

PREDICTION RELIABILITY OF THE TRANSPORT SIMULATION MODELS: A BEFORE AND AFTER STUDY IN NAPLES

Ennio Cascetta
Andrea Papola
University of Naples "Federico II" - Italy

Armando Carteni
University of Salerno - Italy

1. INTRODUCTION

A transportation system can be defined as the combination of elements and their interactions, which produce the demand for travel within a given area and the supply of transportation services to satisfy this demand. The relevant interactions among the various elements of a transportation system can be simulated with mathematical models. Supply models simulate the performances of the transportation services available among the different zones and demand models simulate the relevant aspects of travel demand as a function of the activity system and of the supply performances. Typically, the characteristics of travel demand simulated include the number of trips in the reference period and their distribution among the different zones, the different transport modes, and the different paths. The interaction between demand and supply can be simulated through assignment models which allow the calculation of link flows.

In literature there is a great number of applications of these models. They are generally estimated so as to reproduce the observed actual system conditions and then applied to predict the future system conditions due to possible changes in the supply system (Ashley D.J., 1987; Daly A. et al, 1989; Koppelman F.S., 1989; Algiers S., 1994; Cascetta E., 2001; Coppola P. and Carteni A., 2001; de Luca S. and Papola A., 2001). But what is the reliability of these predictions? Very few studies have been conducted in order to verify the reliability of these models by comparing the real future system conditions with those predicted by the model.

This paper investigates the prediction reliability of the transport system models, through a *before and after study*, taking advantage of the opening of a new part of a Neapolitan subway in December 2002. Moreover a new procedure for the OD demand vector updating is proposed able to identify possible macro mistakes in the supply/assignment model. The paper is divided into three sections; in the first the methodology is presented; in the

second the results of a *before* and *after* application are described; in the third conclusions of the study are reported.

2. METHODOLOGY

All the elements of a transport model system (zoning, graph, cost functions, demand, assignment model, etc.) are generally affected by some approximations. The latter can be certainly reduced through an accurate and experienced “simulation work”; however a further model improvement can be reached by using some observed link flows (which are the output of a transport model system) in order to correct/update the model. Normally the demand vector is considered the most crucial and problematic element to be simulated and, therefore, traffic counts are generally used in order to correct it (generally obtained through demand models and/or surveys) so that the whole model system is able to reproduce the observed link flows.

This procedure generally produce offsetting mistakes since the difference between assigned and observed link flows, as just said, is due to a number of approximations. In other words, by updating only the transport demand there is the risk of identifying the wrong demand vector that assigned to a wrong supply with a wrong assignment model reproduces correctly the observed link flows. To reduce these offset mistakes, traffic counts should be used to update the whole model system. Obviously, the variables are many and infinite new configurations of the model able to perfectly reproduce the observed link flows can be found.

But how this estimated model will be able to forecast link flows in a scenario different from that used to estimate it? Which is the best way to use this few “tuning” data (traffic counts)? Which are the most efficient updating procedure and which the most important data to be corrected? The main aim of this study is just to give some first answers to some of these questions. In order to do it, the methodology proposed and used in the application (see also figure 3) is:

1. to plan a traffic count survey with respect to two different supply scenarios (which in the following will be called *before* and *after* scenarios) so to estimate the *before* and the *after* traffic count vectors;
2. to update progressively a base simulation model through the before traffic counts;
3. to validate the forecasting capability of all the different configurations of “base” and “updated” models by comparing the simulated and the observed link flows in the *after* scenario

The complete updating procedure proposed and used in the application (see also figure 3) consists in:

1. a first correction of the supply/assignment model on the basis of the before traffic counts and of some defined performance indicators;
2. application of a new estimator which use the before traffic counts both to update the initial *origin-destination* (OD) *demand vector* and to identify possible macro mistakes in the supply/assignment model;
3. application of the *pivoting method* to estimate the OD vector referred to the *after* scenario through a mode choice model.
4. use of also *after* traffic counts to update the mode choice model parameters. In other words a new vector of coefficients of the mode choice model is estimated in order to reduce as much as possible the distance between modelled and observed link flows in the after scenario.

2.1 Procedure for the updating of the OD demand vector

In this section a procedure for the *updating of the OD demand vector* based on a new estimator is described. As already said, this procedure is also able to identify possible macro mistakes in the supply/assignment model. The more used estimator in literature is the *Generalized Least Squares* (GLS) (Cascetta E., 1984; Cascetta E., 1986; Cascetta E. and Nguyen S., 1986, Bell M.G.H., 1991):

$$d_{GLS} = \underset{x \geq 0}{arg \min} \left[\sum_i \frac{(\hat{d}_i - x_i)^2}{var[\eta_i]} + \sum_l \frac{(\hat{f}_l - \sum_j \hat{m}_{l,j} \cdot x_j)^2}{var[\varepsilon_l]} \right] \quad (1)$$

where:

\hat{d}_i is the generic element of the base OD demand vector \hat{d} ;

x_i is the generic element of the unknown OD demand vector x ;

\hat{f}_l is the generic element of the observed link flow vector \hat{f} ;

$\hat{m}_{l,j}$ is the generic element of the assignment matrix \hat{M} ;

$\eta_i = \hat{d}_i - x_i$ is the difference between the initial estimate and the real figure of the generic OD demand;

$\varepsilon_l = \hat{f}_l - \sum_j \hat{m}_{l,j} \cdot x_j$ is the difference between the observed and the assigned figure of the generic link flow.

Generally much more confidence is given to the estimates of link flows rather than to those of the OD demand vector that is:

$$\text{var}[\eta_i] \gg \text{var}[\varepsilon_l] \quad \forall i, l$$

This assumption has suggested to develop an estimator in which equality constraints between observed and assigned link flows are explicitly considered (that is equivalent to consider $\text{var}[\varepsilon_l]=0 \quad \forall l$). This estimator called *Constrained Generalized Least Squares (CGLS)* has the following expression:

$$d_{CGLS} = \arg \min_{\substack{Mx = \hat{f} \\ x \geq 0}} \left[\sum_i \frac{(\hat{d}_i - x_i)^2}{\text{var}[\eta_i]} \right] \quad (2)$$

The (2) is a quadratic and convex optimisation problem where the constraints are defined by linear equation/disequation which define a convex set and the objective function is strictly convex. Therefore, the solution of problem (2), if exist, is global and unique. The variances of the differences η_i are considered different for each OD pair and equal to the initial demand estimate:

$$\text{var}[\eta_i] = \hat{d}_i \geq 0 \quad \forall i$$

The main difference with the *GLS* is that the *CGLS* estimator could generate an inconsistent problem (no solution because of possible incompatibility between \hat{M} and \hat{f} : $\hat{M} \cdot x \neq \hat{f}$). In this case, in the hypothesis of uncongested network and of absence of measurement errors, the error must be in the assignment matrix \hat{M} (i.e. in the supply and/or assignment model). Interestingly, these incompatibilities can easily be identified through the resolution of the following linear optimization problem (which is the first phase of the two phases simplex algorithm (Bazaraa M.S. et al, 1989)):

$$w^* = \arg \min_{\substack{Mx + Ih = \hat{f} \\ x \geq 0; h \geq 0}} I^T h \quad (3)$$

where I is the identity matrix and h is the artificial variables vector. Obviously it occur:

- $w^* = 0 \Rightarrow$ problem (2) is consistent;
- $w^* > 0 \Rightarrow$ problem (2) is inconsistent (no solutions).

Moreover, if $w^* > 0$, the “ $h_i > 0$ ” variables allow to identify the links on which there is incompatibility between observed and assigned link flows. Through a systematic analysis of the paths alternative to those using the “ $h_i > 0$ ” links it is generally possible the identification of possible macro mistakes in the supply/assignment model.

Note that, by applying a standard *GLS* estimator, these mistakes would be ignored and very unreal OD demand could be estimated to fit with observed link flows. When all possible mistakes in the supply/assignment model are corrected so that $w^* = 0$, the optimization problem (2) can be solved. In literature there are different algorithms to solve problem (2) like the *reduced gradient* method, the *convex simplex* or the *interior point method (IPM)* (Bazaraa M.S. et al, 1993).

2.2 Pivoting method

The pivoting method is a forecasting technique where models are used to estimate the variations with respect to a current (*before*) demand, rather than directly the future (*after*) demand (Cascetta E., 2001). This approach is generally used when a better estimate of the OD demand vector with respect to that obtainable with just demand models (e.g. an OD demand vector estimated through models and then corrected through observed link flows) is available. In this case, modeling approximations can be reduced by using demand models as simulators of demand variations: the *after* demand estimates are obtained as follows:

$$d_{after}^{PIVOT} = d_{before} \cdot \frac{d_{after}^{Mod}(SE^{after}, T^{after}, \beta)}{d_{before}^{Mod}(SE^{before}, T^{before}, \beta)} \quad (4)$$

were:

d_{after}^{PIVOT} is the pivoting estimation of the *after* scenario demand vector;

d_{before} is an updated estimation of the *before* scenario demand vector;

d_{before}^{Mod} and d_{after}^{Mod} are the model estimation of the demand vectors for respectively the *before* and the *after* scenario;

$SE^{before}, SE^{after}, T^{before}, T^{after}$ are the vectors of socio economic and level of service variables (which influence the model choice probabilities) with respect to both *before* and *after* scenarios;

β is the model parameter vector.

2.3 Updating of the model parameters

Link flows can be used also to update the parameters of the demand models through estimators analogous to those used for the OD demand vector updating (Hogberg P., 1976; Cascetta E. and Russo F., 1997; Cascetta E., 2001; Cascetta E. and Postorino M.N., 2001). The GLS, in particular, assume the following expression (compare with (1)):

$$\beta_{GLS} = \underset{\mathbf{b}}{\operatorname{argmin}} \left[\sum_k \frac{(\hat{\beta}_k - b_k)^2}{\operatorname{var}[\tau_k]} + \sum_l \frac{(\hat{f}_l - \sum_j \hat{m}_{l,j} \cdot d_j(\mathbf{b}))^2}{\operatorname{var}[\varepsilon_l]} \right] \quad (5)$$

where:

\mathbf{b} is the vector of unknown parameters of the demand models;

$\hat{\beta}_k$ is the initial estimate of the generic parameter (generally obtained through disaggregate data);

τ_k represent the generic difference $\hat{\beta}_k - b_k$

Problem (5) is generally unconstrained and can easily be solved through the gradient algorithm. Note that the objective function is not convex, so it would be necessary to use numerical analysis techniques to found a set of local minimum points and then choose the best solution among them.

3. APPLICATION

As said before, the application proposed has been possible thanks to the opening of the second track of line 1 (Naples underground) between Vanvitelli and Dante stations (December 2002). Before that, only one track was operative in that part of the line (see figure 1) with an average time interval of 30 minute. The line 1 users, between Piscinola and Vanvitelli stations, availed of a service with an average time interval of 6 minutes, had to change train at Vanvitelli station, to continue to Dante station.

Note that between Vanvitelli and Dante stations there is an important interchange node between line 1 (Museo station) and line 2 (Cavour station). With the opening of the second track, it is possible to effect a continuous underground service between Piscinola and Dante stations, with no interruption and interchange and with an average time interval of 6 minutes. This context was used to carry out the methodological procedure described in section 2.0. Therefore, firstly a traffic count survey was carried out in both *before* and *after* scenarios (section 3.1). Then the available “base” model (consisting in a supply model ¹, an estimation of the public transport OD demand vector, a mode and an hyper-path choice model) was progressively updated through the before traffic counts so as to upgrade some performance

indicators (section 3.2), Precisely, the supply and the hyper-path choice model have been updated first (section 3.3) and then the proposed CGLS estimator has been applied for updating the *before* OD demand.

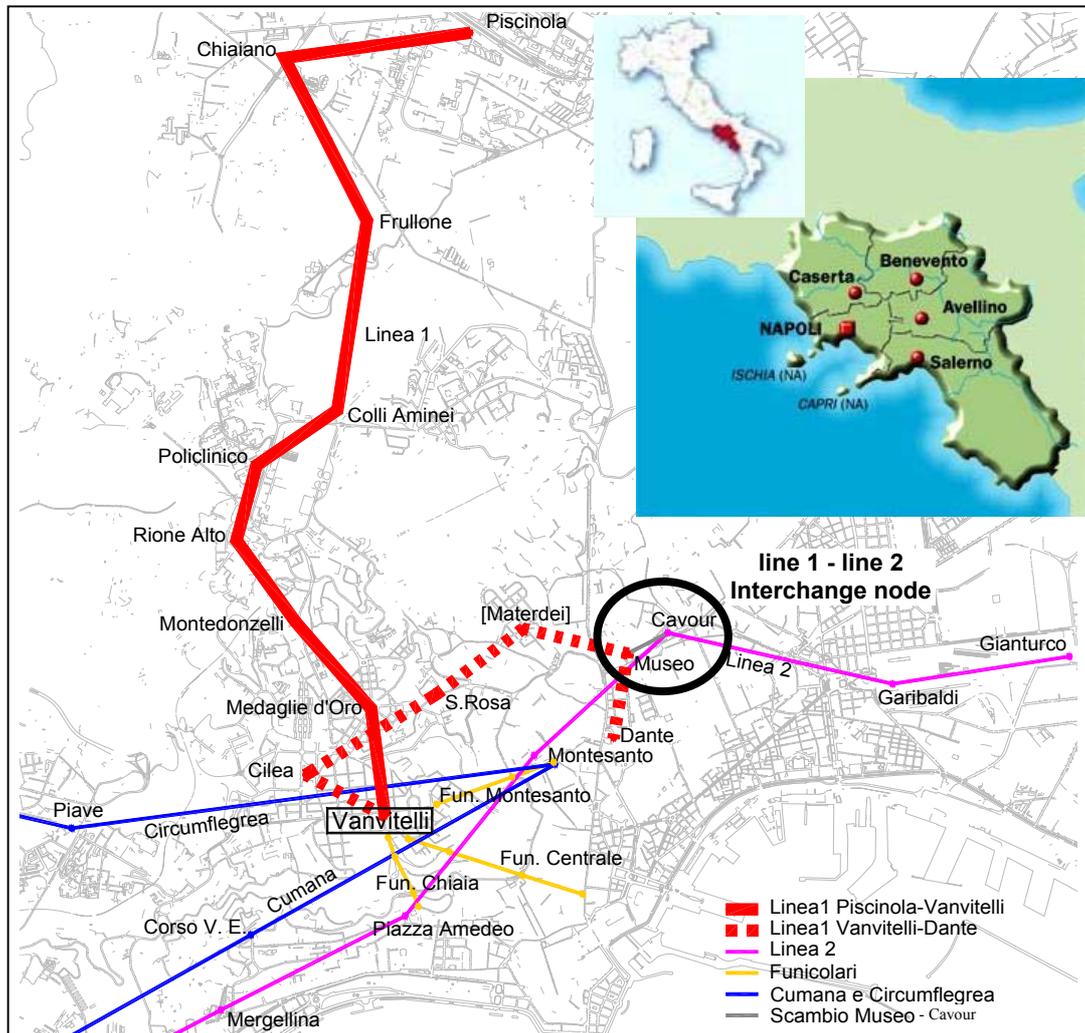


Figure 1 – The layout of line 1 of the Naples underground

Finally, the model system in all its configurations (“base”, with the supply/assignment model updated, with the OD demand updated, with the OD demand forecasted through the pivoting method, etc.) has been used to forecast *after* link flows and validated by comparing the latter with the observed link flows.

3.1 Traffic counts survey

The correct planning of the traffic counts survey appears to be critical in a before and after study. In this application the identification of the traffic count sections has been performed through the model system implemented. In more detail the initial demand estimate was assigned to the transport network of the *after* scenario so as to identify the OD pairs which might be captured by the new underground service. The “reduced” OD vector defined by these OD pairs has then been assigned to the transport network concerning the *before* scenario thus identifying the transport services (and therefore the infrastructural links) currently used by such demand. In some of these links, as representative of the latter services, count sections have been identified so as to define a cordon line. On this sections, three traffic survey were carried out both before (November 2002) and after (May 2003 and May 2005) the opening of the new underground connection.

The observing period includes both peak and off-peak hours in the morning and in the afternoon (of the average business day). Counts have involved: (a) vehicle flows of the private transport network; (b) on board passenger flows of public transport lines; (c) passenger flows getting in and out at stops of public transport lines. The number of traffic count sections is 47 and 55 respectively in the before and in the after scenarios.

From data collected it has resulted that the morning peak hour falls between 8:00 and 9:00 when takes place approx. the 48% of the total flow measured in the 7:00-9:30 peak period. Moreover, in the morning peak hour traffic flows exceed by approx. the 75% those of the morning off-peak hour and by approx. the 55% those of the afternoon peak hour; the gap between the afternoon peak and off-peak hour is, instead, approx. the 20%.

In the following, the analysis has concerned only the morning peak period (7:00-9:30).

In table 1 demand values referred to the area mostly affected by the new railway connection in the three lapses of time considered (November 2002, May 2003 and May 2005) are reported. From this table it is possible to observe a first increase of underground trips (+3.368 trips: +8,5%) coming from both bus and car mode after six months from the opening of the new underground connection. In May 2005, a much greater increase of metro trips is observed with respect to the *before* scenario (+9.321 trips: +23,6%), coming with approx. the same percentages from bus and car mode.

<i>trips</i>	<i>BEFORE</i>		<i>AFTER</i>	<i>diff</i>	<i>diff</i>	<i>diff %</i>	<i>diff %</i>
	<i>November 2002</i>	<i>May 2003</i>	<i>May 2005</i>	<i>2003 – 2002</i>	<i>2005 – 2002</i>	<i>2003 – 2002</i>	<i>2005 – 2002</i>
metro	39.573	42.941	48.894	3.368	9.321	8,5%	23,6%

bus	20.849	18.776	14.533	-2.073	-6.316	-9,9%	-30,3%
car	55.524	52.550	46.819	-2.974	-8.705	-5,4%	-15,7%
<i>Total</i>	<i>115.946</i>	<i>114.267</i>	<i>110.246</i>	<i>-1.679</i>	<i>-5.700</i>	<i>-1,4%</i>	<i>-4,9%</i>

Table 1 – Demand values referred to the area mostly affected by the new railway connection (7:00-9:30)

By analysing variations in the modal split (table 2) in the area mostly affected by the new underground connection, a 2% increase of public transport (from 52% to 54%) is noted in the six months following the realization of the new service and a further 3,5% increase after two years, with public transport covering approx. the 58% of total trips.

In the following May 2005 will be taken as the *after* scenario.

<i>mode</i>	<i>BEFORE</i>		<i>AFTER</i>
	<i>November 2002</i>	<i>May 2003</i>	<i>May 2005</i>
car	47,9%	46,0%	42,5%
public transport (bus+metro)	52,1%	54,0%	57,5%

Table 2 – Modal split in the area mostly affected by the new railway connection (7:00-9:30)

3.2 Performance indicators

The Mean Square Error (MSE), the waited Mean Absolute Percentage Deviation ($MAPD_w$) and the R^2 are the performance indicators used to value the quality of the estimations for both *before* and *after* scenario:

$$MSE = \frac{\sum_{i=1}^{nc} (f_i^a - f_i^c)^2}{n_c - 1}; \quad MAPD_w = 100 \cdot \frac{\sum_{i=1}^{nc} f_i^c \cdot \frac{|f_i^a - f_i^c|}{f_i^e}}{\sum_{i=1}^n f_i^c}; \quad R^2 = 1 - \frac{\sum_{i=1}^{nc} (f_i^a - f_i^c)^2}{\sum_{i=1}^{nc} (f_i^a - \bar{f}^c)^2};$$

where:

f_i^c is the generic traffic count on link i ;

f_i^a is the generic assigned traffic flow on link i ;

n_c is the number of traffic count sections;

$\bar{f} = \frac{\sum_{i=1}^{nc} f_i^c}{n_c}$ is the average traffic count.

In table 3 $MAPD_w$ value intervals generally adopted are reported.

$MAPD_w$	Quality of the estimation
$\leq 10 \%$	very good

> 10 %, ≤ 20 %	good
> 20 %, ≤ 30 %	no good
> 30 %	Bad

Table 3 – Quality of the estimation with respect to the $MAPD_w$.

3.3 Correction of the supply/assignment model

As previously said, the first phase of this before and after analysis has been the correction of the supply/assignment model through the performance indicators introduced in the previous section. Some further corrections on these models were made after the application of the CGLS estimator proposed and described in section 2.0 to the updating of the OD demand vector. Indeed, as already said, the proposed updating procedure allow the identification of incompatibilities between observed and assigned link flows and, consequently, of possible mistakes in the supply/assignment models. In particular, in the described application, some incompatibilities were found and solved through the correction of some unreal link travel times (pedestrian times on stairs, on board travel times, etc.)

In detail the following correction were carried out:

- supply model,
 - network configuration;
 - link travel times.
- assignment model,
 - hyper-path choice model parameters.

Network configuration

Particular attention was made in the correct location of stop nodes in order to correctly identify the available lines of each stop node. That means, for example, to avoid locating a stop in the intersections between two “line” links or identifying with a same stop node the stop nodes relative to the two ways of a same line.

Link travel times

As already said, some pedestrian and on board travel times were corrected as a feedback of the OD correction procedure.

Concerning the waiting times, different formulations were adopted as a function of both frequency and regularity of the line. In detail for low regular lines (bus, tram, etc.), the standard “frequency approach” was adopted:

$$t_w = \theta \text{Int} = 60 \cdot \theta / f \quad 0.5 \leq \theta \leq 1$$

where Int is the time interval between two successive services of a same line, f is the line frequency and θ is the parameter of the line regularity. For high

regular lines (metro, funiculars, railways, etc.) the following formulation was adopted as a function of the frequency line (figure 2):

$$\begin{aligned} t_w &= 60 \cdot \theta f & \theta &= 0.5 & \text{for } f \geq 4 \\ t_w &= 60 \cdot 0.2/f + 15(\theta - 0.2) & \theta &= 0.5 & \text{for } f < 4 \end{aligned}$$

to take into account of the different user behaviour in high regular and low frequency service choice (schedule based approach) and of the consequent generally lower waiting time for these services.

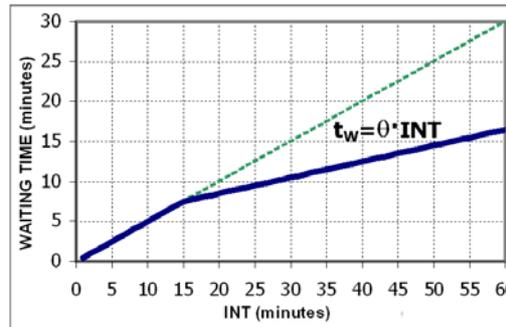


Figure 2 – t_w as a function of the time interval for high regular lines.

Furthermore, during the surveys, the arrival instants of all services were recorded and, therefore, the average time intervals between two successive services of the same line and the corresponding frequency computed through the following simple expression:

$$\hat{INT} = \frac{t_n - t_1}{n - 1} \Rightarrow \hat{f} = \frac{60}{\hat{INT}}$$

where n is the number of services of a same line observed during the survey period and t_1, t_n are the arrival instants of respectively the first and the last service of that line in the survey period.

Moreover an estimate of the line regularity parameter θ was also obtained through the following expression which assumes constant user arrival rate at the stop:

$$\hat{\theta} = \frac{n-1}{2} \frac{\sum_{i=1}^{n-1} \Delta_i^2}{\left[\sum_{i=1}^{n-1} \Delta_i \right]^2}$$

where $\Delta_i = t_{i+1} - t_i$ is the time interval between two successive services of the line. The estimated θ vary from 0,50 to 0,99 with an average of 0,61. In tab. 4

the regularity parameters and frequencies adopted in the corrected supply model are synthesized.

Parameters	# in the updated supply model
θ bus (observed lines)	measured value (0,50-0,99)
θ bus (non observed lines)	average measured value 0,61
θ railway	0,5
Frequency bus (observed lines)	measured value

Table 4 – Regularity parameters and frequencies in the updated supply model

Hyper-path choice model parameters

The systematic utility specification V_h assumed for the generic iper-path h in the iper-path choice model is the following:

$$V_h = \beta_{wait} \cdot Time_{wait,h} + \beta_{walk} \cdot Time_{walk,h} + \beta_{onboard} \cdot Time_{onboard,h} + \beta_{transfer} \cdot Numbransfer_h$$

Also these coefficients β have been corrected in order to minimize the performance indicators introduced in section 3.2 getting the values reported in tab. 5.

$\beta_{waiting}$	β_{walk}	$\beta_{onboard}$	$\beta_{transfer}$
-2.1	-1.9	-1	-7.8

Table 5 – Coefficients β in the corrected iper-path choice model

In tab. 6 the performance indicator before and after the correction of the supply/assignment model are reported.

Model	Scenario	MSE	MAPD _w	R ²
base	before	1.134.876	33 %	0,82
Supply/assignment updated	before	710.611	26 %	0,89

Table 6 – Performance indicator before and after the correction of the supply/assignment model.

3.4 Pivoting method and mode choice model

In the proposed application, after applying the estimator (2) to update the initial demand vector, d_{before} (which after the updating is called d_{before}^{CGLS}) the pivoting method has been applied to estimate the *after* demand vector, $d_{after}^{CGLS PIVOT}$. The general expression (4) (section 2.0) has been adapted to the available data. In more detail, to forecast the public transport OD vector by using the only mode choice model, the expression (4) becomes:

$$d_{after}^{CGLS\ PIVOT} = d_{before}^{CGLS} \cdot \frac{P_{afetr}^{Mod}(SE^{after}, T^{after}, \hat{\beta})}{P_{before}^{Mod}(SE^{before}, T^{before}, \hat{\beta})} \quad (6)$$

were P_{before}^{Mod} and P_{afetr}^{Mod} are the vectors of public transport choice probabilities simulated by the mode choice model.

The used mode choice models is a Multinomial Logit:

$$p^c(m / oshd) = \frac{\exp(V^c_{odmh})}{\sum_{j=1}^m \exp(V^c_{odmh})}$$

mode	Attributes
car	T_{car}^{od} is the network computed travel time (hours) for the <i>OD</i> car trip; C_{car}^{od} is the network computed travel cost (€) for the <i>OD</i> car trip; N_{car}^o , for each origin <i>O</i> , is the ratio between the average car number and the average driving licence persons in a household.
motorcycle	T_{motor}^{od} is the network computed travel time (hours) for the <i>OD</i> motorcycle trip (uncongested network); $Centre^{od}$ is a dummy variable equal to one if both the origin and the destination of the trip are inside the centre of the city; <i>motorcycle</i> is an alternative specific attribute (ASA).
public transport	T_{PT}^{od} is the network computed total travel time (hours) for the <i>OD</i> public transport trip; C_{PT}^{od} is the cost (€) for the <i>OD</i> public transport trip; NT^{od} is the number of transfer for the <i>OD</i> public transport trip; $Rail_Acc^{od} \in [0,1]$ is a railway transport accessibility attribute; it has been calculated by multiplying the origin accessibility $Rail_Acc^o$ and the destination accessibility $Rail_Acc^d$. These attribute have been calculated as ratio between the portion of the considered zone affected by the railway stations (the intersection between the zone area and the 500 metre radius circles with centre in the railway stations) and the total area of the zone. <i>P T</i> is the public transport ASA.
walking	T_{walk}^{od} is the network computed walking time (hours) for the <i>OD</i> walking trip; <i>walking</i> is the walking ASA.

Table 7 – Attributes used in the systematic utilities of the mode choice model.

3.5 Validation and results analysis

As mentioned in section 2 and described in detail in table 8, the whole model system have been validated in four different configurations with respect to the *after* traffic counts. In other words, the base model system have been

progressively updated by using *before* traffic counts and then used to forecast *after* link flows to be compared with the corresponding observed flows. In this way, the relative importance of each updating activity have been evaluated. A fifth configuration has been simulated to complete the updating of the whole model system by using all available data (both before and after traffic counts). In figure 3, the whole methodological architecture followed to simulate after traffic flows in all five configurations is described.

In detail, the first two analysis concern the model in its base configuration. The difference among the two concerns the identification of the after demand that in the first case coincides with the before estimate and in the second case is computed through the pivoting method. This comparison allow to evaluate the relative importance of using models to forecast a future demand when dealing with considerable changes in the supply services. The third analysis differs from the second in virtue of the updating of the supply/assignment model while the fourth also introduce the updating of the OD demand in the before scenario. As mentioned, in the fifth analysis, *after* traffic counts are also used to update the mode choice model parameters.

<i>Id.</i>	<i>system configuration</i>	<i>before scenario</i>	<i>after scenario</i>
1	- "base model" - <i>after</i> demand equal to <i>before</i> demand.	\hat{d}_{before} M_{before}	\hat{d}_{before} M_{after}
2	- "base model"; - <i>after</i> demand by pivoting <i>before</i> demand.	\hat{d}_{before} M_{before}	\hat{d}_{after}^{PIVOT} M_{after}
3	- supply and assignment models updated; - <i>after</i> demand by pivoting <i>before</i> demand.	\hat{d}_{before} M_{before}^{upd}	\hat{d}_{after}^{PIVOT} $M_{after}^{updated}$
4	- supply and assignment models updated; - <i>before</i> demand updated; - <i>after</i> demand by pivoting the updated <i>before</i> demand.	\hat{d}_{before}^{CGLS} $M_{before}^{updated}$	$\hat{d}_{after}^{CGLS PIVOT}$ $M_{after}^{updated}$
5	- supply and assignment models updated; - <i>before</i> demand updated; - mode choice model parameters updated; - <i>after</i> demand by pivoting the updated <i>before</i> demand with the updated mode choice model.	\hat{d}_{before}^{CGLS} $M_{before}^{updated}$	$\hat{d}_{after}^{CGLS \beta_{GLS} PIVOT}$ $M_{after}^{updated}$

Table 8 – The different configurations simulated

In table 9 the performance indicator figures for each configuration of the model system are reported in both scenarios. The first two show as the base

model is not accurate enough for reproducing the transport system working. Comparisons between the first and the second configuration in the after scenario give also clear evidence of the importance of simulating demand variations between the *before* and *after* scenario when dealing with considerable changes in the supply system. In the third configuration, the relative increase in the model performances due to the correction of the supply/assignment model can be valued. In the fourth configuration, the *before* traffic counts are used at their best by updating the whole model system and, consequently the best forecasting capability of the model system can be valued. In the last scenario, the further increasing obtainable in term of performance indicators when also *after* traffic counts are used to update the mode choice model parameters can be valued.

Id.	<i>before scenario</i>			<i>after scenario</i>		
	MSE	MAPD _w	R ²	MSE	MAPD _w	R ²
1	1.134.876	33 %	0,82	1.742.393	42 %	0,74
2	1.134.876	33 %	0,82	1.435.764	36 %	0,79
3	710.611	26 %	0,89	836.866	30 %	0,86
4	0	0 %	1,0	476.834	19 %	0,89
5	0	0 %	1,0	396.345	16 %	0,92

Table 9 – Performance indicator values for each configuration.

Lastly, the difference between the *before* and the *after* scenario in term of public transport demand level simulated by the model was evaluated and reported in table 10. The figures represent the sum over all OD pairs interested in the new service (see section 3.1) of the model demand output.

Id.	Public transport demand level (7:00-9:30)	Percentage variation with respect to the updated <i>before</i> demand
4 <i>before</i>	13.602	-
4 <i>after</i>	16.046	18%
5 <i>after</i>	16.281	20%

Table 10 – Public transport demand level (7:00-9:30) simulated by the model.

4. CONCLUSIONS

This study has investigated the forecasting reliability of transport model systems through a study *before* and *after* the realization of a new underground connection; moreover, an updating methodology of the OD demand has been proposed able also to identify possible macro mistakes in the supply/assignment model.

Figure 3 – Methodological architecture of the before and after study

Bibliography

Algers S. (1993), *Long distance travel behaviour models in Sweden*, **6th WCTR Proceedings**, Lione 1992.

Ashley D.J. (1987), *Forecasting passenger travel demand – international aspects*, **Transportation**, **14**.

Bazaraa M.S., Jarvis J.J., Sherali H.D. (1989), **Linear Programming and Network Flows**, Wiley.

Bazaraa M.S., Shetty C.M., Sherali H.D. (1993), **Non linear Programming: theory and algorithms**, Wiley.

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

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

Cascetta E. (1986), *A class of travel demand estimators using traffic flow*, **CRT Publication 375**, Universite de Montreal, Montreal, Canada.

Cascetta E. and Nguyen S. (1986), *A unified framework for estimating or updating Origin-Destination matrices from traffic counts*, **Transportation Research**, **22B**.

Cascetta E. and Russo F. (1997), *Calibrating aggregated travel demand models with traffic counts: estimators and statistical performance*, **Transportation**, **24/3**.

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

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

Coppola P. and Carteni A. (2001), *A study on the elasticity of long-range travel demand for passenger transport*, **Trasporti Europei**, **19**.

Daly A., Gunn H., Hungerink G. and Vrolijk P. (1989), *Application of the Netherlands national model to support investigation of transport policy*, **17th PTRC Proceedings**.

de Luca S. and Papola A. (2001), *The evaluation of Travel demand management policies in the urban area of Naples*, **8th “Urban Transport and the Environment” Proceedings**, Lemnos, Grecia, WIT Press, pp.185-194.

Hogberg P. (1976), *Estimation of parameters in models for traffic prediction: a non-linear regression approach*, **Transportation Research**, **10B**, pp. 263-265.

Koppelman F.S. (1989), *Multidimensional model system for intercity travel choice behaviour*, **Transportation Research Record**, **1241**.

¹ The area involved in the case study includes the Naples metropolitan area. In this area 344 traffic zones (255 belonging to Naples municipality and 89 external) and 11 external centroids have been identified. The network consists of 3.161 nodes and 14.175 links (7.504 of which for pedestrian); services considered are all lines on rail and road operated by ANM, CIRCUMVESUVIANA, METRONAPOLI, SEPSA companies.