Available online at www.sciencedirect.com
ScienceDirect
Computers & Operations Research 35 (2008) 436–444
computers & operations research
www.elsevier.com/locate/cor

Using gravity models for the evaluation of new university site locations: A case study

Giuseppe Bruno*, Gennaro Improta

Dipartimento di Ingegneria Economico Gestionale, Universitá di Napoli Federico II, Piazzale Tecchio, 80, 80125 Naples, Italy

Available online 23 May 2006

Abstract

A fundamental aspect of competitive spatial models is the choice behaviour of potential customers to patronize a facility. Most of the models used to describe this phenomenon are essentially based on the adaptation of Newton's law of gravitation to the economic case (gravity model).

This paper shows an application of this model to describe the behaviour of potential students in the choice of a university site. The results provided by the model have been compared with the actual data and show that the gravity model can describe the behaviour of potential students with good approximation.

Keywords: Location; Gravity model

1. Introduction

Over the last few decades, there has been a general growth in the demand for higher education in industrialized countries [1–3]. An increase in the perceived value of education [4], a wide expansion in jobs requiring higher skills [5] and the development of governmental policy strategies oriented to promote participation in higher education and universities are some of the fundamental reasons behind this phenomenon.

In many cases, this process has been accompanied by a devolution of the national educational systems, which strongly enlarged the power of the local regional authorities and the autonomy of each university (see, for instance, [6,7]).

In this context, regional authorities and/or single university institutions have been entitled to organize and manage the entire system and, hence, to decide on the number and the location of their own facilities.

In order to rationally decide optimal locations of new sites, a fundamental aspect to take into account is the choice behaviour of potential "customers" to patronize a facility where, in this application, customers correspond to students and facilities to available university sites. This aspect has been studied in literature within the context of spatial models in which consumer choice models play a key role. These models are essentially based on the adaptation of *Newton's law of gravitation* to the economic case [8–10]. This so-called *gravity model* and its variants have subsequently been widely used in regional science, transportation planning, marketing, and facility location studies (see, for instance, [11–16]).

* Corresponding author.
E-mail address: giuseppe.bruno@unina.it (G. Bruno).

0305-0548/$ - see front matter
doi:10.1016/j.cor.2006.03.008

Table 1
Distribution of population of Campania over the provinces

Province	Resident population	Squared kilometres	Density per squared kilometres
Avellino	429,178	2791	153
Benevento	287,042	2070	138
Caserta	852,872	2639	323
Salerno	1,073,643	4922	218
Napoli	3,059,196	1171	2612
Total	5,701,931	13,595	419

This paper shows an application of the gravity model to describe the behaviour of potential students in the choice of a university facility. The research started from a study that the operations research group of the University of Naples Federico II was asked to carry out in order to forecast the effects on student distribution caused by the location of a further site of the Faculty of Engineering.

The paper is organized as follows: in the next section the case study is described, while the methodology used to approach the problem is illustrated in Section 3. In Section 4, the gravity model used to describe the behaviour of the users in their choice is described and commented on. Finally, Section 5 shows the analysis of the results obtained with some conclusions in Section 6.

2. The case study

The case study focuses on the problem of the location of a further University Faculty of Engineering in the Campania region of Italy.

Campania is the second largest Italian region for number of inhabitants (5,702,000 in 2001) representing about 10% of the total national population. Its territory is divided into five provinces (Naples, Avellino, Benevento, Caserta, and Salerno), each with very different population distributions. Table 1 shows the population and the density for each province ([17]). Although the province of Naples covers less than 10% of the entire territory, it represents more than half of the total regional population. Map 1 shows the region with the subdivision in provinces.

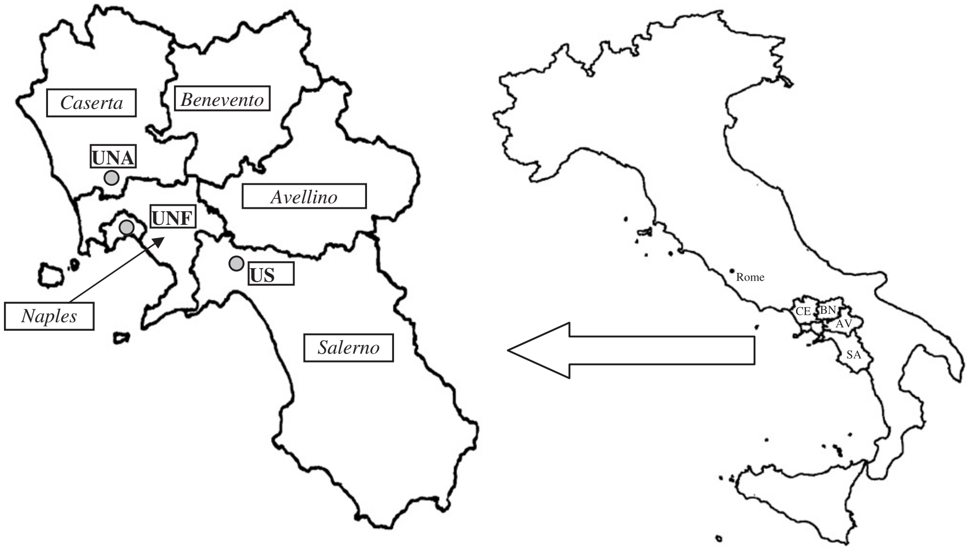
By the end of the 1980s, there were five university sites in the region: four located in the city of Naples (University of Naples Federico II—UNF, Oriental Institute, Università Navale, and Suor Orsola Benincasa) and one located in the city of Salerno (University of Salerno—US). The total number of enrolments was about 160,000, about 65% concentrated at the UNF site which is the third largest Italian University.

In the same period, the number of enrolments at the Faculty of Engineering were distributed between the UNF (70%) and the US (30%). In this context, considering the general policy tendency to increase the number of sites in order to achieve a distribution more equal across the territory, the UNF administration decided to locate a new Engineering Faculty site in the city of Aversa (in the following indicated for simplicity UNA) for three courses (Aeronautics, Civil, and Electronics). The position in the region of these faculties is indicated in Map 1. As these kinds of courses were offered only by the Engineering Faculty of UNF, the presence of the new site (UNA) should have fostered a distribution of the enrolments on these courses between the existing “old” site (UNF) and the “new” site to be opened (UNA). In this context, the operations research group performed a study [18] whose aim was to estimate the effects of the presence of the new site on the future distribution of enrolments. As the new site “competes” with the old site in attracting demand, the problem can be studied in the context of the competitive spatial models [19,20]. The next section illustrates the methodology used to carry out this research.

3. Research methodology

As stated above, this research aims to define a possible scenario regarding the distribution of enrolments on some courses of the Faculty of Engineering in the Campania region, after the opening of a new faculty site location, performing qualitative and quantitative evaluations on the effects of the new location decision.

The study carried out can be conceptually divided into three phases according to the classic approach used to solve transportation problems. The approach started by considering *the design of a zoning system* and the collection of



Map 1. The Campania region with the subdivision in provinces and the location of the university sites.

statistical data associated to each zone of the study area. This data was then used to estimate the demand generated by each zone (*generation phase*). Afterwards, this estimated demand was allocated to the available destinations; in other words the demand was distributed over the space, thus producing a demand matrix (*distribution phase*). In our application, as the aim was to forecast the distribution of enrolments between the old site (UNF) and the new site (UNA), the demand was the number of enrolments while the available destinations were represented by the two available sites. The following paragraphs provide a more detailed description of each phase.

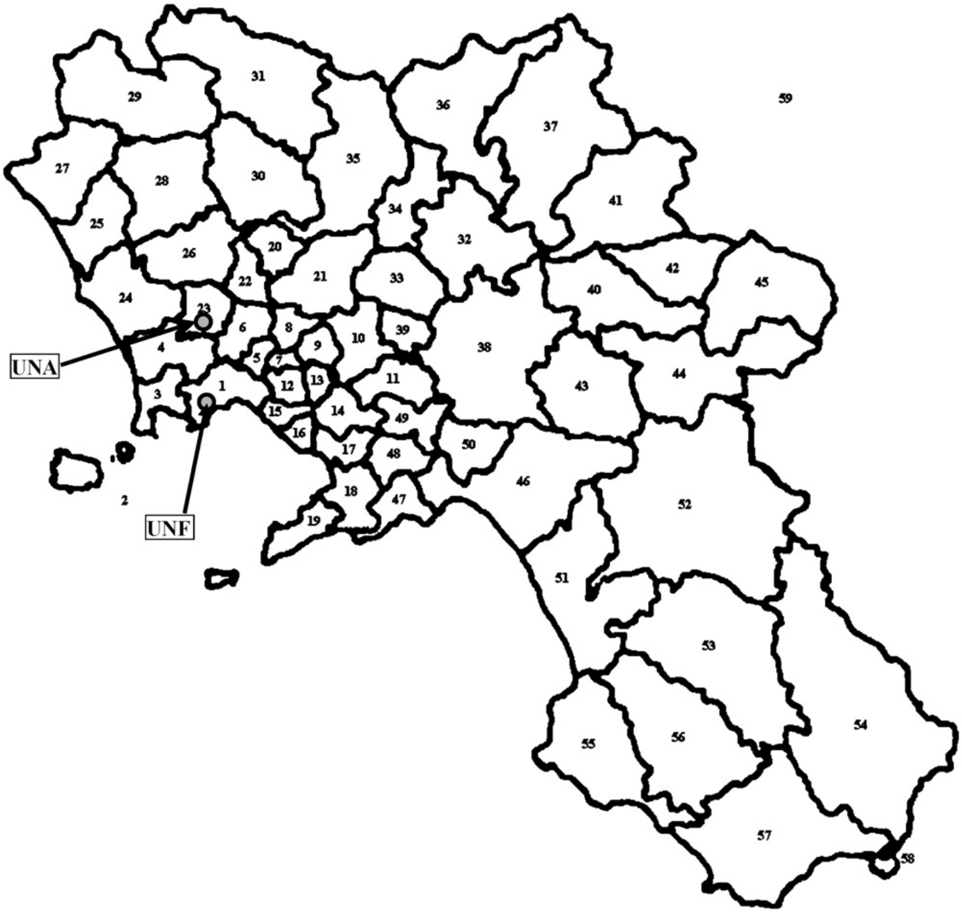
Zoning system design: A zoning system was used to aggregate available data in a manageable way. In our context, the Campania region was a natural choice for the study area, as almost all the demand for university sites was concentrated in the regional context. The study area was divided into *internal zones*. The number and the size of these zones were defined in such a way that each zone had a sufficient number of enrolments to avoid "randomized" numbers in the data. Zones have been formed as an aggregation of census zones as outlined by the Italian Institute of Statistics and were defined on the basis of land-use composition and similarity of zone characteristics. In practice considering the existing transportation networks and facilities, enrolments of the same zones reasonably use the same paths to reach the available sites. As a result, 58 internal zones were determined; a further zone (59) represents the rest of the world outside the region. Map 2 shows the zoning of the study area. UNF and UNA sites are located, respectively, in zone 1 and 23.

Generation phase: The phase aimed to predict the total demand generated by each zone and attracted to the available destinations. In our study, we aimed to determine the new distribution of enrolments after the presence of a new site or, in practice, the perturbation that the presence of a new facility caused to the existing distribution of enrolments in each zone. For this reason, we quantified the demand associated to each zone as the average number of enrolments during the last three years (1989–1991) before the opening of the new facility. The allocation of the enrolments to each zone was carried out on the basis of their permanent home address. Map 3 shows the enrolment rate in terms of number of students over 10,000 inhabitants and pictures the situation before the opening of the new site.

Distribution phase: This phase focused on the distribution of the enrolments generated by each zone over each available facility. A number of methods have been proposed over the years to describe this problem for various application fields. The best known of these methods is the gravity model which is described in the next section.

4. The gravity model

The use of the law of gravitation to describe consumer choice is based on the assumption that the probability that a customer chooses a facility that is proportional to its attractiveness and inversely proportional to a power of the distance



Map 2. Zoning of the study area.

to it. In practice, the so-called gravity model is based on the following formula:

$$G_{ij} = k_{ij}(P_i)^{\alpha_i}(Q_j)^{\beta_j}F(d_{ij}), \tag{1}$$

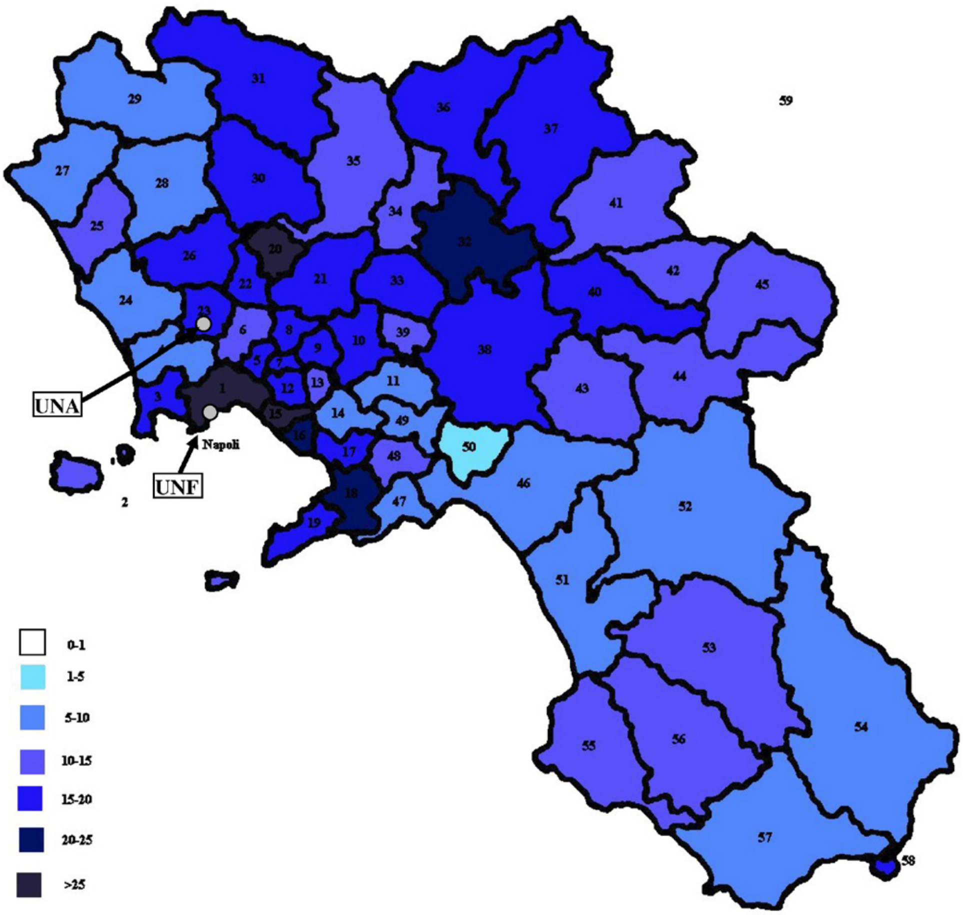
where G_{ij} is the flow from node i to node j. It depends on the activity (P_i) being considered at node i, the nodal attraction of node j (Q_j), while the friction between i and j is measured as a function $F(.)$ of the distance d_{ij} from i to j; k_{ij}, α_i and β_j are calibration parameters.

A frequently used expression for the disutility function is $F(d_{ij})$ is $F(d_{ij}) = (d_{ij})^{-n}$; in most applications, $1 \leqslant n \leqslant 3$.

The meaning of the weights P_i and Q_j depends on the specific application. Their magnitudes can be affected by various attributes; in practice $P_i = P_i(a_1, a_2, \ldots, a_p)$ and $Q_j = Q_j(a_1, a_2, \ldots, a_p)$, where $a_1, a_2, \ldots, a_p$ are attributes related to nodes. Both P_i and Q_j depend on land use factors associated to i and j such as employment, business type, and public facilities. In many versions of the model we assume $\alpha_j = \beta_j = 1$, $k_{ij} = k_{ji} = k$ and, hence, the flow G_{ij} is equal to

$$G_{ij} = kP_iQ_jF(d_{ij}). \tag{2}$$

Let N be the number of generations and M the number of attractions, the result of the application of a gravity model is a *demand matrix* G_{ij} where row i contains the demand originating from that zone i and attracted to each site j. The



Map 3. Distribution of the enrolments in terms of number of students over 10,000 inhabitants before the opening of the new site.

sum of the demand in a row should equal the demand emanating from that zone; the sum of the demand in a column corresponds to the total demand attracted to that zone. These conditions can be written as

$$\sum_j G_{ij} = O_i, \tag{3}$$

$$\sum_i G_{ij} = D_j. \tag{4}$$

In some cases there is information about only one of these constraints. If the generations (O_i) are known, the model is called "*origin constrained*". On the other hand, if attractions (D_j) are available, the model is "*destination constrained*".

In our study the total number of enrolments associated to each zone is taken from historical data. Assuming this data as constant, we wanted to determine the distribution of the enrolments from each zone to each of the two available facilities. For this reason, the model is *origin constrained* according to the following equations:

$$G_{i1} + G_{i2} = O_i, \quad i = 1, \ldots, N. \tag{5}$$

Using (2), the expressions of G_{i1} and G_{i2} are

$$G_{i1} = k\frac{P_i Q_1}{(d_{i1})^n}, \tag{6}$$

$$G_{i2} = k\frac{P_i Q_2}{(d_{i2})^n} \tag{7}$$

so that:

$$\frac{G_{i1}}{G_{i2}} = \frac{Q_1}{Q_2}\left(\frac{d_{i2}}{d_{i1}}\right)^n. \tag{8}$$

Substituting in (8)

$$G_{i1} = O_i\frac{(Q_1/Q_2)(d_{i2}/d_{i1})^n}{1+(Q_1/Q_2)(d_{i2}/d_{i1})^n}, \tag{9}$$

$$G_{i2} = \frac{O_i}{1+(Q_1/Q_2)(d_{i2}/d_{i1})^n}. \tag{10}$$

The percentages $G_{i1\%}$ and $G_{i2\%}$ of enrolments attracted, respectively, from nodes 1 and 2 are then given by

$$G_{i1\%} = \frac{G_{i1}}{O_i} \times 100 = \frac{(Q_1/Q_2)(d_{i2}/d_{i1})^n}{1+(Q_1/Q_2)(d_{i2}/d_{i1})^n} \times 100, \tag{11}$$

$$G_{i2\%} = \frac{G_{i2}}{O_i} \times 100 = \frac{1}{1+(Q_1/Q_2)(d_{i2}/d_{i1})^n} \times 100. \tag{12}$$

5. Analysis of the results

In our case study, in order to apply expressions (4) and (5) to evaluate the distribution of enrolments between the two available destinations, we had to quantify the parameters O_i, Q_1, Q_2, d_{i1}, d_{i2} $(i = 1, N)$ included in the model.

In the following, we let node 1 represent the old location of UNF while node 2 represents the new location of UNA.

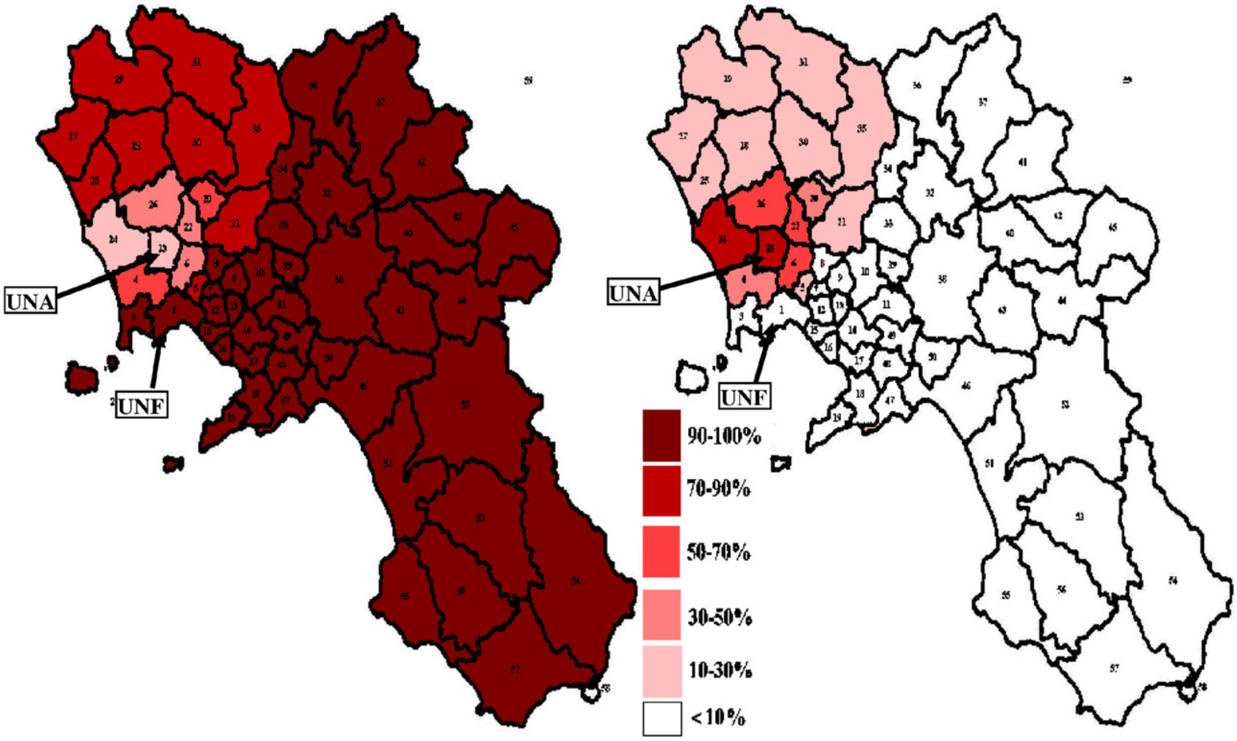
The values O_i $(i = 1, N)$ represent the total number of enrolments generated by each node i as provided by the generation phase. In order to predict the effects of the new location presence, we assumed O_i $(i = 1, N)$ equal to the average number of enrolments during the last three years (1988–1991) before the opening of the new site.

The weights Q_1, Q_2 should measure the ability of the two sites to attract student demand. For this reason, we set these values equal to the resident population of the cities hosting the two sites, as the dimension of the population can be adopted as a measure of availability of services and opportunities. In particular, $Q_1 = 1,004,500$ and $Q_2 = 53,369$ which represent, respectively, the population of Naples and Aversa in 2001 [17].

The distances d_{i1}, d_{i2} $(i = 1, N)$ should represent the distances that a potential student perceives to cover the path from zone i to location 1 and 2, respectively. These distances were defined on the basis of the results of about 500 questionnaires delivered to the students and represent the average values of the answers regarding the time (in minutes) that students perceive necessary to reach each of the locations during home-work rush hour by public transport.

The value n in the expression of the deterrence function is a calibration parameter. Using these assumptions, it was possible to apply the gravity model to the case study described. Map 4 shows the percentage distribution of enrolments from each zone to the two sites as provided by the application of expressions (5) with $n = 2$. In particular, the percentage of the total number of enrolments that should have chosen the new location was estimated equal to about 20%. According to these results, the presence of the new location should have significantly modified the distribution of enrolments only in proximity to the new location.

As the aim of the study was to forecast a possible distribution of enrolments in the presence of two available competitive sites, the model used could be considered effective if it is able to simulate the real phenomenon with a good approximation. In order to verify the effectiveness of the model, the results provided by its application were compared to the actual number of the enrolments at the two faculties (UNF and UNA) 10 years after the opening of the new site in Aversa. After 10 years of operation of the facility, we can indeed assume that the transient evolution of the phenomenon



Map 4. Percentage distribution of enrolments from each zone to the two sites provided by the gravity model.

has reached a steady condition. Map 5 shows the actual percentage distribution of enrolments between the two sites for the academic year 2001–2002. A comparison with Map 4 indicates that the model provided a result which is quite similar to the actual distribution.

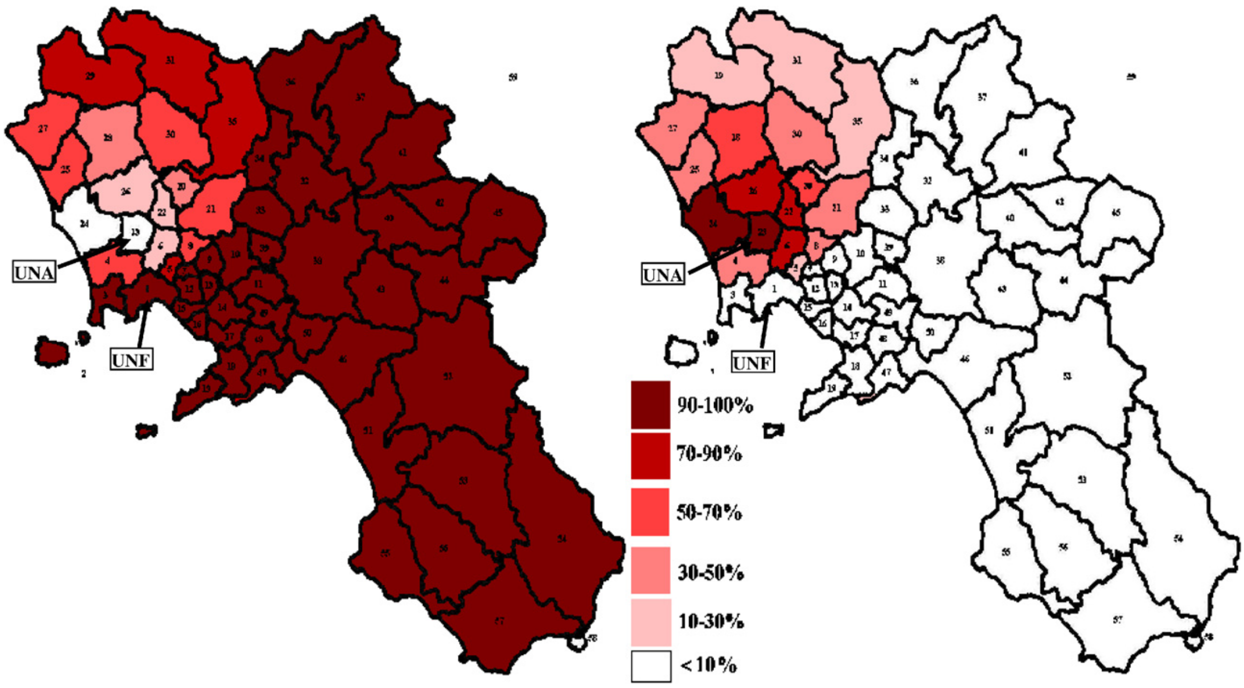
In order to realize a quantitative evaluation of the effectiveness of the model, we simulated the distribution of the actual enrolments after the opening of the new site, i.e., enrolments for the academic year 2001–2002. In other words, we supposed that the data for the total actual number of enrolments from each zone i was available and known; on the basis of this data, we determined the distribution of enrolments between the two competitive sites. According to the notation introduced above, we indicate with

$O'_i (i = 1, N)$: the total actual number of enrolments from each zone i after the opening of the “new” site;
G'_{i1} : the actual number of enrolments from each zone i which opted for the “old” site;
G_{i1}: the predicted number of enrolments from each zone i which opted for the “old” site according to the application of the gravity model;
G'_{i2} : the actual number of enrolments from each zone i which opted for the “new” site. Of course $G'_{i2} = O'_i - G'_{i1}$;
G_{i2} : the predicted number of enrolments from each zone i which opted for the “new” site according to the application of the gravity model. Of course $G_{i2} = O'_i - G_{i1}$;
e_{i1} : the difference $(G'_{i1} - G_{i1})$ between the actual and the predicted enrolments from each zone i which opted for the “old” site.

In order to define a concise measure of the effectiveness of the model, we introduced an average error index:

$$e_{\%} = \frac{\sum_{i=1,N} |e_{i1}|}{\sum_{i=1,N} G_{i1}} \times 100 = \frac{\sum_{i=1,N} |G'_{i1} - G_{i1}|}{\sum_{i=1,N} G_{i1}} \times 100. \tag{13}$$

Fig. 1 shows the values $e_{\%}$ over the parameter n. The error provided by the model is quite limited and varies between 4.62% ($n = 2.2$) and 16.47% ($n = 1.0$). Furthermore, the degree of errors e_{i1} in terms of number of students is very low;



Map 5. Actual percentage distribution of enrolments from each zone to the two sites.

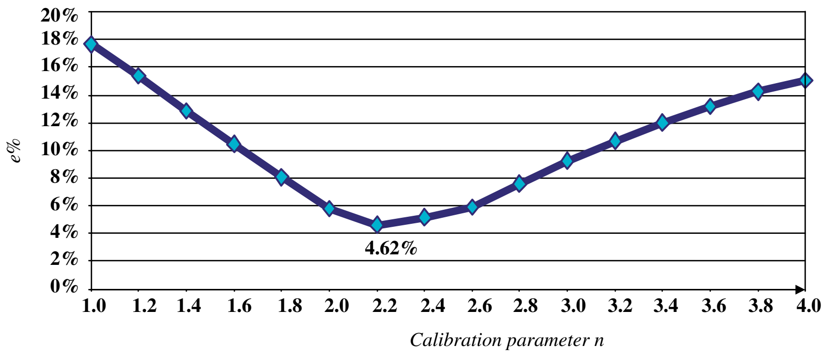


Fig. 1. Error provided by the model over the calibration parameter n.

it assumes quite significant values in the neighbourhood of the new location site (zones 20–26). This is due to the fact that the presence of the new location produced an increase in demand toward the new site that the gravity model cannot consider. These results indicate that the gravity model is quite effective in describing the distribution of enrolments in the presence of two competitive sites. The model also appears quite well-constructed as the error is limited for a wide range of the calibration parameter value.

The aspects illustrated suggest that the gravity model could be used with a good approximation to describe user behaviour in decisional location problems for the definition of optimal positions for university sites. In this context, the degree of error determined by the application of the model, if confirmed, can be considered acceptable and interesting.

6. Conclusions

This paper shows an application of the gravity model to describe the behaviour of potential students in the choice of a university site. The research started from a study whose aim was to forecast the effects on student distribution caused by the location of a further faculty site in a spatial context in which only another competitive site was present.

In order to verify the effectiveness of the model in describing the behaviour of potential students in the choice of their own faculty sites, the results of the model were compared with the actual data gathered by observation of the evolution of the phenomenon, 10 years after the opening of the new site.

The obtained results suggest that such an approach may be used to evaluate the effects of the presence of new locations on the distribution of the future demand. For this reason, it should be useful to apply the model to other case studies; in this way it should be possible to verify its efficacy and its robustness together with the role of the various elements of the model (i.e., the calibration parameters or the disutility function).

Acknowledgments

The authors are grateful to the anonymous referees whose comments and suggestions significantly improved both clarity and precision of the paper.

References

[1] Craig J. The expansion of education. Review of Educational Research 1981;9:151–213.
[2] Robinson R, Ralph J. Technical change and the expansion of the schooling in the United States, 1890–1970. Sociology of Education 1984;57(3):134–52.
[3] Garnier MA, Hage J. Class, gender and school expansion in France: a four-systems comparison. Sociology of Education 1991;64:229–50.
[4] Ramirez F, Boli J. The political construction of mass schooling: European origins and worldwide institutionalisation. Sociology of Education 1987;60:1–17.
[5] Walters P. Occupation and labor market effects on secondary and post-secondary expansion in the United States, 1890-1970. American Sociological Review 1984;49:659–71.
[6] Mora J, Villareal E. Financing for quality: a new deal in Spanish higher education. Higher Education Policy 1996;9:175–88.
[7] Ludahl L. Sweden: decentralization, deregulation, quasi-market—and then what?. Journal of Educational Policy 2002;17(6):687–97.
[8] Hoteling H. Stability in competition. Economic Journal 1929;39:41–57.
[9] Reilly WJ. The law of retail gravitation. New York: Knickerbocker Press; 1929.
[10] Huff DL. Defining and estimating a trading area. Journal of Marketing 1964;28:34–8.
[11] Converse PD. New laws of retail gravitation. Journal of Marketing 1949;14:379–84.
[12] Wilson AG. A statistical theory of spatial distribution models. Transportation Research 1967;1:253–69.
[13] Erlander S. Optimal spatial interaction and gravity model. Berlin: Springer; 1980.
[14] Fotheringham AS, O'Kelly ME. Spatial interaction models: formulation and applications. Dordrecht: Kluwer Academic; 1989.
[15] Sen A, Smith TE. Gravity models of spatial interaction behaviour. Berlin: Springer; 1995.
[16] Serra D, Colomé R. Consumer choice and optimal location models: formulations and heuristics. Papers in Regional Science 2001;80(4): 425–38.
[17] ISTAT Popolazione e movimento anagrafico dei comuni, Anno 2001. Annuari ISTAT; 2002 [in Italian].
[18] Bruno G, De Marsanich C, Improta G, Sforza A. Analisi della distribuzione territoriale degli iscritti alla Facoltà di Ingegneria dell'Università di Napoli e valutazioni sulla localizzazione di una seconda sede della Facoltà. Internal Report; 1992 [in Italian].
[19] Eiselt HA, Laporte G. Competitive spatial models. European Journal of Operational Research 1989;39:231–42.
[20] Eiselt HA, Laporte G, Thisse JF. Competitive location models: a framework and bibliography. Transportation Science 1993;27:44–54.