A generation constrained approach for the estimation of O/D trip matrices from traffic counts

Domenico Iannò
Urban Transport Company, Reggio Calabria, Italy
Maria Nadia Postorino
"Mediterranea" University of Reggio Calabria, Italy

1. INTRODUCTION

The analysts of the transportation systems know that one of the most important problems in the simulation and planning of these systems is the evaluation of the origin/destination (OD) trip matrix, or in other words of the number of trips made by users in a pre fixed area and in a reference time interval by using one or more transportation modes.

The accuracy of the OD estimate is a crucial aim in most practical applications because significant variations with respect to the true, unknown, trip matrix could result in incorrect forecasting of the traffic flow values on the transportation network links and then in incorrect evaluation of the effects of possible modifications of the system.

The methods proposed in the literature for the evaluation of the OD trip matrix can be grouped in three classes: direct estimation, model estimation and estimation from traffic counts. Among these, the estimation from traffic counts has received a great attention in the last years, thanks to the cheapness of the information sources. In fact, while the first two methods require a great effort in terms of time and economic resources, because the estimate of the OD matrix is obtained by using directly or indirectly the data collected by interviews to the users, the latter method requires only the use of a set of traffic flow values, counted on some suitably chosen links of the transportation network, in order to correct and improve an initial, old estimate of the OD matrix.

For this reason, different models and different applications have been proposed in the literature to resolve the problem of estimating the OD trip matrix by using link traffic counts, named in the following OD Count Based Estimation (ODCBE) problem. Although the proposed approaches are different in the mathematical formulation, almost all the methods present in the literature need an initial estimate of the OD matrix (called *target matrix*) and a set of link traffic flows measured on the transportation network considered. Most of them are formulated as optimisation or mathematical programming problems with an objective function and a set of constraints.

The aim of the models is to reproduce the counted traffic flows by updating an old OD estimate until a prefixed threshold test is satisfied. Then, a part of the mathematical formulation of the proposed models refers to the reproduction of the traffic flows by using an assignment procedure, or, alternatively, an assignment matrix, whose elements are the OD flow percentages that use each link of the network for which the traffic counts are available. These percentages depend on the generalised costs for all the links of the network, but for congested networks generally these costs are not known by the analyst and they are computed on the basis of the link flows forecasted through the assignment model, that again depend on the OD matrix to be assigned.

Different approaches have been proposed in the literature for resolving the ODCBE problem, depending on the nature of the obtained estimates and the structure of the assignment matrix. From this latter point of view, two approaches can be identified: ODCBE problems with constant link costs (uncongested transportation networks) and ODCBE problems with flow-function link costs (congested transportation networks). Both the first and the second approach have been largely used in the literature; for uncongested networks, some estimators are based on the maximum entropy principle (Van Zuylen and Willumsen, 1980; Willumsen, 1984), other methods are derived from classical statistical approaches such as the Generalised Least Square (GLS) estimator (Cascetta, 1984; Bell, 1991), or Bayesian inference (Maher, 1983), while Brenninger-Göthe *et al*., 1989, following a GLS approach, proposed a multiobjective programming formulation. A systematic analysis of the statistical foundations of the different methods for uncongested networks is proposed by Cascetta and Nguyen, 1988. For congested networks, contributions are given by Nguyen (1977), Fisk and Boyce (1983); the bi-level approach is used by Fisk (1988), Yang (1995) and Chen and Florian (1994), while Cascetta and Postorino (2001) propose a fixed point approach.

The second distinction refers to the nature of the obtained estimates, and two approaches can be identified: static or dynamic, depending on the hypotheses made about the target matrix and the corresponding updated OD matrix. In the static case, only the mean value of the OD matrix is estimated by using the information contained in the mean traffic flow values measured on some transportation links in the same reference time period. In the dynamic case, the information contained in the link traffic flows measured in consecutive time periods are used to update the corresponding OD trip matrix reproducing those flows for sequential time periods. Then, the OD trip matrix obtained is referred to consecutive time periods and it is variable in the time (Nguyen et al., 1989; Cascetta et al., 1993; Ashok and Ben-Akiva, 1993; Wong and Tong, 1998).

A particular aspect of the ODCBE problem is that the same counted traffic flow values can be obtained by different OD matrices, as it will be explained in the following, so even if the model allows to reproduce the traffic flow values measured on the transportation network, however the traffic flows on the remaining, no counted links of the transportation network could be reproduced badly.

In order to take into account this important aspect, in this paper some other sources of external information have been used to obtain more reliable estimates of the OD trip matrix. Particularly, the emission data from each trip origin traffic zone have been introduced in the model in order to make congruent the OD trip estimate with the trip emission from each zone.

The problem has been resolved by using a GLS approach and a stochastic assignment model, by modifying the objective function in order to take into account the emission data.

The paper is organized as follows: section 2 reports the general formulation of the ODCBE models, and focuses on the GLS approach; section 3 describes the proposed approach and the modification of the GLS function in order to take into account the emission data; section 4 reports the results obtained with the proposed methodology and finally section 5 reports the conclusion of the work.

2. GENERAL FORMULATION OF THE ODCBE PROBLEM

Most of the ODCBE problems can be represented by the following general formulation:

$$\min_{\mathbf{v},\mathbf{t}} g_1 F_1(\mathbf{t}, \hat{\mathbf{t}}) + g_2 F_2(\mathbf{v}, \hat{\mathbf{v}}) \qquad (1)$$

s.t.

$$\mathbf{v} = \mathrm{A}(\mathbf{t})$$

where:

- $\hat{\mathbf{t}} = (\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_n)^T$ is an initial estimate of the OD matrix (the so-called target matrix), organized in a vector so that the i-th component corresponds to the i-th OD pair, for each $i \in I$, I being the set of the n OD pairs;
- $\hat{\mathbf{v}} = (\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_m)^T$ is the vector of the counted link flows on a subset S of the whole network link set L;
- **t** and **v** represent respectively the current estimate of the OD matrix to be updated and the link traffic flows reproduced on the transportation network links of the subset S, corresponding to **t**; they are obtained under some hypotheses of the assignment model;
- F_1 and F_2 are two functions that measure the generalized distance between the current estimate of the OD matrix and the target matrix, and between the simulated traffic flow values and the measured ones; F_1 and F_2 can be variously specified, for example as entropic function or Euclidean distance;
- **v**=A(**t**), called assignment map, is the relationship linking the simulated traffic flow values **v** to the OD matrix **t**; A(**t**) can be variously specified, following the hypotheses made on the transportation network (congested or uncongested) and the path choice model that has to be used to compute the path choice probability and then the link use percentages (stochastic or deterministic path choice models);
- g_1 and g_2 are weights that measure the different confidence level in the initial observations ; they can be constant for all the initial data or can vary for each of them. In this latter case, they are represented by a variance-covariance matrix.

The solution of the problem (1) is an OD matrix **t*** that assigned to the transportation network by using the assignment map A(**t**), reproduces the traffic flows **v** whose distance from the counted ones, $\hat{\mathbf{v}}$, measured by F_2, is the smallest.

As regards the function F_1 and F_2, they can be variously specified leading to different models. Most of them, as Cascetta and Nguyen (1988) showed, can be cast in a general statistical framework; particularly, the Generalized Least Square (GLS) model is robust with respect to distribution assumptions and it has been used by different authors to construct estimators for the OD trip matrix under different hypotheses about the assignment map (among the others, Cascetta, 1984; Yang, 1995; Cascetta and Postorino, 2001).

According to the hypotheses on the assignment map, two groups of models can be identified, based on the hypotheses about the cost functions defined on the transportation network links, as already described in the section 1.
The assignment map A(**t**) can be generally specified as:

$$\mathbf{v} = A(\mathbf{t}) = \mathbf{M}(\mathbf{t})\mathbf{t} \qquad (2)$$

where **M**(**t**) is the assignment matrix whose elements m_{li}'s are the link use percentages of link l, connecting the i-th OD pair. Generally, **M** could be a function of **t**, and it depends on the hypotheses made about the cost functions.
If the network is considered uncongested, then:

$$\mathbf{M}(\mathbf{t}) = \mathbf{M} = \mathbf{AP(C)}$$

where **A** is the link-path incidence matrix, whose elements have value 1 if the link l belongs to the path k, or value 0 if the link l does not belong to the path k; **P** is the path choice probability matrix, whose elements assume values equal to the probability of choosing the path k connecting the i-th OD pair, if k connects i, and 0 otherwise; **C** is the vector of path costs, linked to the link costs by the relationship:

$$\mathbf{C} = \mathbf{A}^T\mathbf{c}$$

In the case of uncongested networks, the path cost, that in turn is related to the link cost functions, does not depend on the path flow and then **M** can be estimated at the beginning and the elements m_{li} are constant values.
If the hypothesis of congested network is considered more realistic (as in the case of individual transport networks for urban areas), then:

$$\mathbf{M}(\mathbf{t}) = \mathbf{AP(C(h))}$$

where **h** is vector of the path flow, that is the percentage of the OD demand, for a given OD pair, that uses the path k connecting the pair itself; path flows and link flows are linked by the relationship:

$$\mathbf{v} = \mathbf{Ah}$$

The network being congested, the cost functions depend on the traffic flows and then the **P** matrix is a function of the OD demand. **M** cannot be estimated only once, because its values vary with **t**, the current estimate of the unknown OD trip matrix **t***. In this case, the ODCBE problem can be resolved by using a bi-level approach (Fisk, 1988, 1989; Bell, 1991; Chen and Florian, 1994; Yang, 1995) or a fixed point approach (Cascetta and Postorino, 2001).
Another aspect referred to the assignment matrix is that **M** is not known but it can only be estimated; in this latter case, the values of its elements depend on the specific hypotheses about the path choice model (deterministic or stochastic) other than the cost functions (constant or flow-dependent); generally, the following relationship between the true, unknown assignment

matrix **M** and the estimated one $\hat{\mathbf{M}}$ can be written (Cascetta and Postorino, 2001):

$$\mathbf{M(c)}=\hat{\mathbf{M}}(\mathbf{c})+\mathbf{E}^{sim}$$

where $\mathbf{E}^{sim}$ is a matrix of unknown assignment or simulation errors, one for each element of the assignment matrix; the estimated matrix **M*** is given by:

$$\hat{\mathbf{M}}(c)=\mathbf{A}\hat{\mathbf{P}}(\mathbf{c}) \tag{3}$$

$\hat{\mathbf{P}}$ being the estimate of the path choice probability matrix **P**.
From (2), link flows can then be simulated by using the assignment hypotheses as:

$$\mathbf{v}=\mathbf{Mt}=(\hat{\mathbf{M}}(\mathbf{c})+\mathbf{E}^{sim})\mathbf{t}=\hat{\mathbf{M}}(\mathbf{c})\mathbf{t}+\mathbf{E}^{sim}\mathbf{t}=\hat{\mathbf{M}}(\mathbf{c})\mathbf{t}+e^{sim} \tag{4}$$

where e^{sim} is a vector of assignment errors related to each link flow.
On the other hand, the counted link flows are different from the actual average flows (because of measurement errors, variability of user choices, and so on) and then the following relationship can be written:

$$\hat{\mathbf{v}}=\mathbf{v}+e^{meas} \tag{5}$$

where e^{meas} is the measurement error vector.
If equations (4) and (5) are combined together, the relationship linking the counted flows to the unknown OD flow vector is:

$$\hat{\mathbf{v}}=\hat{\mathbf{M}}(\mathbf{c})\mathbf{t}+e \tag{6}$$

where $e=e^{sim}+e^{meas}$ includes both the assignment and measurement errors and has a variance-covariance matrix **W**.
Furthermore, the relationship between the true, unknown OD matrix $\mathbf{t}_{true}$ and the target one can be written as:

$$\hat{\mathbf{t}}=\mathbf{t}_{true}+h \tag{7}$$

where h is a vector of errors depending on the sampling or model nature of $\hat{\mathbf{t}}$ and whose variance-covariance matrix is **Z**.
The GLS specification of the general formulation (1) is (Cascetta, 1984):

$$\mathbf{t}^{GLS}=\arg\min_{\mathbf{t}\in T}(\mathbf{t}-\hat{\mathbf{t}})\mathbf{Z}^{-1}(\mathbf{t}-\hat{\mathbf{t}})+(\hat{\mathbf{v}}-A(\mathbf{t}))\mathbf{W}^{-1}(\hat{\mathbf{v}}-A(\mathbf{t})) \tag{8}$$

where **Z** and **W** are the variance-covariance matrices referred respectively to the target OD matrix and the counted flow vector, while T is the feasibility space of the solution **t**. Particularly, the constraint is that the feasible solution **t** must be greater than or equal to zero as it has to be expected because **t** represents the number of trips in the reference time periods.

The formulation (8) specified under the hypotheses of uncongested networks can then be written as (Cascetta, 1984):

$$\mathbf{t}^{GLS} = \arg\min_{\mathbf{t}\geq 0} (\mathbf{t}-\hat{\mathbf{t}})^T \mathbf{V}^{-1}(\mathbf{t}-\hat{\mathbf{t}}) + (\hat{\mathbf{v}}-\hat{\mathbf{M}}\mathbf{t})^T \mathbf{W}^{-1}(\hat{\mathbf{v}}-\hat{\mathbf{M}}\mathbf{t}) \qquad (9)$$

3. THE PROPOSED APPROACH

The most appealing aspect of the ODCBE approach is its cheapness, as already recalled in the introduction, because it does not require expensive interviews that are the basis for both the direct estimation and the model approaches. In the first case, a set of collected data being available, the OD demand is estimated by using the results of the statistical methods; in the second case the collected data are used to calibrate the parameters of the models in order to apply them to the pre fixed study area. The nature of the information is the same in both cases, but the number of the required items could be different; generally, a greater number of interviews needs for the direct estimation, while for the model calibration more information for interviewed user generally has to be collected.
The ODCBE approach, on the contrary, just requires the use of on old OD trip matrix (the target matrix) and the values of some traffic counts suitably located on the network (generally, a subset of about 18-20% of the whole link set could be considered satisfactory).The use of only traffic count values cannot be considered sufficient to update correctly the OD trip matrix: in fact, given only a set of counted traffic flows, there could be more OD matrices that assigned to the network reproduce the same observed flow values, because the number of unknown variables, i.e. the elements of the OD trip matrix, is generally really greater than the number of independent relationships linking the counted flows to the OD trip matrix, this latter relationship expressed by eq. (6).
This is one of the reason why the target matrix is introduced in the ODCBE formulation: in this way, in fact, the number of information increases and a more reliable OD matrix estimate could be obtained. Even if the further information contained in the target matrix is taken into account, however the OD estimate could be unsatisfactory, because there could be some other sources of incorrectness. Particularly, some elements of the OD trip matrix could be greater than the expected ones, for example the trip emission from one zone could be greater than the actual one or the number of trips at destination could be greater than the number of workplaces, students-at-schools and so on.
Because of the traffic flows are the result of the contribution of different OD pairs, as the following relationship shows:

$$v_l = \sum_i m_{li} t_i$$

it could happen that the reproduction of the traffic flows is correct, but the single values of the OD trip matrix could be unrealistic.
Furthermore, nothing can be said about the reproduction of the no-counted traffic flows, that is of the remaining set of links on which no counts are

available; on these links the traffic flows reproduction could be unreliable due to the possible incorrectness of the OD estimate.

In order to take into account this aspect, another sources of information can be considered, that works as an emission constraint and explicitly considers that the estimated trips generated by each traffic zone cannot be greater than the maximum actual emission from the zone itself (similar considerations could be made at destination).

The trip emission depends on the time period considered; e.g., for systematic trips, inside the study area, in the early morning period, it could not be greater than the number of inhabitants that live in the zone. Particularly, the emission is a percentage of the living population in a given zone, under the previous conditions. Similar considerations can be made for other time periods, e.g. in the afternoon (return-to-home) where the emission data could refer to the actual population present in the zone at that time. However, the nature of the estimate also depends on the considered approach, static or dynamic. For the static case, considering a study area and systematic trips, the most useful reference emission data are the population living in the zone. Obviously, the emission data have to be referred to the transport mode considered, coherently with the information contained in the counted traffic flows.

Let $\mathbf{t}_{true,o\cdot}$ be the true, unknown trip emission from zone o, for a given transport mode. A similar relationship as (7) can be written between the estimated zone emission $\hat{\mathbf{t}}_{o.}$ and the true one:

$$\hat{\mathbf{t}}_{o.} = \mathbf{t}_{true,o} + \chi \qquad (10)$$

where χ is a vector of errors whose variance-covariance matrix is **U**.

The estimated zone emission $\hat{\mathbf{t}}_{o.}$ represents a further source of information obtained by model assumptions or statistical analyses, and it cannot be trivially obtained by the target matrix (in this case there should not be another different source of information). The values $\hat{\mathbf{t}}_{o.}$ could be obtained by available, revised emission data, that could be more recent in comparison with the target values, and then can be used to improve the OD matrix final estimate.

The values of the variance-covariance matrix **U** depends on the specific assumptions made on $\hat{\mathbf{t}}_{o.}$, according to the fact that it can be obtained by sampling estimate or by model estimator. In the first case the results of the sampling theory allow to obtain an estimate of the elements of **U**; in the second case, the elements of **U** are obtained as a function of the specific hypotheses made on the model underlying the estimate of $\hat{\mathbf{t}}_{o.}$.

From an analytical point of view, the general ODCBE problem, taking into account the emission data as further source of information, can be expressed as:

$$\mathbf{t}^* = \arg\min_{\mathbf{t}\geq 0} F_1(\hat{\mathbf{t}},\mathbf{t}) + F_2(\hat{\mathbf{M}}\mathbf{t},\hat{\mathbf{v}}) + F_3(\hat{\mathbf{t}}_{o.},\mathbf{t}_{o.}) \qquad (11)$$

where $\mathbf{t}_{o.}$ is the current estimate of the trip emission from the zone o, $\hat{\mathbf{t}}_{o.}$ is the maximum forecasted emission from the zone o, F_3 is a measure of the

distance between $\mathbf{t}_{o.}$ and $\hat{\mathbf{t}}_{o.}$ The other quantities are the same as in the expression (1).
If the GLS approach is used, the problem can be specified as:

$$\mathbf{t}^* = \arg\min_{\mathbf{t}\geq 0}(\mathbf{t}-\hat{\mathbf{t}})^T \mathbf{V}^{-1}(\mathbf{t}-\hat{\mathbf{t}}) + (\hat{\mathbf{v}}-\hat{\mathbf{M}}\mathbf{t})^T \mathbf{W}^{-1}(\hat{\mathbf{v}}-\hat{\mathbf{M}}\mathbf{t}) + (\mathbf{t}_{o.}-\hat{\mathbf{t}}_{o.})^T \mathbf{U}^{-1}(\mathbf{t}_{o.}-\hat{\mathbf{t}}_{o.}) \tag{12}$$

where **U** is the variance-covariance matrix of the error vector x.
The model (12) considers three distinct sources of information that contribute to the estimate of the actual OD matrix and are related to the target OD matrix, the counted traffic flows and the trip emission from each zone.
From an algorithmic point of view, the problem (12) can be easily resolved by using standard constrained optimisation techniques, where the constraints refer to the OD demand values that, as already said, have to be greater than or equal to zero.

4. NUMERICAL RESULTS

In order to test the model (12), some numerical examples have been carried out on a test network. Figure 1 shows the network considered for the experiment, formed by five centroids (20 OD pairs), 15 nodes and 54 links.
The aim of the experiment was to compare on the same network the models (9) and (12), that differ from the set of information used, in order to verify the expected improvements with model (12). The data used are the same for the two models, but model (12) requires also the trip emission additional data.
The results have been evaluated with reference to the statistical index RRMSE (Relative Root Mean Square Errors), that if **X** is the reference variable (e.g., the true OD matrix or the counted flows), **X*** the estimated one, and N the number of observations is defined as:

$$RRMSE(X^*, X) = \frac{\sqrt{\sum_i (X_i^* - X_i)^2 / N}}{\dfrac{\sum_i X_i}{N}}$$

A first set of RRMSE values is computed for the target matrix, $RRMSE(\hat{\mathbf{t}}, \mathbf{t}_{true})$, and the estimated OD matrix solution of the problem, $RRMSE(\mathbf{t}^*, \mathbf{t}_{true})$, as regards a true OD matrix, generated for the experiment, $\mathbf{t}_{true}$. A second set of RRMSE values is computed for the traffic flows estimated with the target matrix, $RRMSE\ (\mathbf{v}(\hat{\mathbf{t}}), \hat{\mathbf{v}})$, and the solution OD matrix, $RRMSE(\mathbf{v}(\mathbf{t}^*), \hat{\mathbf{v}})$, $\hat{\mathbf{v}}$ being the counted flows. Furthermore, the capability to reproduce flows can be evaluated not only as regards the counted links, but also the total link flow vector, and particularly the no-counted links (hold-out-sample), that is $RRMSE\ (\mathbf{v}(\mathbf{t}^*)_{no\_counted}, \mathbf{v}_{no\_counted})$. This latter indicator is useful in order to verify the reliability of the OD matrix for the whole network.

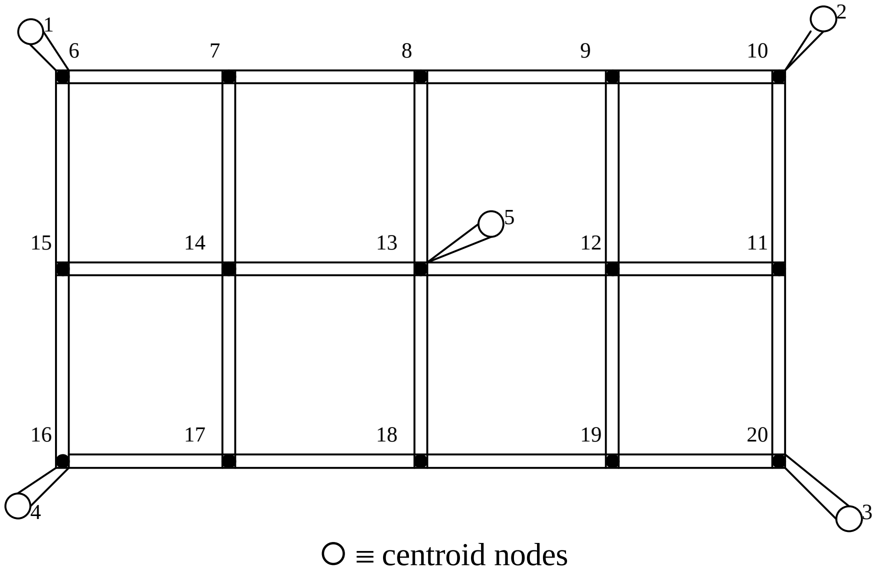

Figure 1 – The test network

As initial data, a true OD trip matrix has been fixed in order to compute the RRMSE value of the estimated OD trip matrix **t***, with respect to the true one, $\mathbf{t}_{true}$ (unknown in the reality) and it is reported in table 1.

Table 1 – The true OD matrix used for the experiment

	1	2	3	4	5
1	0	800	500	1000	300
2	800	0	1000	500	300
3	500	1000	0	800	300
4	1000	500	800	0	200
5	200	200	200	200	0

The target matrix has been generated by the true one, by considering a relationship of the kind:

$$\hat{t}_i = t_{i,true} + \varepsilon_i \qquad (13)$$

where ε is a random term extracted by a normal distribution, with a standard deviation/mean ratio (variation coefficient) cv_t=0.4. Different OD target matrices have been generated, in order to test more times the effects of the introduction of the emission data in the model formulation, and to verify the results from a statistical point of view.
Other data to be used refer to the set of counted traffic flows; in this case they have been simulating by assigning the true OD matrix to the network; the obtained flows can be considered as “true” flows, under the conditions of the assignment model; to take into account the errors that could be made during the detection of the traffic flows, a similar expression as (13) has been considered:

$$\hat{v}_l = v_l(\mathbf{t}_{true}) + \eta_l \qquad (14)$$

where $v_l(\mathbf{t}_{true})$ are the traffic flows resulting from the assignment of the true OD matrix and η is random term extracted by a normal distribution with a variation coefficient cv_v=0.05.
The hypotheses on the assignment model refer to the link cost functions and the path choice model. The link cost function are not flow-dependent, then the network is considered uncongested; the path choice model considered in this experiment is a stochastic one, specifically the logit model (Ben-Akiva and Lerman 1984; Cascetta, 2001). To assign the OD demand by using the logit model the paths between each OD pair have been explicitly computed (given the test network dimensions this is not a problem, while for real size networks the explicit enumeration could be relevant, but in the literature different algorithms exist that allow implicit enumeration of the feasible path, see for example Dial, 1971). The paths considered in this paper are those whose cost is no greater than 25% of the minimum path cost (the cost is reduced to only the time term); they have been ranked by increasing order of cost and for each OD pair the first 5 paths, if there are, have been considered.
The cost function is of the type:

$$t_r=s/L$$

where t_r is the running time on link l, s is the speed allowed on link l and L is the length of link l. For each link, both s and L have been specified and two sets of links (set I and set II) having the same uncongested running time have been identified, while the waiting time is constant under the hypothesis of uncongested network and it has been assumed equal for each link (see table 3).

Table 3 – Cost function for the links of the test network

Set **I**	Set **II**
7-14; 8-13; 9-12; 11-12; 12-9; 12-11; 12-13; 12-19; 13-8; 13-12; 13-14; 13-18; 14-7; 14-13; 14-15; 14-17; 15-14; 17-14; 18-13; 19-12	6-7; 6-15; 7-6; 7-8; 8-7; 8-9; 9-8; 9-10; 10-9; 10-11; 11-10;11-20; 15-6; 15-16; 16-15; 16-17; 17-16; 17-18;18-17; 18-19; 19-18; 19-20; 20-19; 20-11
t_r=123, t_w=7.56	t_r=144, t_w=7.56

Table 4 – The set of counted links selected for the experiment

Counted links set	Counted values
8 - 9	806
9 - 8	954
12 - 13	1198
13 - 12	1245
13 - 18	332
14 - 17	563
15 - 16	1284
16 - 15	1250
17 - 14	460
18 - 13	472

The set of counted flows has been randomly chosen, by selecting some links on the whole set, even if generally some criteria could be used as suggested

by Lam and Lo (1991), Yang and Zhou (1998), Postorino (2000, 2001). The set of the selected counted links and the generated counted values are reported in table 4.
The assignment matrix has been evaluated by using the results of the assignment model, particularly, the elements $\hat{m}_{li}$ of $\hat{\mathbf{M}}$ are given by the sum of the path choice percentages for those paths linking the i-th OD pair and using the same link l. Given the characteristics of the test network the estimate of $\hat{\mathbf{M}}$ is not a problem, because all paths for each OD pair have been explicitly enumerated, but for real size network alternative methods could be identified.
In order to apply the model (12) another source of information needs, i.e. the emission data for each traffic zone. Different sets of emission data have been generated from the true emission data, these latter obtained by summing the elements of the true OD matrix for each row. Following a similar procedure as for the generation of the target matrix and the counted traffic flow, the operational emission data have been obtained by summing to the true emission values a random term extracted by a normal distribution, with different values of cv_e. (0.1, 0.2, 0.4).
Finally, the variance-covariance matrices have been considered diagonal and have been easily evaluated from the hypotheses made about the variation coefficient values, because, given cv, the variances values can be expressed as: $\sigma^2 = cv \times \mu$, μ being the mean value of the considered variable (the OD matrix, the traffic flows, the emission data).
The results are reported in table 5; they refer to the average values of the statistical indicators on 50 tests carried out for each type of model, as specified in the first row.

Table 5 – The results obtained (average values on 50 experiments)

	RRMSE ($\mathbf{t}^*$, $\mathbf{t}_{true}$)	*RRMSE* ($\mathbf{v}(\mathbf{t}^*)$, $\hat{\mathbf{v}}$)	*RRMSE* ($\mathbf{v}(\mathbf{t}^*)_{no\_counted}$, $\mathbf{v}_{no\_counted}$)
ODCBE without emission data	0.38	0.06	0.25
ODCBE with emission data (cv_e=0.1)	0.28	0.04	0.12
ODCBE with emission data (cv_e=0.2)	0.31	0.05	0.15
ODCBE with emission data (cv_e=0.4)	0.35	0.05	0.20

As it can be seen, the results obtained with the introduction of the additional emission data are better than those provided by the standard model (9), even if, as it could be expected, the results obtained with the ODCBE model (12) by increasing the cv_e tend to be worst, but still better than those obtained with the ODCBE model (9). The two approaches give similar results when the cv_e assumes a values of 0.4.
All the statistics referred to the results of the model (12), in terms of RRMSE, are good; but if the RRMSE value with respect to the counted flows are practically the same for the two approaches, the most important result is that the RRMSE value with respect to the no-counted flows are significantly better, and then this confirms the improvement that can be obtained by using additional sources of data in resolving the ODCBE problem because a better reproduction of the traffic situation on the whole transportation network can be obtained. This is confirmed also by the better values obtained in terms of RRMSE($\mathbf{t}^*$, $\mathbf{t}_{true}$), that is the estimated OD matrix by model (12) is closer to the true one than that estimated by model (9).

The additional data introduced in the model (12) could be considered as an emission constraint that allows to select the more reliable OD matrix in the feasibility space of solution, i.e., the OD matrix that not only reproduces the counted traffic flows but is more coherent with the emission data.

5. CONCLUSIONS

The estimate of the OD trip matrix from traffic counts is a very appealing topic and methods and models to improve the existing techniques are present in the literature. One important aspect in the application of most of the proposed models is the nature of the required data; the information is often referred to on old OD matrix to be updated, called target matrix, and the set of counted traffic flows. However, some problems can arise when only this two sources of information are used, particularly because some unreliable values for single OD pairs could be obtained. In order to overcome this problem another source of information has been considered that allow to take into account the maximum trip emission from each traffic zone.
The introduction of this new source of information in a modified GLS model led to interesting results, because some improvements were obtained both in the reproduction of a true OD matrix, generated for the experiment, and particularly in the no-counted traffic flows. These latter are the traffic flows on the network links of the transportation network for which no counts are available. The set of no-counted flows can be considered as an hold-out sample that represents a landmark for verifying the reliability of the estimated OD trip matrix.
The results obtained being really encouraging, some other experiments could be made on real networks and under the hypotheses of congested networks, by applying some fixed point or bi-level approaches.

REFERENCES

Ashok K, Ben-Akiva M. (1993) Dynamic origin/destination matrix estimation and prediction for real-time traffic management systems, *Proceedings of the 12th International Symposium on the Theory of Traffic Flow and Transportation*, ed. C. F. Daganzo.

Bell M.G.H. (1991) The estimation of origin-destination matrices by constrained generalised least squares, **Transportation Research**, **25B** (1),13-22.

Ben-Akiva M., Lerman (1985) *Discrete choice analysis*, MIT Press, Cambridge, Mass.

Brenninger-Göthe M., Jornsten K.O., Lundgren J.T. (1989) Estimation of origin/destination matrices from traffic counts using multiobjecive programming formulation, **Transportation Research 23B**(4), 257-269.

Cascetta E. (1984) Estimation of trip matrices from traffic counts and survey data: a generalised least squares estimator, **Transportation Research**, **18B** (4/5), 289-299.

Cascetta E., Nguyen S. (1988) A unified framework for estimating or updating origin/destination matrices from traffic counts, **Transportation Research**, **22B** (6), 437-455.

Cascetta E., Inaudi D., Marquis G. (1993) Dynamic estimators of origin/destination matrices using traffic counts, **Transportation Science 27**, 363-373.

Cascetta E., Postorino M.N. (2001) Fixed point models for the estimation of O/D matrices using traffic counts on congested networks, **Transportation Science,** N. **35**(2), 134-147.

Cascetta E. (2001) *Transportation systems engineering: theory and methods*, Kluwer.

Chen Y., Florian M. (1994) O/D demand adjustment problem with congestion: part I. Model analysis and optimal conditions, Centre de Recherche sur les transports, **Publication CRT-94-56**.

Dial R.B. (1971) A probabilistic multipath traffic assignment model which obviates path enumeration", **Transportation Science** (5), 83-111.

Fisk C.S. (1988) On combining maximum entropy trip matrix estimation with user optimal assignmen, **Transportation Research, 22B** (1), 66-79.

Fisk C.S. (1989) Trip matrix estimation from link traffic counts: the congested network case, **Transportation Research, 23B** (2), 331-356.

Fisk C.S., Boyce D.E. (1983) A note on trip matrix estimation from link traffic counts data, **Transportation Research***,* **17B** (2), 245-250.

Lam W.H.K., Lo H.P. (1990) Accuracy of O/D estimates from traffic counts, **Traffic Engineering + Control** (31), 358-367.

Maher M.J. (1983) Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach, **Transportation Research 17B** (5), 435-447.

Nguyen S. (1977) Estimating an O/D matrix from network data: A network equilibrium approach, *Publication 87, Centre de Recherche sur les Transports, Universitè de Montreal,* Montreal, Canada*.*

Nguyen S., Morello E., Pallottino S. (1989) Discrete time dynamic estimation model for passengers origin/destination matrices in transit networks, **Transportation Research**, **22B** (2), 251-260.

Postorino M.N. (2000) Estimation of OD trip matrices from traffic counts: the selection of the counted links, in *Metodi e Tecnologie dell'Ingegneria dei trasporti Seminario 1999*, FrancoAngeli editore., 238-252.

Postorino M.N. (2001) An analysis of the counted links selection criteria for the estimation of the OD trip matrix from traffic counts, **Trasporti e Trazione** (5), 148-157.

Van Zuylen J.H., Willumsen L.G. (1980) The most likely trip matrix estimated from traffic counts, **Transportation Research 14B**(2), 281-293.

Willumsen L.G. (1984) Estimating time-dependent trip matrices from traffic counts, *Ninth International Symposium on Transportation and Traffic Theory*, Utrecht: VNU Science Press, 397-411.

Yang H. (1995) Heuristic algorithms for the bi-level origin-destination matrix estimation problem, **Transportation Research**, **29B** (4), 231-242.

Yang H., Zhou J. (1998) Optimal traffic counting locations for Origin/Destination matrix estimation, **Transportation Research**, **32B**(2), 109-126.

Wong S.C., Tong C.O. (1998) Estimation of time-dependent origin/destination matrices for transit networks, **Transportation Research**, **32B** (1), 35-48.