1. INTRODUCTION

The problem of the estimation or updating of origin/destination (O/D) matrices from traffic counts both for road and transit networks have been analysed and many methods have been presented in literature during the last 20 years; they can be mainly classified in static and dynamic estimation methods.

The static estimation problem refers to the estimation of the O/D matrix relative to a reference period from traffic counts measured in the same period over some links of the network. Static estimation methods differ among them in relation to the used estimator and if they consider congested networks or otherwise. Many estimation methods have been presented for uncongested transit networks on the basis of the maximum entropy principle (Van Zuylen and Willumsen, 1980), the Bayesian inference (Maher, 1983), the Maximum likelihood estimator (Spiess, 1983) and the Generalised Least Squares one (Cascetta, 1984; Bell, 1991). Estimation methods which consider congestion are those proposed by Fisk (1988), Florian and Chen (1995) and Cascetta and Postorino (2001).

The dynamic estimation can be defined as the estimation of a time-varying O/D matrix from time-varying traffic counts. For road networks the theoretical framework of the dynamic O/D estimation was defined in Cascetta et al. (1993). For transit networks only few works have been presented both using a maximum entropy estimator, but considering different approaches in transit modelling: the frequency-based approach and the schedule-based one. One of the frequency-based methods is that of Nguyen et al. (1989), while in the framework of the schedule-based approach Wong and Tong (1998) presented an estimation method for urban transit networks that considers a deterministic path choice model in which the time-dependent least cost paths, one for each od pair, are defined.

In the sphere of low-frequency (extraurban) uncongested transit networks, this paper presents a methodology to estimate dynamic (time-varying) transit O/D matrices from time-varying traffic counts using a Generalised Least Squares estimator and a schedule-based dynamic path choice model of stochastic type. This methodology is actually used in the Decision Support System of the Italian Railways to estimate the railway O/D matrices at national scale.

Section 2 defines the general formulation of the problem. Section 3 describes the schedule-based dynamic approach in transit modelling which is used in the estimation method presented in section 4 in order to estimate a time-varying O/D matrix for extraurban transit networks. Finally, section 5 reports an application example to a realistically-sized transit network.
2. THE GENERAL FORMULATION

The dynamic estimation problem can be specified extending the relationships between flows, counts and demand, formalised in the static approach, in order to explicitly consider the time dependencies. In particular, it is necessary to describe the relationship between time-varying link counts and demand referring to the time interval \( i \) in which demand relative to the \( od \) pair leaves the origin and the time interval \( j \) in which link counts \( f_j \) are measured.

The reference period \( T \) is divided into \( n \) time slices (e.g. of 15 minutes each) and each time slice \( i \) is represented by its middle point \( \tau_{Di} \), which represents the origin departure time for users with departure time in time interval \( i \). In the following \( d_{\tau_{Di}} \) indicates the demand vector with origin departure time \( \tau_{Di} \), whose generic element \( d_{\tau_{Di}}^{od} \) represents the demand relative to the \( od \) pair and origin departure time \( \tau_{Di} \).

Hence, the dynamic estimation can be defined as the estimation of a set of time-varying trip vectors \( d_{\tau_{Di}} \) (one for each time interval) from time-varying traffic counts (see fig.1), whose generic element \( d_{\tau_{Di}}^{od} \) represents the number of trips on the \( od \) pair leaving the origin during time interval \( i \).

\[ \text{Figure 1 – Example of dynamic (time-varying) O/D trip matrices.} \]

In the following we indicate: \( h_i \) as the path flow vector, whose generic element \( h_{i,k} \) represents flow on path \( k \) relative to the \( od \) pair and departing in the time interval \( i \); \( P_i \) as the path choice matrix relative to time interval \( i \), whose generic element \( p[k/od, \tau_{Di}] \) represents the fraction of \( d_{\tau_{Di}}^{od} \) using path \( k \); \( B_{ij} \) as the crossing fraction matrix, whose generic element \( b_{i,j}^{k,l} \) represents the fraction of \( d_{\tau_{Di}}^{od} \) using path \( k \) and contributing to flow on link \( l \) in time interval \( j \); \( M_{ij} \) as the assignment matrix, whose generic element \( m_{i,j}^{od,\tau_{Di}} \) represents the fraction of demand \( d_{\tau_{Di}}^{od} \) contributing to flow on link \( l \) during time interval \( j \).

The flow on link \( l \) during the time slice \( j \) \((f_{i,j})\) can be written as

\[ f_{i,j} = \sum_{i=1}^{J} \sum_{k} b_{i,j}^{k,l} \cdot h_{i,k} \quad (1) \]

and considering all link counts in a matrix form, we have
The link flow vector $f_j$ can be calculated through a dynamic assignment model and is usually different from the vector of link counts relative to time interval $j$, $\hat{f}_j$, whose generic element $\hat{f}_{i,j}$ represents counts on link $l$ during time interval $j$.

The estimation of the O/D trip demand vector $d = (d_{\tau_D_1}, d_{\tau_D_2}, \ldots, d_{\tau_D_n})$ can be carried out by “efficiently” combining traffic counts with all other available information. Estimators can be classified in classic estimators (like the Maximum Likelihood or the Generalised Least Squares), if they combine experimental information (sample surveys) with traffic counts, and Bayesian estimators, if they combine non-experimental information (“a priori” information) with traffic counts.

As formalised in the traditional static approach, the estimation of time-varying O/D matrices from time-varying on-board counts can be written as

$$ d = (d_{\tau_D_1}, d_{\tau_D_2}, \ldots, d_{\tau_D_n}) = \arg \min \{ z_1[(x_1, \ldots, x_n); (\hat{d}_{\tau_D_1}, \ldots, \hat{d}_{\tau_D_n})] + z_2[(\hat{f}_1, \ldots, \hat{f}_n); (\hat{\hat{f}}_1, \ldots, \hat{\hat{f}}_n)] \}$$

Eqn(3) represents an optimisation problem made of the minimisation of the “distance” between the unknown time-varying demand $x = (x_1, \ldots, x_n)$ and the a priori time-varying demand $\hat{d} = (\hat{d}_{\tau_D_1}, \ldots, \hat{d}_{\tau_D_n})$, represented by the function $z_1[.]$, and between time-varying link flows $\hat{f} = (\hat{f}_1(x), \ldots, \hat{f}_n(x))$, obtained using a dynamic assignment model, and the observed link flows (time-varying on-board counts) $\hat{\hat{f}} = (\hat{\hat{f}}_1, \ldots, \hat{\hat{f}}_n)$, represented by the function $z_2[.]$. Functions $z_1[.]$ and $z_2[.]$ can be differently specified according to the chosen estimator.

In the sphere of the dynamic O/D estimation a further classification in simultaneous and sequential estimators can be done (Cascetta et al., 1993). 

Simultaneous estimators allow the joint estimation of O/D matrices, one for each time slice, to be obtained as reported in eqn(3). Sequential estimators provide a sequence of estimated O/D matrices, in which the estimation of the O/D matrix for the time slice $j$ depends on the estimation of the O/D matrices relative to the previous time slices $j-1, j-2, \ldots$. Hence the estimation of the demand vector $d_{\tau_D_j}$, given the estimates $(d_{\tau_D_1}, d_{\tau_D_2}, \ldots, d_{\tau_D_{j-1}})$ relative to previous time intervals, can be written as

$$ d_{\tau_D_j} = \arg \min \{ z_1[x_j; \hat{d}_{\tau_D_j}] + z_2[\hat{f}(x_j / d_{\tau_D_1}, \ldots, d_{\tau_D_{j-1}}); \hat{\hat{f}}_j] \}$$

Whatever the formulation and used estimator are, the core of the problem is the computation of the link flow vector $f_j$ of eqn(2) through the estimation of the assignment matrix $M_{i,j}$.

Assignment matrix $M_{i,j}$ can be obtained through an estimation of the crossing fraction matrix $B_{i,j}$ and the path choice matrix $P_i$, using one of the two possible
approaches in transit modelling: the frequency-based approach and the schedule-based one.

The frequency-based approach considers services in terms of lines. In this case run scheduled times are not considered explicitly, but we refer to the service frequency, from which the name of the approach derives. Path choice is simulated on the basis of the concept of hyperpath or optimal strategies (Nguyen and Pallottino, 1988; Spiess and Florian, 1989). The main hypothesis of the hyperpath model is that users choose the line in an indifferent adaptive way, that is they board the first arriving run belonging to their line choice set, defined as the set of lines that minimises the expected total travel time. The underlying hypotheses that allow such results are that vehicles arrive at stops in a completely random way and that users arrive at stops with constant rate. The above described hypotheses are fully acceptable for services with high frequency, very low punctuality and low user information, but can generate considerable approximations when used in different contexts, in particular when ITS (Intelligent Transportation Systems) are present, or in the case of low-frequency services. In fact in the case of high-frequency and not completely random services, and in particular with real time information at stops about vehicle arrival times, user behaviour can be assumed intelligent adaptive (i.e. the choice of the run at the stop, in the attractive choice set, is made by comparing the attributes of alternatives, particularly the waiting time), so the real arrival and departure times of transit vehicles have to be taken into account. In addition, for low-frequency services, the difference between user desired departure/arrival times and real run departure/arrival times have to be explicitly considered as a further disutility component in user path choice. Hence, the use of the frequency-based approach is not the best, since it does not let us take into account these aspects, producing errors in the prediction of on-board flows, waiting times, transfer times and transfer stops. Moreover, as in frequency-based approach we can compute only average line on-board loads, we are unable to take into account peaks of loads on vehicles that can occur inside the reference period. Finally, the use of a constant average rate of user arrivals at stops, when this rate changes significantly during the period of analysis, could lead to relevant errors in vehicle load calculations.

In order to overcome these problems and to consider more coherent assumptions in relation to user and service characteristics, in recent years the schedule-based approach has been developed and applied. This approach refers to services in terms of runs, using the real vehicle arrival/departure times at stops to obtain attributes that can be explicitly considered in run choice. As will be deepened in the following section, this approach allows us to take into account the evolution in time both of supply and demand, as well as run loads and level of service attributes.

3. THE SCHEDULE-BASED APPROACH TO TRANSIT SIMULATION

The schedule-based dynamic approach to transit services is more complex than the frequency-based one, as transportation system time-dependencies have to be explicitly considered both during a single day period (within-day dynamic) and in following periods (day-to-day dynamic). For this reason, this approach requires an explicit treatment for the origin/destination matrix (as user’s departure or arrival target time distribution, both within-day and day-to-
day, has to be considered), the supply model (as service supply variations during the reference period have to be taken into account) and the path choice model (which has to consider both within-day and day-to-day time-dependencies). Models presented in literature differs among them in relation to the way in which they specify these aspects both for low-frequency (Nuzzolo and Russo, 1994; Carraresi et al, 1996; Cascetta et al, 1996; Florian, 1998; Nielsen and Jovicic, 1999; Nuzzolo et al, 2000) and high-frequency transit services (Hickman and Wilson, 1995; Nuzzolo and Russo, 1998; Wong and Tong, 1999; Nuzzolo et al., 2001).

In the following sections the main characteristics of the schedule-based dynamic approach for transit networks proposed by Nuzzolo et al. (2001) will be described in details.

3.1 The Diachronic Representation of Services

Scheduled services can be represented through a space-time or diachronic graph $\Omega$ (Nuzzolo and Russo, 1994). The diachronic graph $\Omega$ consists of three different sub-graphs (fig.2) in which each node has an explicit time coordinate: a service sub-graph $\Omega_g$ in which each run of each line is defined both in space, through its stops, and in time, according to its arrival/departure times at stops; a demand sub-graph $\Omega_d$, in which each node represents a temporal centroid, in order to simulate the space-time characteristics of the trip; an access/egress sub-graph $\Omega_{ae}$ which allows the connection of the demand sub-graph with access/egress stops and stops between them.

The service sub-graph $\Omega_g$ consists, in turn, of different sub-graphs $\Omega_{g,r}$ (one for each run of the transit services). Referring to the generic run $r$, the relative sub-graph $\Omega_{g,r}$ consists of nodes representing the arrival and departure times at stops, and links representing travel from one stop to another (run section) or the dwelling of the vehicle at a given stop. Other nodes represent the time in which users board or alight from each run at the stop. These nodes are connected to the nodes representing run arrival and departure through boarding and alighting links. Finally, the whole sub-graph $\Omega_g$ is built by connecting all the sub-graphs $\Omega_{g,r}$ through links representing the user transfer from one run to the next one at the same stop (stop axis).

The demand sub-graph $\Omega_d$ representing temporal demand segmentation is made up, in turn, by the same number of sub-graphs $\Omega_{d,c}$ as centroids. For each spatial centroid $c$, the sub-graph $\Omega_{d,c}$ consists of temporal centroids, that are nodes located spatially in the position of the spatial centroid $c$ and temporally according to the user origin departure times $\tau_{Di}$.

The access/egress sub-graph $\Omega_{ae}$ is made up of: links connecting origin temporal centroids to nodes on the first boarding stop axis, to represent the access to the services; links connecting alighting nodes of the stop axis to the destination spatial centroids, to represent the egress from the transit system; links connecting stops, to represent possible interchanges between different stops.

Finally, the global diachronic graph $\Omega$ is obtained linking properly the three sub-graphs and the supply configuration $b_j$ is represented through the relative graph $\Omega_j$. 

© Association for European Transport 2001
In the dynamic approach, a path $k$ is defined both in space and in time and includes the space-time sequence: origin $o$ with origin departure time $\tau_{Di}$, access stop $s$ with relative arrival time $\tau_{Dis}$, run (or sequence of runs) with run departure time $\tau$ from access stop, egress stop $s'$ and destination $d$ with relative arrival time at destination $\tau_d$. A path can be defined through a sequence of links of the $\Omega$ graph as reported in fig.2. The diachronic graph is very useful as it allows efficient use of standard network algorithms (as least-cost paths) and allows more straightforward treatment of time-varying flows and counts.

### 3.2 The Dynamic Stochastic Path Choice Model

Schedule-based dynamic path choice models, that are the core of the dynamic assignment models, can be classified according to transit system characteristics they consider and namely the service frequency: medium-low frequency, which is typical of regional or low-demand areas, and high frequency, which usually refers to urban transit networks.

As in the following the proposed estimation method refers to a low-frequency (extraurban) transit network, stochastic path choice and assignment models will be specified for low-frequency transit services following the approach described in Nuzzolo et al. (2000). The specification for high-frequency transit services can be deepened on Nuzzolo et al. (2001).
Low-frequency transit services are characterised by regular service functioning and it is possible to assume that users have all information at their disposal before starting the trip, so path choice can be assumed fully pre-trip and concerns run (or sequence of runs) choice. In the framework of the random utility theory users choose the path minimizing the perceived disutility $U_r$, sum of a systematic utility $V_r$ and a random term $\varepsilon_r$, expressed as

$$U_r = V_r + \varepsilon_r = \sum_j \beta_j X_{jr} + \varepsilon_r$$  \hspace{1cm} (5)$$

Systematic utility $V_r$ is a linear combination, through $\beta_j$ parameters, of attributes $X_{jr}$, which mainly regard access time, on-board time, transfer time, number of transfers, monetary cost, comfort, in addition to the disutilities that occur because of the difference (that can be considerable) between desired user departure time and vehicle scheduled departure time or between desired user arrival time at destination and run scheduled arrival. In the literature, this difference is called early schedule penalty or late schedule penalty (Nuzzolo et al., 2000) and an example is reported in fig. 3.

The probability $p[r|od, \tau_{Di}]$ of choosing run $r$, given the od pair and origin departure time $\tau_{Di}$, which represents the elements of the path choice matrix $P_i$ of eqn(2), can be expressed as

$$p[r|od, \tau_{Di}] = \text{prob}(U_r > U_r') = \text{prob}(V_r + \varepsilon_r > V_{r'} + \varepsilon_{r'}) \quad \forall r \neq r'$$  \hspace{1cm} (6)$$

Different path choice models can be specified according to the service structure. In particular we can refer to two types of transit systems: single class-single service transit systems (e.g. regional bus services) and multi-class, multi-service transit systems (e.g. railways services). For single class-single service transit systems, according to the hypotheses on random residuals $\varepsilon_r$, one of the random utility models can be specified. Path choice models for multi-class, multi-service transit systems (e.g. railways) can be defined in the same way, except for the model specification that should consider the existing correlation between alternatives through the use of nested-logit or probit models.

Figure 3 – Example of early/late schedule penalty.
4. THE PROPOSED ESTIMATION METHOD

This section presents a method to estimate time-varying O/D matrices from time-varying on-board counts for uncongested transit networks. In this case the use of a simultaneous estimator is fully acceptable under the assumption of user origin departure time fixed. Using a Generalised Least-Squares (GLS) estimator (Cascetta, 1984), the estimation problem (3) can be written as

\[
\mathbf{d} = \arg\min_{\mathbf{x}} \left\{ \sum_{j=1}^{n} (x_j - \hat{d}_j)^T M_j^{-1} (x_j - \hat{d}_j) + \sum_{i=1}^{n} (\sum_{j=1}^{n} M_{i,j} \cdot x_i - \hat{I}_j)^T W^{-1} (\sum_{i=1}^{n} M_{i,j} \cdot x_i - \hat{I}_j) \right\}
\]

where \( \mathbf{W} \) is the variance-covariance matrix relative to vector \( \mathbf{\epsilon} = (\xi_1, \ldots, \xi_j, \ldots, \xi_n) \) of assignment model and counting errors, characterised by \( E(\mathbf{\epsilon}) = 0 \) and \( \text{var}[\mathbf{\epsilon}] = \mathbf{W} \); \( \hat{d}_j \) is an available estimation of the unknown demand vector \( \mathbf{x}_j \) obtained through a random sample such as

\[
\hat{d}_j = \mathbf{x}_j + \eta_j
\]

in which \( \eta_j \) is a random vector that considers demand sample errors, characterised by \( E(\eta_j) = 0 \) and \( \text{var}[\eta_j] = \mathbf{Z}_j \).

In order to solve problem (7), it is necessary to calculate the assignment matrix \( M_{i,j} \) defined as

\[
M_{i,j} = B_{i,j} \cdot P_i
\]

This paper proposes the use of the schedule-based dynamic transit modelling approach described in section 3 to calculate in a more precise and direct way the crossing fraction matrix \( B_{i,j} \) and the path choice matrix \( P_i \).

In fact, as each link of the diachronic graph represents a precise location in both space and time, it is possible to calculate in a simpler way the crossing fraction matrix \( B_{i,j} \), made of elements \( b_{i,j}^{k,l} \), because each path \( k \) is individuated on the graph by a sequence of links that allow to define directly the fraction of path load \( h_{i,k} \) crossing link \( a \) in time interval \( j \).

Elements of path choice matrix \( P_i \) can be calculated by eqn(6), which represents the probability of choosing path \( k \) (individuated by run \( r \)) for users travelling on the \( od \) pair and departing in time interval \( i \).

Hence, using the schedule-based dynamic approach described in section 3, eqn(9) can be easily calculated and, considering the diachronic network model, eqn(7) can be rewritten through the same formulation of the static estimation problem as

\[
\mathbf{d} = \arg\min_{\mathbf{x}} \left\{ ((\mathbf{x} - \hat{\mathbf{d}})^T \mathbf{Z}^{-1} (\mathbf{x} - \hat{\mathbf{d}}) + (\hat{\mathbf{M}} \cdot \mathbf{x} - \hat{\mathbf{I}})^T \mathbf{W}^{-1} (\hat{\mathbf{M}} \cdot \mathbf{x} - \hat{\mathbf{I}})) \right\}
\]

where it is better to highlight that both demand and flow vectors have a space-time characterisation which allows time-varying demand, flows and counts to be considered.
This approach allows problem (10) to be easily solved by using a traditionally project gradient algorithm. An application of the proposed estimation method is reported in the following section.

5. AN APPLICATION EXAMPLE
This section reports an application example of the proposed methodology to test its applicability to real transit networks. The application regards the estimation railway O/D matrices at national scale, which are made of 13420 od pairs for each service (high speed and slow speed trains) and class (first and second). The estimation is carried out on the basis of about 868 time-varying counts on 257 of the 334 trains of the railway services over several sections. The reference period (one day) was divided into 15 time intervals of one hour each in the rush hours, while wider intervals for other periods were defined. An origin departure time for each time interval was considered. In order to calculate path choice probabilities (6), the path choice model structure, attributes and parameters are those reported in Nuzzolo et al. (2000).

The starting O/D matrices are the railway O/D matrices of sold tickets, which are available by automatic station reports. These O/D matrices differ from the real ones as they do not consider on-board sold tickets, railway passes and free tickets, in addition to the fact that tickets users buy could be used in one of the next 60 days. For this reason, in order to simulate in a more precise way railway services, the updating of railway O/D matrices of sold tickets from traffic counts is crucial.

The accuracy of a solution is evaluated by the difference between the generic O/D matrix \( \hat{d} \) and the real one \( d \). Different statistics like the root mean square error (rmse), the relative mean error (rme) and the mean square error (mse) are used; they are defined as follows:

\[
\begin{align*}
\text{rmse} &= \sqrt{\frac{\sum \text{od} \sum |d_{j}^{\text{od}} - \overline{d}_{j}^{\text{od}}|^2}{N}} \\
\text{rme} &= \frac{\sum \text{od} |d_{j}^{\text{od}} - \overline{d}_{j}^{\text{od}}|}{\sum \text{od} d_{j}^{\text{od}}} \\
\text{mse} &= \frac{\sum \text{od} (d_{j}^{\text{od}} - \overline{d}_{j}^{\text{od}})^2}{N}
\end{align*}
\]

(11)

where \( N \) represents the number of observation (i.e. if \( n \) is the number of considered time intervals, it is \( n \) times the number of od pair). Performances of the considered estimator have been considered by computing average values of statistics (11) on the basis of different starting matrices \( \hat{d} \) obtained through random perturbations of the real O/D matrix \( d \) as

\[
\hat{d}_{j}^{\text{od}} = \overline{d}_{j}^{\text{od}} + \phi (u - 0.5) \overline{d}_{j}^{\text{od}}
\]

(12)

where \( u \) is a value extracted by an uniform \((0,1)\) random variable and \( \phi \) is a parameter that is 0.7, 1.4, 2.0 to which correspond variation coefficients of 0.35, 0.70 and 1, respectively.

© Association for European Transport 2001
Estimated O/D matrices have been carried out by using a project gradient algorithm, in which the optimal solution is reached when the difference between two successive iterations does not exceed the 1% ($\varepsilon=0.01$).

For the above described application, using a Personal Computer based on a PentiumIII700MHz processor, the algorithm converges in less than two hours. Values of the objective function carried out during the optimisation phase are pictured in fig.4, while statistics for the three different estimations ($\phi = 0.7, 1.4, 2.0$) are reported in table 1. Analysing statistics results of table 1 we can see that the method succeeded in improving the starting matrix, showing that the absolute reductions are larger as higher the perturbation is. Different applications with different number of time intervals and perturbations have been considered showing that the percentage of reduction are approximately constant and improvements are slightly larger for higher number of intervals (i.e. smaller interval widths).

![Figure 4 – Estimation algorithm: obj. function values.](image)

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>rme</th>
<th>mse</th>
<th>rmse</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>0.03</td>
<td>794.0</td>
<td>0.26</td>
</tr>
<tr>
<td>1.4</td>
<td>0.05</td>
<td>1216.7</td>
<td>0.32</td>
</tr>
<tr>
<td>2.0</td>
<td>0.09</td>
<td>2300.4</td>
<td>0.44</td>
</tr>
</tbody>
</table>

6. CONCLUSIONS
In this paper a method to estimate dynamic (time-varying) O/D matrices for public transport networks from time-varying on-board counts has been presented. The estimation problem has been formalised for low-frequency transit services assuming that the network is not congested and using a schedule-based transit modelling approach to carry out in a more precise and coherent way model variables (flows) to be used in the adopted GLS estimator. An application example to test the applicability of the proposed methodology for operations planning has been described. This methodology is actually used in the Decision Support System of the Italian Railways to estimate the railway O/D matrices at national scale.
Further developments of this research are underway and mainly regard the specification of the dynamic estimation problem for congested transit network and the problem formalisation for high-frequency services to be used for the simulation of urban public transport networks.

Bibliography
Maher, M.J. (1983) Inference on trip matrices from observations on link volumes. A Bayesian statistical approach, Transportation Research, 17B.


