

An assessment of the accuracy of models in predicting railway traffic flows: a before and after study in Naples

A. Carteni, G. Galante & I. Henke

*Department of Civil, Construction and Environmental Engineering,
University of Naples Federico II, Italy*

Abstract

The relevant interactions among the various elements of a transportation system can be simulated with mathematical models. 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. 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. Starting from these considerations, this paper investigates the prediction reliability of the transport system models, developed for the Campania (Italy) Regional Metro System (RMS), through a before and after study, taking advantage of the opening of new parts of railways system occurred over the last twenty years. Some interesting remarks can be made from the results analysis allowing to avoid unreal demand estimations. For example, a correction procedure of transport model system based on aggregate data (traffic counts) seems to be necessary to obtain a model with an acceptable capability to reproduce/forecast transport systems. Otherwise, when dealing with considerable changes in the supply system (for example the opening of new metro lines), considerable changes in the OD demand might also be expected and, therefore, a model is absolutely needed to forecast these changes.

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. 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. But what is the reliability of these predictions? Different studies, in many field of transportation system underline the importance of performing suitable calibration and validation procedure for estimate mathematical models; example are:

- demand estimation (e.g. the seminal works of [1–11]);
- travel demand mobility policies (e.g. [12–15]);
- traffic flows estimation (e.g. [16, 17]);
- driving behaviour models (e.g. [18]);
- road evacuation problems (e.g. [19]);
- freight transport (e.g. [20–22]).

Even if all these study perform accurate calibration and validation procedures, 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.

Starting from these considerations, this paper investigates the prediction reliability of the transport system models, developed for the Campania (Italy) *Regional Metro System (RMS)*, through a before and after study, taking advantage of the opening of new parts of railways system occurred over the last twenty years. This paper is organised as follows. Section 2 describes the simulation methodology applied. In section 3 the analysis of the before and after results on the new rail lines are reported, while in section 4 conclusions are reported.

2 Estimation methodology for transportation system performances

As said, 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. 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 transportation 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 approach may produce offsetting mistakes since the difference between assigned and observed link flows, as previously said, and is due to a number of approximations in all modelling components. In other words, by updating only the transport demand the analyst might run the risk of identifying the wrong demand vector which, once it is 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 can be found. The latter are able to perfectly reproduce the observed link flows.

How could this estimated model 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 objective of this study is to provide answers to some of these questions. In order to do it, the methodology proposed and used in the application was:

- traffic counts survey with respect to two different supply scenarios (which in the following will be called *before* and *after* scenarios) in order to estimate the *before* and the *after* traffic count vectors;
- to update progressively a base mode choice model (calibrated through disaggregate RP surveys) through the before traffic counts;
- 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 consists of:

- a first correction of the supply/assignment model on the basis of the before traffic counts;
- an application of a *Generalized Least Squares (GLS)* estimator to update the initial *origin-destination (OD) demand vector* [23–25];
- an application of the *pivoting method* to estimate the OD vector referred to the *after* scenario through a mode choice model;
- the use of *after* traffic counts to update the mode choice model parameters. In other words a new vector of coefficients of the mode choice model was estimated in order to reduce as much as possible the distance between modelled and observed link flows in the after scenario.

In our case study, direct estimation of origin-destination flows was performed starting from a preexisting origin-destination flows vectors [26], and available traffic counts were used in order to correct (update) these demand estimations so that the whole model system was able to reproduce the observed aggregate data.

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 (for details see [27]). 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.

Finally link flows were used to update the parameters of the demand models through estimators analogous to those used for the OD demand vector updating [27–30].

3 The before and after application

3.1 The case study

The new transportation planning approach (RMS), started in 1996 in Napoli and extended in 2001 to the whole Campania region (the second largest region in Italy), is based on the idea that only a highly integrated and extended railway system can provide sustainable mobility in an area with densities such as those of the central area of Campania.

The starting point of the RMS project is the recognition that transportation systems are made up of several interacting elements of very different kind, both physical and organizational. Infrastructures, services, prices and regulations interact with each other and they can be used simultaneously to reach policy objectives. Each of them has been included as part of the RMS project [31–33].

The main concept of the RMS project is to provide high frequencies, even-spaced timetable; interchanges between different lines; integrated fares, integration with other modes and land-use.

During the RMS project, two significant new rail lines have been open to operation, the *Arcobaleno* line, 10.5 km long with 5 stations, linking Napoli with the municipality of Aversa (opened in April 2009) and the *Monte del Vesuvio* line, linking Napoli with Salerno (opened in June 2008). The line Monte del Vesuvio, (54.7 km long with 2 stations) represents, as extension of the High Speed/High Capacity (HS/HC) line Roma-Napoli, the north-south fast link, making the metropolitan service more reliable due to a reduction of road traffic flows.

3.2 Traffic counts survey

Traffic counts survey provide useful information on users' actual choices relevant for the model to be calibrated. As stated in the previous section, an RP survey was carried out by Trenitalia with the objective of collecting traffic counts before and after the opening of the new lines of the RMS project.

Line *Monte del Vesuvio* has been mainly conceived for commuters residing in the Municipality of Salerno. Users are workers that every day reach their workplace, but also students that at least once per week reach the universities in the towns of Napoli and Salerno. The opening of this line has brought a change in the behaviour of these users who make systematic trips; users that have increased trip frequency are 12.9% and a new generated demand is estimated equal to 10.2%. Moreover among those already making this trip, 30.5% of the users has abandoned the bus, while 8.5% has abandoned car and motorbike.

The analysis of the trip purpose indicates that users of this line make mainly systematic trips (72.1%) and specifically the great majority (48.9%) makes trips to reach their workplace and 23.2% of the users goes to school, to university or, in general, it makes trips for study purpose.

The total number of users/day registered on this line is 1,346. Among them a new generated demand is registered equal to 11.1%; users who have chosen this line because of the increase in trip frequency represent the 7.3.

With respect to *Arcobaleno* Line, *Piscinola* station is the node towards which the users of the metropolitan area are directed to get in Napoli and, thanks to the interchange with metro Line 1, users can reach the other urban destinations in Napoli. Passengers towards the *Piscinola* station represent the 43.4% of the total traffic during the whole day and are concentrated in the morning when this percentage becomes equal to 55.7%.

Users are mainly university students of the metropolitan area who reach Napoli for studying purpose. The opening of this line has completely changed the behaviour of these users who make mainly systematic trips: users that have increased trip frequency are 29.8% generating a new demand equal to 10.7%. Moreover among those already making this trip, the 84.2% have abandoned buses; specifically 49.9% before the introduction of the new service used car or motorbike, while 34.3% used bus.

The analysis of the trip purpose indicates that the users of Line “*Arcobaleno*” make mainly systematic trips (66.5%) and specifically the majority equal to 37.7% goes to school, to university or in general to study and the 28.8% goes to work.

3.3 Forecasting approaches simulated

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*, *TRENITALIA* companies.

A Multinomial Logit model specification has been used to model mode choices with four mode or service alternatives: *car*, *motorcycle*, *public transport*, *walking*. For each mode, the systematic utility includes both level-of-service and socio-economic attributes (for all models details see [34, 35]).

As mentioned in section 2 and described in detail in next figure and table, the whole model system has been validated in four different forecasting approaches with respect to the *after* traffic counts. In this way, the relative importance of each updating activity has been evaluated.

The *base model system* has been progressively updated by using *before* traffic counts and then used to forecast *after* link flows to be compared with the corresponding observed flows.

The *base model* consists in a supply model, an estimation of the public transport OD demand vector, a mode and an hyper-path choice model; the supply and the hyper-path choice model have been updated (corrected) in order to reproduce before traffic counts. This update activity has consisted in the identification

of incompatibilities between observed and assigned link flows and, consequently, of possible mistakes in the supply/assignment models. Specifically, in this application, some incompatibilities were found and solved through the correction of some unreal link travel times related to: network configuration, link travel times and hyper-path choice model parameters.

Attention was paid in the correct location of stop nodes in order to correctly identify the available lines of each stop node. That avoided 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.

Furthermore, pedestrian and on board travel times were corrected. Concerning the waiting times, different formulations were adopted as a function of both frequency and regularity of the line.

A fifth forecasting approach has been simulated to complete the updating of the whole model by using all available data (both before and after traffic counts).

Specifically, the first two analysis concern the model in its base forecasting approach. The difference between 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 allows the evaluation of 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.

The fourth analysis introduces the updating of the OD demand in the before scenario.

In the fifth analysis, *after* traffic counts are also used to update the mode choice model parameters. It is obvious that for a forecasting approach of this type the *after* traffic counts have lost their hold out sample role and then they have been used as calibration sample. This procedure would be useful whether a further after scenario were available characterized by other changes in the transport supply and referred to a time period subsequent to the previous analyzed ones (this will be one of the further research perspectives considered by the authors).

The performance indicators used to value the quality of the estimations for both *before* and *after* scenario are the *weighed Mean Absolute Percentage Deviation* ($MAPD_w$) and the *average Daily Rail Lines Demand Level Variation* (DLV):

In Table 1 the performance indicator values for each forecasting approach are reported in both scenarios. The first two show as the base model is not accurate enough for reproducing the transport system also in the before scenario (average percentage estimation errors of about 29%). Comparisons between the first and the second forecasting approach 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 (reduction of the average percentage estimation errors of about 5%). In the third forecasting approach, the relative increase in the model performances due to the correction of the supply/assignment model can be valued (average percentage estimation errors of 18% for the before scenario and about the 31% for the after scenario). In the

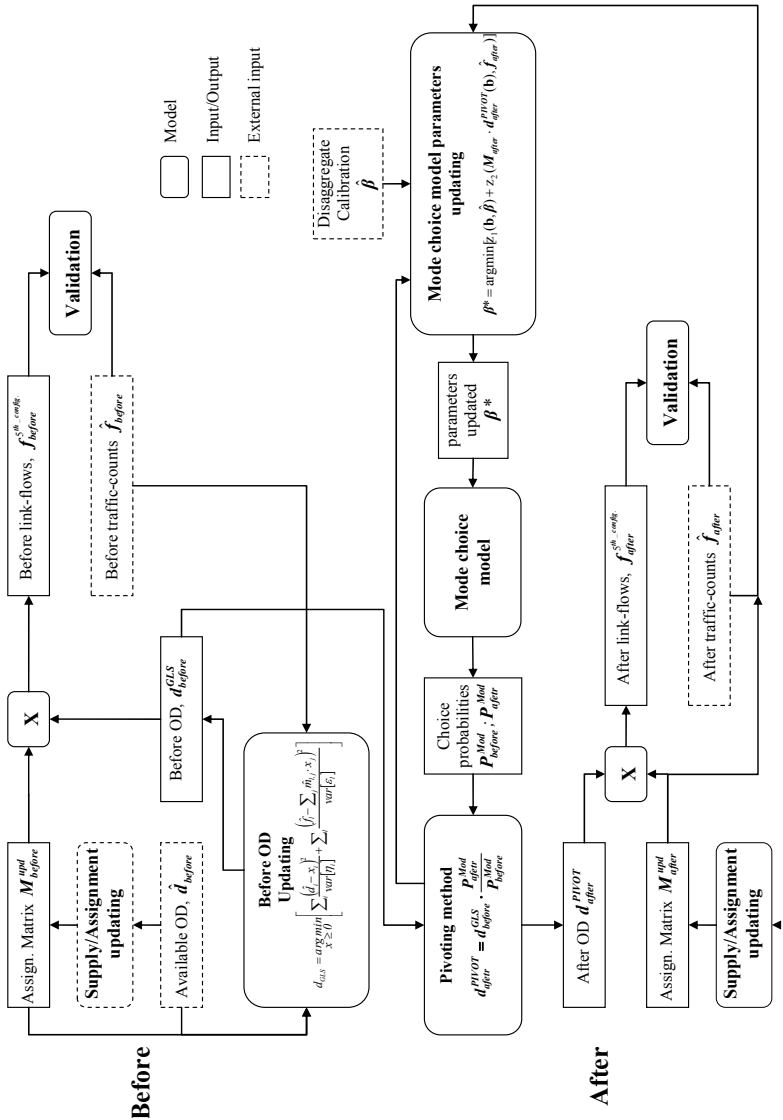


Figure 1: The before and after validation methodology.



fourth forecasting approach, the *before* traffic counts are used at their best by updating the whole model system (average percentage estimation errors of about 6% for the before scenario) and, consequently the best forecasting capability of the model system can be valued (average percentage estimation errors of about 18% for the after scenario).

As said before, a change in the transport supply, such as that of the after scenario, has brought significant changes in users' choice behaviour. Specifically, the path choice (i.e. the service), the mode choice and all the choices connected with trip frequency and destination. The first two types of choices are in general connected with short-medium changes; those referred to the third type are, on the other hand, connected with medium-long term changes. The RP survey carried out for the after scenario (7–10 months after the opening of the new lines) has “caught” the medium-term effects (changes in the path choice, in the mode choice and a 10% increase of demand level).

In the light of the above considerations, in the application of the model, elastic only with respect to the mode choice and path choice, a systematic error was expected for the estimates of the changes of the total demand level. This outcome is confirmed by the values of the average Daily Rail Lines Demand Level Variation (*DLV*) indicator which, as it is reported in Table 1, underestimates systematically the rail demand level also in scenario 4, which is considered the best in terms of forecasting ability (16% underestimation of the rail demand). Such result highlights how models considered elastic, only on the mode and path choices dimensions, can be used for short-term forecasts or for small changes of the transportation supply system (e.g. changes in public transport frequency; new bus routes, etc.).

As said above, the *after* traffic counts has been also used to update the mode choice model parameters. In this last forecasting approach a further increasing has been obtained in terms of performance indicators (average percentage estimation errors of about 14% for the after scenario).

Table 1: Before and after: forecasting approaches simulated and performance indicator values.

Id.	forecasting approach simulated	before scenario		after scenario	
		<i>MAPD_w</i>	<i>DLV</i>	<i>MAPD_w</i>	<i>DLV</i>
1	- “base model” - after demand equal to before demand (rigid demand)	29%	-9%	40%	-24%
2	- “base model” - after demand by pivoting before demand (elastic demand)	29%	-9%	35%	-24%
3	- supply and assignment models updated - after demand by pivoting before demand	18%	-6%	31%	-21%
4	- supply and assignment models updated - before demand updated - after demand by pivoting the updated before demand	6%	0%	18%	-16%

4 Conclusions

This study has investigated the forecasting reliability of transport model systems through a before and after study taking advantage of the opening of new rail lines of the Campania *Regional Metro System*. The estimation methodology proposed has been applied to the simulation of the Napoli Metropolitan Area public transport system. Some errors of the supply/assignment models were detected and corrected in order to avoid possible unreal OD demand estimation. Some interesting remarks can be made from the results analysis. First of all, a correction procedure of transport model system based on aggregate data (traffic counts) seems to be necessary to obtain a model with an acceptable capability to reproduce/forecast transport systems (see forecasting approaches 1 and 2 in Table 1). Secondly, when dealing with considerable changes in the supply system (for example the opening of two new metro lines), considerable changes in the OD demand might also be expected and, therefore, a model is absolutely needed to forecast these changes (compare forecasting approaches 1 and 2 in Table 1). Thirdly, errors can be found in the whole model system. Consequently not only the demand vector, but also the supply/assignment model, should be updated through these aggregate data: model improvements that can be reached in that way are not trivial (compare forecasting approaches 2 and 3 in Table 1). Finally, once the whole model system has been updated, the best forecasting capability of the model system can be valued (see the performance indicators of the after scenario in forecasting approaches 4 in Table 1).

The application of the model, considered elastic only on the mode and path choices dimensions, can be used for short-term forecasts or for small changes of the transportation supply system. More sophisticated models should be developed and applied in the case of frequency and destination changes.

References

- [1] Algiers S., Long distance travel behaviour models in Sweden. *6th WCTR Proceedings*, Lione 1992. pp. 147–157, 1993.
- [2] Ashley D.J., Forecasting passenger travel demand – international aspects. *Transportation*, 14, 1987.
- [3] Daly A., Gunn H., Hungerink G., Vrolijk P., Application of the Netherlands national model to support investigation of transport policy. *17th PTRC Proceedings*, 1989.
- [4] Koppelman F.S., Multidimensional model system for intercity travel choice behavior. *Transportation Research Record*, 1241. pp. 1–8, 1989.
- [5] Cacetta E., Papola A., Marzano V., Simonelli F., Vitiello I, Quasi-dynamic estimation of o-d flows from traffic counts: formulation, statistical validation and performance analysis on real data. *Transportation Research B*, vol. 55, pp. 171–187, 2013.
- [6] Ben-Elia E., Di Pace R., Bifulco G.N., Shiftan Y., The impact of travel information's accuracy on route-choice. *Transportation Research Part C: Emerging Technologies*, 26, pp. 146–159, 2013.



- [7] Carteni A., Punzo, V., Travel time cost functions for urban roads: A case study in Italy. *WIT Transactions on the Built Environment*, 96, pp. 233–243. DOI: 10.2495/UT070231, 2007.
- [8] Ercolani M., Placido A., D’Acierno L., Montella B., The use of microsimulation models for the planning and management of metro systems. *WIT Transactions on the Built Environment* 135, pp. 509–521, 2014.
- [9] D’Acierno L., Gallo M., Montella B., Placido A., Analysis of the interaction between travel demand and rail capacity constraints. *WIT Transactions on the Built Environment* 128, pp. 197–207, 2012.
- [10] Quattrone A., Vitetta A., Random and fuzzy utility models for road route choice. *Transportation Research Part E: Logistics and Transportation Review*, Volume 47, Issue 6, Pages 1126–1139, 2011.
- [11] Simonelli F., Papola A., Marzano V., Vitiello I., A methodology for locating link count sensors taking into account the reliability of prior o-d matrix estimates. *Transportation Research Record*, No. 2263 pp. 182–190, 2011.
- [12] Biggiero G., A survey on impacts of bike oriented policies on residents and economic activities. *WIT Transactions on Ecology and the Environment*, 191, 2014.
- [13] Biggiero L., The impact of transport management on local activities system: The role of limited traffic zones. *WIT Transactions on the Built Environment*, 138, pp. 669–678, 2014.
- [14] Galante G., Russo Spina M., Advanced Travel Demand Management policies. *WIT Transactions on the Built Environment*, 2014.
- [15] de Martinis V., Pagliara F., Wilson A., The evolution and planning of hierarchical transport networks. *Environment and Planning B: Planning and Design*, 41(2), pp. 192–210, 2014.
- [16] Punzo V., Ciuffo B., No free lunch theorems applied to the calibration of traffic simulation models. *IEEE Transactions on Intelligent Transportation Systems (in press)*, DOI: 10.1109/TITS.2013.2287720, 2013.
- [17] Punzo V., Borzacchiello M., Ciuffo B., On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data. *Transportation Research Part C* 19 (2011), 1243–1262, DOI: 10.1016/j.trc.2010.12.007, 2011.
- [18] Bifulco, G.N., Pariota, L., Brackstone, M., McDonald, M., Driving behaviour models enabling the simulation of Advanced Driving Assistance Systems: Revisiting the Action Point paradigm. *Transportation Research Part C: Emerging Technologies*, 36, pp. 352–366, 2013.
- [19] Polimeni A., Vitetta A., Joint network and route optimization in road evacuation. *WIT Transactions on Ecology and the Environment* 155, pp. 1053–1065, 2011.
- [20] Carteni A., de Luca, S., Analysis and modeling of container handling equipment activities. *European Transport – Trasporti Europei*, (46), pp. 52–71. ISSN: 18253997, 2010.
- [21] Carteni A., de Luca S., Tactical and strategic planning for a container terminal: Modelling issues within a discrete event simulation approach.



- Simulation Modelling Practice and Theory*, 21(1), pp. 123–145. DOI: 10.1016/j.simpat.2011.10.005, 2012
- [22] de Luca S., Cantarella G.E., Carteni A., A macroscopic model of a container terminal based on diachronic networks. *Operations Research/ Computer Science Interfaces Series*, 46, pp. 285–310, DOI: 10.1007/978-0-387-84812-9, 2009.
- [23] Cascetta E., Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. *Transportation Research*, 18B (4/5), pp. 289–299, 1984.
- [24] Cascetta E., Nguyen S., A unified framework for estimating or updating Origin-Destination matrices from traffic counts, *Transportation Research*, 22B, pp. 437–455, 1986.
- [25] Bell M.G.H, The estimation of origin-destination matrices by constrained generalized least squares. *Transportation Research*, 22B, pp. 13–22, 1991.
- [26] Carteni A., Updating demand vectors using traffic counts on congested networks: A real case application. *WIT Transactions on the Built Environment*, 96, pp. 211–221, DOI: 10.2495/UT0702112007, 2007.
- [27] Cascetta E., *Transportation System Modeling: Theory and Applications*. Springer, 2009.
- [28] Hogberg P., Estimation of parameters in models for traffic prediction: a non-linear regression approach. *Transportation Research*, 10B, pp. 263–265, 1976.
- [29] Cascetta E., Russo F., Calibrating aggregated travel demand models with traffic counts: estimators and statistical performance. *Transportation*, 24/3, pp. 271–293, 1997.
- [30] Cascetta E., Postorino M.N., Fixed point models for the estimation of O-D matrices using traffic counts on congested networks. *Transportation Science*, 35/2, pp. 134–147, 2001.
- [31] Cascetta E., Carteni A., Henke I. Stations quality, aesthetics and attractiveness of rail transport: empirical evidence and mathematical models [Qualità delle stazioni, estetica e attrattività del trasporto ferroviario: evidenze empiriche e modelli matematici]. *Ingegneria Ferroviaria*, 69 (4), pp. 307–324 ISSN: 0020-0956, 2014.
- [32] Cascetta E., Carteni A., The hedonic value of railways terminals. A quantitative analysis of the impact of stations quality on travellers behaviour. *Transportation Research Part A* vol. 61, pp. 41–52; DOI:10.1016/j.tra.2013.12.008, 2014.
- [33] Cascetta E., Carteni A., A quality-based approach to public transportation planning: theory and a case study. *International Journal of Sustainable Transportation*, Taylor & Francis, Vol. 8, Issue 1. pp. 84–106. DOI: 10.1080/15568318.2012.758532, 2014.
- [34] Bifulco G.N., Carteni, A., Papola, A., An activity-based approach for complex travel behaviour modelling. *European Transport Research Review*, 2(4), pp. 209–221. DOI: 10.1007/s12544-010-0040-3, 2010.
- [35] De Luca S., Carteni A., A multi-scale modelling architecture for estimating of transport mode choice induced by a new railway connection: The



Salerno-University of Salerno-Mercato San Severino Route [Un'architettura modellistica multi-scala per la stima delle ripartizioni modali indotte da un nuovo collegamento ferroviario: il caso studio della tratta Salerno-Università di Salerno-Mercato San Severino]. *Ingegneria Ferroviaria*, 68(5), pp. 447–473. ISSN: 0020-0956, 2013.

