POINT MODELS

\[
\delta (\hat{M}) = \arg \min_{x \in S} [Z_1(X, \hat{d}) + Z_2(\hat{M}X, \hat{f})]
\]

\[
\hat{M} = M(c(v(d)))
\]

\[
d^* = \delta (\hat{M}(d^*))
\]

\[
d^* = \arg \min_{x \in S} [Z_1(X, \hat{d}) + Z_2(c(v(d^*)), \hat{f})]
\]

⇒ BILEVEL OPTIMIZATION MODELS

\[
d^* = \arg \min_{x \in S} [Z_1(X, \hat{d}) + Z_2(c(v(x)), \hat{f})]
\]

where \( v(x) = \arg \min_{f \in A(x)} z(\hat{f}) \)

taken from:
STATIC O-D ESTIMATION

Solution methods

• LINK COSTS UNKNOWN ⇒ FIXED POINT SOLUTION ALGORITHMS

GENERAL STRUCTURE:
✓ CALCULATION OF ASSIGNMENT MATRIX $\hat{M}$ CORRESPONDING TO DEMAND $d^{k-1}$.
  ➢ ASSIGNMENT OF DEMAND VECTOR $d^{k-1}$: $\hat{v}^k = \hat{v}(d^{k-1})$
  ➢ ESTIMATION OF LINK COSTS AND ASSIGNMENT MATRIX: $c^k = c(\hat{v}^k)$

✓ ESTIMATION OF DEMAND SUPPORT VECTOR $Y^k$:

$$y^k = \arg\min_{x \in S} \{z_1(x, \hat{d}) + z_2(M^k x, \hat{f})\}$$

✓ DEMAND VECTOR UPDATING (MSA):

$$d^k = \frac{k-1}{K} d^{k-1} + \frac{1}{K} y^k$$

✓ STOP TEST:

$$y^k \approx d^{k-1}$$

---

STATIC O-D ESTIMATION

Solution methods

• APPLICATION

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>2,903</td>
</tr>
<tr>
<td>Centroids</td>
<td>167</td>
</tr>
<tr>
<td>Road Links</td>
<td>5,102</td>
</tr>
<tr>
<td>Connector links</td>
<td>646</td>
</tr>
<tr>
<td>O-D pairs</td>
<td>27,889</td>
</tr>
</tbody>
</table>

Counts (7:30-8:30 a.m.) 82
Hold-out counts 20

\(\hat{d}\) is an out-dated estimate of the O-D table
STATIC O-D ESTIMATION

Solution methods

• APPLICATION

Results

\[
MSE(f^*, \hat{f}) = \frac{1}{n} \sum_{i} (f_i^* - \hat{f}_i)^2
\]

\[
RMSE\% = \frac{MSE(f^*, \hat{f})^{1/2}}{\hat{f}} \sum_n f_i n_i
\]

Before updating

After updating

<table>
<thead>
<tr>
<th>Before updating</th>
<th>After updating</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE before</td>
<td>1,409,442</td>
</tr>
<tr>
<td>MSE after</td>
<td>110,822</td>
</tr>
<tr>
<td>% reduction of MSE</td>
<td>-92%</td>
</tr>
<tr>
<td>RMSE% before</td>
<td>0.52</td>
</tr>
<tr>
<td>RMSE% after</td>
<td>0.21</td>
</tr>
<tr>
<td>% reduction of RMSE</td>
<td>-59%</td>
</tr>
</tbody>
</table>
STATIC O-D ESTIMATION
Solution methods

• APPLICATION → RESULTS (B)

Validation with hold-out counts

Validation with hold-out counts

<table>
<thead>
<tr>
<th></th>
<th>MSE before</th>
<th>MSE after</th>
<th>% reduction of MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assigned Flows</td>
<td>1,581,023</td>
<td>68,456</td>
<td>-96%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>RMSE% before</th>
<th>RMSE% after</th>
<th>% red. of RMSE%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assigned Flows</td>
<td>1.10</td>
<td>0.25</td>
<td>-77%</td>
</tr>
</tbody>
</table>

Cross validation (Before updating)

Cross validation (After updating)

<table>
<thead>
<tr>
<th>INITIAL O/D</th>
<th>INT</th>
<th>EX</th>
<th>TOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>INT</td>
<td>39,355</td>
<td>31,338</td>
<td>70,693</td>
</tr>
<tr>
<td>EX</td>
<td>43,942</td>
<td>13,454</td>
<td>57,396</td>
</tr>
<tr>
<td>TOT</td>
<td>83,297</td>
<td>44,792</td>
<td>128,089</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ESTIMATED O/D</th>
<th>INT</th>
<th>EX</th>
<th>TOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>INT</td>
<td>47,971</td>
<td>29,497</td>
<td>77,467</td>
</tr>
<tr>
<td>EX</td>
<td>41,256</td>
<td>8,688</td>
<td>49,944</td>
</tr>
<tr>
<td>TOT</td>
<td>89,226</td>
<td>38,185</td>
<td>127,411</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% VAR</th>
<th>INT</th>
<th>EX</th>
<th>TOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>INT</td>
<td>22%</td>
<td>-6%</td>
<td>10%</td>
</tr>
<tr>
<td>EX</td>
<td>-6%</td>
<td>-35%</td>
<td>-13%</td>
</tr>
<tr>
<td>TOT</td>
<td>7%</td>
<td>-15%</td>
<td>-1%</td>
</tr>
</tbody>
</table>

Department of Transportation Engineering University of Naples “Federico II”
OUTLINE

• PRELIMINARY CONSIDERATIONS

• PART I: STATIC O-D ESTIMATION

• PART II: DYNAMIC O-D ESTIMATION

• PART III: QUASI-DYNAMIC OD ESTIMATION

DYNAMIC O-D ESTIMATION

Extension of previous results to the case of time-varying (within day dynamic) demand and link flows

• Notation and terminology

• Relationship between within-day Dynamic Traffic Counts and O-D demand flows

• Simultaneous dynamic estimators of O/D trip matrices
DYNAMIC O-D ESTIMATION

Notation and terminology

• Observation period
  \( n_j \) : number of intervals 1, 2, ..., \( n_j \)
  \( T \) : duration of each interval
  \( (n_j \cdot T) \) = total study period

• O-D trip vectors \( d_{od}[t] \) \{ \( d_{od}[t] \) \}
  \( d_{od}[t] \) : number of trips between O-D pair \( o,d \) leaving the origin during interval \( t \)

[Diagram of O-D flow vectors]

• Link flow vectors
  \( f_j \equiv \{ f_{ijj} \} \)
  \( f_{ijj} \) : average flow which can be counted on link \( i,j \) during interval \( j \)

• Within-day dynamic link flow definition
  – counting section
    \( f_{ijj} = \frac{n(s_c, j)}{T} \)
  – path flow vector \( h_{ijj} = \{ h_{ik}[t] \} \)
    \( h_{ik}[t] \) : path flow, average number of trips per time unit on path \( k \) leaving during period \( t \)
DYNAMIC O-D ESTIMATION

Notation and terminology

- Path choice fraction matrix $P_t = \{p(k/t)\}$
  $p(k/t)$: fraction of O/D flow $d_{od}(t)$ following path $k$, given departing interval $t$
  $h_k[t] = p(k | t) \cdot d_{od}[t]$

- Dynamic incidence matrix $\Delta_{t,j} = \{\delta_{l,k}[t,j]\}$
  path flow $h_{t}[t]$ departing at time $t$ crossing link $l$ during interval $j$
  $f_{l}[j] = \sum_k \sum_{t \in j} \delta_{l,k}[t,j] \cdot h_k[t]$

- Dynamic assignment matrix $M_{t,j} = \{m_{l,od}[t,j]\}$
  fraction of O/D flow $d_{od}(t)$ contributing to flow on link $l$ during period $j$ ($f_{l}[j]$)
  $f_{l}[j] = \sum_{od} \sum_{k \in t \in j} \delta_{l,k}[t,j] \cdot p(k/t) \cdot d_{od}[t] = \sum_{od} \sum_{l \in j} m_{l,od}[t,j] \cdot d_{od}[t]$

Relationship between within-day dynamic traffic counts and O-D demand flows

Estimates of variables can be obtained through within-day dynamic assignment models:

- Estimate of path choice fraction matrix $\hat{P}_t$
  estimate of departure time/path choice probabilities (from demand model)

- Estimate of the dynamic incidence matrix $\hat{\Delta}_{t,j}$
  function of link performances (e.g. average speed) on intervals comprised between $t$ and $j$ (see Dynamic Network Flow Propagation Models)

- Estimate of demand assignment matrix $\hat{M}_{t,j}$
  fractions of O/D flow $d_{od,t}$ contributing to flow on link $l$ during period $j$
  $\hat{M}_{t,j} = \hat{\Delta}_{t,j} \cdot \hat{P}_t$

ESTIMATED ASSIGNMENT MAP

$f_{ij} = \sum_{t \in j} \hat{M}_{t,j} \cdot d_{ij} + \epsilon_{ij}^{SIM}$
DYNAMIC O-D ESTIMATION

Relationship between within-day dynamic traffic counts and O-D demand flows

EXAMPLE: DISCRETE PATH FLOW REPRESENTATION

$d_{1-4}[t=1]=5$ vehicles

Path choice fraction matrix:

\[
R_k = \begin{bmatrix}
2/5 \\
3/5 \\
0
\end{bmatrix}
\]

DYNAMIC O-D ESTIMATION

Relationship between within-day dynamic traffic counts and O-D demand flows

EXAMPLE: DISCRETE PATH FLOW REPRESENTATION

Incidence matrix: $\Delta_{t,j}$
**DYNAMIC O-D ESTIMATION**

*Relationship between within-day dynamic traffic counts and O-D demand flows*

**EXAMPLE: DISCRETE PATH FLOW REPRESENTATION**

Assignment matrices: $M_{[j]} = \Delta_{[j]} \cdot P_t$

- **Observed link flows (counts)** $\hat{f}_{[j]} = \{\hat{f}_{[j]}\}$
  - flow counted on link $l$ in time interval $j$
    - $f_{[j]} = \sum_{t=1}^{t_{[j]}} \hat{M}_{[t,j]} \cdot d_{[t]} + \epsilon_{[j]}$
    - $f_{[j]} = \hat{f}_{[j]} + \epsilon_{[j]}$
    - $\hat{f}_{[j]} = \sum_{t=1}^{t_{[j]}} \hat{M}_{[t,j]} \cdot d_{[t]} + \epsilon_{[j]}$

where $\epsilon_{[j]} = \epsilon_{SIM_{[j]}} - \epsilon_{OBS_{[j]}}$

- $\epsilon = $ vector of random error terms due to:
  - assignment model errors
  - counting errors

---

*Department of Transportation Engineering University of Naples “Federico II”*
DYNAMIC O-D ESTIMATION
Dynamic Estimators of O-D trip matrices

• SIMULTANEOUS ESTIMATORS
Computing the whole set of time-dependent O-D matrices by using counts over all intervals simultaneously

\[ d_{ij}^{\hat{}} \ldots d_{nj}^{\hat{}} \] \[ f_{ij}^{\hat{}} \ldots f_{nj}^{\hat{}} \]

\[ d_{ij}^{*} \ldots d_{nj}^{*} \]

DYNAMIC O-D ESTIMATION
Dynamic Estimators of O-D trip matrices

• SIMULTANEOUS ESTIMATORS
Estimation of the whole set of demand vectors \((d_{ij}^{*}, \ldots d_{nj}^{*})\):

\[(d_{ij}^{*}, \ldots d_{nj}^{*}) = \arg\min_{x_{[t]}=0} x_{[n]} \geq 0 z_1(x_{[t]}, \ldots x_{[n]}; \hat{d}_{ij}^{\prime}, \ldots \hat{d}_{nj}^{\prime}) + z_2(x_{[t]}, \ldots x_{[n]}; \hat{f}_{ij}^{\prime}, \ldots \hat{f}_{nj}^{\prime})\]

\(\hat{d}_{ij}^{\prime}\): initial information about O/D vector \(d_i\)
\(z_1(.)\) \(z_2(.)\): functions to be defined, depending on the chosen estimator

GLS estimator

\[ z_1(.) = \sum_{j=1}^{n} (x_j - \hat{d}_j)^T Z_j^{-1} (x_j - \hat{d}_j) \]

\[ z_2(.) = \sum_{j=1}^{n} \left( \sum_{t=1}^{T} M_{[t,j]} x_t - \hat{f}_j \right)^T W^{-1} \left( M_{[t,j]} x_t - \hat{f}_j \right) \]
DYNAMIC O-D ESTIMATION
Dynamic Estimators of O-D trip matrices

• COMPUTATION OF THE ASSIGNMENT MAP
  ✓ link performances \( (r_{11},...,r_{jj}) \) known (observed)
  Dynamic Network Flow Propagation Model (DNFP) (tracking)
  \[
  \hat{M}_{ij} = M_{ij}(r_{11},...,r_{jj})
  \]
  \[
  v_{ij} = (d_{11},...,d_{jj},...) = \sum_{t=1}^{T} \hat{M}_{ij} d_t
  \]
  ✓ link performances unknown
  Dynamic traffic assignment Model (DTA)
  \[
  \hat{M}_{ij} = M_{ij}(r_{11}(d),...,r_{jj}(d))
  \]
  \[
  v_{ij}(d) = \sum_{t=1}^{T} M_{ij}(r_{11}(d),...,r_{jj}(d)) d_t
  \]

OUTLINE

• PRELIMINARY CONSIDERATIONS

• PART I: STATIC O-D ESTIMATION

• PART II: DYNAMIC O-D ESTIMATION

• PART III: QUASI-DYNAMIC OD ESTIMATION
QUASI DYNAMIC O-D ESTIMATION in within-day dynamic contexts

\[ d_{od}^j = g_{jo}^j \cdot p_{d|o}^j \]

Given \( n_c \) centroids, \( n_l \) counted links and a period \( \tau \) including \( n_j \) time slices, the generic o-d flow for the time slice \( j \) may be expressed as the product between the demand generated by \( o \) during the time slice \( j \), \( g_{jo}^j \), and the fraction going to destination \( d \) moving from \( o \) within the time slice \( j \), \( p_{d|o}^j \).

The factors affecting \( g_{jo}^j \) are inherently within-day time varying, while the factors affecting \( p_{d|o}^j \) are more stable across different time slices.

\[ \frac{\partial g_{jo}^j}{\partial j} \gg \frac{\partial p_{d|o}^j}{\partial j} \]

The distribution probability \( p_{d|o}^j \) of the \( n_j \) time slices \( j \) within \( T \) may be reasonably approximated by its average \( p_{d|o}^{(j)} \) over \( T \).

\[ d_{od}^j = g_{jo}^j p_{d|o}^j \cong g_{jo}^j p_{d|o}^{(j)} = d_{od}^{j,qd} \]

Intrinsic Error

The intrinsic bound is the difference between the demand flow \( d_{od}^j \) and the corresponding quasi-dynamic flow \( d_{od}^{j,qd} \).

\[ i e_{od}^j = d_{od}^j - d_{od}^{j,qd} \]
QUASI DYNAMIC O-D ESTIMATION in within-day dynamic contexts

The main problems in O-D estimation with traffic counts depends on the balance between the nr of equation and the nr of unknown

The quasi-dynamic assumption allows reducing the number of unknowns:
- generation profiles for each origin and time slice: $n_i \times n_j$ unknowns;
- average (within $T$) distribution shares: $n c^2$ unknowns while the equations are $n_i \times n_j$

Playing on the lenght of $T$ equations and unknown can be balanced

Counted Flows

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_1$</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>$d_{1[1]} = 9$</td>
<td>$d_{2[1]} = 8$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_2$</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>$d_{1[2]} = 12$</td>
<td>$d_{2[2]} = 12$</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>9</td>
<td>$p(3/1) = 2/3$</td>
<td>$p(3/2) = 1/4$</td>
<td></td>
</tr>
<tr>
<td>$f_3$</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>$d_{1[3]} = 15$</td>
<td>$d_{2[3]} = 16$</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3 time slices
3 sensors
9 independent equations

3 time slices
2 origins
2 distribution shares
6 generation unknowns
2 independent distribution unknowns
8 total unknowns
QUASI DYNAMIC O-D ESTIMATION in within-day dynamic contexts

GLS-based quasi-dynamic o-d flows estimator

\[
d^* = \arg \min_{x \in S} \left[ z(x, \hat{d}) + z(f(x, \hat{f})) \right]
\]

\[
\{g^1, ..., g^n : p^1, ..., p^n \} = \arg \min_{x \in S} \left[ z(x, \pi^x, \pi^0, d^0, a^0) + z(f(x, \pi^x, \pi^0, d^0, a^0)) \right]
\]

\[
\{g^1, ..., g^n, p^1, ..., p^n \} = \arg \min_{x \in S} \left( \sum_{j=1}^{n} \left( x_j - \pi_{j,O}^{(i)} - \hat{d}_j \right)^2 \right)^{1/2} \sum_{j=1}^{n} \left( \sum_{i=1}^{n} \left( \pi_{j,i}^{(O)} - \hat{f}_j \right) \right)^{1/2}
\]

Department of Transportation Engineering University of Naples “Federico II”

QUASI DYNAMIC O-D ESTIMATION

Real test site of A4 - A23 motorways in North-Eastern Italy

- Full test site with 17 origins, 124 links and 272 o-d pairs
- The A4 branch between Palmanova and Trieste was eliminated and replaced by a virtual junction close to the A4-A23 intersection
- Two independent closed systems (one per carriageway) made by 13 origins, 91 o-d pairs and 49 links each
- The entrance/exit toll system allows to obtain the true 10 minutes o-d flows (cars)

Department of Transportation Engineering University of Naples “Federico II”
QUASI DYNAMIC O-D ESTIMATION in within-day dynamic contexts

Scheme of the experiment

Quasi-dynamic O-D estimation for different durations of the sub-periods $\tau$ of constant distribution percentages:

<table>
<thead>
<tr>
<th>vehicle class</th>
<th>indicator</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>6</th>
<th>12</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>MSE</td>
<td>2.42</td>
<td>3.15</td>
<td>3.80</td>
<td>4.05</td>
<td>4.62</td>
<td>5.15</td>
<td>6.01</td>
</tr>
<tr>
<td></td>
<td>CV_RMSE</td>
<td>0.363</td>
<td>0.415</td>
<td>0.456</td>
<td>0.470</td>
<td>0.503</td>
<td>0.531</td>
<td>0.573</td>
</tr>
<tr>
<td>passenger</td>
<td>MSE</td>
<td>1.38</td>
<td>1.77</td>
<td>2.11</td>
<td>2.20</td>
<td>2.51</td>
<td>2.66</td>
<td>2.81</td>
</tr>
<tr>
<td></td>
<td>CV_RMSE</td>
<td>0.424</td>
<td>0.481</td>
<td>0.525</td>
<td>0.536</td>
<td>0.573</td>
<td>0.589</td>
<td>0.605</td>
</tr>
<tr>
<td>freight</td>
<td>MSE</td>
<td>0.83</td>
<td>1.05</td>
<td>1.23</td>
<td>1.29</td>
<td>1.42</td>
<td>1.67</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td>CV_RMSE</td>
<td>0.603</td>
<td>0.679</td>
<td>0.734</td>
<td>0.753</td>
<td>0.789</td>
<td>0.858</td>
<td>0.961</td>
</tr>
</tbody>
</table>

$CV_{RMSE} = \frac{RMSE}{d_{true}}$
QUASI DYNAMIC O-D ESTIMATION

Real test site of A4 - A23 motorways in North-Eastern Italy

Pearson’s chi-squared test (Kendall and Stuart, 1979)

✓ Null hypothesis $H_0$: the observed frequencies $d_{ij}$ and the expected frequencies $d^*_{ij}$ come from the same distribution

$$
X^2 = \sum \frac{\left( d_{ij} - d^*_{ij} \right)^2}{d^*_{ij}} = \sum \frac{\left( g_i^j p_{ij}^o - g_i^j p_{ij}^{o(i)} \right)^2}{g_i^j p_{ij}^{o(i)}} = \sum g_i^j \left( p_{ij}^o - p_{ij}^{o(i)} \right)^2
$$

Likelihood ratio (LR) test (Kendall and Stuart, 1979)

✓ Null hypothesis $H_0$: the observed distribution shares $p_{ij}^o$ and the corresponding quasi-dynamic distribution shares $p_{ij}^{\omega(j)}$ come from the same distribution

✓ The quasi-dynamic distribution shares represent the null model and the observed distribution shares the alternative model

$$
L' = -2 \ln \frac{L_{null}}{L_{alternative}} = -2 \ln \frac{\prod_i p(d|o \tau(j))^e_i}{\prod_i p(d|o)^e_i} = -2 \sum_i d_{ij} \ln \frac{p(d|o \tau(j))}{p(d|o)}
$$

Probability of acceptance larger than 80% exhibited by almost 60% of tests in the most favourable situation ($t_\tau=0.5$ h) and by about 40% of tests in the worst case ($t_\tau=24$ h)

Department of Transportation Engineering University of Naples "Federico II"
Comparison of the probability of acceptance of the null hypothesis using the two different statistics ($\chi^2$ and LL tests), for $t = 1$ hour

- Perturbation of the observed o-d flows to obtain the seed o-d flows
  - the observed generation profiles are perturbed using a coefficient of variation of 0.3
  - uniform distribution shares across all destinations are assumed

- Estimation/updating on the basis of a subset of 15 link counts
  - the subset is chosen by means of the max flow method proposed by Yang and Zhou (1998) amongst the observed link flows
  - the observed link flows are calculated through the dynamic network loading of the observed o-d flows
  - error-free link counts, error-free assignment map

- The quality of the tested estimator is measured by comparing the updated o-d flows with the observed o-d flows
  - standard goodness-of-fit measures are used
QUASI DYNAMIC O-D ESTIMATION

Experiments on the real test site of A4 - A23 motorways in North-Eastern Italy

Tested estimators

✓ QD-GLS estimator
  • the distribution shares are kept constant for the whole day

✓ Static GLS estimator
  • 24 daily static estimate/updates (a duration $T_s=1$ hour is assumed)

✓ Simultaneous dynamic estimator

✓ Kalman filter estimator
  • three different experiments are carried out depending on the type of seed o-d flows

QUASI DYNAMIC O-D ESTIMATION

Experiments on the real test site of A4 - A23 motorways in North-Eastern Italy

Kalman Filter (KF)

✓ The KF is a recursive algorithm that uses a series of measurements observed over time to produce estimates of unknown state variables

✓ The main assumption of the KF is that the underlying system is a linear dynamical system and that all error terms and measurements have a Gaussian distribution

✓ The KF is the best linear unbiased estimator (BLUE)
QUASI DYNAMIC O-D ESTIMATION

Experiments on the real test site of A4 - A23 motorways in North-Eastern Italy

Kalman Filter (KF)

- Typically applied in on-line estimation/prediction but suitably adaptable to off-line contexts
- State variables: o-d estimates
- Measurement equation: dynamic network loading equations

\[ \hat{x}_j' = \sum_{j=1}^{n} m_{jj} x_j + \mu_j \]

- Transition equation: o-d flows related to a time slice j are expressed as a result of the update of an historical estimate by means of a within-day autoregressive process of order p based on the deviations from the historical estimates observed for the p time slices prior to j

\[ x_{j+1} = d_{j+1} + \sum_{j=1}^{p} \phi_{j+1} (x_j - d_{j}) + \alpha_{j+1} \]

Department of Transportation Engineering University of Naples “Federico II”

QUASI DYNAMIC O-D ESTIMATION

Experiments on the real test site of A4 - A23 motorways in North-Eastern Italy

The simultaneous estimator provides a noteworthy reduction but is outperformed by the QD

Results

<table>
<thead>
<tr>
<th>Day</th>
<th>Seed matrix</th>
<th>Simultaneous updating</th>
<th>Quasi-dynamic (p=7,36)</th>
<th>Kalman filter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>absolute error</td>
<td>relative error</td>
<td>absolute error</td>
</tr>
<tr>
<td></td>
<td></td>
<td>estimation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>intrinsic error</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>updating</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>true seeds</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>simultaneous seeds</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>quasi-dynamic seeds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MO</td>
<td>April 14th</td>
<td>32.9</td>
<td>20.9</td>
<td>-37%</td>
</tr>
<tr>
<td></td>
<td>April 15th</td>
<td>36.1</td>
<td>21.0</td>
<td>-40%</td>
</tr>
<tr>
<td>TH</td>
<td>April 16th</td>
<td>36.7</td>
<td>21.6</td>
<td>-41%</td>
</tr>
<tr>
<td></td>
<td>April 17th</td>
<td>38.7</td>
<td>21.9</td>
<td>-36%</td>
</tr>
<tr>
<td>TH</td>
<td>April 18th</td>
<td>48.2</td>
<td>23.3</td>
<td>-52%</td>
</tr>
</tbody>
</table>

Effectiveness of the QD-GLS estimator: significant reduction with respect to the initial perturbed seed o-d flows

The performances of the Kalman filter are entirely dependent on the quality of the seed o-d flows
QUASI DYNAMIC O-D ESTIMATION

Experiments on the real test site of A4 - A23 motorways in North-Eastern Italy

Results of hourly o-d estimates obtained with the static GLS estimator and by aggregating simultaneous, QD-GLS and Kalman filter estimates

<table>
<thead>
<tr>
<th>Day</th>
<th>All links</th>
<th>All links</th>
<th>All links</th>
<th>All links</th>
<th>All links</th>
<th>All links</th>
<th>All links</th>
<th>All links</th>
<th>All links</th>
<th>All links</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intrinsics</td>
<td>Intrinsics</td>
<td>Intrinsics</td>
<td>Intrinsics</td>
<td>Intrinsics</td>
<td>Intrinsics</td>
<td>Intrinsics</td>
<td>Intrinsics</td>
<td>Intrinsics</td>
<td>Intrinsics</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.10</td>
<td>0.20</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.60</td>
<td>0.70</td>
<td>0.80</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.10</td>
<td>0.20</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.60</td>
<td>0.70</td>
<td>0.80</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
</tr>
<tr>
<td></td>
<td>QD-GLS</td>
<td>QD-GLS</td>
<td>QD-GLS</td>
<td>QD-GLS</td>
<td>QD-GLS</td>
<td>QD-GLS</td>
<td>QD-GLS</td>
<td>QD-GLS</td>
<td>QD-GLS</td>
<td>QD-GLS</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.10</td>
<td>0.20</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.60</td>
<td>0.70</td>
<td>0.80</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
</tr>
<tr>
<td></td>
<td>Simultaneous seeds</td>
<td>Simultaneous seeds</td>
<td>Simultaneous seeds</td>
<td>Simultaneous seeds</td>
<td>Simultaneous seeds</td>
<td>Simultaneous seeds</td>
<td>Simultaneous seeds</td>
<td>Simultaneous seeds</td>
<td>Simultaneous seeds</td>
<td>Simultaneous seeds</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.10</td>
<td>0.20</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.60</td>
<td>0.70</td>
<td>0.80</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
</tr>
<tr>
<td></td>
<td>Quasi-dynamic seeds</td>
<td>Quasi-dynamic seeds</td>
<td>Quasi-dynamic seeds</td>
<td>Quasi-dynamic seeds</td>
<td>Quasi-dynamic seeds</td>
<td>Quasi-dynamic seeds</td>
<td>Quasi-dynamic seeds</td>
<td>Quasi-dynamic seeds</td>
<td>Quasi-dynamic seeds</td>
<td>Quasi-dynamic seeds</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.10</td>
<td>0.20</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.60</td>
<td>0.70</td>
<td>0.80</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
<td>RMSE</td>
</tr>
</tbody>
</table>

The QD-GLS estimator outperforms both the static and the simultaneous estimators.

Department of Transportation Engineering University of Naples “Federico II”

QUASI DYNAMIC O-D ESTIMATION

Experiments on the real test site of A4 - A23 motorways in North-Eastern Italy

Distances between the observed link flows and the link flows obtained assigning the updated o-d flows

The intrinsic bound is the distance between the observed link flows and the flows obtained assigning the “true” quasi-dynamic o-d flows. The QD-GLS estimator is very robust on the hold-out sample, outperforming the simultaneous estimator and allowing the Kalman filter to obtain very effective results.

The simultaneous estimator outperforms others for counted links.

Department of Transportation Engineering University of Naples “Federico II”
CONCLUSIONS

STATIC

✓ easier to implement (static assignment matrix)
✓ faster to compute

DYNAMIC

✓ better estimators of static flows!
✓ needed for dynamic assignment
✓ quasi-dynamic outperforms simultaneous
✓ quasi-dynamic computationally more time-demanding

CONCLUSIONS

Innovative vehicles tracking system
Example on board units in Italy
80’000 vehicle trajectories (two months)
CONCLUSIONS
Innovative mobility surveys

CONCLUSIONS
Innovative info sources to estimate O-D flows