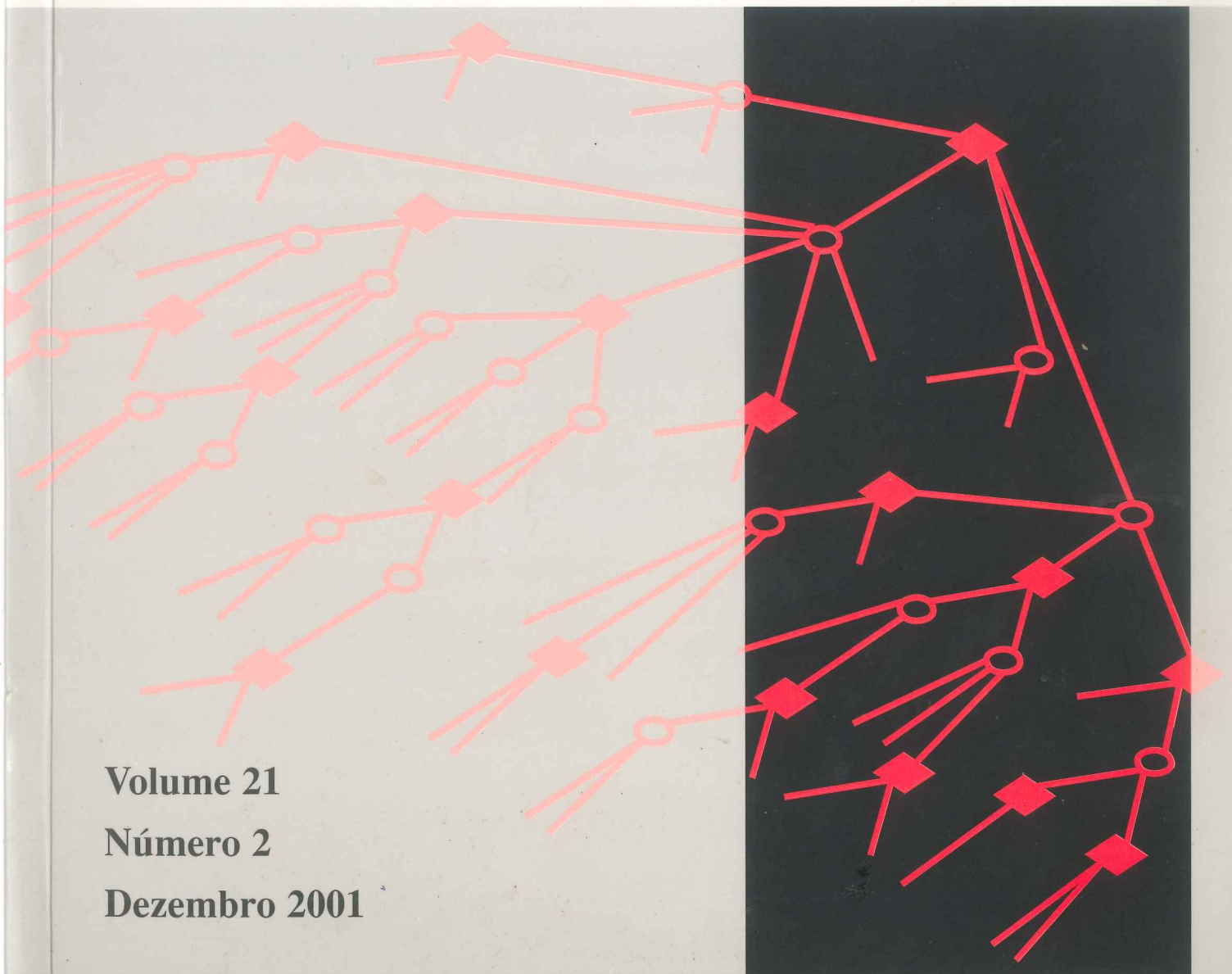


Investigação Operacional



Volume 21

Número 2

Dezembro 2001

INVESTIGAÇÃO OPERACIONAL

Propriedade:

APDIO — Associação Portuguesa de Investigação Operacional

ESTATUTO EDITORIAL

<<Investigação Operacional>>, órgão oficial da APDIO cobre uma larga gama de assuntos reflectindo assim a grande diversidade de profissões e interesses dos sócios da Associação, bem como as muitas áreas de aplicação da I. O. O seu objectivo primordial é promover a aplicação do método e técnicas da I.O. aos problemas da Sociedade Portuguesa.

A publicação acolhe contribuições nos campos da metodologia, técnicas, e áreas de aplicação e software de I. O. sendo no entanto dada prioridade a bons casos de estudo de carácter eminentemente prático.

Patrocinadores

Fundação Calouste Gulbenkian

FCT

Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA E DA TECNOLOGIA

Apoio do Programa Operacional Ciência, Tecnologia,
Inovação do Quadro Comunitário de Apoio III.

ISSN nº 0874-5161

Dep. Legal nº 130 761 / 98

Execução Gráfica: J. F. Macedo - Astrografe

700 Ex.

2001/12

INVESTIGAÇÃO OPERACIONAL

Volume 21 - nº 2 - Dezembro 2001

Publicação Semestral

Editor Principal: Joaquim J. Júdice
Universidade de Coimbra

Comissão Editorial

M. Teresa Almeida Inst. Sup. Economia e Gestão	Laureano Escudero IBM, Espanha	Rui Oliveira Inst. Superior Técnico
C. Henggeler Antunes Univ. de Coimbra	Edite Fernandes Univ. do Minho	J. Pinho Paixão Univ. de Lisboa
Jaime Barceló Univ. de Barcelona	J. Soeiro Ferreira Univ. do Porto	M. Vaz Pato Inst. Sup. Economia e Gestão
C. Bana e Costa Inst. Superior Técnico	J. Fernando Gonçalves Univ. do Porto	A. Guimarães Rodrigues Univ. do Minho
M. Eugénia Captivo Univ. de Lisboa	Luís Gouveia Univ. de Lisboa	António J. L. Rodrigues Univ. de Lisboa
Domingos M. Cardoso Univ. de Aveiro	Rui C. Guimarães Univ. do Porto	J. Pinho de Sousa Univ. do Porto
João Clímaco Univ. de Coimbra	J. Assis Lopes Inst. Superior Técnico	Reinaldo Sousa Univ. Católica, Rio Janeiro
J. Dias Coelho Univ. Nova de Lisboa	Carlos Luz Inst. Polit. Setúbal	L. Valadares Tavares Inst. Superior Técnico
João P. Costa Univ. de Coimbra	Virgílio P. Machado Univ. Nova de Lisboa	Isabel H. Themido Inst. Superior Técnico
Ruy Costa Univ. Nova de Lisboa	Manuel Matos Univ. do Porto	B. Calafate Vasconcelos Univ. do Porto
J. Rodrigues Dias Univ. de Évora	N. Maculan Univ. Fed., Rio Janeiro	Luís N. Vicente Univ. de Coimbra
	José F. Oliveira Univ. do Porto	

A Revista “INVESTIGAÇÃO OPERACIONAL” está registada na Secretaria de Estado da Comunicação Social sob o nº 108335.

Esta Revista é distribuída gratuitamente aos sócios da APDIO. As informações sobre inscrições na Associação, assim como a correspondência para a Revista devem ser enviadas para a sede da APDIO - Associação Portuguesa para o Desenvolvimento da Investigação Operacional - CESUR, Instituto Superior Técnico, Av. Rovisco Pais, 1049 - 001 Lisboa.

Este Volume foi subsidiado por:

Fundação para a Ciência e Tecnologia (FCT)

Fundação Calouste Gulbenkian

Para efeitos de dactilografia e composição, foram utilizados equipamentos gentilmente postos à disposição pelo Centro de Investigação Operacional (DEIO - FCUL).

Assinatura: 25 Euros

A GENERAL GENERATOR OF PROJECT NETWORKS IN TERMS OF THEIR MORPHOLOGICAL FEATURES

J. Antunes Ferreira

L. Valadares Tavares

J. Silva Coelho

CESUR, Instituto Superior Técnico,
Av. Rovisco Pais
1000 Lisboa, Portugal

Abstract

Most methods of project scheduling are based on algorithms (such as branch and bound or a large variety of heuristics) and their comparative study implies their application to large sets of generated networks. Usually, the performance of these algorithms depends on the features of the project networks under study and such features are a function of the application domain (for instance, project networks of the construction industry and of banking services tend to belong to different types).

Unfortunately, the proposed generators have multiple shortcomings and do not allow the generation of project networks in terms of their morphology which is a critical condition to relate the performance of the studied algorithm with the features of the network, for each domain of application.

A new generator is proposed in this paper based on the concept of progressive level and producing networks in terms of six morphologic indicators.

The generation of large networks (≈ 1000 activities) is achieved with very low computing times.

Keywords

project network, generation, morphology

Introduction

The study of important problems of project scheduling such as the optimal scheduling of a project under resource constraints or the estimation of the project risk has been based on alternative algorithms (such as the exact methods or the heuristics proposed by Patterson, 1984, Patterson et al, 1990, Demeulemeester et al, 1992, Tavares, 1995, Kolisch and Sprecher, 1996).

The number of different algorithms presented to optimize problems of project scheduling is quite large and it is steadily increasing as it has been shown by the International Workshops of the EURO Working Group on Project Management and Scheduling organized since 1988 (See the Editorials of the Special Issues on Project Management and Scheduling published by EJOR with the selected papers of these Workshops: Tavares and Weglarz, 1990; Carlier and Tavares, 1993; Tavares and Vercellis, 1994). This increase may be explained by the variety of problems under study, by the diversity of project networks and also by the uncertain performance of each procedure.

Actually, most often, the performance of each algorithm depends on the features of the studied project network and such features depend on the studied domain (for instance, project networks of the construction industry and of banking services tend to have different features).

Therefore, a better understanding about the validity, the scope and the scoring of each algorithm implies its application to a simulated set of instances which should be generated preserving the features of the available real sample of project schedules belonging to the studied domain.

Then, the development of methods to generate project networks preserving the type of networks identified in a sample of real examples becomes a key condition to allow the scientific assessment of so many different procedures and to help project managers to find the most appropriate decisions for the type of network under study.

Unfortunately, this objective has been hardly pursued in the literature and the number of published generators seems to extremely small (≤ 5) which is in sharp contrast with the number of published solution procedures for most problems.

Actually, the major generators already published were proposed by Demeulemeester et al, 1993, by Kolisch et al, 1995 and by Agrawal et al, 1996.

The analysis of the simulation method has to give special attention to the features which are supposed to be preserved by the generated sample of instances in order that the real sample of studied networks can be reproduced by simulation. Therefore the features preserved by these three methods should be considered:

A - Demeulemeester, Dodin and Herroelen, 1993

This method just preserves the number activities and nodes. The AoA (Activity-on-Arc) representation has to be adopted as the problem of redundancy is not considered.

Actually, each generated arc is an activity and hence the problem of redundancy does not exist. However, there are networks requiring the introduction of dummy activities to be represented by AoA (See Figure 1) but this type of activities is not considered by the authors. Checking redundancy is required if dummy activities are included as it is also shown in Fig. 1. Therefore, this method of generation excludes all networks requiring such type of activities.

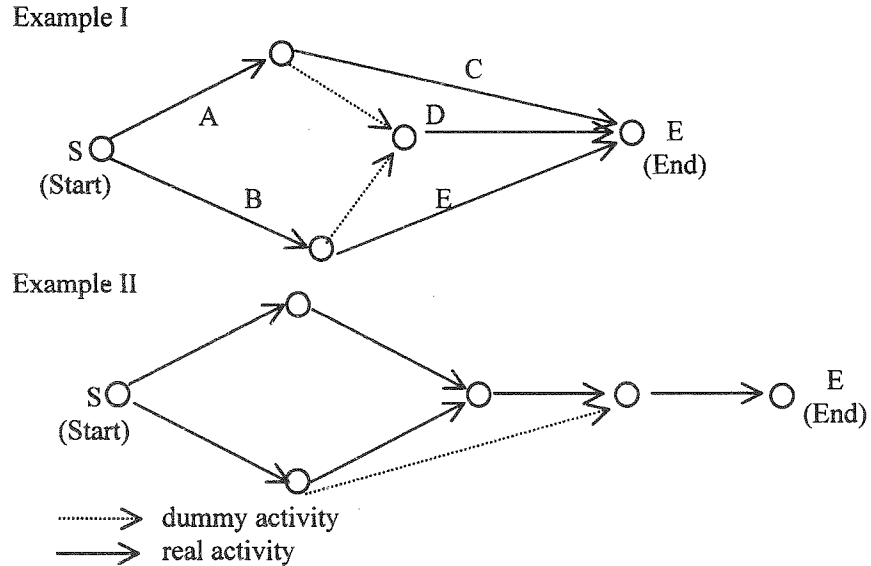


Figure 1 - Example of a network (I) requiring a two dummy activities to be represented by AoA and an example with a redundant dummy activity (II)

B - Kolisch, Sprecher and Drexler, 1995

This generator, Pro Gen, adopts the AoN (Activity-on-Node) representation and attempts to preserve the number of non-redundant arcs per node (network complexity). Also, the consideration of another parameter has been discussed: the network restrictiveness (Thesen, 1977). This parameter measures the number of precedence-feasible activity sequences as a fraction of the total number of all activity sequences.

This parameter is hard to be computed and Schwindt, 1996, has adopted an estimator (Thesen, 1977) which can be computed in polynomial time.

C - Agrawal, Elmaghraby and Herroelen, 1996

This procedure, DAGEN, generates Activity-On-Arc networks. DAGEN is based on the "reduction complexity" index proposed by Bein, Kamburowski and Stallman (Bein et al, 1992) measuring the non-conformity of a network to the series parallel case. DAGEN receives from the user this index, the number of nodes and arcs and tries to get a feasible solution. There is no checking of redundancy on the arcs which means that this generator has the same drawback of the generator A.

Several remarks can be made about these three contributions:

- Methods A and C imply the AoA representation and dummy activities are not considered. However, as it is recognized by one of the authors of C (Elmaghraby, 1995, page 386 and 388), "the AoA representation of the same project, commonly referred to as the graph D is not unique, and is cluttered with dummy activities and dummy events..." and "Undoubtedly, the AoN representation is the more direct, the more frugal, and is unique". Thus, this drawback which is common to A and C may be quite a serious one.

- The adoption of the AoA representation has another serious drawback if used for simulation: it requires receiving the number of the nodes from the user. Unfortunately, the number of nodes in AoA is more an output as it is a function of the number of the activities and of the adopted precedence structure than an input. Thus, this is not an easy input to be given.
- The first method just preserves the number of activities and nodes and the others simply ignore all the variety of shapes.

Therefore, these methods cannot guarantee the reproduction of different types and shapes of project networks described by appropriate indicators. Furthermore, quite often, the size of the discussed generated networks is below 120: the results obtained in A are discussed with no more than 60 activities and the program of DAGEN does not accept more than 100 nodes (AoA). The generator B has no upper limit for the number of nodes and examples with 120 activities were discussed (Kolisch et al, 1999) but the study of different generators has adopted no more than 50 activities to compare alternative heuristics (Ahn and Erenguc, 1998 and Sprecher and Drexler, 1998).

The method presented in this paper adopts a radically different approach which is based on:

- the use of AoN representation to avoid the drawback of requiring from the user the input of the number of nodes (events, using AoA) which is unknown, most often. Thus, the number of dummy activities does not have to be specified;
- the adoption of the hierarchical level as the basic concept to study the morphology of the network;
- the construction of a framework of six morphological indicators (instead of just two as it was proposed in generator B) to synthesize the morphologic type of the network;
- the construction of a generator based on these indicators;

Obviously, the generation of the specific features of each activity (duration, cost, etc.) can easily be carried out using standard procedures once the network has been generated.

The proposed generator can have an easy computational implementation producing large networks with 1000 or more activities. The generation time of a network with 1000 activities is less than 15 seconds. The generated test set includes 802 networks with a number of activities from 25 to 1000 and so it is the first test set including large problems. Actually, this is in contrast with the most common test set adopted until now which is the collection of 110 instances proposed by (Patterson, 1984) with a number of activities ranging from 5 to 49.

2. On the morphology of project networks

Project networks have been intensively studied since the seminal work of (CLARK, 1962; BATTERSBY, 1967) and they are usually represented using the notation AoN (Activity - on - Node) or AoA (Activity - on - Arc). The authors in this paper adopt the former to avoid the need to specify the number of nodes in the AoA representation due to the reason discussed.

In the following figures each node is represented by a circle and the start or the end of the network is denoted by a dummy node represented by a solid circle.

These representations just describe the structure of the precedences of the project and they can follow an extremely wide variety of morphological types as these networks only have to comply with the conditions of acyclicity and connectivity. The concept of "morphological type" is not well defined in the literature but any experienced project manager knows quite well that the representation of each project network can be compared to one of several typical morphologic types such as those presented in Figure 2 (serial, parallel or triangular shapes) and that the problems and the risks associated with the management of the project depend heavily on the specific type of each project.

The most appropriate concept to describe these structures of precedences is the concept of hierarchical level (Elmaghraby, 1977) but unfortunately it has been completely ignored in the literature devoted to the development of generators of project networks (See Demeulemeester et al, 1993, Kolisch et al, 1995 and Agrawal et al, 1996). This concept includes the progressive level and the regressive level.

The progressive level (a_i), the regressive level (b_i) and the concept of level float (Δ_i) for any generic activity, i , (with $i=1, \dots, N$ being N the number of activities of the project network) can be defined as described below. The full set of activities is denoted by Ω .

$$a) \quad a_i = \text{Max}_{j \in J(i)} a_j + 1$$

where a_i is the progressive level of activity i and $J(i)$ is the set of activities directly preceding activity i (direct predecessors of i). The maximal progressive level is denoted by M . The precedence link connecting any pair of activities, j and i , is called a direct precedence link. If $J(i)$ is empty, then $a_i = 1$ by definition (the dummy node representing the beginning of the network corresponds to $i = 0$ and it has level zero).

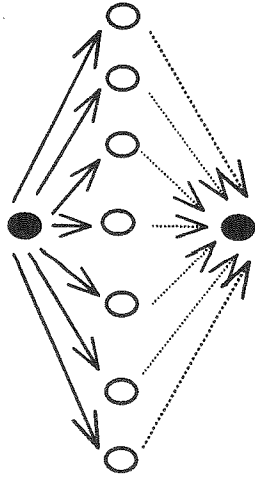
$$b) \quad b_i = \text{Min}_{k \in K(i)} b_k - 1$$

where b_i is the regressive level of activity i and $K(i)$ is the set of activities including activity i as a direct precedent activity (direct successors of i). If $K(i)$ is empty, then $b_i = M$. It should be noted that no level is associated to the dummy node representing the end of the project and therefore the links received by this node are not counted (See Figure 3 illustrating this rule). For the example of Figure 3, the progressive and regressive levels of each activity are also computed and presented in this Figure.

Serial case:



Parallel case:



Triangular case:

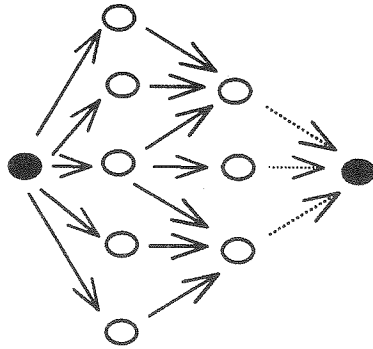
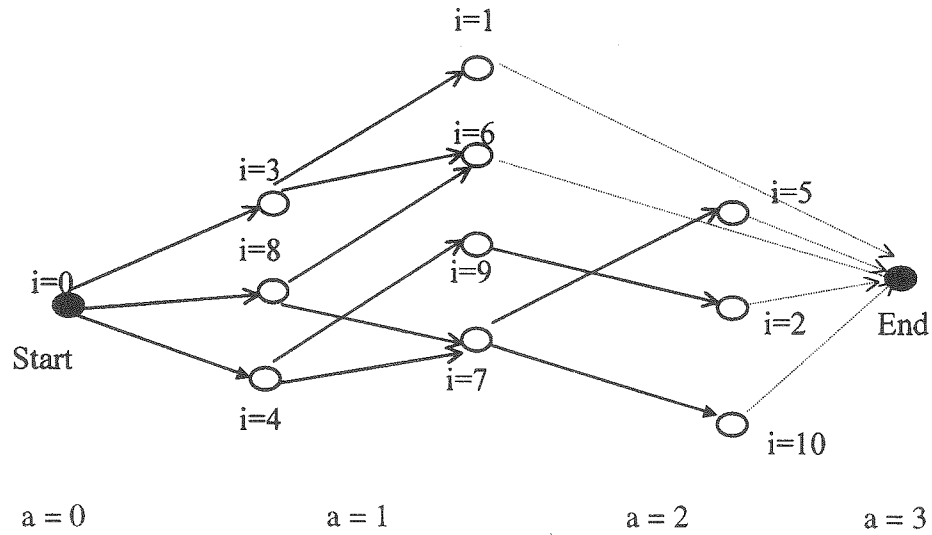


Figure 2 - Examples of different morphological types of project networks



N° of activities: 10

N° of links: 12

Activity i	a_i	b_i
0	0	0
1	2	3
2	3	3
3	1	2
4	1	1
5	3	3
6	2	3
7	2	2
8	1	1
9	2	2
10	3	3

Figure 3 Example of a project network and of the progressive and regressive levels of its activities

c) $\Delta_i = b_i - a_i$ as is shown in Figure 4

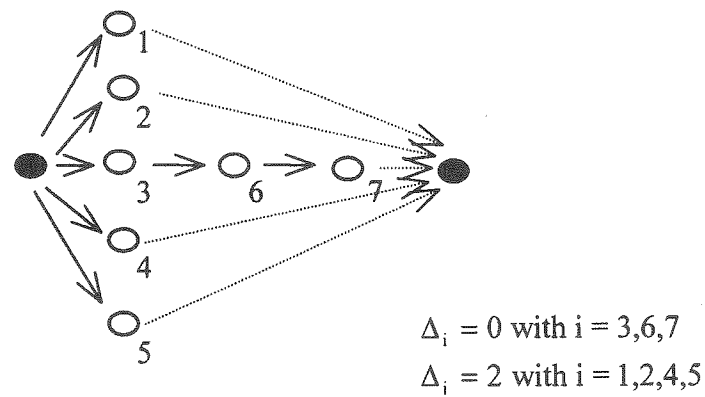
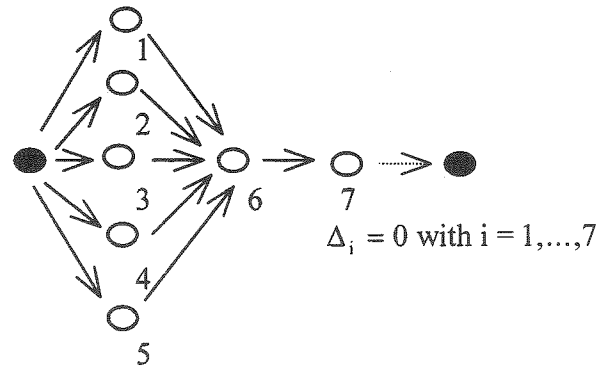


Figure 4 - Examples of the level float

The proposed concept of "morphological type" of a project network is herein defined in terms of three major perspectives.

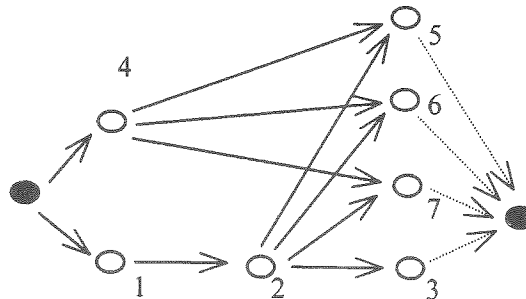
A. the graphical shape of the network's representation using the AoN rule and representing each precedence by an arrow with the same length and each activity by a circle of the same size. (See previous examples in Figure 2).

B. the number of non-redundant direct precedence links.

The concept of precedence is always a restriction to the freedom of any project manager because the number of alternative feasible project schedules does not increase and, usually, it is reduced every time a restriction is introduced. Actually, the easiest project is the parallel one and the hardest one is the serial one. Therefore, a measure of such difficulty is the number of non-redundant precedence links, D , which will be called by project "hardness".

The project hardness is completely different from the concept of computational difficulty to achieve an optimal schedule. Actually, this last type of difficulty tends to be lower when there are fewer alternative feasible schedules and therefore it will tend to be negatively correlated with project hardness (the parallel case has the maximal number of feasible schedules and the serial case has the minimal number).

C. the level length, L , of the non-redundant direct precedence links of the network. The level length (L_{ij}) of each non-redundant precedence relation from i to j is defined by the difference between the progressive level of the two linked activities (Figure 5).



Level length of the precedence links:

$$(4,5); (4,6); (4,7) = 2$$

$$(1,2); (2,3); (2,5); (2,6); (2,7) = 1$$

Figure 5 - Level length of the precedence links

A significant number of precedence links with level length greater than 1 means that the project network is far from being decomposable into a set of sequential stages to be performed according to Tavares, 1994. Obviously, in this case, the level floats tend also to be positive.

The sequence of the events defined by the start of each activity of a project network is not uniquely determined if the float level of one or more activities is positive. Therefore, the project schedule is less determined if such floats are higher.

3. The proposed set of morphological indicators

Each of the three previously presented perspectives can be described in terms of a few topological features which will be used to construct a set of morphological indicators.

These indicators will be very useful to categorize the networks into several classes and they will provide the basic framework for the generation of networks.

The perspective A can be synthetized by:

- Size

The corresponding indicator I_1 , is defined by the number of activities:

$$I_1 = N$$

- Level Dimension

The absolute level dimension is defined by the dimension of the longest path measured in terms of the hierarchical levels:

$$I_2 = \text{Max}_{i \in \Omega} a_i$$

The corresponding indicator I_2 is the relative dimension and measures how serial is the shape of the network (for $N > 1$):

$$I_2 = \frac{M-1}{N-1}$$

where $0 \leq I_2 \leq 1$ being 0 the parallel case and 1 the serial one (Figure 2).

- Width

This magnitude can be defined by the number of activities at each progressive level, $W(a)$, with $a = 1, \dots, M$ being, $W = 1, 2, \dots$.

The maximal width is denoted by MW and several shapes can be adopted to describe the function $W(a)$ (see Figure 6).

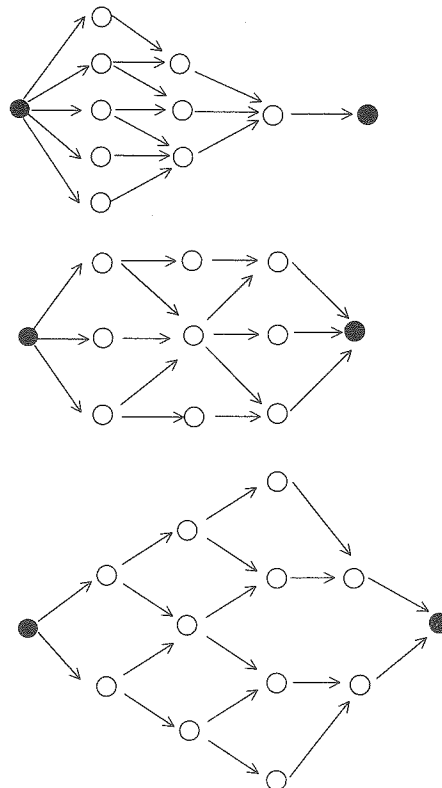


Figure 6 - Several shapes for the width of the network described by $W(a)$

Each level implies at least one activity and therefore the width of each level should be measured by $[W(a)-1]$. Obviously, the relative width should be then defined by $[W(a)-1]/[N-M]$ because $(N-M)$ is the number of activities which are free to be allocated to each level.

Thus, the proposed indicator, I_3 , is only meaningful if $N > M$ and it is defined by the function $I_3(a) = \frac{W(a)-1}{N-M}$ with $1 \leq a \leq M$. Obviously, one should have $\sum_{a=1}^M I_3(a) = 1$.

Often a linear and symmetrical variation of $W(a)$ can be assumed. In these cases it means that:

- a) The maximum width (MW) occurs in the $M/2$ level or in the two nearest integer values;
- b) The value of $W(a)$ for the first level and for the latest one must be equal: $W(1) = W(M)$.

Obviously, if $W(1) = MW$, it means that the width in the first level is the maximum one and equal to the width in the centre of the network and it will have a rectangular shape. If $W(1)$ has a value less than MW then a triangular occurs but the shape is so more close to the rectangular one as $W(1)$ is nearer MW . The slope of the triangular shape increases when the ratio $W(1)/MW$ decreases. By these reason, and in the cases described above, a simplified indicator can be proposed:

$$I_3^* = W(1)/MW$$

The perspective B (the number of non-redundant direct precedence links) can be described in terms of the number of non-redundant direct precedence links per activity, as it has been already discussed (Kolisch et al, 1995).

The minimal number of these links for a network with N activities and M progressive levels is equal to N if the links between the initial node and the activities of level one will be also counted but ignoring the connections between the activities of the last level and the final node. This is the convention adopted in this paper.

Obviously, many other direct precedence links can be received by an activity i , with level length equal to $1, 2, \dots, a_i$ where a_i is the progressive level of i .

The maximal number of these links occurs when all activities of a progressive level are directly linked with all the activities of the level $i-1$.

The maximal number, D , of the non-redundant precedence links with level length equal to one is obtained by the connection of each activity, i , belonging to the level a ($a = 0, 1, \dots, M-1$) with all the activities belonging to the level $(a+1)$. Therefore, D is given by:

$$D = W(1) + \sum_{a=1}^{M-1} W(a) \cdot W(a+1)$$

and the proposed indicator, I_4 , expressing the precedence index for links with a length equal to 1, is defined by:

$$I_4 = \frac{n(1)-N}{D-N}$$

where $n(1)$ is the number of links with level length equal to 1, which will satisfy $0 \leq I_4 \leq 1$ as $n(1) \geq N$.

This indicator expresses how dense is the precedence structure of the network for the level length equal to one.

The perspective C can be synthetized by the distribution of the level float of the activities or by the distribution of the level length of each direct precedence link. A project network with longer links has higher floats and these magnitudes are strongly correlated.

Usually, the number of links of level length L, n(L), decreases with their level length L and therefore the following assumption may often be adopted

$$n(L+1) = n(L) \cdot p$$

with $L = 1, 2, \dots, V-1$ where is V the maximal level length and where p is a factor between 0 and 1. This assumption implies an exponential decrease of n(L) with L being p the constant decreasing rate. It should be noted that V cannot be higher than (M-1).

The proposed indicators to express this perspective are

$$0 \leq I_5 = p \leq 1$$

and

$$0 \leq I_6 = \frac{V-1}{M-1} < 1.$$

The average level length, \bar{L} , of the direct non-redundant precedence links can easily be determined by:

$$\bar{L} = \frac{Vp^V(p-1) + (1-p^V)}{(1-p^V)(1-p)}$$

and the total number, T, of non-redundant direct precedence links is given by:

$$T = n(1) \cdot \frac{1-p^V}{1-p}$$

Furthermore, the maximal value which can be given to V is related with the number of links with length equal to 1 because:

$$n(V) = n(1) \cdot p^{V-1} \geq 1$$

or

$$\ln n(1) + (V-1) \ln p \geq 0$$

and therefore

$$V \leq 1 + \frac{\ln n(1)}{(-\ln p)}$$

4. The generation method

The project network will be generated in terms of the defined indicators, I_1, I_2, I_3, I_4, I_5 and I_6 :

$$I_1 = N > 1$$

$$0 \leq I_2 = \frac{M-1}{N-1} \leq 1$$

$$0 \leq I_3(a) = \frac{W(a)-1}{N-M} \leq 1 \text{ with } 1 \leq a \leq M \text{ and } \sum_{a=1}^M I_3(a) = 1$$

$$0 \leq I_4 = \frac{n(1)-N}{D-N} \leq 1$$

$$0 \leq I_5 = p \leq 1$$

$$0 \leq I_6 = \frac{V-1}{M-1} \leq 1$$

N concerns the size of the network and I_2 describes how much it has a serial morphology. A significant number of project networks of the construction industry was analysed and I_2 was found to be lower than 0.3.

The selection of I_3^* is easier to be made than the choice of I_3 as I_3^* is a scalar. Two typical cases are $I_3^* = 0.8$ (almost rectangular shape) and $I_3^* = 0.2$ (almost triangular shape).

The selection of I_4 concerns the density of links with level length equal to one and, obviously, the number of possible non-redundant links with $L > 1$ diminishes when I_4 increases.

Therefore, this difficulty is particularly severe if $I_5 (=p)$ is high because then we should also have a large number of links with $L > 1$. The analysed networks have often $I_5 < 0.3$ and the numerical experiments carried out have shown that I_4 should be also $I_4 < 0.5$ to avoid a significant probability of generating an unfeasible network.

The importance of I_6 grows when $I_5 > 0.1$ and its recommended range is between 0.05 and 0.3.

The user should specify a value for each indicator of the set I_1, \dots, I_6 and the feasibility of such selection is tested by the program. The feasibility test is done in the generation procedure that will return no network if the indicator set is not feasible. Then, the following procedure to generate a network is adopted.

- 1 - Generation of N and M given I_1, I_2 .
- 2 - Allocation of the activities to the levels, $a = 1, \dots, M$, in terms of I_3 (a) and introduction of precedences with level length equal to one between arbitrarily chosen activities, one from each level.
- 3 - Generation of the population of possible links with level length L, $P(L)$, starting with $L = 1$, and the link from the initial node to the activities of level 1 are created deterministically. Generation of the population of activities of level $a = 2, \dots, M$ not yet receiving a precedence link from level $(a - 1)$ (population, \bar{P}).
- 4 - Random selection from $P(L)$ of a link and eventual updating of \bar{P} . Repetition of this step until the number prescribed by I_4 is reached.

Every time a link is selected, a test of feasibility is performed to verify that the number of activities not yet receiving a link from the previous level is not higher than the number of links with level length equal to one that should be added to the network to obtain $n(1)$.

If the answer to this test is negative, then the last link generated is deleted and the test is repeated.

If the answer remains negative, another deletion is introduced. After obtaining a positive answer, the process of introducing a new link goes on.

It should be noted that there is no risk of redundancy with links having $L = 1$ and therefore the redundancy check is just carried out for $L > 1$.

5 - After completing the generation of links with level length equal to one, the links with level length equal to two, three, ... V are generated using a similar procedure:

- generation of $P(L)$
- random selection of a link from $P(L)$
- checking of non-redundancy until the prescribed number of links for $L = 2, \dots, V$ is reached.

If no set of non-redundant links of level length L is found fulfilling $n(L)$, a recursive search is developed starting at the previous level length $(L-1)$ until a feasible solution is achieved or until a maximal time limit is reached

A software program - NETGEN - was developed in Microsoft Visual C++ 2.2 and most examples with 50-100 activities are generated in less than one second using a Pentium 166 MHz. Networks with 1000 activities are generated in less than 15 seconds. The full set of generated nets, can be found in the site <http://members.nbci.com/jcoelho72/psp/index.html> in the Kolish format, (Kolish and Sprecher, 1996).

Also, another program - NETVIEW - was developed to produce a graphical representation of the generated network having the objective of being similar to a network drawn by using human intuition.

The NETVIEW software is based on the following procedures:

- the activities are represented by nodes and they are located in terms of their progressive level being equally spaced in each level along a central strip (S) of the plane. Every node is represented by a circle except the starting node and those nodes not preceding any other activity which are represented by a small square. The final node is not represented.
- a specific indicator, π_i , measuring for each activity, i , the degree of difficulty for drawing its link is computed by: $\pi_i = \sum_{L=1}^V n_i(L) \cdot 2^{2 \cdot (L-1)} \cdot i$ where $n_i(L)$ is the number of links including i with length $L = 1, 2, \dots, V$. (excepting those connecting the activity to the final node). This function is adopted because a longer link is more difficult to be drawn without confusion with other links than a shorter one.
- at each level, the nodes are ordered in terms of π_i allocating an activity to each node from bottom up. This procedure guarantees that the most difficult nodes will be located near the upper border which is convenient to reduce the number of long links crossing most of the network.
- the links with length equal to one are represented by bold full straight lines;
- the links longer than one are represented by curves distributed on strips of space outside S . A link with length $L' > L$ will go through a strip more peripheral than the

curve representing the link with length equal to L . These curves are described by B-splines (FOLEY, et al, 1982) and the following conventions are used:

- full line ($L = 2$)
 - thin full line ($L \geq 3$)
- the links connecting each activity with the final node are not represented as they are not required to understand the network.

NETGEN and NETVIEW are now included in a more powerful software to study project networks, RISKNET (Tavares and Coelho, 1998).

An example is given for

$$I_1 = 14; I_2 = 0.38; I_3 = 0.14; 0.22; 0.28; 0.22; 0.14; \text{ for } a = 1; 2; 3; 4; 5, \text{ respectively};$$

$$I_4 = 0.00; I_5 = 0.42; I_6 = 0.60.$$

It should be noted that the computed $W(a)$ have to be rounded to have an integer number of activities of each level and to keep $\sum W(a) = N$.

The generation time was less than one second. The obtained results are presented in Table 1 and in Figure 7.

A collection of 802 test problems was generated with $I_1 = 25; 50; 100; 250; 500; 1000;$
 $I_2 = 0,05; 0,1; 0,2; I_3^* = 0,8; 0,5; 0,2; I_4 = 0,1; 0,2; 0,4; I_5 = 0,1; 0,15; 0,2; I_6 = 0,1; 0,5$

Activity	Progressive Level	Regressive Level	Immediate Precedent Activities
0	0	0	
1	1	1	0;
2	1	2	0;
3	2	2	1;
4	2	3	1;
5	2	3	2;
6	3	3	2; 3;
7	3	4	4;
8	3	5	3;
9	3	3	3;
10	4	4	4; 5; 6;
11	4	4	9;
12	4	5	2; 7;
13	5	5	7; 9; 10;
14	5	5	2; 4; 11;

Table 1 - Features of the Generated Network with 14 activities

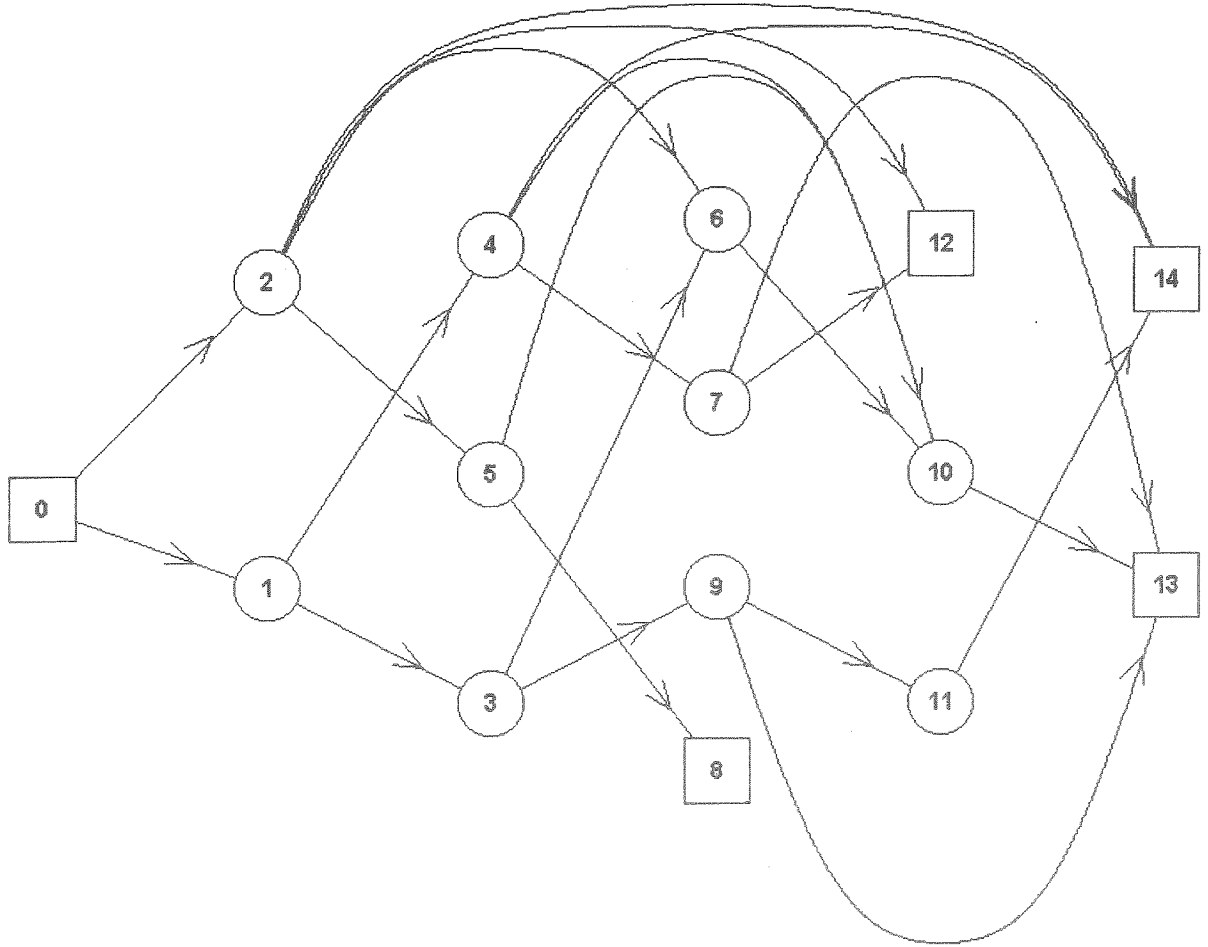


Figure 7 - The Network of Table 1

Conclusions

The concept of hierarchical level was shown to be particularly useful to build up a set of six indicators describing the morphological features of the project network.

The proposed method to generate networks preserves such features and it is easily implemented. The major advantages of this method over the previous proposals are the following ones:

- not requiring as input the number of nodes in the AoA representation which is not easily given by the user;
- avoiding the risk of generating cyclical networks;
- allowing the specification of six indicators describing the morphology of the network instead of just two;
- being easily implemented and producing large networks (N=1000) in less than 15 seconds (Pentium 166 MHz).

The first test set including large networks (from $N = 25$ to $N = 1000$) was produced and it is available from the authors.

The software developed to produce graphical representations is also particularly efficient to draw complex networks in small areas.

The presented examples illustrate the use of the proposed method.

References

- [1] Agrawal, M. K., Elmaghraby, S. E. and Herroelen, W. S., DAGEN a generator of testsets for project activity nets, *European Journal of Operational Research* 90 (1996) 376-382.
- [2] Ahn, T. and Erenguc, S. S., The resource constrained project scheduling problems with multiple crashable modes: a heuristic procedure, *European Journal of Operational Research* 107 (1998) 250-259.
- [3] Battersby, A., *Network analysis for planning and scheduling*, Macmillan, London (1967).
- [4] Bein, W. W., Kamburowski, J. E. and Stallmann, M. M., Optimal reduction of two-terminal directed acyclic graph, *SIAM Journal on Computing* 21 (1992) 1112-1129.
- [5] Carlier, J. and Tavares, L. V., Editorial, *European Journal of Operational Research* 64 (1993) 153-155.
- [6] Clark, C. E., The PERT model for the distribution of an activity time, *Operations Research* 10 (1962) 405-406.
- [7] Demeulemeester, E. and Herroelen, W., A branch-and-bound procedure for the multiple resource-constrained project scheduling problem, *Management Science* 38 (1992) 1803-1818.
- [8] Demeulemeester, E., Dodin, B. and Herroelen, W., A random activity network generator, *Operations Research* 41 (1993) 972-980.
- [9] Elmaghraby, S. E., *Activity networks: project planning and control by network models*, Wiley (1997).
- [10] Elmaghraby, S. E., Activity nets: A guided tour through some recent developments, *European Journal of Operational Research* 82 (1995) 383-408.
- [11] Foley, J. D. and Van Dams, A., *Fundamentals of Interactive Computer Graphics*, Addison Wesley (1982).
- [12] Kolisch, R. and Sprecher, A., PSPLIB - a project scheduling problem library, *European Journal of Operational Research* 96 (1996) 205-216.
- [13] Kolisch, R., Sprecher, A. and Drexel, A., Characterization and generation of a general class of resource-constrained project scheduling problems, *Management Science* 41 (1995) 1693-1703.
- [14] Kolisch, R., Schwindt, C. and Sprecher, A., Benchmark instances for project scheduling problems, in J. Weglarz, ed., *Handbook on recent advances in project scheduling*, Kluwer, Amsterdam (1999).
- [15] Patterson, J. H., Slowinski, R., Talbot, B., and Weglarz, J., Computational experience with a backtracking algorithm for solving a general class of precedence and resource-constrained scheduling problems, *European Journal of Operational Research* 49 (1990) 68-79.
- [16] Patterson, J.H., A comparison of exact approaches for solving the multiple constrained resource, project scheduling problem, *Management Science* 30 (1984) 854-867.
- [17] Schwindt, C., Generation of resource-constrained project scheduling problems with minimal and maximal time lags, Technical Report n° 489, Institut für Wirtschaftstheorie und Operations Research, Universität Karlsruhe (1996).
- [18] Sprecher, A. and Drexel, A., Multi-mode resource-constrained project scheduling by a simple, general and powerful sequencing algorithms, *European Journal of Operational Research* 107 (1998) 431-450.
- [19] Tavares, L. V., A multistage non-deterministic model for project scheduling under resources constraints, *European Journal of Operational Research* 49 (1994) 92-101.
- [20] Tavares, L. V., A review on the contributions of Operational Research to Project Management, *EURO XIV* (1995) 67-82.
- [21] Tavares, L. V. and Weglarz, J., Editorial, *European Journal of Operational Research* 49 (1990) 1-2.
- [22] Tavares, L. V. and Vercellis, C., Editorial, *European Journal of Operational Research* 78 (1994) 145.
- [23] Tavares, L. V. and Coelho, J. S., Risk Net, research report, CESUR Lisbon (1998).
- [24] Thesen, A., Measures of the restrictiveness of project networks, *Networks* 7 (1977) 193-208.

MODELO DE SIMULAÇÃO PARA A GESTÃO DE UM SISTEMA MULTI-ESCALÃO DE ARTIGOS REPARÁVEIS

Maria Elisa Cunha

Universidade Autónoma de Lisboa
R. de Santa Marta, 56
1169-023 Lisboa

José Álvaro Assis Lopes

Ana Paula Barbosa Póvoa

Centro de Estudos de Gestão
SAEG - Instituto Superior Técnico
Av. Rovisco Pais
1049-001 Lisboa

Abstract

A lot of multi-echelon inventory models for repairable items have been developed for three decades ago. However, even the most recent models have restrictions to their application. To focus some of those restrictions a simulation model is developed in this paper. The model is based on Díaz and Fu (1997) and represents a two-echelon inventory system, one base and one depot, for repairable items whose repair takes place, only, at the depot. At the base, a limited number of working items and a Poisson failure process are assumed. At the depot, a limited repair capacity and an exponential distributed repair time is considered.

Two random variables are characterized, the number of failure items at the base waiting for the transport to the depot, N_1 , and the number of failure items in the repair facility, N_2 . A negative binomial distribution is used to describe the variables $N_1 + N_2$. The simulation model results are compared to the ones present by Díaz and Fu (1997).

Resumo

A gestão de artigos reparáveis em estruturas multi-escalão tem sido alvo de estudo desde há cerca de três décadas. Todavia, nos modelos desenvolvidos, até à data, persistem algumas limitações, abordadas no presente trabalho através do desenvolvimento de um modelo de simulação.

Tendo por base o artigo de Díaz e Fu (1997) desenvolveu-se um modelo de simulação representativo de um sistema de gestão com dois escalões, uma base e um depósito, para artigos reparáveis, cuja reparação só ocorre no depósito. Os artigos em utilização, em número limitado, avariavam segundo uma distribuição de Poisson. No centro de reparação, com capacidade limitada, o tempo de reparação segue uma distribuição exponencial.

Pretende-se determinar as características de duas variáveis aleatórias, número de artigos à espera de transporte para o depósito e número de artigos no centro de reparação, bem como ajustar uma distribuição Binomial Negativa à variável aleatória obtida pela soma das duas anteriores. É resolvido um exemplo ilustrativo e os resultados obtidos, pelo modelo de simulação, são comparados com os apresentados por Díaz e Fu (1997).

Keywords

Simulation, Stochastic Models, Decision Aid Systems

1. Introdução

Pelo interesse científico e pela aplicação prática demonstrada, a gestão de artigos reparáveis em estruturas multi-escalão tem sido alvo de estudo desde há cerca de três décadas. Estas estruturas são consideradas cada vez mais importantes pelas organizações, quer industriais quer de prestação de serviço, devido ao realismo com que permitem tratar as interacções existentes entre as entidades que as constituem. Além disso, a operacionalidade das estruturas é dominada pelos artigos reparáveis que, no seu conjunto, são responsáveis pela maior parte do investimento imobilizado em "stock" na organização.

A gestão destas estruturas envolve um número de variáveis de decisão vasto. Contudo, pode-se dizer que o seu objectivo principal consiste na determinação do número de artigos a armazenar, em cada um dos locais da estrutura, de modo a satisfazer determinado nível de serviço e condicionantes de ordem económica, política ou estrutural da organização.

O primeiro modelo a ser desenvolvido, para dar resposta a este propósito, foi o METRIC (Multi-Echelon Technique for Recoverable Item Control, Sherbrooke (1968)), com imediata aplicação na Força Aérea dos E.U.A.. O objectivo do modelo era expresso por uma função, a minimizar, representando o número esperado de unidades cuja procura não pode ser satisfeita de imediato. Argumentando que este critério de optimização está mais relacionado com o reaprovisionamento do que com a operacionalidade da estrutura, objecto real do problema, Sherbrooke (1971) considerou um novo objectivo, a minimização de uma função que traduzia o número esperado de aviões não operacionais devido à falta de unidades armazenadas. Posteriormente, outros modelos foram desenvolvidos, uns relaxando hipóteses formuladas no METRIC outros com uma abordagem diferente ao problema. Simon (1971), assumindo que os tempos de transporte são determinísticos, propõe expressões para as funções de distribuição do "stock" físico, do número de unidades em reparação e do número de unidades cuja procura não pode ser imediatamente satisfeita. Mais tarde Muckstadt (1973), relaxando a hipótese do METRIC de que os artigos são artigos finais, apresenta o modelo MOD-METRIC. Este modelo determina os valores a armazenar em cada um dos locais da estrutura considerando o relacionamento hierárquico dos artigos finais e das suas componentes. Slay (1984), alegando que, nos casos práticos, a variância da procura é muito superior à média, relaxa a hipótese do METRIC de que a procura segue uma distribuição de Poisson. Desenvolve o modelo VARI-METRIC, generalizando a distribuição de probabilidade da procura. Tendo por base o METRIC, Graves (1985) propõe uma nova distribuição para o número de artigos encomendados. Gross, Kioussis e Miller (1987) modelizam a estrutura multi-escalão recorrendo à teoria das cadeias de Markov considerando um número finito de estados. Para determinar as distribuições de probabilidade estacionárias consideram o método de decomposição, da estrutura inicial em subestruturas. Díaz e Fu (1997) relaxam a hipótese formulada no METRIC de que o centro de reparação tem capacidade ilimitada e propõem a utilização de resultados gerais das filas de espera.

Todavia, apesar do trabalho desenvolvido, ainda hoje se procuram modelos alternativos que melhor se adaptem aos sistemas reais. Esta pesquisa deve-se sobretudo a dois tipos de limitações nos modelos existentes. A primeira diz respeito às características demasiado restritivas dos pressupostos, que impedem que os sistemas reais sejam adequadamente modelizados. A segunda relaciona-se com a natureza aplicacional do modelo pois, para que seja facilmente aplicável, é fundamental que o modelo seja pouco complexo quer em termos de compreensão quer em termos de implementação. Não nos podemos esquecer que, devido à dependência existente entre os elementos que constituem a estrutura multi-escalão (por mais simples que seja), o modelo matemático associado tende a ser complicado pois tem de determinar, simultaneamente, todos os parâmetros de controlo.

O presente trabalho tem como objectivo principal superar algumas dessas limitações. Para isso, tendo por base o modelo proposto por Díaz e Fu (1997), desenvolveu-se um modelo de simulação no qual se modeliza um sistema de gestão com dois escalões, uma base e um depósito, para artigos reparáveis, cuja reparação só ocorre no depósito. Os artigos em utilização, em número limitado, avariam segundo uma distribuição de Poisson. No centro de reparação, com capacidade limitada, o tempo de reparação segue uma distribuição exponencial. Os resultados provenientes do modelo de simulação são comparados com os propostos por Díaz e Fu (1997), constatando-se alguma divergência.

Este documento está organizado da seguinte forma. Na secção 2 é descrito o problema multi-escalão e é feita uma breve apresentação dos modelos que se consideraram mais relevantes para o desenvolvimento deste trabalho. Na secção 3 faz-se a caracterização do modelo de simulação desenvolvido. Na secção 4 faz-se a comparação dos resultados obtidos através do modelo de simulação com os propostos por Díaz e Fu (1997). Na secção 5 apresentam-se as conclusões e indica-se o desenvolvimento futuro.

2. O Problema da Gestão de Artigos Reparáveis em Estruturas Multi-Escalão

As estruturas multi-escalão cuja operacionalidade é dominada pelos artigos reparáveis são constituídas, na maioria dos casos práticos, por dois escalões. Por este motivo, os modelos existentes foram desenvolvidos sob o pressuposto de que o sistema é composto por dois escalões. O escalão de nível mais baixo, designado por base, é composto por n elementos; o escalão de nível mais alto, designado por depósito, é composto por um só elemento.

Atendendo à natureza dos artigos reparáveis (procura pouco frequente, custo unitário elevado, tempo de reaprovisionamento longo), a política de gestão utilizada, quer para o envio de artigos para reparação quer para a encomenda de artigos a reutilizar, é do tipo $(s - 1, s)$. Este conceito simplifica a análise do problema permitindo que o processo da procura seja idêntico nas bases e no depósito.

Genericamente, um sistema com dois escalões pode ser descrito como se segue. Quando um artigo em utilização avaria é levado para a base, à qual está associado, e é solicitado ao armazém outro para o substituir. Se existirem, no armazém da base, artigos prontos a reutilizar,

procede-se à satisfação do pedido; caso contrário, o pedido fica a aguardar até que na base existam artigos nessa condição, dizendo-se que o pedido é "backorder". A reparação dos artigos pode ser efectuada na base, com uma probabilidade r , ou no depósito, com uma probabilidade $(1 - r)$. Sempre que um artigo é enviado para o depósito, a base encomenda-lhe um outro, equivalente, que possa ser imediatamente utilizado. Se o depósito possuir, em armazém, artigos para satisfazer o pedido, procede de imediato ao seu envio; caso contrário, o pedido fica a aguardar até que no depósito existam artigos reparados, sendo "backorder". Terminada a reparação, os artigos são colocados no armazém do local onde se efectuou a reparação ou, caso existam pedidos suspensos nesse local, alguns irão de imediato ser satisfeitos. Como objectivo final, pretende-se determinar o número de artigos a manter em armazém, quer na base quer no depósito, de modo a atingir determinada medida de desempenho do sistema, satisfazendo condicionantes de ordem económica, política ou estrutural.

Da pesquisa bibliográfica referida na secção 1., vamos focar o nosso artigo nos trabalhos de Sherbrooke (1968), Graves (1985) e Díaz e Fu (1997). O modelo proposto pelos últimos autores, base do modelo de simulação por nós desenvolvido, representa uma evolução no modelo de Graves que, por sua vez, constitui uma evolução do modelo proposto por Sherbrooke.

Os três modelos partem dos seguintes conceitos genéricos: a procura, quer nas bases quer no depósito, segue uma distribuição de Poisson; os artigos são sempre reparáveis, mas a sua reparação só se efectua no depósito; os artigos são considerados artigos finais, isto é, não é considerado o relacionamento dos artigos com as suas componentes. Admitem ainda o pressuposto, simplificador e fundamental para a forma como tratam o problema, de que os processos de transporte das bases para o depósito, de reparação no depósito e de transporte do depósito para as bases são independentes.

Sherbrooke (1968) aplica o teorema de Palm das filas de espera, sob os pressupostos de que o número de artigos é tão grande que pode ser considerado infinito e que o centro de reparação tem capacidade ilimitada. Nestas condições, as quantidades encomendadas são idênticas ao nível de ocupação duma fila de espera do tipo $M/G/\infty$. Assim, para determinar o número de artigos a armazenar no depósito ajusta uma distribuição de Poisson ao número de artigos em reparação no depósito; para determinar o número de artigos a armazenar na base i ajusta uma distribuição de Poisson à variável aleatória número de artigos "backorders" no depósito pertencentes à base i .

Graves (1985) utiliza o mesmo ajustamento que Sherbrooke (1968) para calcular o número de artigos a armazenar no depósito mas para a determinação do número de artigos a armazenar nas bases propõe o ajustamento duma distribuição Binomial Negativa. Desde que o depósito satisfaça as encomendas utilizando a filosofia "primeira a chegar, primeira a ser satisfeita", os resultados obtidos a partir deste ajustamento são mais favoráveis do que os obtidos utilizando a distribuição de Poisson.

Mais tarde, Díaz e Fu (1997) utilizam o mesmo ajustamento que Graves (1985) para determinar o número de artigos a armazenar nas bases mas propõem que, também, se ajuste uma distribuição Binomial Negativa para o cálculo dos artigos a armazenar no depósito. Os autores sugerem que os dois primeiros momentos desta distribuição sejam obtidos recorrendo a resultados gerais da teoria das filas de espera, fazendo o seguinte raciocínio:

Modelizam o processo de transporte de cada base para o depósito como sendo uma fila do tipo $M/G/\infty$. Deste modo, o número de artigos a serem transportados, representados por N_1 , coincide com a taxa de ocupação da fila, verificando-se a igualdade seguinte:

$$E[N_1] = \text{Var}[N_1] = \sum_i \lambda_i L_{i0} \quad (1)$$

onde λ_i representa a média de avarias na base i e L_{i0} o tempo de transporte da base i para o depósito. O número de clientes que saem deste sistema constitui um processo de Poisson independente da taxa de ocupação.

Numa das situações consideradas no documento, Díaz e Fu (1997) admitem que o centro de reparação, apenas para um tipo de artigo, é modelizado por uma fila do tipo $M/M/k$. Representando por N_2 o número de artigos no centro de reparação, aplicando resultados gerais da teoria das filas de espera tem-se que:

$$E[N_2] = k\rho + \frac{\rho(k\rho)^k}{(1-\rho)^2 k!} p_0 \quad (2)$$

$$\text{Var}[N_2] = k\rho \left(1 + \frac{(k\rho)^k}{(1-\rho)k!} p_0 \right) + \frac{\frac{\rho(k\rho)^k}{(1-\rho)k!} p_0 \left[1 + \rho \left(1 - \frac{(k\rho)^k}{(1-\rho)k!} p_0 \right) \right]}{(1-\rho)^2} \quad (3)$$

com

$$p_0 = k\rho \frac{1}{\sum_{n=0}^{k-1} \frac{(k\rho)^n}{n!} + \frac{(k\rho)^k}{(1-\rho)k!}} \quad (4)$$

$$\rho = \frac{\lambda}{k\mu} \quad (5)$$

Finalmente, sugerem que o primeiro e o segundo momento da distribuição, que permite determinar o número de artigos a armazenar no depósito, sejam aproximados por $E[N_1] + E[N_2]$ e $\text{Var}[N_1] + \text{Var}[N_2]$, respectivamente.

As vantagens dos ajustamentos propostos por Díaz e Fu (1997) tornam-se evidentes quando a taxa de ocupação do centro de reparação é elevada.

É com base neste último trabalho que vamos iniciar o nosso estudo.

3. Descrição do Modelo de Simulação

De forma a ultrapassar algumas das desvantagens ainda persistentes no tratamento do problema da gestão de artigos reparáveis vamos neste trabalho, e com base no sistema descrito por Díaz e Fu (1997), desenvolver um modelo de simulação. Assim, considerou-se que:

- existem dois escalões; no escalão de nível mais baixo existe uma base e no escalão de nível mais elevado existe um depósito;
- ambos os escalões têm armazenados artigos reutilizáveis;
- todos os artigos em utilização estão associadas à base;
- os artigos em utilização avariam segundo uma distribuição de Poisson;
- só existe reparação no depósito;
- o tempo de reparação tem distribuição exponencial;
- quando um artigo acaba de ser reparado, é colocado no armazém do depósito;
- a política de gestão utilizada quer no envio de artigos avariados para o depósito quer na encomenda de artigos reutilizáveis é do tipo $(s - 1, s)$;
- quando um artigo em utilização avaria são desencadeados três procedimentos:
 - * primeiro, o artigo é retirado de operação sendo substituído por outro retirado do armazém da base; se neste não existirem artigos o pedido é definido como "backorder" e o número de artigos em operação fica reduzido de uma unidade;
 - * segundo, o artigo é enviado para o depósito, para aí ser reparado;
 - * terceiro, solicita-se um artigo ao armazém do depósito, para reabastecer o armazém da base; se no depósito não existirem artigos o pedido é definido como "backorder".

Na Figura 1 está representado o modelo de simulação que se desenvolveu utilizando o "software" EXTEND.

Optou-se por um modelo de simulação discreto no qual o tempo avança em função da ocorrência dos acontecimentos. Estes podem ser a ocorrência de avarias, o final da reparação de um artigo ou o seu transporte, entre a base e o depósito ou vice-versa. O modelo é composto por componentes interligadas entre si, designadas por blocos. O movimento dos artigos no sistema é sugerido pelas linhas orientadas que ligam os blocos. A cada bloco, identificado pelo seu nome (localizado na parte inferior do mesmo), está associada uma tarefa que se descreve em seguida.

Executivo: Bloco fundamental a todo o modelo de simulação discreto; permite controlar o tempo de duração da simulação.

PeçasOperação: Representa o conjunto de artigos em operação. Os artigos quando avariam "saem" pelo conector da direita (que liga este bloco ao bloco TiraPeçasOperação). Quando é solicitado um artigo ao armazém da base, ele é colocado em operação através do conector superior esquerdo (que estabelece a ligação deste bloco com o bloco DiminuiStockBase).

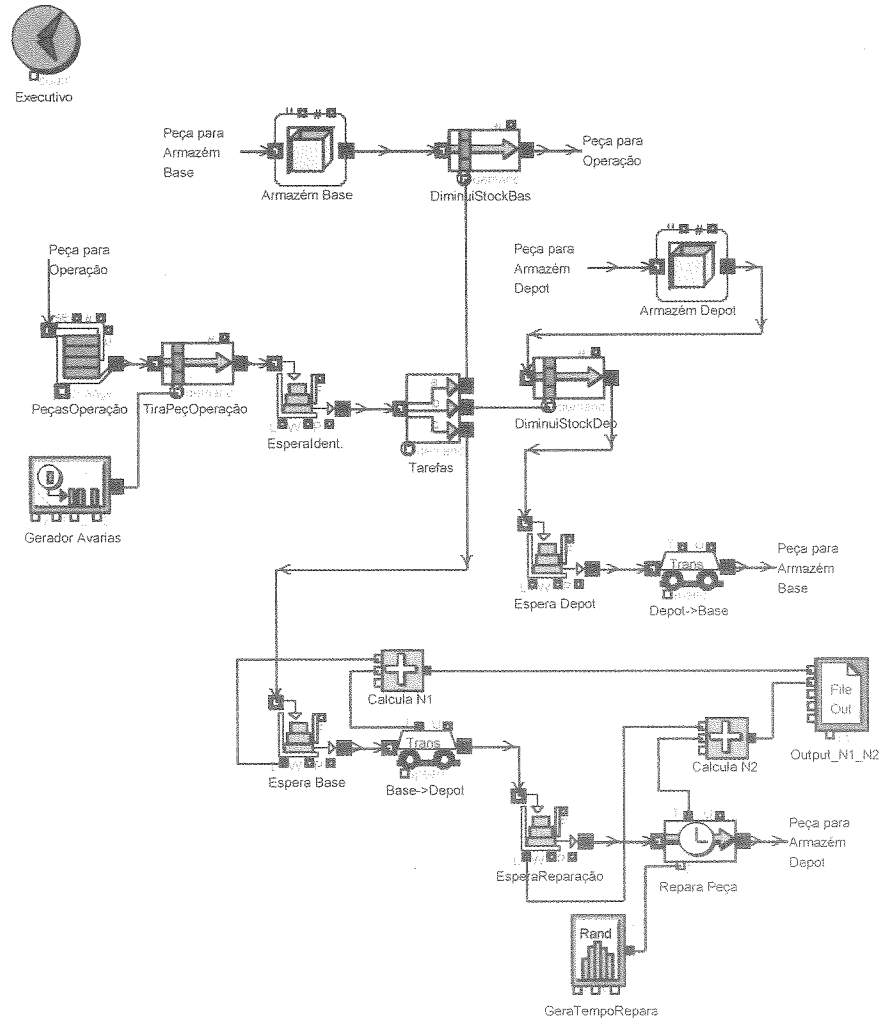


Figura 1 - O modelo de simulação

Gerador Avarias: Gera avarias, aleatoriamente, nos artigos em operação, segundo uma distribuição de Poisson. A ordem é dada a partir do conector direito.

TiraPeçaOperação: Sempre que é gerada uma avaria, a informação entra pelo conector inferior (que liga este bloco ao bloco Gerador Avarias). Este bloco diminui o número de artigos em operação, através do conector esquerdo. O artigo avariado "sai" pelo conector direito deste bloco sendo transferido para o bloco seguinte EsperaIdent.

EsperaIdent: Representa uma fila de espera onde são colocados os artigos avariados. Estes "entram" na fila pelo conector superior esquerdo e "saem" pelo conector da direita sempre que o bloco Tarefas solicite um artigo. Isso é feito automaticamente após o bloco Tarefas ter finalizado o encaminhamento do artigo anterior. Esta fila com a disciplina FIFO, por hipótese, é necessária porque o bloco Tarefas não tem capacidade para armazenar artigos.

Tarefas: Representa os três procedimentos executados sempre que existe uma avaria, reconhecida pela entrada de informação através do conector da esquerda. O conector superior direito corresponde ao pedido, efectuado ao armazém da base, para que um artigo seja colocado

em operação; o conector do meio corresponde ao pedido de um artigo ao armazém do depósito; finalmente, o conector inferior direito corresponde ao envio do artigo avariado para o depósito, para aí ser reparado.

Armazém Base: Representa o armazém da base ao qual são solicitados artigos para substituir os que avariaram em operação. O artigo solicitado "sai" pelo conector da direita (que liga este bloco ao bloco *DiminuiStockBas*). Quando um artigo é solicitado ao armazém do depósito ele é colocado no armazém da base pelo conector da esquerda.

DiminuiStockBas: Este bloco controla os pedidos ao armazém da base. Se existe pelo menos um artigo armazenado, informação obtida através do conector esquerdo, o bloco diminui uma unidade ao número de artigos armazenados. O artigo é colocado em operação através do conector direito, com ligação ao bloco *PeçasOperação* por meio do bloco "conector" *Peça para Operação*. Se não existem artigos armazenados o bloco regista o pedido, incrementando de uma unidade as "backorders" existentes na base. Neste caso, quando tiver sido recebido um artigo enviado pelo depósito, um dos pedidos suspensos é satisfeito, sendo o número de unidades "backorders" reduzido de uma unidade.

Armazém Depot: Representa o armazém do depósito ao qual são solicitados artigos para substituir aqueles enviados para reparação. O artigo solicitado "sai" pelo conector da direita (que liga este bloco ao bloco *DiminuiStockDep*). Quando um artigo é reparado, é colocado no armazém do depósito pelo conector da esquerda.

DiminuiStockDep: Este bloco controla os pedidos ao armazém do depósito. Se existe pelo menos um artigo armazenado, informação obtida através do conector esquerdo, este bloco decreta, de uma unidade, o número de artigos armazenados. O artigo será colocado numa fila de espera, *Espera Depot*, enquanto aguarda transporte para o armazém da base. Se não existem peças armazenadas no armazém do depósito, o bloco regista o pedido, incrementando de uma unidade as "backorders" existentes no depósito. Neste caso, quando a reparação de um artigo estiver terminada, um dos pedidos suspensos é satisfeito, sendo o número de unidades "backorders" diminuído de uma unidade.

Espera Depot: Representa uma fila de espera onde são colocados os artigos a transportar para a base. Admite-se que a disciplina da fila é FIFO. Os artigos "entram" na fila pelo conector superior esquerdo e "saem" pelo conector da direita quando o bloco *Depot→Base* está disponível para efectuar o transporte.

Depot→Base: Representa o veículo que transporta os artigos do depósito para a base. Sempre que o veículo está livre, é retirado um artigo da fila de espera, *Espera Depot*. Admite-se que o veículo só tem capacidade para transportar um artigo. Este "entra" para o veículo pelo conector esquerdo e, depois de chegar à base, "saí" pelo conector direito, sendo colocado no armazém da base.

Espera Base: Representa uma fila de espera onde são colocados os artigos avariados que esperam ser transportadas para o depósito. Admite-se que a disciplina da fila é FIFO. Os artigos

"entram" na fila pelo conector superior esquerdo e "saem" pelo conector da direita quando o bloco Base→Depot está disponível para efectuar o transporte.

Base→Depot: Representa o veículo que transporta os artigos avariados da base para o depósito. Sempre que o veículo está livre, é retirado uma peça da fila de espera, Espera Base. Admite-se que o veículo só tem capacidade para transportar um artigo. Este "entra" para o veículo pelo conector esquerdo e, depois de chegar ao depósito, "saí" pelo conector direito sendo colocado numa fila de espera, EsperaReparação, até que exista um técnico disponível para o reparar.

EsperaReparação: Representa uma fila de espera onde são colocados os artigos avariados que esperam ser reparados. Os artigos "entram" na fila pelo conector superior esquerdo e "saem" pelo conector da direita quando o bloco ReparaPeça solicita algum. Admite-se que a disciplina da fila é FIFO.

ReparaPeça: Representa o centro de reparação no qual existe apenas um técnico. Quando o técnico está livre, retira um artigo da fila EsperaReparação e inicia a sua reparação. O artigo fica retido, no bloco, o tempo necessário para que seja reparado. O tempo de reparação segue uma distribuição exponencial. Quando a reparação termina, o artigo "saí" do bloco pelo conector direito para ser colocado no armazém do depósito através do bloco "conector" Peças para Armazém Depot.

Calcula N1: Calcula a soma do número de artigos à espera de serem transportados para o depósito, indicado pelo conector L do bloco Espera Base, com o número de artigos a serem transportados da base para o depósito (zero ou um), informação obtida através do conector T do bloco Base→Depot.

Calcula N2: Calcula a soma do número de artigos à espera de serem reparados no depósito, indicado pelo conector L do bloco EsperaReparação, com o número de artigos a serem reparados no centro de reparação do depósito (zero ou um), informação obtida através do conector T do bloco Repara Peça.

Output_N1_N2: Representa um ficheiro onde são escritos, no final de cada passo de simulação, o número de artigos a serem transportados ou à espera de transporte da base para o depósito, que passaremos a designar por N_1 , e o número de artigos a serem reparados ou à espera de reparação no depósito, designado por N_2 .

Díaz e Fu (1997) apenas faz referência à elevada taxa de utilização do centro de reparação, não apresentando valores para os parâmetros do sistema. Desta forma, para o modelo de simulação foram arbitrados os seguintes valores:

- unidade de tempo: mês;
- unidade de comprimento: pé (imposta pelo software);
- número inicial de artigos em operação: 40 unidades;
- número inicial de artigos no armazém do depósito: 5 unidades;
- número inicial de artigos no armazém da base: 4 unidades;

- o número de avarias por unidade de tempo segue uma distribuição de Poisson de parâmetro 2;
- o tempo de reparação é Exponencial de parâmetro 0,4;
- distância entre a base e o depósito: 10.000 pés;
- distância entre o depósito e a base: 10.000 pés;
- velocidade de transporte: 142.105.263,2 pés/unidade de tempo

os quais foram utilizados para resolver o modelo.

4. Apresentação e Análise de Resultados

4.1. Resultados obtidos por simulação

Os resultados a seguir apresentados foram obtidos registando, ao fim de cada passo da simulação, o número de peças à espera de transporte ou a serem transportados para o depósito, N_1 , e o número de peças à espera de serem reparadas ou a serem reparadas, N_2 . A simulação foi efectuada durante um intervalo de tempo de 1800 meses, amplitude esta que resultou numa análise previamente efectuada à estacionariade da série das médias dos valores médios das referidas variáveis.

Dos 19.092 valores registados, quer para a variável número de peças avariadas na base à espera de transporte ou a serem transportadas para o depósito quer para a variável número de peças avariadas no depósito, para o cálculo dos valores esperados e das variâncias de N_1 , N_2 e N_1+N_2 , apenas foram considerados os 10.000 últimos valores. Admitiu-se que durante os primeiros valores o sistema tenderia a atingir o estado de equilíbrio. Os valores obtidos foram os seguintes:

$$E[N_1] = 0,1884$$

$$\text{Var}[N_1] = 0,391306^2 = 0,1531$$

$$E[N_2] = 4,3090$$

$$\text{Var}[N_2] = 3,545924^2 = 12,5739$$

$$E[N_1+N_2] = 4,4974$$

$$\text{Var}[N_1+N_2] = 3,526732^2 = 12,4378$$

Como $\text{Var}[N_1] + \text{Var}[N_2] = 12,7270 \neq \text{Var}[N_1+N_2]$ pode-se imediatamente concluir que as variáveis aleatórias N_1 e N_2 não são linearmente independentes.

Tente-se caracterizar a variável aleatória N_1+N_2 para se inferir sobre a sua distribuição de probabilidade. Foi utilizado o programa SPSS, para determinar as estatísticas relevantes à caracterização da distribuição, elaborar a tabela de frequências, desenhar o histograma e o "box plot".

Na Tabela 1 encontram-se anotadas as frequências observadas da variável, na Tabela 2 estão registadas as estatísticas que caracterizam a variável, no Gráfico 1 o seu histograma e no Gráfico 2 o "box plot".

N ₁ +N ₂			
Valores Observados	Frequência Absoluta	Frequência Relativa	Freq. Relativa Acumulada
0	396	4,0	4,0
1	1882	18,8	22,8
2	1460	14,6	37,4
3	1204	12,0	49,4
4	889	8,9	58,3
5	971	9,7	68,0
6	730	7,3	75,3
7	542	5,4	80,7
8	453	4,5	85,3
9	394	3,9	89,2
10	324	3,2	92,5
11	260	2,6	95,1
12	171	1,7	96,8
13	108	1,1	97,8
14	81	,8	98,7
15	69	,7	99,3
16	45	,5	99,8
17	18	,2	100,0
18	3	,0	100,0
Total	1000	100,0	---

Tabela 1 - Tabela de frequências para a variável aleatória N_1+N_2

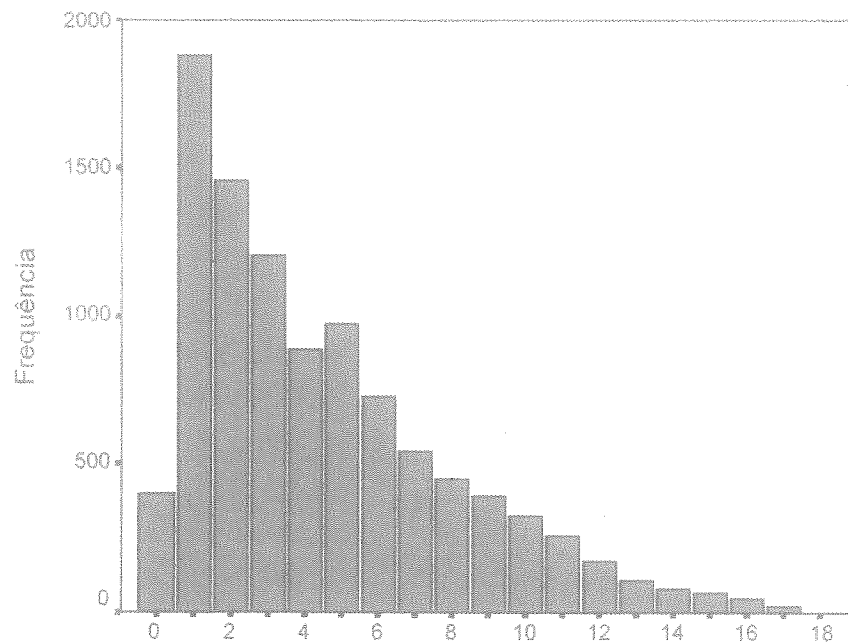
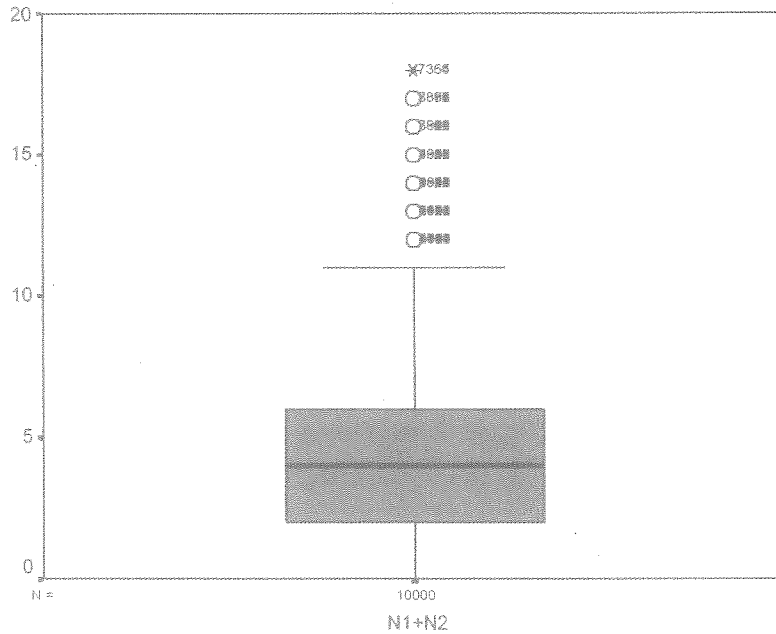


Gráfico 1 - Histograma da variável aleatória N_1+N_2

Gráfico 2 - "Box Plot" da variável aleatória N_1+N_2

Estatísticas da variável N_1+N_2	Estatísticas	Desvio Padrão
Média	4,50	3,53E-02
95% Intervalo de Limite Inf. Confiança para a média	4,43	
Limite Superior	4,57	
5% "Trimmed Mean"	4,23	
Mediana	4,00	
Variância	12,438	
Desvio Padrão	3,53	
Valor Mínimo	0	
Valor Máximo	18	
Intervalo	18	
Intervalo Interquartil	4,00	
"Skewness"	1,021	,024
"Kurtosis"	,529	,049

Tabela 2 - Estatísticas para caracterizar a distribuição da variável aleatória N_1+N_2

Analisando as tabelas e os gráficos conclui-se que a distribuição da variável em estudo:

- é assimétrica positiva (coeficiente de assimetria, "skewness", maior do que zero);
- é platicúrtica (curtose inferior a três);
- a moda é maior do que zero;

- tem variância muito superior à média, logo a medida designada por "lexis ratio", dada por $\tau = \sigma^2/\mu$ e equivalente ao coeficiente de variação nas distribuições contínuas, é superior a um.

Atendendo a estas características, ajustou-se uma distribuição Binomial Negativa à amostra. Utilizando o método da máxima verosimilhança, obtiveram-se os valores $\hat{r} = 3$ e $\hat{p} = 0,4$ para os parâmetros, r e p , da distribuição.

Utilizou-se o teste do χ^2 para testar a aderência entre a distribuição teórica, BN(3; 0,4), e a distribuição observada na amostra. O valor obtido para a estatística, dada a sua elevada ordem de grandeza, permitiu concluir que o ajustamento não é o mais adequado. Os valores da variável aleatória N_1+N_2 para os quais existe maior discrepância entre a frequência observada e a teórica são os primeiros, como se pode verificar no gráfico seguinte.

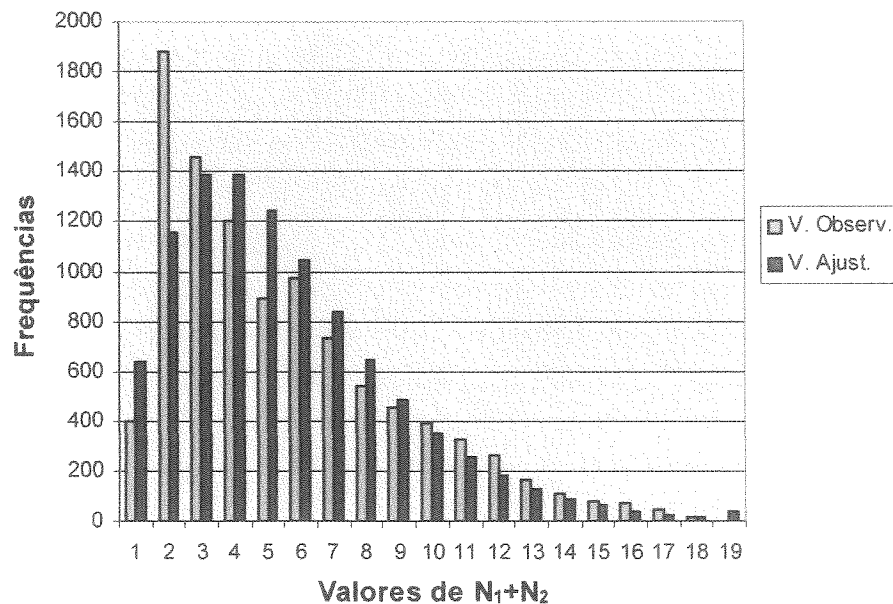


Gráfico 3 - Distribuição dos valores observados vs. distribuição BN(3; 0,4)

4.2 Resultados obtidos utilizando as expressões de Díaz e Fu (1997)

Considerando os valores arbitrados para os parâmetros do modelo e efectuando a sua substituição nas expressões (1), (2) e (3) tem-se que:

$$E[N_1] = 2 \cdot 0,00007037 = 0,00014074$$

$$\text{Var}[N_1] = 2 \cdot 0,00007037 = 0,00014074$$

$$E[N_2] = 4$$

$$\text{Var}[N_2] = 20$$

4.3 Análise dos resultados

O valor de $E[N_1] = 0,00014074$ obtido pela expressão (1), coincide com a taxa de ocupação da fila de espera, $M/G/\infty$, que permitiu modelizar o processo de transporte. Pode-se assim dizer que representa uma percentagem do serviço total oferecido no sistema. De facto, atendendo aos valores estabelecidos para a distância entre a base e o depósito, 10.000 pés, e a

velocidade de transporte, 142.105.263,2 pés/mês, pode-se determinar o número médio de viagens, mensais, que cada veículo tem possibilidade de efectuar. Este valor, igual a 14.210,5263 viagens/mês, corresponde à taxa do serviço de transporte, ou seja, μ . O seu inverso, cujo valor é 0,00007037, representa o tempo médio de transporte. Como o número médio de avarias mensais é igual a 2, valor de λ , então a taxa de ocupação do sistema é igual a $\rho = \lambda/\mu = 0,000140741$.

O valor de $E[N_1] = 0,1885$, obtido por simulação, representa o número médio de artigos a serem transportados da base para o depósito. Atendendo a este valor, conclui-se que em média, por mês, se faz uma viagem entre a base e o depósito, logo, $\lambda = 1$. Uma viagem, no modelo, corresponde a 20.000 pés, pois contabiliza-se a viagem de ida e a de regresso ao ponto de partida. Como em cada mês o veículo pode percorrer 142.105.263,2 pés, então poderá fazer 7.105,26316 viagens de ida e volta. Logo, $\mu = 7.105,26316$ e $1/\mu = 0,000140741$. Desta forma, a taxa de ocupação do sistema é igual a $\rho = \lambda/\mu = 0,000140741$.

Conclui-se, assim, que os valores obtidos para $E[N_1]$ são diferentes por representarem medidas distintas do sistema a que se referem. Contudo, do valor de $E[N_1]$ calculado por simulação pode-se obter o valor obtido através da expressão (1).

A expressão (1) traduz a igualdade entre a média e a variância. No modelo de simulação tem-se que $E[N_1] \neq \text{Var}[N_1]$, como se concluiu efectuando um teste ao valor esperado.

Os valores para o número médio de artigos no centro de reparação, $E[N_2]$, calculados pela expressão (2) e por simulação, 4 e 4,3090, respectivamente, são diferentes mas da mesma ordem de grandeza. O mesmo não acontece para os valores de $\text{Var}[N_2]$.

Relativamente à distribuição da variável N_1+N_2 , concluiu-se que não é adequadamente ajustada pela distribuição $BN(3; 0,4)$.

5. Conclusões e Pesquisa Futura

Díaz e Fu (1997) sugeriram que a variável aleatória obtida pela soma do número de artigos a serem transportados ou à espera de transporte da base para o depósito com o número de artigos no centro de reparação do depósito, fosse ajustada por uma distribuição Binomial Negativa de parâmetros determinados utilizando resultados genéricos das filas de espera do tipo $M/G/\infty$ e $M/M/k$, através das expressões (1), (2) e (3).

No presente trabalho, desenvolveu-se um modelo de simulação representativo do sistema de gestão descrito em Díaz e Fu (1997) para determinar a média e a variância das variáveis, número de artigos a serem transportados ou à espera de transporte da base para o depósito, N_1 , e número de artigos no centro de reparação do depósito, N_2 , bem como da variável aleatória obtida pela soma das duas anteriores.

Os resultados obtidos por simulação foram comparados com os propostos por Díaz e Fu (1997) permitindo concluir o seguinte:

Os valores de $E[N_1]$ são diferentes por representarem medidas distintas do sistema; contudo, do valor obtido por simulação pode-se obter o valor calculado utilizando a expressão

(1). Os valores de $\text{Var}[N_1]$ são diferentes. Além disso, efectuado um teste ao valor esperado rejeita-se a hipótese da igualdade dos valores $E[N_1]$ e $\text{Var}[N_1]$ obtidos por simulação. Os valores de $E[N_2]$ e $\text{Var}[N_2]$ obtidos por simulação e pelas expressões (2) e (3), respectivamente, são diferentes embora se possa dizer que os valores médios são da mesma ordem de grandeza.

Relativamente ao ajustamento da distribuição Binomial Negativa à variável N_1+N_2 , obtida por simulação, concluiu-se que não é o mais adequado.

As diferenças encontradas podem ser motivadas pela dependência dos processos de avarias, de transporte e de reparação verificada no modelo de simulação. Lembremo-nos que em todos os modelos baseados no METRIC desenvolvidos até à data, e particularmente no de Díaz e Fu (1997), se parte da suposição de que estes processos são independentes.

Em futuros desenvolvimentos pretende-se definir o tipo de distribuições de probabilidade que melhor se ajustem à variável aleatória N_1+N_2 .

Paralelamente, as condições iniciais do modelo de simulação desenvolvido devem ser generalizadas, construindo modelos mais generalizados que melhor espelhem as situações práticas. A título de exemplo, pode-se admitir maior número de bases, que a reparação se efectue no depósito e nas bases, a possibilidade de transferência de artigos duma base com artigos em armazém para outra que não os possua. De forma complementar, também se pode admitir que a reparação dos artigos seja efectuada atendendo a regras de prioridades pré-estabelecidas, quer para as bases quer para os próprios artigos.

Referências

- [1] Díaz, A., Fu, M. C., *Models for multi-echelon repairable item inventory systems with limited repair capacity*, European Journal of Operational Research 97 (1997) 480-492.
- [2] EXTEND, *Imagine That Inc.*, Versão 4.1.1. (1998).
- [3] Feeney, G. J., Sherbrooke, C. C., *The (s-1, s) inventory policy under compound Poisson demand*, Management Science 12 (1966) 391-411.
- [4] Graves, S. C., *A multi-echelon inventory model for repairable item with one-for-one replenishment*, Management Science 31 (1985) 1247-1256.
- [5] Gross, D., Kioussis, L. C., Miller, D. R., *A network decomposition approach for approximating the steady state behaviour of Markovian multi-echelon repairable item inventory systems*, Management Science 33 (1987) 1453-1468.
- [6] Law, A. M., Kelton, W. D., *Simulation Modeling & Analysis*, Second Edition, McGraw-Hill International Editions, Singapore (1991).
- [7] Muckstadt, J. A., *A model for multi-item, multi-echelon, multi-indenture inventory system*, Management Science 32 (1973) 1567-1581.
- [8] Ross, S. M., *Introduction to probability models*, Sixth Edition, Academic Press, New York (1997).
- [9] Sherbrooke, C. C., *Optimal inventory modelling of systems: Multi-echelon techniques*, John Wiley & Sons, Inc, New York (1992).
- [10] Sherbrooke, C. C., *An evaluator for the number of operationally ready aircraft in a multi-level supply system*, Operations Research 19 (1971) 618-635.
- [11] Sherbrooke, C. C., *METRIC: A multi-echelon technique for recoverable item control*, Operational Research 16 (1968) 122-141.
- [12] Simon, R. M., *Stationary properties of a two-echelon inventory model for low demand items*, Operations Research 19 (1971) 761-773.
- [13] Slay, F. M., *VARI-METRIC: An approach to modelling multi-echelon resupply when the demand process is Poisson with a Gamma Prior*, Logistics Management Institute, Washington, D. C. Report AF301-3 (1984).
- [14] SPSS, *SPSS Inc. Versão 10.0.5* (1999).

DESIGUALDADES REGIONAIS EM PORTUGAL CONTINENTAL: UMA ANÁLISE ESTATÍSTICA MULTIVARIADA

Maria Manuela Lourenço Marquês

Departamento de Matemática
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa

João Oliveira Soares

CEG-IST/Secção Autónoma de Economia e Gestão
Instituto Superior Técnico

Maria Cristina V. Macedo da Cunha Coutinho

Departamento de Engenharia Civil
Instituto Superior de Engenharia de Lisboa

Abstract

The aim of this article is to find a coherent socio-economic classification of the Portuguese continental territory, in order to support regional development policy.

This classification is obtained through the use of a multivariate statistical method - cluster analysis -, and is based on a wide number of demographic, economic, health, education, employment and culture indicators, concerning the 275 municipalities into which the territory was divided in 1995.

The results lead to the identification of 4 regions with differing degrees of development, reflecting the well-known asymmetry between coastal and inland zones. This classification is much more useful for policy making than the usual division in administrative regions (NUT-2), which are geographically contiguous but clearly heterogeneous.

Resumo

Este artigo pretende encontrar uma classificação sócio-económica do território continental português que possa servir de base a uma política coerente de desenvolvimento regional.

Esta classificação é obtida com recurso a uma técnica de estatística multivariada - a análise de clusters -, e é baseada num alargado número de indicadores demográficos, económicos, de saúde, educação, emprego e cultura, respeitantes aos 275 concelhos em que se dividia o território em 1995.

Os resultados apontam para a identificação de 4 grandes zonas de diferente nível de desenvolvimento, reflectindo a conhecida assimetria litoral-interior. Esta classificação revela-se particularmente útil para a condução de políticas correctoras, quando confrontada com a habitual divisão em regiões administrativas (NUT-2), contíguas geograficamente mas claramente heterogéneas.

Keywords

regional studies; multivariate statistical analysis, clusters analysis

1. Introdução

Em Portugal Continental, os dois níveis fundamentais de autoridade político-administrativa são de âmbito nacional e concelhio. Entre estes encontramos as regiões administrativas - Norte, Centro, Lisboa e Vale do Tejo, Alentejo e Algarve -, que constituem a base da política regional e correspondem a 5 das 206 NUT-2 da União Europeia.

De acordo com os números publicados pela Comissão Europeia em 1999 [2], três destas regiões - Norte, Centro e Alentejo - estão entre as 25 mais pobres da União, com um PIB per capita (já calculado em termos de paridade do poder de compra) correspondente respectivamente a 62, 60 e 58 por cento da média comunitária de 1995. Esta percentagem eleva-se ligeiramente para o Algarve (70), mas só Lisboa e Vale do Tejo (89) excede o limiar de 75, que delimita as regiões prioritárias na atribuição de Fundos Estruturais (as chamadas regiões do 'objectivo 1' - ver [3]).

Esta questão da desclassificação da região de Lisboa e Vale do Tejo como região prioritária motivou justamente uma discussão durante a negociação do presente Quadro Comunitário de Apoio, tendo em conta a heterogeneidade de desenvolvimento na mesma região. Ora, tal como referem Lipzshitz e Ravch [11], a investigação sobre desigualdades regionais e as políticas para as combaterem, tem prestado pouca atenção às diferenças sócio-económicas intra-regionais, centrando-se sobretudo na análise das diferenças inter-regionais. Essa análise deverá passar pela consideração de um conjunto alargado de indicadores sócio-económicos, extravasando a mera referência ao PIB per capita, usual medida padrão da dimensão e do desempenho das economias regionais [2].

É esta, pois, a motivação do trabalho que aqui se apresenta, sendo de referir como estudos anteriores que utilizam a mesma ferramenta de análise estatística, os estudos de Ozimek[14] para os Estados Unidos da América, de Openshaw[13] para o Reino Unido, e de Lema e Mather [9] [10], Brandão et al [1] e Ferrão e Jansen-Butler [5] para Portugal. Contudo, estes trabalhos foram levados a cabo com objectivos diferentes, diferindo também nas variáveis analisadas.

Na análise utilizámos indicadores dos 275 concelhos em que Portugal Continental se encontrava dividido à data de início deste trabalho, englobando os domínios da demografia, actividades económicas, saúde, educação, emprego e cultura. A metodologia utilizada recorre à análise de Clusters.

O resto do artigo desenvolve-se como segue:

- No ponto dois começa-se por apresentar, duma forma sucinta, a informação estatística de base local. São ainda referidos os critérios utilizados para seleccionar os indicadores sócio-económicos, os quais são aqui objecto de análise.
- No ponto três procede-se à identificação de grupos homogéneos de concelhos em termos económicos e sociais, obtida a partir da análise de clusters.
- Finalmente, o ponto quatro apresenta as principais conclusões e pistas para um desenvolvimento futuro deste estudo.

2. Principais Indicadores Económico-Sociais de Base Local

A principal fonte dos dados utilizados no desenvolvimento deste trabalho foi o INE - Instituto Nacional de Estatística [6] [7] [8].

Na tarefa de selecção dos dados a uma escala concelhia encontraram-se disponíveis quatro tipos de informação estatística:

- *Inventário Municipal*, que apresenta e analisa dados recolhidos através de um inquérito efectuado a todos os concelhos, procurando traçar o quadro da vida local.
- *Anuário Estatístico Regional*, que agrega um conjunto de informação estatística de âmbito regional, com uma desagregação até ao nível concelhio e está organizada em três grandes capítulos: Território e População, Actividade Económica e Indicadores Sociais.
- *Estudo Sobre o Poder de Compra Concelhio*, que apresenta uma hierarquização dos concelhos que constituem o todo nacional, em termos do respectivo poder de compra.
- *Censos*, que contém dados sobre edifícios, alojamentos, famílias, núcleos familiares e indivíduos, caracterizados com base em variáveis demográficas e sócio-económicas.

O Inventário Municipal é da responsabilidade da respectiva Direcção Regional do INE, sendo elaborado um por cada região. Apenas os Inventários Municipais das regiões de Lisboa e Vale do Tejo e do Algarve se referiam a dados do mesmo ano (1995); para a região Norte o Inventário disponível referia-se ao ano 1992; para a região Centro ao ano de 1994; e para a região do Alentejo a 1993. Por esta razão decidiu-se desprezar este tipo de informação.

Os Censos demográficos são publicados com uma periodicidade decenal, sendo a última publicação referente a 1991.

Os últimos Anuários Estatísticos disponíveis no início deste trabalho apresentavam dados relativos ao ano 1995. Este tipo de informação, por ser a mais actualizada e a mais diversificada, foi a que assumiu maior relevo.

Concluída a recolha de dados, contaram-se 55 indicadores concelhios, alguns dos quais com pouco interesse para este estudo. Assim, procedeu-se subsequentemente a uma selecção destes indicadores, assente no princípio de que a situação económico-social pode ser caracterizada através dos seguintes referenciais: população, actividades económicas, saúde, educação, emprego e cultura.

Perante este critério, subsistiram os 33 indicadores descritos no Quadro 1 (código e descrição). À excepção da percentagem da população activa na população total e da percentagem da população desempregada na população activa, que se referem ao ano de 1991, todos os outros indicadores se reportam ao ano de 1995.

Código	Descrição
DENSPOP	Densidade Populacional (Habitantes por Km ²)
POP 0-24	Percentagem da população dos 0 aos 24 anos na população total
POP 25-64	Percentagem da população dos 25 aos 64 anos na população total
TNAT	Número de nados-vivos por 1000 habitantes
TMORT	Número de óbitos por 1000 habitantes
TMIG	Diferença entre o número de entradas e saídas por migração, por 1000 habitantes
IMPORT	Valor das importações por habitante (10 ⁴ Esc.)
EXPORT	Valor das exportações por habitante (10 ⁴ Esc.)
ALOJAM	Número máximo de camas em estabelecimentos hoteleiros por 1000 habitantes
EMP ABC	Número de Empresas do sector primário por 1000 habitantes
EMP DEF	Número de Empresas do sector secundário por 1000 habitantes
EMP GQ	Número de Empresas do sector terciário por 1000 habitantes
DEPOSITO	Total de depósitos por habitante (10 ³ Esc.)
DESPCAM	Despesas da Câmara por habitante (10 ³ Esc.)
ELECDOM	Consumo de Electricidade doméstico por habitante (10 Kwh)
ELECIND	Consumo de Electricidade industrial por habitante (10 Kwh)
AGUA	Água distribuída pela Câmara por habitante (10 Esc.)
TELEFONE	Postos Telefónicos por 100 habitantes
COMBUST	Venda de Combustíveis por habitante (Kg)
HOSP	Número de Hospitais por 1000 habitantes
CENTSAU	Número de Centros de Saúde por 1000 habitantes
CAMAS	Número de camas hospitalares por 1000 habitantes
MEDICOS	Número total de médicos por 1000 habitantes
TMORTINF	Número de óbitos com menos de 1 ano por 1000 nados-vivos
ENSBAS	Número de estabelecimentos de ensino Básico por Km ²
ENSSEC	Número de estabelecimentos de ensino Secundário por Km ²
ENSSUP	Número de estabelecimentos de ensino Superior por Km ²
ESPECT	Número de espectadores em espectáculos públicos por habitante
BIBLIOT	Bibliotecas por 1000 habitantes
AMB/TOT	Relação entre as despesas da Câmara em Ambiente e o total das despesas
TACTIV	percentagem da população activa na população total
TDES	Percentagem da população desempregada na população activa
PODCOMP	Indicador do Poder de Compra per capita

Quadro 1 - Código e descrição dos indicadores

Análise das variáveis

Apresenta-se no Quadro 2 uma descrição das variáveis, em termos de média, desvio padrão, coeficiente de variação e valores mínimos e máximos verificados para todos os concelhos.

Variáveis	Média	Mínimo	Máximo	Desvio Padrão	Coeficiente de Variação
DENSPOP	281,059	7,152	7786,250	849,345	3,022
POP_0_24	32,201	20,588	43,073	4,578	0,142
POP_25_6	48,899	39,486	56,805	3,437	0,070
TNAT	9,299	3,648	16,592	2,479	0,267
TMORT	12,808	5,607	34,167	3,857	0,301
TMIG	0,240	-26,000	22,200	6,434	26,796
IMPORT	19,781	0,000	511,362	46,318	2,342
EXPORT	19,633	0,000	412,617	44,298	2,256
ALOJAM	19,091	0,000	1369,477	93,216	4,883
EMP_ABC	17,107	0,326	81,971	11,583	0,677
EMP_DEF	26,271	3,115	294,471	21,963	0,836
EMP_GQ	56,156	4,783	649,760	43,205	0,769
DEPOSITO	1307,197	0,000	12652,752	1108,685	0,848
DESPCAM	82,117	34,892	222,521	34,773	0,423
ELECDOM	25178,585	45,800	492354,139	54331,007	2,158
ELECIND	37045,332	6,812	581223,910	82369,457	2,223
AGUA	197862,636	159,049	6483982,000	585056,233	2,957
TELEFONE	31,767	6,406	68,700	7,995	0,252
COMBUST	972,534	104,352	58210,119	3536,131	3,636
HOSP	0,011	0,000	0,287	0,025	2,296
CENTSAU	0,498	0,064	2,174	0,389	0,782
CAMAS	2,312	0,000	34,384	4,091	1,769
MEDICOS	1,241	0,000	18,707	1,652	1,332
TMORTINF	7,208	0,000	56,604	8,657	1,201
ENSBAS	0,267	0,015	4,893	0,506	1,898
ENSSEC	0,027	0,000	1,224	0,110	4,080
ENSSUP	0,010	0,000	1,080	0,086	8,599
ESPECT	0,282	0,000	5,728	0,629	2,226
BIBLIOT	0,171	0,000	0,613	0,113	0,663
AMB_TOT	9,066	0,753	41,464	6,607	0,729
TACTIV	40,340	25,700	53,000	5,744	0,142
TDES	6,461	1,700	24,400	3,355	0,519
PCOMP	60,944	22,765	314,093	33,935	0,557

Quadro 2 - Estatísticas descritivas

Pela observação dos valores mínimos e máximos pode-se constatar que os 275 concelhos apresentam uma grande diferenciação para as diferentes variáveis. Em particular a densidade populacional apresenta uma amplitude de mais de 7000 habitantes por Km², sendo ainda de destacar o elevado coeficiente de variação da taxa de migração.

Passemos agora a uma análise da interdependência entre as variáveis, atentando na respectiva matriz de correlações (Quadro 3):

	DENSPOP	POP_0_24	POP_25_6	TNAT	TMORT	TMIG	IMPORT	EXPORT	ALOJAM
DENSPOP	1.00								
POP_0_24	0.08	1.00							
POP_25_6	0.39	0.08	1.00						
TNAT	0.21	0.77	0.37	1.00					
TMORT	-0.25	-0.72	-0.51	-0.67	1.00				
TMIG	-0.09	0.22	0.38	0.27	-0.38	1.00			
IMPORT	0.40	0.09	0.38	0.19	-0.23	0.14	1.00		
EXPORT	0.14	0.21	0.34	0.29	-0.31	0.23	0.79	1.00	
ALOJAM	-0.01	0.00	0.11	0.18	-0.01	0.11	-0.04	-0.04	1.00
EMP_ABC	-0.34	-0.46	-0.23	-0.45	0.48	-0.18	-0.20	-0.23	-0.04
EMP_DEF	-0.02	-0.07	0.15	0.10	-0.04	0.13	0.01	0.02	0.06
EMP_GQ	0.10	-0.14	0.17	0.05	0.02	0.07	0.05	0.00	0.17
DEPOSITO	0.17	-0.08	0.01	-0.02	0.03	-0.05	0.07	-0.02	0.04
DESPCAM	-0.12	-0.54	-0.31	-0.44	0.61	-0.19	-0.10	-0.16	0.23
ELECDOM	0.78	0.12	0.46	0.30	-0.31	0.01	0.41	0.19	0.05
ELECIND	0.35	0.22	0.50	0.36	-0.42	0.20	0.41	0.43	-0.03
AGUA	0.79	0.02	0.40	0.20	-0.21	-0.05	0.42	0.13	0.06
TELEFONE	0.39	-0.26	0.45	0.09	0.02	0.07	0.28	0.15	0.36
COMBUST	0.00	0.01	0.14	0.01	-0.04	0.11	0.26	0.40	0.03
HOSP	0.20	0.02	0.22	0.14	-0.09	-0.08	0.12	0.05	-0.02
CENTSAU	-0.25	-0.59	-0.43	-0.57	0.61	-0.37	-0.22	-0.23	-0.08
CAMAS	0.26	-0.03	0.20	0.10	-0.05	-0.18	0.09	0.01	-0.01
MEDICOS	0.57	0.01	0.40	0.21	-0.21	-0.08	0.28	0.10	0.04
TMORTINF	0.01	0.10	-0.09	0.06	0.01	-0.06	-0.03	0.00	0.03
ENSBAS	0.97	0.17	0.36	0.27	-0.30	-0.11	0.41	0.15	-0.01
ENSSEC	0.87	0.00	0.27	0.12	-0.13	-0.22	0.38	0.12	0.00
ENSSUP	0.70	-0.06	0.14	0.03	-0.02	-0.29	0.31	0.06	0.02
ESPECT	0.48	-0.02	0.29	0.17	-0.11	-0.03	0.23	0.08	0.16
BIBLIOT	0.06	-0.32	-0.10	-0.30	0.36	-0.24	-0.02	-0.12	-0.09
AMB_TOT	0.15	-0.02	0.44	0.18	-0.22	0.31	0.26	0.25	0.13
TACTIV	0.37	0.34	0.79	0.59	-0.61	0.38	0.35	0.37	0.13
TDES	0.01	-0.28	0.08	-0.23	0.20	-0.13	-0.04	-0.08	-0.09
PCOMP	0.69	0.00	0.66	0.31	-0.30	0.07	0.45	0.25	0.18

	EMP_ABC	EMP_DEF	EMP_GQ	DEPOSITO	DESPCAM	ELECDOM	ELECIND	AGUA
EMP_ABC	1.00							
EMP_DEF	0.37	1.00						
EMP_GQ	0.36	0.90	1.00					
DEPOSITO	0.28	0.75	0.81	1.00				
DESPCAM	0.34	0.05	0.18	0.13	1.00			
ELECDOM	-0.39	-0.01	0.10	0.17	-0.18	1.00		
ELECIND	-0.40	0.02	0.02	0.00	-0.29	0.58	1.00	
AGUA	-0.31	-0.02	0.12	0.22	-0.08	0.94	0.47	1.00
TELEFONE	-0.12	0.12	0.28	0.22	0.14	0.43	0.20	0.49
COMBUST	-0.03	0.02	0.03	0.00	0.07	0.00	0.05	0.00
HOSP	-0.17	-0.03	0.02	0.06	-0.08	0.21	0.13	0.22
CENTSAU	0.37	-0.01	0.01	-0.03	0.46	-0.30	-0.31	-0.23
CAMAS	-0.20	-0.06	0.02	0.12	-0.07	0.33	0.16	0.34
MEDICOS	-0.30	-0.02	0.12	0.21	-0.13	0.70	0.40	0.68
TMORTINF	-0.09	-0.03	-0.03	-0.02	0.02	0.00	0.00	0.00
ENSBAS	-0.39	-0.01	0.09	0.21	-0.16	0.82	0.35	0.83
ENSSEC	-0.26	-0.03	0.12	0.28	-0.01	0.77	0.24	0.82
ENSSUP	-0.15	-0.03	0.10	0.29	0.05	0.73	0.17	0.78
ESPECT	-0.24	-0.03	0.13	0.24	0.03	0.56	0.24	0.63
BIBLIOT	0.22	0.05	0.13	0.23	0.40	-0.01	-0.13	0.09
AMB_TOT	-0.08	0.02	0.08	-0.03	-0.18	0.28	0.28	0.24
TACTIV	-0.31	0.11	0.08	-0.02	-0.42	0.45	0.51	0.36
TDES	0.20	-0.06	0.01	-0.06	0.19	-0.07	-0.08	0.00
PCOMP	-0.32	0.05	0.21	0.25	-0.14	0.73	0.45	0.75

	TELEFONE	COMBUST	HOSP	CENTSAU	CAMAS	MEDICOS	TMORTINF	ENSBAS
TELEFONE	1.00							
COMBUST	0.11	1.00						
HOSP	0.20	-0.01	1.00					
CENTSAU	-0.02	-0.05	-0.06	1.00				
CAMAS	0.24	-0.02	0.78	-0.05	1.00			
MEDICOS	0.48	0.02	0.36	-0.21	0.57	1.00		
TMORTINF	0.00	-0.06	0.11	0.05	0.08	-0.02	1.00	
ENSBAS	0.36	-0.02	0.22	-0.29	0.29	0.61	0.02	1.00
ENSSEC	0.40	0.00	0.26	-0.16	0.34	0.61	0.02	0.92
ENSSUP	0.34	0.00	0.23	-0.09	0.35	0.59	0.02	0.78
ESPECT	0.49	0.10	0.31	-0.17	0.40	0.56	-0.02	0.53
BIBLIOT	0.12	0.04	0.06	0.24	0.15	0.21	0.03	0.05
AMB_TOT	0.37	0.01	0.19	-0.18	0.17	0.28	-0.05	0.12
TACTIV	0.28	0.10	0.18	-0.54	0.13	0.35	-0.07	0.37
TDES	0.04	0.06	0.03	0.24	0.02	-0.02	-0.04	-0.07
PCOMP	0.73	0.16	0.31	-0.32	0.41	0.72	-0.01	0.69

	ENSSEC	ENSSUP	ESPECT	BIBLIOT	AMB_TOT	TACTIV	TDES	PCOMP
ENSSEC	1.00							
ENSSUP	0.92	1.00						
ESPECT	0.59	0.59	1.00					
BIBLIOT	0.18	0.22	0.17	1.00				
AMB_TOT	0.09	0.05	0.19	-0.07	1.00			
TACTIV	0.25	0.12	0.28	-0.20	0.35	1.00		
TDES	0.01	0.02	0.05	0.27	0.01	-0.08	1.00	
PCOMP	0.71	0.61	0.69	0.08	0.41	0.56	0.01	1.00

Quadro 3 - Matriz de correlação

Esta matriz evidencia a existência de fortes relações entre algumas das variáveis, sendo algumas destas interdependências expectáveis e outras não tanto.

Sem procurar fazer uma leitura completa dos valores obtidos, a qual extravasaria o objectivo deste artigo, podem-se, contudo, sublinhar alguns aspectos:

- A densidade populacional está obviamente muito correlacionada com os estabelecimentos dos diferentes tipos de ensino por Km². A sua forte correlação com os consumos per capita de electricidade doméstica e de água deverá ter por explicação a também elevada correlação com o índice de poder de compra concelhio.
- A taxa de natalidade, muito associada à percentagem de população jovem (0 a 24 anos), está inversa e fortemente relacionada com a taxa de mortalidade, o que é preocupante já que indicia o envelhecimento progressivo de algumas zonas. De forma óbvia, a percentagem de população dos 25 aos 64 anos está fortemente correlacionada com a taxa de actividade e o poder de compra.
- Os valores das importações e das exportações têm um coeficiente de correlação elevado (79%), denotando a dependência de fornecimentos externos por parte do nosso sector exportador. O número de empresas do sector secundário, do terciário, e o valor dos depósitos bancários por habitante apresentam também correlações elevadas, com valores entre 75% e 90%.
- O poder de compra tem um coeficiente de correlação elevado com o consumo de electricidade doméstico, com o consumo de água, com o número de postos telefónicos, com o número de médicos e também com o número de indivíduos que frequentam os espectáculos.
- Curiosamente, o número de centros de saúde por 1000 habitantes está significativamente correlacionado com a taxa de mortalidade. A explicação estará no facto de esta última ser mais acentuada nas zonas de baixa densidade populacional.

- Curiosa, ainda, a ausência de correlação entre o número de estabelecimentos do ensino secundário por Km² e a percentagem de população jovem (dos 0 aos 24 anos), mas compreensível se atentarmos na elevada correlação do número de estabelecimentos com a variável populacional.

3. Agrupamento dos Concelhos

Com o objectivo de efectuar o agrupamento dos concelhos em níveis similares de desenvolvimento sócio-económico, aplicou-se a análise de clusters a todas as variáveis estandardizadas.

Para determinação dos grupos homogéneos de concelhos é utilizado inicialmente o método hierárquico aglomerativo de formação de clusters. Recorre-se, de seguida, a um método não hierárquico (K-means), utilizando-se o número de clusters obtido pelo método hierárquico (veja-se [4]). As sementes iniciais são também escolhidas com base no método hierárquico¹.

Método hierárquico

A análise de clusters pelo método hierárquico foi efectuada utilizando vários critérios de agregação e diversas medidas de distância. Entre os vários critérios de agregação, o critério de Ward [15] foi o que permitiu obter uma melhor caracterização dos concelhos em grupos distintos. No entanto, em todas as análises efectuadas evidenciam-se apenas 3 clusters com um número considerável de concelhos. Quando é sugerida a consideração de mais um ou dois clusters, estes são constituídos por escassos elementos notoriamente diferenciados dos restantes. Correspondem ao cluster que compreende compreensivelmente Lisboa e Porto; bem como, por vezes, a um cluster constituído por Sardoal e Vila Nova da Barquinha, concelhos que apresentam um número de empresas por habitante claramente superior aos dos restantes.

Método não hierárquico

Procedeu-se seguidamente a uma análise pelo método não hierárquico, considerando-se de início 3 clusters. Utilizando vários conjuntos de sementes, obteve-se, repetidamente, um cluster formado simplesmente por Lisboa e Porto. Por esta razão, prosseguiu-se a análise sem estes dois concelhos, considerando-os um quarto cluster.

As sementes iniciais (Arraiolos, Bragança e Almada) foram escolhidas com base nas análises de clusters pelo método hierárquico, com a preocupação de serem concelhos consideravelmente diferentes a nível sócio-económico.

Na fig.1 apresentam-se os contornos dos três clusters (quatro com o de Lisboa e Porto). As três grandes manchas decompõem genericamente o País de forma longitudinal, correspondendo à conhecida assimetria litoral-interior.

¹ Para uma análise mais detalhada, veja-se [12].

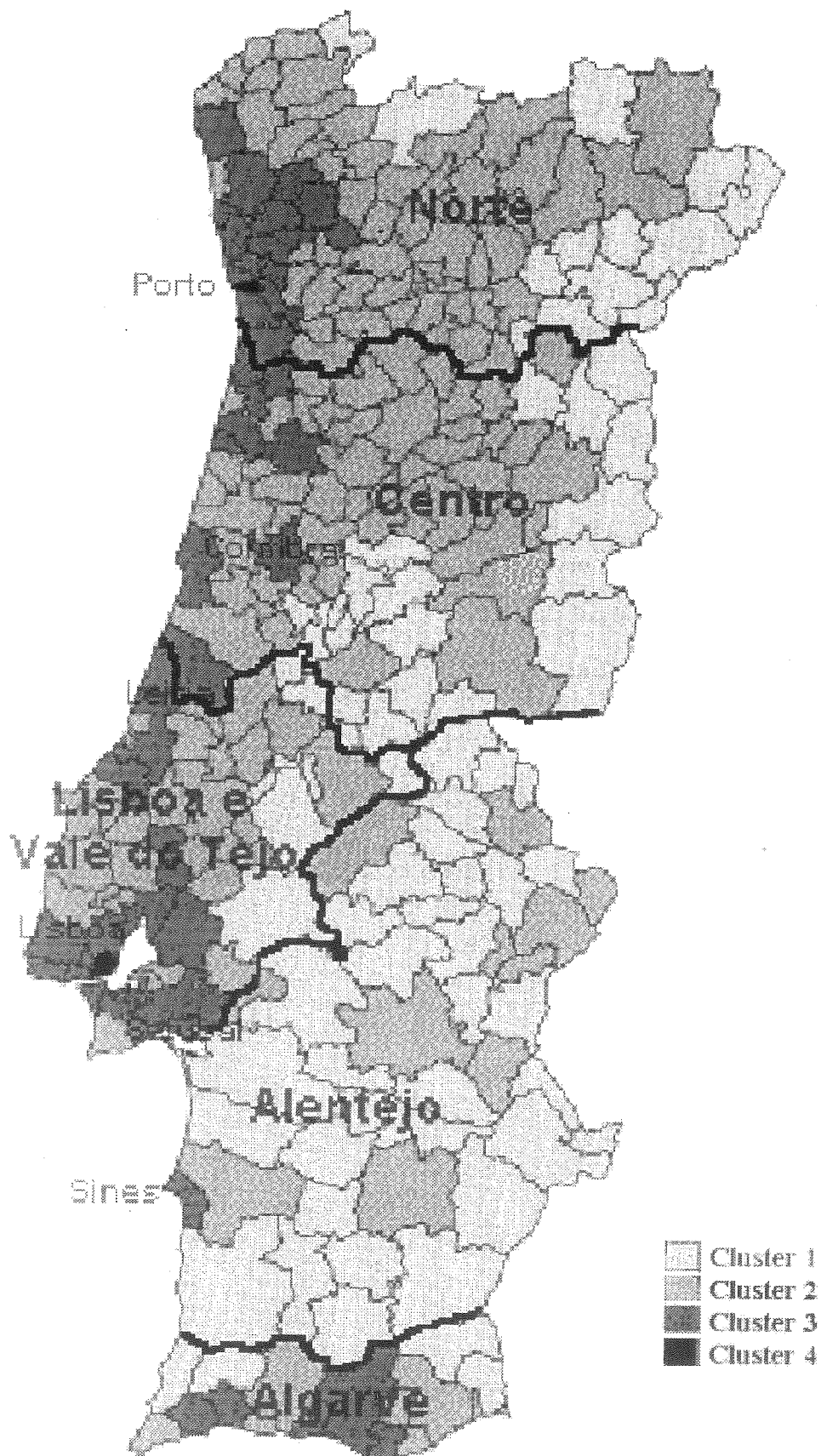


Figura 1 - Agrupamento dos concelhos segundo o nível sócio-económico

Para uma melhor análise dos três grupos, no Quadro 4 são apresentados a média e o desvio padrão das variáveis estandardizadas em estudo.

Variável	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	Média	Desv. Pad.	Média	Desv. Pad.	Média	Desv. Pad.	Média	Desv. Pad.
DENSPOP	-0.30	0.02	-0.20	0.11	0.75	1.49	7.80	0.22
POP_0_24	-0.99	0.64	0.38	0.86	0.52	0.66	-0.80	0.58
POP_25_64	-0.60	0.80	-0.12	0.81	1.26	0.58	1.15	0.18
TNAT	-0.92	0.66	0.17	0.84	0.96	0.68	0.10	0.37
TMORT	0.99	1.00	-0.21	0.60	-0.98	0.53	0.15	0.65
TMIG	-0.61	0.97	0.13	0.74	0.73	0.88	-3.61	0.66
IMPORT	-0.39	0.09	-0.16	0.43	0.93	1.77	3.63	2.85
EXPORT	-0.42	0.08	-0.16	0.37	1.09	1.89	0.64	0.77
ALOJAM	-0.13	0.21	-0.06	0.43	0.37	2.19	0.15	0.12
EMP_ABC	0.73	1.03	-0.07	0.79	-0.89	0.58	-1.38	0.05
EMP_DEF	0.08	1.74	-0.07	0.46	0.08	0.32	-0.29	0.22
EMP_GQ	0.11	1.75	-0.14	0.36	0.18	0.43	1.14	0.28
DEPOSITO	0.11	1.61	-0.09	0.45	-0.04	0.51	3.45	2.06
DESPCAM	0.92	1.07	-0.34	0.61	-0.51	0.81	0.86	1.03
ELECDOM	-0.41	0.06	-0.22	0.21	0.96	1.17	8.15	0.63
ELECIND	-0.43	0.04	-0.28	0.23	1.41	1.66	1.75	1.10
AGUA	-0.32	0.03	-0.20	0.16	0.72	1.15	8.93	2.57
TELEFONE	-0.11	0.55	-0.28	0.85	0.82	1.23	3.63	1.40
COMBUST	-0.09	0.19	-0.09	0.15	0.40	2.29	0.01	0.05
HOSP	-0.23	1.35	-0.01	0.77	0.29	0.75	2.46	0.15
CENTSAU	1.03	1.13	-0.29	0.56	-0.77	0.26	-0.73	0.09
CAMAS	-0.20	0.97	-0.02	0.91	0.23	0.98	3.80	0.24
MEDICOS	-0.36	0.18	-0.15	0.41	0.75	1.62	6.22	0.43
TMORTINF	0.04	1.48	-0.02	0.80	-0.02	0.48	0.24	0.23
ENSBAS	-0.41	0.09	-0.13	0.21	0.68	1.22	8.79	0.50
ENSSEC	-0.22	0.03	-0.16	0.06	0.39	0.90	10.44	0.60
ENSSUP	-0.12	0.00	-0.10	0.05	0.02	0.20	11.57	1.14
ESPECT	-0.19	0.58	-0.20	0.47	0.61	1.38	6.65	2.85
BIBLIOT	0.64	1.26	-0.25	0.62	-0.39	0.80	2.59	1.33
AMB_TOT	-0.37	0.62	-0.10	0.95	0.85	1.17	0.25	0.26
TACTIV	-0.75	0.81	-0.04	0.76	1.27	0.52	1.09	0.22
TDES	0.64	1.35	-0.30	0.62	-0.15	0.78	0.19	0.08
PCOMP	-0.52	0.28	-0.19	0.61	1.09	0.90	6.62	1.19

Quadro 4 - Estatísticas descritivas para cada cluster

O cluster 1, que apelidámos de "Zonas Rurais Menos Desenvolvidas", é formado por concelhos cuja localização incide sobretudo no interior do país, alongando-se para litoral na zona do Alentejo. Abrange três concelhos isolados no norte do país - Melgaço, Montalegre e Vinhais -, para além de duas grandes manchas no norte interior e no Alentejo, com alguns pontos de descontinuidade provocados por concelhos do cluster 2.

Este agrupamento caracteriza-se por uma população muito envelhecida, onde a taxa de natalidade é muito baixa e a taxa de mortalidade muita elevada. As empresas existentes pertencem, quase exclusivamente, ao sector primário. O número de Centros de Saúde por habitante é relativamente elevado, o que se justifica pelo facto destes concelhos terem uma densidade populacional baixa e assim ser mais difícil otimizar estes recursos. A percentagem de população activa é muito reduzida e a taxa de desemprego é muito elevada, comparativamente com os concelhos dos restantes clusters.

O cluster 2, "Zonas Rurais Mais Desenvolvidas", é constituído por uma vasta superfície do território, que revela uma grande concentração no norte e centro do país. Abrange ainda alguns concelhos do Alentejo e do Algarve.

Como se pode concluir pela observação da tabela referente ao cluster 2, todas as variáveis se aproximam da média global. Os concelhos pertencentes a este cluster estão num patamar intermédio a nível socio-económico, sendo menos carenciados que os do cluster 1 mas mais desfavorecidos que os do cluster 3.

O cluster 3, "Zonas Urbanas Litorais", é constituído por concelhos da orla litoral do país. Destaca-se, no norte do país, a mancha constituída pelos concelhos do Grande Porto (excluindo o próprio concelho do Porto). Ao centro é de referir algumas manchas isoladas constituídas pelos concelhos de Coimbra e Figueira da Foz e alguns concelhos do distrito de Leiria. Temos ainda uma mancha contínua formada por alguns concelhos em torno das cidades de Lisboa e Setúbal. No Alentejo, temos o concelho de Sines e duas aglomerações na costa algarvia.

Estes concelhos caracterizam-se por uma densidade populacional muito elevada, uma taxa de natalidade elevada e uma taxa de mortalidade reduzida. Regista-se um grande peso da população pertencente à classe etária dos 25 aos 64 anos e da população activa na população total. Em contraste com os concelhos do cluster 1, estes apresentam uma taxa de migração elevada.

No que toca aos indicadores económicos, observa-se que o volume de importações e exportações é acentuado face à média nacional. A variável referente às empresas do sector primário apresenta um valor muito abaixo dos das empresas dos sectores secundário e terciário. O poder de compra dos habitantes destes concelhos é francamente superior aos dos restantes (excluindo Lisboa e Porto). Os consumos de água e electricidade apresentam ainda valores muito acima da média global. Em comparação com os clusters 1 e 2, também o número de postos telefónicos existentes nestes concelhos é claramente mais elevado.

Nos campos da educação, saúde e cultura, vê-se que o número de estabelecimentos de ensino (básico, secundário e superior) por km² também está bastante acima do encontrado nos

clusters 1 e 2, mas já o número de bibliotecas e centros de saúde por habitante é mais reduzido. Talvez se possa justificar o facto por as populações estarem mais concentradas, permitindo que um grande número de pessoas frequentem o mesmo centro ou a mesma biblioteca. A confirmar, de certa forma, esta suposição, está o valor elevado da variável número de médicos por habitante. Verifica-se, ainda, que nestes concelhos é elevada a relação entre as despesas da Câmara em Ambiente e o total das despesas.

Como foi referido anteriormente, os concelhos de Lisboa e Porto, destacam-se claramente dos restantes, tendo formado o cluster 4 - "As Duas Cidades Principais". Pela observação da média das variáveis, conclui-se imediatamente que estes dois concelhos apresentam características singulares, tendo atingido um grau de desenvolvimento muito superior ao dos restantes.

4. Conclusões e Desenvolvimentos Futuros

A primeira conclusão deste trabalho é a de que as técnicas de estatística multivariada utilizadas permitiram com sucesso uma identificação perfeitamente legível das zonas/regiões do território continental português com diferente grau de desenvolvimento, detalhada até ao nível dos municípios.

A segunda grande conclusão deste estudo é no sentido de realçar um facto sobejamente reconhecido: o da assimetria litoral-interior. Qualquer política de desenvolvimento regional tem nesta realidade um facto incontornável, ilustrado mais finamente nos seus contornos geográficos pelos resultados desta análise.

A conclusão última, e talvez a mais importante, é a de que a classificação efectuada põe em evidência a debilidade de uma análise baseada meramente nas regiões administrativas (NUT-2), as quais evidenciam áreas de desenvolvimento muito desiguais no seu seio.

Quanto a possíveis desenvolvimentos, dois são desde logo óbvios: realizar outras análises para datas diferenciadas no tempo, tão logo os dados estejam disponíveis - avaliação da dinâmica do desenvolvimento territorial; e empreender estudos e interpretações aprofundadas separadamente para cada um dos sectores aqui considerados.

Referências

- [1] Brandão, A., Pires, A., Portugal, J., Agrupamentos de concelhos de Portugal e sua caracterização, *Revista de Estatística* 1 (1998).
- [2] European Commission, Sixth Periodic Report on the Social and Economic Situation and Development of the Regions of the European Union (1999).
- [3] European Council, Regulation (EC) No 1260/1999 of the European Council, *Official Journal of the European Communities* 26.6.1999 (1999).
- [4] Everitt, B. S., *Cluster Analysis* (3rd. ed.), Wiley, Nova Iorque (1993).
- [5] Ferrão, J., Jansen-Butler, C., Existem "regiões periféricas" em Portugal?, *Análise Social* 24 (1978) 355-371
- [6] INE - Instituto Nacional de Estatística, *Censos* (1991)
- [7] INE - Instituto Nacional de Estatística, *Anuário Estatístico Regional* (1995a).
- [8] INE - Instituto Nacional de Estatística, *Estudo Sobre o Poder de Compra Concelhio* (1995b).
- [9] Lema, P.B., Mather, P.M., *Factor Analysis e Cluster Analysis Aplicados a Dados Estatísticos Sobre Portugal*, Nottingham University, Geography Department (1970).
- [10] Lema, P.B., Mather, P.M., *O Norte de Portugal - Ensaio de Análise Multivariada*, *Estudos de Geografia Humana e Regional* 5, Centro de Estudos Geográficos, Universidade de Lisboa (1977).

- [11] Lipshitz, G., Raveh, A., Socio-economic differences among localities: a new method of multivariate analysis, *Regional Studies* 32 (1998) 747-757.
- [12] Marquês, M. M., *Análise Estatística Multivariada para Caracterização Sócio-Económica dos Concelhos do Território Continental*, Tese de Mestrado em Investigação Operacional e Engenharia de Sistemas, IST - UTL (1999).
- [13] Openshaw S. (Org.), *Census Users Handbook*. Geoinformation International And John Wiley & Sons, Cambridge (1995).
- [14] Ozimek J., *Targeting For Success: A Guide to New Techniques for Measurement and Analysis in Database and Direct Response Market*, McGraw-Hill, Berkshire (1993).
- [15] Ward, J., Hierarchical grouping to optimize an objective function, *Journal of the American Statistical Association* 58 (1963) 236-44.

A EFICIÊNCIA TÉCNICA DAS ESTAÇÕES DE CORREIO PORTUGUESAS

UMA APLICAÇÃO DA METODOLOGIA DEA - DATA ENVELOPMENT ANALYSIS

Teresa Sofia de Sousa Fernandes

Estagiária de MAEG/ISEG nos CTT

Maria Cândida Mourão

Centro de IO, FC/UL e
Instituto Superior de Economia e Gestão, Gab. 203Q
Rua do Quelhas, 6
1200-781 Lisboa

Alberto Pimenta

DES/CTT

Abstract

The main purpose of this study is to measure the technical efficiency of the Portuguese Post Offices by applying the DEA methodology. DEA (Data Envelopment Analysis) takes advantage of linear programming techniques that based on the existing data (referred to as *outputs* and *inputs*), establish a frontier. The efficiency of any given firm then relates directly with the distance to that established frontier.

The time sample for this study includes the years from 1997 to 1999.

Labor is the input whereas the various services supplied by the post offices are the outputs.

As a first step the study focuses on sorting the various post offices as a function of their respective technical efficiency. In a second stage, and with a reduced sample where information existed, we were able to incorporate in the model information on the quality of the services provided.

Results achieved with the DEAP program are presented and discussed.

Resumo

Este estudo tem como principal objectivo medir a eficiência técnica das Estações de Correio portuguesas através da aplicação da metodologia DEA.

A DEA usa técnicas da programação linear para construir uma fronteira baseada nos dados existentes (designados por *outputs* e *inputs*), sendo o grau de eficiência das unidades envolvidas medido tendo em conta a distância à fronteira estabelecida.

O período amostral refere-se aos anos de 1997 a 1999. Considerou-se como *input* o trabalho e como *outputs* as diferentes operações realizadas nas Estações de Correio. O principal objectivo, da primeira fase deste estudo, foi o de escalonar as diferentes Estações de Correio observadas em função da sua eficiência técnica. Numa segunda fase, restringiu-se a amostra, de modo a incluir no modelo informação adicional relativa à qualidade do serviço prestado, apenas disponível para algumas das Estações.

Apresentam-se e analisam-se resultados obtidos através do programa DEAP.

Keywords

DEA, technical efficiency (pure), mathematical programming, input-oriented.

1. Introdução

Este trabalho tem como principal objectivo estudar a eficiência técnica das Estações de Correio (EC), tanto a nível nacional como a nível regional. Pretende-se, deste modo, não só avaliar o desempenho de cada EC em particular, mas também comparar a eficiência das diversas estações entre si, visando auxiliar os agentes económicos envolvidos na definição de medidas de melhoria da *performance* e podendo também servir de base à adopção de políticas de melhoria da eficiência das Estações de Correio nacionais.

Para alcançar este objectivo, utiliza-se uma metodologia determinística não-paramétrica designada por *Data Envelopment Analysis* (DEA), aplicada a uma amostra de 910 Estações de Correio, no período de 1997 a 1999.

A DEA permite avaliar o desempenho relativo de um conjunto de unidades, significativamente homogéneas, em que as unidades mais eficientes definem uma fronteira em relação à qual se mede a eficiência das restantes unidades. Uma boa introdução a esta metodologia pode ser encontrada em Santos e Themido (1994).

A utilidade da DEA reside, precisamente, em fornecer indicações quanto à afectação de recursos, estabelecendo, assim, metas para as unidades ineficientes, tendo em vista orientar e melhorar o seu desempenho, com o fundamento de que outras unidades, num contexto idêntico, conseguiram atingir as mesmas metas.

Note-se que o facto de uma *Decision Making Unit* (DMU) ser eficiente não significa necessariamente que ela produz o nível máximo de *output* dado o conjunto de *inputs* e a tecnologia existente, significa sim que ela é uma das DMU's de "melhor prática" da amostra.

Neste trabalho, primeiro, são apresentadas algumas considerações acerca do conceito de eficiência. Em seguida, são identificados os *inputs* e *outputs* de uma Estação de Correio e expõe-se a metodologia DEA. Por fim apresentam-se os resultados mais significativos bem como as principais conclusões do estudo, quer a nível nacional quer a nível regional.

2. Conceito de Eficiência

Segundo Farrell (1957), a **eficiência técnica** (TE) refere-se à capacidade que uma DMU tem de produzir o *output* máximo possível, a partir de um dado conjunto de *inputs* e da tecnologia existente (ou a capacidade de empregar o volume mínimo possível de *inputs*, dado o nível de *output* pretendido e a tecnologia existente). A TE pode ser decomposta em eficiência de escala e eficiência técnica pura.

A **eficiência de escala** (SE) é conseguida quando a DMU produz o nível de *output*, para o qual o custo médio da produção é o mínimo.

A **eficiência técnica pura** (PTE) representa a redução máxima possível na quantidade dos *inputs* para a DMU se tornar tecnicamente eficiente.

Como referido, a eficiência é medida em relação a uma fronteira. No modelo utilizado neste estudo, a fronteira eficiente é caracterizada por uma *Input Requirement Function* (IRF),

extensão do modelo de função de produção proposto por Battese e Coelli (1995), e que foi desenvolvido por Battese, Heshmati e Hjamarsson (1998).

Este modelo de fronteira representa a solução para um problema de minimização da quantidade de trabalho requerida, dado um determinado vector de *outputs*. A envolvente assim obtida representa uma fronteira, na medida em que nenhuma DMU se poderá situar abaixo desse mínimo, podendo, no entanto, situar-se acima dele.

Decisões menos acertadas dos gestores, imperfeições nos mercados ou mesmo a existência de regulamentações governamentais, podem contribuir para que uma DMU empregue uma quantidade de trabalho superior à mínima. Assim, a diferença entre a quantidade de trabalho efectivamente observada para uma determinada DMU e a quantidade dada pela fronteira (eficiente) pode ser interpretada como um indicador de ineficiência.

3. Definição dos *Inputs* e *Outputs* das Estações de Correio

O *input* utilizado neste trabalho representa a quantidade ou o potencial de trabalho (em horas) efectuado pelos empregados das Estações de Correio (EC's).

Quanto aos *outputs*, estes consistem em nove tipos de serviços prestados pelas Estações de Correio. O agrupamento do grande leque de produtos (ou serviços) disponíveis numa EC foi feito com base no tempo que cada um deles leva a ser concluído, obtendo-se, assim, grupos relativamente homogêneos no que diz respeito às quantidades de trabalho exigidas. Ou seja, dentro de cada grupo, as operações (ou serviços) que o compõem necessitam de um período de tempo idêntico para a sua realização, exigindo, assim, o mesmo montante de *input* trabalho.

Foram, então, criados os nove grupos (ou *outputs*) que se apresentam de seguida:

- Corfax e Telegramas;
- Correspondências;
- EMS/Correio Urgente;
- Outras Operações;
- Serviços Financeiros Postais.
- Certificados de Aforro;
- Embalagens e Saquetas;
- Encomendas;
- Serviços Especiais de Correspondências;

Os *outputs* encontram-se medidos em operações unitárias, em que uma operação unitária corresponde a cinco minutos.

Posteriormente é também incluído no modelo, como *output*, o **tempo (total) médio de espera** (TME) por cliente, em cada Estação de Correio, desde que este chega à EC até ser atendido. Este *output* permite medir a qualidade do serviço prestado por cada EC, onde a qualidade do serviço pode ser encarada como uma medida da *performance* da estação.

No entanto, como os *outputs* devem estar positivamente correlacionados com os *inputs* (Doble (1995)), e dado existir uma correlação negativa entre o tempo médio de espera e o *input* (quanto maior for o número de trabalhadores no atendimento numa EC, menos tempo um cliente tem que esperar para ser atendido), o tempo médio de espera tem que ser invertido. A inversão pode ser conseguida pelo seu simétrico ou subtraindo-o ao valor mais alto verificado na amostra. Neste estudo, optou-se pela segunda alternativa, para que se possa trabalhar apenas

com valores positivos. A variável da qualidade representa, assim, os minutos que um cliente não tem que esperar, tendo em conta o tempo máximo que ele poderá ter que esperar. Obviamente, que se a inversão fosse conseguida pelo simétrico a variável, com uma interpretação diferente, originaria os mesmos resultados.

Note-se que o tempo médio de espera é um *output* controlável, ao contrário dos restantes *outputs*, sobre os quais os gerentes das Estações de Correio não têm qualquer controlo, visto estes serem obviamente determinados pela procura.

4. Metodologia DEA (Data Envelopment Analysis)

A eficiência técnica das Estações de Correio é então medida com a metodologia DEA. Recorrendo a métodos de optimização com Programação Matemática, esta metodologia foi introduzida por Charnes, Cooper e Rhodes (1978), que generalizaram as técnicas de medidas de eficiência de Farrell (1957).

4.1 Questões que se Colocam

Quando se pretende estudar a Eficiência Técnica de um conjunto de DMU's utilizando a DEA, é necessário ter em atenção duas questões.

A primeira diz respeito aos objectivos das DMU's em análise: minimização dos *inputs* ou maximização dos *outputs*, conduzindo, respectivamente, a medidas de eficiência técnica orientadas no sentido do *input* ou a medidas de eficiência técnica orientadas no sentido do *output*.

A segunda questão está relacionada com a tecnologia associada à função de fronteira: rendimentos constantes à escala (CRS) ou rendimentos variáveis à escala (VRS).

4.1.1 Minimização dos *Inputs* vs Maximização dos *Outputs*

As medidas de eficiência técnica orientadas no sentido do *input* respondem à questão: "Em quanto é que as quantidades dos *inputs* podem ser (proporcionalmente) reduzidas sem alterar as quantidades dos *outputs* produzidas?".

Alternativamente, pode perguntar-se: "Em quanto é que as quantidades dos *outputs* podem ser (proporcionalmente) aumentadas sem alterar as quantidades dos *inputs* utilizadas?". Obtêm-se assim, as medidas orientadas no sentido do *output*, em oposição às primeiras, orientadas no sentido do *input*.

Nas Estações de Correio o objectivo mais apropriado é, sem dúvida, a minimização do *input* (trabalho). Pois, por um lado, devido à natureza exógena da procura para a maioria dos seus produtos, as EC's não podem optar pela maximização do *output*. Não tendo controlo sobre os seus *outputs*, a única alternativa viável é ajustar as necessidades de *input* à procura esperada. Por outro lado, os salários dos trabalhadores representam uma proporção elevada dos custos totais, devendo então tentar reduzir-se o custo do *input* trabalho e procurar usá-lo mais eficientemente.

De seguida, apresenta-se o conceito de eficiência técnica para uma orientação no sentido do *input*, ou seja, define-se a medida de eficiência relacionada com a redução proporcional nos *inputs* de modo a que uma DMU se torne eficiente.

Exemplo 1: Considerem-se DMU's que utilizam dois *inputs* (x_1 e x_2) para produzir um único *output* (y).

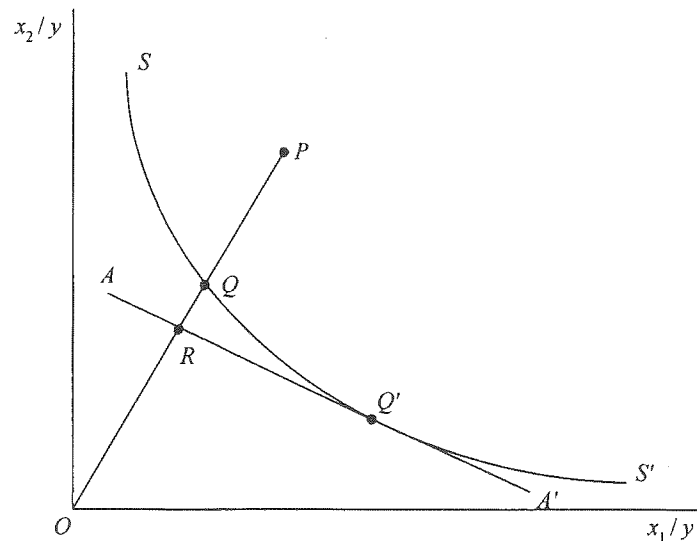


Figura 1 - Eficiência Técnica para uma Orientação no Sentido do *Input*

Na figura 1, SS' representa a isoquanta unitária das DMU's totalmente eficientes.

Se uma DMU utiliza as quantidades de *inputs* definidas pelo ponto P , para produzir uma unidade de *output*, a ineficiência técnica dessa DMU pode ser dada pela distância QP , representando a quantidade em que todos os *inputs* podem ser proporcionalmente reduzidos, sem reduzir o *output*. Geralmente, recorre-se a termos percentuais, ou seja, a QP/OP , para expressar a percentagem em que todos os *inputs* têm de ser reduzidos para obter uma produção tecnicamente eficiente. A **eficiência técnica orientada no sentido do input** (TE_i)¹ de uma DMU situada em P é, então, definida por

$$TE_i = \frac{OQ}{OP} = 1 - \frac{QP}{OP}.$$

Esta medida técnica assume valores entre zero e um, fornecendo informação quanto ao grau de eficiência/ineficiência técnica da DMU em análise. Assim, está-se em presença de uma DMU tecnicamente eficiente se e só se essa DMU verificar $TE_i = 1$. Por exemplo, na figura 1, a DMU situada no ponto Q é tecnicamente eficiente e $TE_i = 1$. ♦

4.1.2 Tecnologia de Referência

Utilizando a DEA, pode medir-se a eficiência técnica quer com rendimentos constantes à escala (CRS) quer com rendimentos variáveis à escala (VRS). A eficiência de uma unidade

¹ O índice "i" indica que se trata de uma medida orientada no sentido do *input*.

medida sob a hipótese de CRS combina eficiência de escala e eficiência técnica pura. Sob a hipótese de VRS apenas é medida a eficiência técnica pura. Deste modo, para além de se estimar a eficiência técnica, é também possível caracterizá-la em termos das suas componentes: eficiência técnica pura e eficiência de escala, como se ilustra na figura 2 em que se considera uma DMU com um só *input* e um *output*.

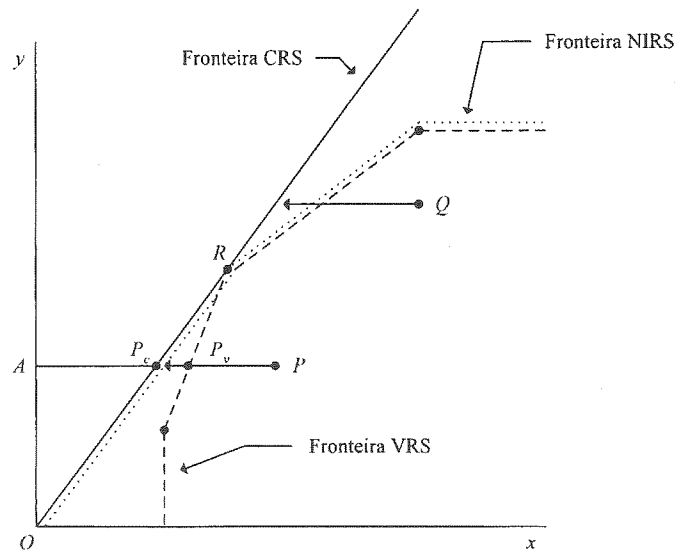


Figura 2 - Cálculo de Economia de Escala

Na figura 2, a fronteira NIRS representa a fronteira de eficiência determinada por rendimentos não crescentes à escala. Com CRS, a ineficiência técnica orientada no sentido do *input* do ponto P é a distância PP_C . Contudo, com VRS, a ineficiência técnica seria apenas PP_V . A diferença entre estas duas medidas de TE, $P_C P_V$, é devida à ineficiência de escala. Assim,

$$TE_{CRS} = \frac{AP_C}{AP}, TE_{VRS} = \frac{AP_V}{AP}, SE = \frac{AP_C}{AP_V}.$$

Neste estudo, a aplicação da metodologia DEA às Estações de Correio é feita quer com rendimentos constantes à escala (CRS) quer com rendimentos variáveis à escala (VRS), o que permite decompor a eficiência técnica total em eficiência técnica pura e em eficiência de escala. No entanto, dar-se-á mais relevância aos resultados referentes à tecnologia VRS, ou seja, à eficiência técnica pura, pois, por um lado, a hipótese de rendimentos constantes à escala (CRS) só deveria ser utilizada caso todas as Estações de Correio funcionassem numa escala óptima, por outro lado, porque se quer ter em conta a escala de operação das estações.

Quando detectada uma ineficiência de escala é também identificada a sua causa: rendimentos crescentes à escala (IRS) ou rendimentos decrescentes à escala (DRS).

4.2 Formulações em Programação Linear

Supõem-se conhecidos dados para K *inputs* e para M *outputs*, para cada uma de N DMU's. Os *inputs* e *outputs* da DMU n (com $n = 1, \dots, N$) são representados, respectivamente, pelos vectores coluna $x_{\bullet n}$ e $y_{\bullet n}$.

A medida de eficiência técnica para a DMU n ($n = 1, \dots, N$) é, então, dada pelo quociente entre a soma ponderada de todos os *outputs* e a soma ponderada de todos os *inputs*. Os ponderadores são os valores que se pretendem determinar, representando o peso que o *input* (*output*) associado deve ter de forma a que a eficiência técnica seja máxima. Assim, pretende determinar-se os valores de:

v_{kn} - peso do *input* k na DMU n ($k = 1, \dots, K$)

u_{mn} - peso do *output* m na DMU n ($m = 1, \dots, M$)

que maximizam a eficiência técnica da DMU n , dada por:

$$\frac{\sum_{m=1}^M u_{mn} Y_{mn}}{\sum_{k=1}^K v_{kn} X_{kn}} \quad (1)$$

Por sua vez, os valores de v_{kn} e u_{mn} devem ser tais que todas as medidas de eficiência não excedam um.

Os vectores coluna $\mathbf{v}_{\bullet n}$ (dimensão $(K \times 1)$) e $\mathbf{u}_{\bullet n}$ (dimensão $(M \times 1)$) representam, assim, os pesos dos K *inputs* e dos M *outputs*, respectivamente.

Os pesos óptimos para cada DMU n (com $n = 1, \dots, N$) podem, então, ser obtidos resolvendo o seguinte problema de programação matemática:

$$(P) \quad \text{Max} \quad \frac{\sum_{m=1}^M u_{mn} Y_{mn}}{\sum_{k=1}^K v_{kn} X_{kn}} \quad (2)$$

$$\text{s.a:} \quad \begin{cases} \frac{\sum_{m=1}^M u_{mn} Y_{mj}}{\sum_{k=1}^K v_{kn} X_{kj}} \leq 1 & j = 1, 2, \dots, N & (3) \\ u_{mn} \geq 0 & m = 1, \dots, M & (4) \\ v_{kn} \geq 0 & k = 1, \dots, K. & (5) \end{cases}$$

O primeiro conjunto de restrições impõe que a eficiência técnica de cada DMU (j), utilizando os diferentes *inputs* e *outputs* nas mesmas proporções da DMU n , não excede um, enquanto as restantes restrições obrigam a que todos os pesos sejam não negativos.

Um contra desta formulação é o de representar um problema de programação não linear com um número infinito de soluções óptimas, pois se $(\mathbf{u}_{\bullet n}^*, \mathbf{v}_{\bullet n}^*)$ for solução óptima, então $(\alpha \mathbf{u}_{\bullet n}^*, \alpha \mathbf{v}_{\bullet n}^*)$ também é solução óptima. Uma tentativa de contornar este problema consiste

em impor a nova restrição $\sum_{k=1}^K v_{kn} x_{kn} = 1$ (Coelli, Rao e Battese (1998)), que tem ainda a vantagem de permitir transformar o problema fraccionário (P) num problema de programação linear (PL).

Obviamente que esta restrição origina um problema diferente. Para evidenciar as diferenças designe-se por v_{kn} e μ_{mn} os pesos associados aos *inputs* e *outputs*, respectivamente, neste novo modelo. Assim, maximizar a eficiência da DMU n , corresponde a resolver o seguinte PL:

$$(TE) \quad TE_i^n = \text{Max} \sum_{m=1}^M \mu_{mn} y_{mn} \quad (6)$$

$$\text{s.a:} \quad \left\{ \begin{array}{l} \sum_{k=1}^K v_{kn} x_{kn} = 1 \quad (7) \\ \sum_{m=1}^M \mu_{mn} y_{mj} - \sum_{k=1}^K v_{kn} x_{kj} \leq 0 \quad j=1, 2, \dots, N \quad (8) \\ \mu_{mn} \geq 0 \quad m=1, \dots, M \quad (9) \\ v_{kn} \geq 0 \quad k=1, \dots, K. \quad (10) \end{array} \right.$$

Seja θ_n a variável dual associada a (7). Então, θ_n representa a proporção mínima dos *inputs* que seriam necessários para atingir os *outputs* de uma DMU eficiente.

Sejam λ_j^n (com $j = 1, \dots, N$) as variáveis duais associadas a (8). Estas variáveis representam os pesos que permitem identificar as DMU's de referência da DMU n . As DMU's de referência da DMU n são DMU's eficientes que utilizam as mesmas proporções de *inputs* na produção de *outputs* em proporções semelhantes às da DMU n . Assim, se n for eficiente, $\lambda_n^n = 1$.

Facilmente se verifica que o dual do PL anterior corresponde a:

$$(DTE) \quad \theta_n^* = \text{Min} \theta_n \quad (11)$$

$$\text{s.a:} \quad \left\{ \begin{array}{l} x_{kn} \theta_n - \sum_{j=1}^N x_{kj} \lambda_j^n \geq 0 \quad k=1, \dots, K \quad (12) \end{array} \right.$$

$$\left\{ \begin{array}{l} \sum_{j=1}^N y_{mj} \lambda_j^n \geq y_{mn} \quad m=1, \dots, M \quad (13) \end{array} \right.$$

$$\left\{ \begin{array}{l} \lambda_j^n \geq 0 \quad j=1, \dots, N \quad (14) \end{array} \right.$$

$$\left\{ \begin{array}{l} \theta_n \text{ livre.} \quad (15) \end{array} \right.$$

Um ponto na fronteira verifica $\theta_n^* = 1$ e representa uma DMU tecnicamente eficiente, de acordo com a definição de Farrell (1957).

Se $\theta_n^* < 1$, a DMU n é ineficiente, sendo portanto possível diminuir o seu *input* trabalho, em $(1 - \theta_n^*)$, sem reduzir nenhum dos seus *outputs*.

O conjunto das DMU's com valores de eficiência iguais a um é, então, considerado o conjunto de referência para as restantes DMU's ineficientes, de eficiência inferior a um ($\theta_n^* < 1$). Nestes casos os pesos $(\lambda_j^n)^*$, determinados pela resolução do problema (DTE), assumem valores diferentes de zero para as DMU's de referência da DMU n . Note-se que, no caso da DMU n ser eficiente, a DMU de referência de n é ela própria, ou seja, $\lambda_n^n = 1$ e $\lambda_j^n = 0$, $\forall j = 1, \dots, N, j \neq n$.

O problema de PL (DTE) deve ser resolvido para cada uma das DMU's da amostra, ou seja, deve ser resolvido N vezes, obtendo-se o valor de eficiência, θ_n^* , de cada DMU n .

Embora com mais variáveis, o problema dual (DTE) impõe menos restrições que o problema primal (TE) ($K+M < N+1$), sendo, neste caso (Coelli, Rao e Battese (1998)), de mais fácil resolução.

Note-se que, mesmo que a medida de eficiência técnica de uma DMU seja um, situando-se portanto na fronteira, não quer dizer que a DMU não possa melhorar a sua *performance*, quer pelo aumento de um ou mais dos seus *outputs* sem aumentar nenhum dos seus *inputs*, quer pela diminuição de um ou mais dos seus *inputs* sem diminuir os seus *outputs*.

As DMU's que podem melhorar o seu desempenho são aquelas que se situam na zona da fronteira eficiente paralela aos eixos. Nestes casos define-se desvio do *input* pela redução na quantidade de determinado *input* que é possível obter sem que para isso a DMU em análise tenha que aumentar outro *input* ou diminuir os *outputs*.

Exemplo 2: Considerem-se de novo DMU's que utilizam dois *inputs* (x_1 e x_2) para produzir um único *output* (y). Na figura 3, C e D são DMU's eficientes e definem a fronteira eficiente (SS'), e as DMU's A e B são DMU's ineficientes.

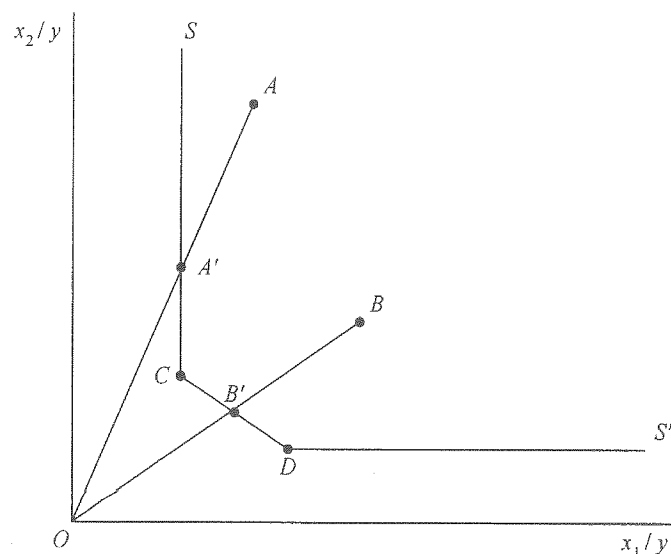


Figura 3 - Medição da Eficiência e Desvios dos *Inputs*

A medida da eficiência técnica de Farrell (1957) para as DMU's A e B é calculada por OA/OA' e OB/OB' , respectivamente. No entanto, é questionável se o ponto A' é um ponto eficiente, visto poder reduzir-se a quantidade do *input* x_2 utilizado (em CA') e continuar a produzir a mesma quantidade de *output*. Este ponto é assim designado como **fracamente eficiente**. Ou seja, a DMU A', embora na fronteira eficiente, pode diminuir x_2 até atingir C sem que para tal tenha que aumentar x_1 ou diminuir o *output*, representando a diminuição de x_2 uma melhoria potencial. A quantidade CA' representa, assim, o desvio do *input*. ♦

Em casos em que existam mais *inputs* e/ou *outputs* que os considerados no exemplo anterior, pode surgir o conceito análogo de **desvio do output**. Este representa o aumento na quantidade de determinado *output* que é possível obter sem que para isso a DMU em causa tenha que diminuir outro *output* ou alterar os *inputs*. Contudo, a informação relativa aos desvios dos *outputs* não é relevante para as Estações de Correio, pois, como frisado anteriormente, os gerentes das EC's não têm qualquer controlo sobre os *outputs*, visto estes serem determinados pela procura, não podendo, por isso, aumentar as quantidades de *outputs* produzidas.

Até agora considerou-se constante a relação entre *output* e *input*, podendo assim, em termos gráficos, esta relação ser representada por uma recta. No entanto, a hipótese de rendimentos constantes à escala (CRS) só é apropriada quando todas as DMU's funcionam numa escala óptima. O uso de CRS quando nem todas as DMU's estão a funcionar numa escala óptima, origina medidas de eficiência técnica (TE) que são confundidas por eficiências de escala (SE).

Banker, Charnes e Cooper (1984) sugeriram uma extensão ao modelo DEA com CRS para explicar situações com rendimentos variáveis à escala (VRS), que permite o cálculo da TE destituída destes efeitos de SE.

Segundo estes autores, o problema de programação linear (DTE) pode ser modificado para explicar VRS adicionando-lhe a restrição de convexidade, $\sum_{j=1}^N \lambda_j^n = 1$.

O modelo VRS pode então ser formalizado como se segue:

$$(VRS) \quad TE_{VRS}^n = \text{Min } \theta_n \quad (16)$$

$$s.a: \quad \left\{ \begin{array}{l} x_{kn} \theta_n - \sum_{j=1}^N x_{kj} \lambda_j^n \geq 0 \quad k=1, \dots, K \quad (17) \\ \sum_{j=1}^N y_{mj} \lambda_j^n \geq y_{mn} \quad m=1, \dots, M \quad (18) \\ \sum_{j=1}^N \lambda_j^n = 1 \quad (19) \\ \lambda_j^n \geq 0 \quad j=1, \dots, N \quad (20) \\ \theta_n \text{ livre.} \quad (21) \end{array} \right.$$

No entanto, o problema de programação linear (VRS) não indica se a DMU está a funcionar numa área de rendimentos decrescentes ou crescentes à escala. De modo a detectar o tipo de rendimentos resolve-se um problema adicional, no qual se impõem rendimentos não crescentes à escala (NIRS). Para isso, obtém-se um novo PL², substituindo a restrição (19) por

$$\sum_{j=1}^N \lambda_j^n \leq 1 \quad (22) \text{ e determina-se } TE_{NIRS}^n, \text{ solução óptima deste novo PL.}$$

A natureza da ineficiência de escala (i.e., devida a rendimentos crescentes à escala (IRS) ou a rendimentos decrescentes à escala (DRS)) para uma determinada DMU pode ser identificada pela análise do resultado da eficiência técnica com rendimentos não crescentes à escala (TE_{NIRS}^n) e do resultado da eficiência técnica com rendimentos variáveis à escala (TE_{VRS}^n).

Assim, se $TE_{NIRS}^n = TE_{VRS}^n$ a ineficiência de escala da DMU n deve-se à existência de rendimentos decrescentes à escala. Se, pelo contrário, $TE_{NIRS}^n \neq TE_{VRS}^n$, então, na DMU n existem rendimentos crescentes à escala.

4.3 Resolução de Modelos DEA

Os modelos DEA, para além de permitirem obter uma medida de eficiência, fornecem informação relativamente aos pares e às metas para cada DMU da amostra considerada.

Os **pares** de uma DMU são as suas DMU's de referência.

As **metas** são as coordenadas do ponto projectado sobre a fronteira eficiente (para a DMU ineficiente), indicando, assim, as quantidades de *input* e *output* que a DMU ineficiente deveria ser capaz de alcançar se estivesse a operar na fronteira eficiente.

Existem várias metodologias que permitem identificar as DMU's que definem a fronteira e o subsequente cálculo dos desvios (Coelli, Rao e Battese (1998)). Neste trabalho foi utilizada a metodologia de uma etapa. Esta metodologia consiste na resolução do problema (DTE), de modo a obter, para cada uma das N DMU's, a medida de eficiência técnica de Farrell (1957), θ_n^* , o que permite identificar as DMU's eficientes ($\theta_n^* = 1$) e as DMU's ineficientes ($\theta_n^* < 1$).

Para cada DMU n ineficiente, a determinação do ponto projectado eficiente é feita através do cálculo residual dos desvios.

Represente-se, então, por $\mathbf{s}_{\bullet n}^+$ o vector dos desvios dos *outputs* associados à DMU n , de dimensão ($M \times 1$), e por $\mathbf{s}_{\bullet n}^-$ o vector dos desvios dos *inputs*, de dimensão ($K \times 1$).

O vector $\mathbf{s}_{\bullet n}^+$ indica, para cada *output* da DMU n , a quantidade em que esta o pode aumentar sem que para isso tenha que diminuir outro *output* ou aumentar algum dos *inputs*.

Os valores de $\mathbf{s}_{\bullet n}^-$, indicam, para cada *input* da mesma DMU, a quantidade em que esta o pode diminuir sem ser necessário aumentar outro *input* ou diminuir algum *output*.

Assim, a metodologia de uma etapa pode resumir-se por:

² No que se segue este PL será referido como o problema (NIRS).

Metodologia E1:

0. Dados: N, K, M, $X_{(K \times N)}$ (matriz dos inputs), $Y_{(M \times N)}$ (matriz dos outputs);
Fazer $n \leftarrow 1$;
1. Determinar θ_n^* e $(\lambda_j^n)^*$ solução óptima (s.o.) de (DTE) ((VRS), ou (NIRS));
2. {Cálculo dos Devios}
 $k \leftarrow 1$;
3. Se $k \leq K$ fazer: $s_{kn}^- \leftarrow x_{kn} \theta_n^* - \sum_{j=1}^N x_{kj} (\lambda_j^n)^*$
 $k \leftarrow k + 1$
voltar a 3
caso contrário (c.c.), ir para 4;
4. $m \leftarrow 1$;
5. Se $m \leq M$ fazer: $s_{mn}^+ \leftarrow \sum_{j=1}^N y_{mj} (\lambda_j^n)^* - y_{mn}$
 $m \leftarrow m + 1$
voltar a 5
c.c., ir para 6;
6. $n \leftarrow n + 1$
Se $n > N$ FIM
c.c., voltar a 1.

No passo 1, determina-se θ_n^* e $(\lambda_j^n)^*$ solução óptima de um só dos problemas (DTE), (VRS) ou (NIRS), consoante se pretenda rendimentos constantes à escala, rendimentos variáveis à escala ou rendimentos não crescentes à escala, respectivamente.

5. Resultados**5.1 A Nível Nacional****Todo o País**

Utilizando o programa DEAP (Coelli (1996)), a DEA aplica-se a cada um dos três anos em separado, quer com CRS quer com VRS, o que permite obter os resultados de eficiência técnica total (TE total), eficiência técnica pura (PTE) e eficiência de escala (SE) para os três anos em estudo. Fornece também informação relativa aos desvios, às DMU's de referência (aos respectivos pesos) e às metas, para cada EC da amostra. Contudo, dada a dimensão da amostra considerada (736³ EC's) serão apenas apresentados resultados em termos gerais.

Com a aplicação da metodologia de uma etapa aos dados dos três anos em separado, obtiveram-se os resultados constantes do quadro 1.

³ Embora a amostra inicial tenha 910 EC's algumas tiveram que ser retiradas quando do estudo a nível nacional, dadas as suas diferentes características.

Ano	Valores Médios			Valores Máximos			Valores Mínimos		
	TE(total)	PTE	SE	TE(total)	PTE	SE	TE(total)	PTE	SE
1997	0.454	0.684	0.708	1.000	1.000	1.000	0.062	0.200	0.118
1998	0.489	0.711	0.724	1.000	1.000	1.000	0.124	0.208	0.124
1999	0.467	0.699	0.708	1.000	1.000	1.000	0.121	0.231	0.121

Quadro 1 - Medidas de Eficiência

Tendo em conta os resultados obtidos com VRS (PTE), pode afirmar-se que, em média, as Estações de Correio portuguesas poderiam ter efectuado uma redução de trabalho em cerca de 32% (1 - 0.684), de 29% (1 - 0.711) e de 30% (1 - 0.699), respectivamente, para os anos de 1997, 1998 e 1999, sem reduzir nenhum dos *outputs*.

Pode também verificar-se que a eficiência se tem mantido relativamente constante ao longo do período considerado (aproximadamente 70%).

A análise dos resultados obtidos com VRS revelou que o número de EC's tecnicamente eficientes, das 736, era de 219 em 1997, 231 em 1998 e de 241 em 1999, representando cerca de 30%, 31% e 33% respectivamente para aqueles anos.

O número de EC's com uma eficiência técnica (pura) menor que 0.5 (metade da eficiência de uma EC de melhor prática) é de 147 em 1997, 112 em 1998 e 139 em 1999, o que representa cerca de 20%, 15% e 19% respectivamente para os três anos.

Repare-se que embora se note um aumento no número de estações eficientes de 1997 para 1999, este não é acompanhado por uma diminuição significativa no número de EC's pouco eficientes.

Cada uma das Estações de Correio foi também avaliada de modo a caracterizar a sua escala de operação.

Considerando, por exemplo, o ano de 1999, tem-se que aproximadamente 5% das EC's exhibe rendimentos constantes à escala (CRS), enquanto 48.5% funciona sob rendimentos decrescentes à escala (DRS) e 46.6% sob rendimentos crescentes à escala (IRS). Conforme se pode verificar no quadro 2, a percentagem de EC's com cada um dos tipos de rendimentos à escala não varia muito de ano para ano.

Ano	IRS	CRS	DRS
1997	53.7%	3.8%	42.5%
1998	46.1%	4.1%	49.9%
1999	46.6%	4.9%	48.5%

Quadro 2 - Rendimentos à Escala

Lisboa, Porto e Resto do País

No que diz respeito a esta amostra, foi analisado em separado a eficiência das Estações de Correio das principais cidades do país: Lisboa e Porto.

A metodologia DEA foi, então, aplicada às Estações de Correio de Lisboa com o objectivo de apurar se se verificavam algumas alterações ao nível dos resultados de eficiência técnica (pura).

Toda a amostra (736)		EC's de Lisboa (142)	
Nº de EC's tecnicamente eficientes	% de EC's tecnicamente eficientes	Nº de EC's tecnicamente eficientes	% de EC's tecnicamente eficientes
34	24%	47	33%

Quadro 3 - EC's de Lisboa Tecnicamente Eficientes

No quadro 3, a 2ª coluna indica a percentagem de EC's de Lisboa tecnicamente eficientes quando se considera toda a amostra, ou seja, para além de serem comparadas entre si, as estações de Lisboa são também comparadas com as EC's do resto do país. Na 4ª coluna apresenta-se a percentagem de EC's de Lisboa tecnicamente eficientes quando se consideram apenas as estações de Lisboa. Se as estações de Lisboa apresentam um maior número de EC's eficientes quando analisadas em separado, isto significa que elas estão a ser comparadas com estações menos eficientes, ou seja, pode concluir-se que muitas das EC's eficientes não se localizam em Lisboa, mas sim no resto do País.

No quadro 4, que se apresenta de seguida, pode constatar-se que o mesmo se verifica com as EC's do Porto.

Toda a amostra (736)		EC's do Porto (88)	
Nº de EC's tecnicamente eficientes	% de EC's tecnicamente eficientes	Nº de EC's tecnicamente eficientes	% de EC's tecnicamente eficientes
27	31%	40	46%

Quadro 4 - EC's do Porto Tecnicamente Eficientes

Qualidade

Posteriormente, restringe-se a amostra (361 EC's), de modo a incluir no modelo informação adicional relativa à qualidade do serviço prestado, apenas disponível para algumas das estações.

Esta informação é de extrema importância para o estudo da eficiência das EC's. Como frisado anteriormente, dado a procura de produtos nas EC's ser exógena, este é o único output sobre o qual os gerentes das estações têm algum controlo, esforçando-se por reduzir o tempo médio de espera (TME), como forma de aumentar a eficiência.

Analisando os desvios da variável TME conclui-se que apenas em 105 EC's da amostra considerada (correspondendo, aproximadamente, a 30%) o tempo médio de espera pode ser reduzido, sendo a redução média de 0.8 minutos, que pode ser considerado um valor relativamente baixo.

Contudo, há que ter atenção à relação existente entre o número de horas de trabalho e o tempo médio de espera, já referida anteriormente. Como se sabe, estes estão inversamente correlacionados, pelo que a melhoria de um deles provocará um declínio no outro. A solução passará por um ajustamento mais exacto entre o número de trabalhadores e a procura de produtos, o que se consegue prevendo melhor o fluxo desta e programando o horário do pessoal em conformidade.

5.2 A Nível Regional

Esta fase do estudo tem como principal objectivo elaborar médias de eficiência a nível regional. No universo das Estações de Correio, a análise regional é feita por Direcção Comercial (DC).

Vaz (2000) apresentou um estudo da eficiência de estações de correio na zona de Bragança. Neste trabalho são estudadas todas as regiões do País, sendo comparadas as diferentes regiões.

O quadro seguinte, indica a percentagem de EC's tecnicamente eficientes em cada DC no número total de estações da respectiva DC, a eficiência mais baixa e a média de eficiência, por DC.

	Nº de EC's	Nº de EC's tecnicamente eficientes	% de EC's tecnicamente eficientes	Eficiência técnica mais baixa	Média de eficiência técnica (%)
DCD	39	8	20.51%	0.298	0.569
DCP	149	38	25.50%	0.295	0.694
DCB	136	55	40.44%	0.223	0.616
DCI	40	12	30.00%	0.176	0.455
DCE	74	24	32.43%	0.245	0.537
DCS	86	18	20.93%	0.180	0.526
DCL	101	14	13.86%	0.200	0.627
DCJ	64	22	34.38%	0.322	0.528
DCG	41	10	24.39%	0.236	0.490
DCM	24	7	29.17%	0.333	0.585
DCA	30	14	46.67%	0.241	0.473

Quadro 5 - Resultados Regionais (por DC)

A simples observação dos resultados constantes do quadro anterior, mostra que as DC's de Lisboa (DCL) e Lisboa/Setúbal (DCS) têm um baixo número de estações eficientes.

De modo a detectar se a eficiência varia a nível regional pode ser útil examinar a proporção das EC's de cada região que são tecnicamente eficientes.

Pode verificar-se que, por exemplo, a DCA (Açores), que apresenta nos três anos a proporção mais elevada de EC's eficientes, apresenta nos mesmos anos uma das piores médias de eficiência. Enquanto, por exemplo, a DCL (Lisboa), apesar de apresentar a mais baixa proporção de EC's eficientes, apresenta também a segunda melhor média de eficiência.

As DC do Porto (DCP), da Beira Litoral (DCB) e de Lisboa (DCL) são as que apresentam uma média de eficiência mais elevada. Enquanto as DC da Beira Interior (DCI) e dos Açores (DCA) apresentam as piores médias.

Estas disparidades verificadas a nível regional poderão ser consequência, para além do potencial de mercado, indutor de menores ou maiores níveis de *output*, de práticas de gestão diferenciadas, do número de pessoal afecto por região, ou ainda devido às condições de trabalho, em termos de organização e utilização de novas tecnologias, de cada área.

6. Conclusões

Neste estudo, faz-se uma análise da eficiência técnica das Estações de Correio Portuguesas, com dados para o período de 1997 a 1999. Esta análise é realizada com base na metodologia DEA que, para além de permitir avaliar a *performance* de cada estação, fornece também indicações quanto à afectação de recursos, estabelecendo metas para as estações ineficientes, com o objectivo de orientar e melhorar o seu desempenho, para que possam atingir a "melhor prática".

A nível nacional verifica-se que a eficiência técnica poderia ser melhorada significativamente, optimizando o *input* trabalho. Conclui-se que a eficiência técnica pura é apenas 70% daquela que é possível, o que torna necessário efectuar alterações no uso dos recursos.

Quanto ao tempo médio de espera por cliente, e apenas para as estações consideradas nesta fase do estudo, conclui-se que este poderia ser reduzido, em média, em 0.8 minutos. O que pode ser conseguido programando melhor o horário do pessoal e em conformidade com a procura prevista.

A nível regional verificam-se diferenças significativas entre Direcções Comerciais distintas, admitindo-se que tal se possa relacionar com práticas profissionais diferentes ou com as condições de trabalho específicas de cada área.

Contudo, e como é usual em metodologias para o estudo da eficiência de DMU's, os resultados aqui obtidos devem ser encarados com algum cuidado. Isto porque a metodologia utilizada (DEA) caracteriza as diferenças entre os valores observados e os valores da fronteira eficiente identificada, quando poderão existir alguns factores aleatórios, entre outros, que escapam ao controlo dos agentes decisores e que se traduzirão na impossibilidade de localização, na fronteira eficiente, de EC's tão eficientes como as da fronteira. Note-se ainda que a DEA identifica a fronteira com base nas DMU's mais eficientes da amostra, sendo por vezes necessário a introdução de DMU's de referência fictícias para que as conclusões não sejam falseadas por amostras pouco eficientes.

No entanto, apesar das limitações apontadas à metodologia DEA, esta pode revelar-se um instrumento bastante útil à Gestão, pois para além de permitir avaliar o desempenho de cada estação fornece também indicações quanto à afectação de recursos, estabelecendo metas para as estações ineficientes, com o objectivo de orientar e melhorar o seu desempenho.

Referências

- [1] Banker, R.D., Charnes, A., Cooper, W.W., *Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis*, Management Science 30 (1984) 1078-1092.
- [2] Battese, G.E., Coelli, T., *A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data*, Empirical Economics 20 (1995) 325-332.
- [3] Battese, G.E., Heshmati, A., Hjalmarsson, L., *Efficiency of Labor - Use in the Swedish Banking Industry: A Stochastic Frontier Approach*, SEE/EFI Working Paper Series in Economics and Finance (1998) 289.
- [4] Charnes, A., Cooper, W.W., Rhodes, E., *Measuring the Efficiency of Decision Making Units*, European Journal of Operational Research 2 (1978) 429-444.
- [5] Coelli, T., *A Guide to DEAP Version 2.1.: A Data Envelopment Analysis (Computer) Program*, CEPA Working Papers 96/08, Department of Econometrics, University of New England, Armidale (1996).
- [6] Coelli, T., Rao, D.S.P., Battese, G.E., *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishers (1998).
- [7] Doble, M., *Measuring and Improving Technical Efficiency in UK Post Office Counters Using Data Envelopment Analysis*, Annals of Public and Cooperative Economics 66 (1995) 31-64.
- [8] Farrell, M.J., *The Measurement of Productive Efficiency*, Journal of the Royal Statistical Society, Series A, CXX, Part 3 (1957) 253-290.
- [9] Santos, J.T., Themido, I., *Introdução à Técnica de DEA - Data Envelopment Analysis*, Investigação Operacional 14 (1994) 159-181.
- [10] Vaz, M.A., *Aplicação DEA à Análise da Eficiência de Estações de Correio*, in Casos de Aplicação da Investigação Operacional, Antunes, C.H. e Tavares, L.V. (eds), Mc Graw Hill (2000) 364-407.

UM CASO DE ESTUDO DE INTEGRAÇÃO SIG-DEA-MCDA: A INFLUÊNCIA DE UMA INSTITUIÇÃO DE ENSINO SUPERIOR EM VÁRIOS MUNICÍPIOS DO ESTADO DO RIO DE JANEIRO

João Carlos Correia Baptista Soares de Mello

Dept. de Eng. de Produção - Univ. Federal Fluminense
Escola de Engenharia, Campus da Praia Vermelha
São Domingos, CEP: 24000-000 Niterói
RJ, Brasil

Eliane Gonçalves Gomes

Marcos Pereira Estellita Lins

Programa Eng. Produção - COPPE - Univ. Federal do RJ
Cidade Universitária, Ilha do Fundão
Centro de Tecnologia, Bloco F, Sala F-105
CEP: 21945-970 Rio de Janeiro
RJ, Brasil

Luiz Alberto Maron Vieira

Coordenadoria de Seleção Acadêmica - Univ. Federal Fluminense
Rua Doutor Celestino, 76-6º andar, Centro
CEP: 24000-000 Niterói
RJ, Brasil

Abstract

Widespread use of Data Envelopment Analysis is hampered by the difficulty felt by non-specialist decision-makers to understand its processes. The present paper outlines a method that helps to find a solution for this problem in such a case when DMU's are spread over a given space, integrating Geographical Information Systems and DEA models. The method is applied to the real case study of Universidade Federal Fluminense's geographical expansion, as measured by indicators such as population and presence of universities in the cities concerned (inputs) and number of students enrolled (outputs). Finally, the influence of the University in each city is mapped, using the MACBETH approach.

Resumo

A difusão da Análise Envoltória de Dados esbarra na dificuldade que tomadores de decisão não especialistas têm em entender seus modelos. Este trabalho apresenta um método para auxiliar na solução deste problema, no caso em que as DMU estão espacialmente distribuídas, integrando Sistemas de Informação Geográfica (SIG) e modelos DEA. O método é aplicado ao caso de estudo real da expansão geográfica da Universidade Federal Fluminense, medida por indicadores relativos à população e presença da universidade nas cidades (*inputs*) e ao número de alunos admitidos (*outputs*). Finalmente, é mapeada a influência da Universidade em cada cidade, usando-se para tal a metodologia MACBETH.

Keywords

Data Envelopment Analysis, Geographical Information Systems, SIG-DEA Integration, Education, MACBETH.

1. Introdução

A dificuldade de aceitação dos modelos de Investigação Operacional (IO) por decisores que não tenham sólida formação em métodos quantitativos é uma das causas da falta de aplicabilidade desses modelos, que pode conduzir à estagnação da IO (Clímaco, 1999) [15].

A criação de interfaces visuais pode ser utilizada para promover uma melhor aceitação de modelos e métodos de IO, em particular os que envolvem localização, ou análise de unidades geograficamente distribuídas. Os Sistemas de Informação Geográfica (SIG) tornam-se uma ferramenta poderosa de visualização, quando é possível o geo-referenciamento das unidades de decisão em análise.

O presente trabalho estuda um caso real, e de crucial importância profissional de um dos autores, usando a integração entre os SIG e o modelo DEA. As unidades de decisão (DMU) são as cidades do Estado do Rio de Janeiro. Pretende-se mostrar a eficiência da Universidade Federal Fluminense (UFF) em motivar jovens destas cidades a tentarem ingressar em seus cursos, bem como a eficiência das cidades em conseguir que seus habitantes ingressem no ensino superior. Os *inputs* são a população de cada cidade e uma variável que será denominada "vagas equivalentes". Os *outputs* são o número de ingressantes nos cursos da UFF (óptica do município) e o número de inscritos no exame de admissão, chamado de concurso vestibular (óptica da UFF). Todas estas variáveis, e as eficiências calculadas (usando o programa *Frontier Analyst*) são mostradas sob a forma de mapas produzidos em ambiente SIG.

2. Sistemas de Informação Geográfica

Segundo a definição de Burrough (1986) [12], os SIG são um poderoso conjunto de ferramentas para colecta, armazenamento, recuperação, transformação e visualização de dados do mundo real, devendo ser vistos, entretanto, como um instrumento capaz de realizar acções que vão além de codificar, armazenar e visualizar dados sobre aspectos da superfície terrestre. Os SIG permitem reunir uma grande quantidade de dados convencionais de expressão espacial, estruturando-os adequadamente, de modo a otimizar o tratamento de seus três componentes (posição, topologia e atributos). Desta forma, permitem a execução de análises e aplicações gráficas complexas, através de rápida formação e alteração de cenários, que propiciam apoio para a tomada de decisão (Barcellos e Santos, 1997) [7].

A chave da diferenciação de um SIG e outros sistemas de informação é seu foco em entidades espaciais e seus relacionamentos, com atenção especial às operações de modelagem e análise espacial (Gomes e Lins, 1999) [25]. Sua habilidade em organizar e integrar conjuntos diferentes de dados através de uma geografia, torna o SIG um poderoso instrumento, onde as buscas espaciais e as operações de sobreposição de camadas se apresentam como suas questões funcionais chave, como destacado por Maguire (1991) [30].

Ao contrário de outros tipos de dados manipulados em diferentes sistemas de informação, os dados geográficos são complexos por incluírem informações sobre posição, possíveis conexões topológicas, e atributos dos objectos em questão (Gomes e Lins, 1999) [25].

Burrough (1986) [12] afirma que os aspectos topológicos e espaciais do processamento de dados geográficos, distinguem sistemas projectados para gráficos e mapas daqueles que utilizam sistemas de processamento de dados. Segundo o mesmo autor, os SIG descrevem os objectos do mundo real em termos de:

- posição em relação a um sistema de coordenadas conhecido;
- inter-relações espaciais (relações topológicas), que descrevem como estes elementos estão unidos, ou como é possível percorrer estes elementos;
- atributos, não relacionados à posição (como cor, custos, incidência de certa doença, etc.).

Em ambiente SIG, para a representação de dados gráficos que descrevem a localização, as forma geográficas e os relacionamentos espaciais entre estas formas, são utilizados basicamente dois modelos de dados: o modelo *raster* ou matricial (representado por um conjunto de células localizadas por coordenadas), e o modelo vectorial (representado por três entidades geográficas principais, a saber, pontos, linhas e áreas). Conforme observa Burrough (1986) [12], ambos os modelos de representação do espaço são estruturas de dados válidas. A escolha do modelo mais adequado depende basicamente da aplicação à qual se destina, dos dados e do *software* disponíveis.

Assim, os SIG oferecem uma maneira potente de apoio à tomada de decisão local, espacialmente referenciada, em um contexto de planeamento local.

3. Análise Envoltória de Dados (*Data Envelopment Analysis - DEA*)

A Análise de Envoltória de Dados surgiu formalmente em 1978 a partir do trabalho de Charnes, Cooper e Rhodes [14], com o objectivo de medir a eficiência de unidades tomadoras de decisão, designadas por DMU (*decision making units*), na presença de múltiplos *inputs* (factores de produção) e múltiplos *outputs* (produtos). Por usar várias variáveis, que poderiam ser entendidas como critérios, tem sido estudada a relação entre DEA e Análise Multicritério (Doyle e Green 1993, [21], Stewart, 1996 [34], Gonzales-Araya et al., 1999 [26]). Alguns autores (Cook e Kress, 1990 [16], Sook e Kress, 1991 [17], Cook et al., 1992 [18], Belton e Vickers, 1993 [8], Sinyany-Stern et al., 1994 [33], Friedman e Sinuany-Stern, 1998 [24], Farinaccio, 1998 [22]) chegam a usar DEA como se fosse uma ferramenta de Multicritério, principalmente nas problemáticas de escolha e ordenação. Esta abordagem tem como característica a independência em relação à opinião do decisor na atribuição de pesos, o que, se por um lado pode levar à atribuição de pesos irrealistas, acaba por dar soluções mais independentes de eventuais manipulações. Esta característica torna os modelos DEA de grande utilidade para ordenar alternativas que possam, de alguma forma, ser consideradas como unidades de produção (isto é, possam-se identificar *inputs* e *outputs*) e onde se queira garantir a não influência de opiniões pessoais.

Podem ser destacadas as seguintes características da abordagem DEA (Lins e Meza, 2000) [29]:

- difere dos métodos baseados em avaliação puramente económica, que necessitam converter todos os *inputs* e *outputs* em unidades monetárias;
- os índices de eficiência são baseados em dados reais (e não em fórmulas teóricas);
- generaliza o método de Farrel (1954) [23], construindo um único *output* virtual e um único *input* virtual;
- é uma alternativa e um complemento aos métodos da análise da tendência central e análise custo benefício;
- considera a possibilidade de que os *outliers* não representem apenas desvios em relação ao comportamento "médio", mas possíveis *benchmarks* a serem estudados pelas demais DMUs (*Decision Making Unit*, definido como a alternativa cuja eficiência está sendo avaliada; o conjunto de DMUs adoptados em uma análise DEA deve ter em comum a mesma utilização de *inputs* e *outputs*, ser homogêneo e ter autonomia na tomada de decisões);
- a abordagem DEA, ao contrário das abordagens paramétricas tradicionais, otimiza cada observação individual com o objectivo de determinar uma fronteira linear por partes (*piece-wise linear*) que compreende o conjunto de DMUs Pareto-Eficiente. Além destas características, pode-se ainda destacar que a abordagem:
 - caracteriza cada DMU por uma única medida de eficiência relativa;
 - realiza projecções de melhoria de eficiência baseada nas melhores práticas;
 - prevê informações complementares sobre rendimento de escala (crescente ou decrescente).

Estas características conferem à abordagem uma potencialidade de resgatar a natureza essencialmente aplicada, presente na origem da IO. Em particular, os modelos DEA fornecem um indicador quantitativo e comparativo, o que são as características desejadas para indicadores em educação (Boclin, 1999) [11].

Foi utilizado neste trabalho o modelo CCR básico, (também conhecido por CRS ou *constant return to scale*), que trabalha com retornos constantes de escala (Charnes et al., 1978) [14], e óptica do *input*. Considera-se que cada DMU k é uma unidade de produção que utiliza n *inputs* y_{ik} , $i = 1, \dots, n$, para produzir m *outputs* x_{jk} , $j = 1, \dots, m$. Este modelo maximiza o quociente entre a combinação linear dos *outputs* e a combinação linear dos *inputs*, com a restrição de que para qualquer DMU este quociente não pode ser maior que 1. Em termos matemáticos, tem-se o seguinte problema de programação fraccionária (Charnes et al., 1996) [13]:

$$\text{Max } ef_k = \frac{\sum_{j=1}^m u_{jk} \cdot x_{jk}}{\sum_{i=1}^n v_{ik} \cdot y_{ik}}, \text{ com as restrições de } \frac{\sum_{j=1}^m u_{jk} \cdot x_{jk}}{\sum_{i=1}^n v_{ik} \cdot y_{ik}} \leq 1, \forall k; u_{jk} \text{ e } v_{ik} \text{ não negativos.}$$

Mediante alguns artifícios matemáticos, este modelo pode ser linearizado, transformando-se em um Problema de Programação Linear.

Apesar de todas as vantagens já enunciadas dos modelos DEA, o facto do seu resultado ser obtido através da resolução de vários problemas de programação linear, pode torná-lo de difícil compreensão (e portanto com pouca aceitação) para os gestores e tomadores de decisão, utilizadores finais dos modelos de IO. A tendência de sofisticar cada vez mais os modelos DEA, embora os torne mais completos, acaba por, como é óbvio, dificultar ainda mais o seu entendimento. Assim, quando não houver necessidades indiscutível de usar modelos mais sofisticados, o uso de modelos mais simples é vantajoso, principalmente se o problema em análise não for um exercício académico, pretendendo antes ser aceite e compreendido por terceiras pessoas. Existem na literatura alguns casos de modelos DEA bastante simples, aplicados com grande sucesso na análise de problemas com dados reais (Tavares e Antunes (1999) [35]).

4. Integração SIG - Modelagem Matemática

Uma das motivações à integração entre SIG e os modelos matemáticos está centrada na visualização como um importante auxílio à decisão. Enquanto tabelas estatísticas fornecem informações em várias dimensões, usadas na avaliação de soluções, esta mesma informação é mais facilmente compreendida quando é complementada com mapas e outros gráficos (Densham, 1996) [20]. É através desta análise detalhada de cenários alternativos que se obtém, no planeamento das acções e na performance dos resultados, uma visualização mais acurada dos possíveis impactos das decisões tomadas.

Para Birkin et al. (1996) [9] é através da integração entre *software* de SIG e a análise baseada em modelos, que se torna possível a obtenção de resultados oriundos da sinergia entre estas duas partes. É criada assim, uma ferramenta inteligente proveniente de duas correntes. A primeira é o desenvolvimento e uso de modelos matemáticos ou estatísticos, vitais para a geração de informações úteis nos mais variados ambientes de tomada de decisão e planeamento. A segunda é a obtenção de saídas mais informativas e úteis, que podem assumir a forma de indicadores de performance. Assim, é possível para planeadores e tomadores de decisão explorar uma série de cenários possíveis, obtendo uma ideia das consequências do curso da acção, antes que possíveis error se tornem irreversíveis.

Para Jankowski (1995) [27], o papel do SIG na implementação de modelos de tomada de decisão espacial não se restringe a encontrar alternativas viáveis; é também auxiliar o tomador de decisão a designar pesos de prioridade aos critérios de decisão, avaliar as alternativas viáveis e visualizar os resultados de sua escolha. A busca geralmente resulta na selecção de um certo número de alternativas que satisfazem valores limiares mínimos. A redução do conjunto de alternativas e a selecção da melhor alternativa geralmente requer o uso de técnicas de análise multicritério. Assim, a melhoria das capacidades do SIG na tomada de decisão pode ser alcançada pela introdução de técnicas multicritério no ambiente SIG. Esta é uma prática já

bastante aceite, e vários *softwares* de SIG já incorporam módulos de multicritério. O estado da arte da integração SIG-multicritério é apresentado por Gomes e Lins (1999) [25].

Devido às várias semelhanças de DEA com alguns modelos multicritério é bastante natural fazer uma integração semelhante entre SIG e DEA, centrada principalmente na visualização de dados e resultados, e com aproveitamento parcial dos cursos de acção já existentes para a integração SIG-Multicritério.

A integração SIG-DEA pode gerar uma potente ferramenta de auxílio à análise espacial através da modelagem, em especial para a localização/afecção de actividades considerando diversos critérios. Neste caso específico, os SIG permitem fornecer uma visão da localização das unidades de decisão em análise (DMUs), dos *inputs* e *outputs* (em separado ou conjuntamente), bem como do resultado final da análise (eficiências). A metodologia aqui proposta (Figura 1) é baseada naquela proposta por Jankowski e Richard (1994) [28] para a integração SIG-Multicritério, adaptada para o caso da integração SIG-DEA.

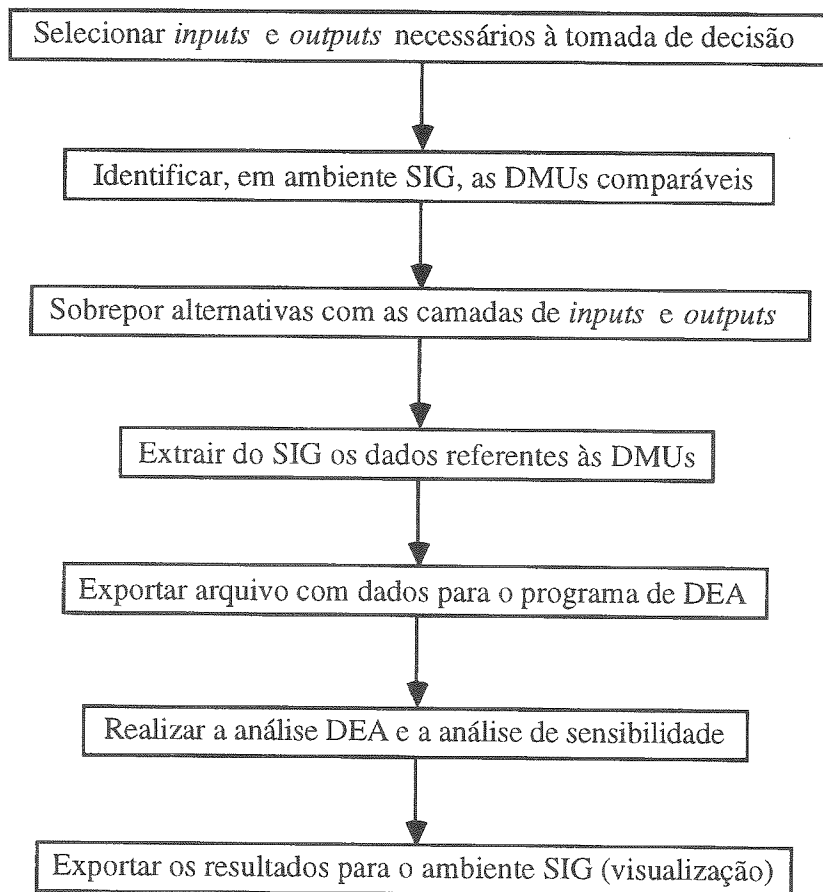


Figura 1 - Metodologia proposta para a integração SIG-DEA

5. Caso de Estudo

A Universidade Federal Fluminense (UFF) foi criada a partir da fusão de várias faculdades isoladas do município de Niterói, RJ, onde é sua sede (Figura 2). Posteriormente foram

agregados uma escola de Engenharia em Volta Redonda, uma escola de Serviço Social em Campos dos Goytacazes, e dois colégios agrícolas (COSEAC, 1999) [19].

A partir de 1984, através de um convénio com a Prefeitura de Santo António de Pádua, a UFF começou um projecto de interiorização de cursos de graduação, que hoje já se expandiu para os municípios de Angra dos Reis, Cabo Frio, Iteperuna, Macaé, e Miracema. A localização geográfica destes municípios pode ser vista na Figura 3.

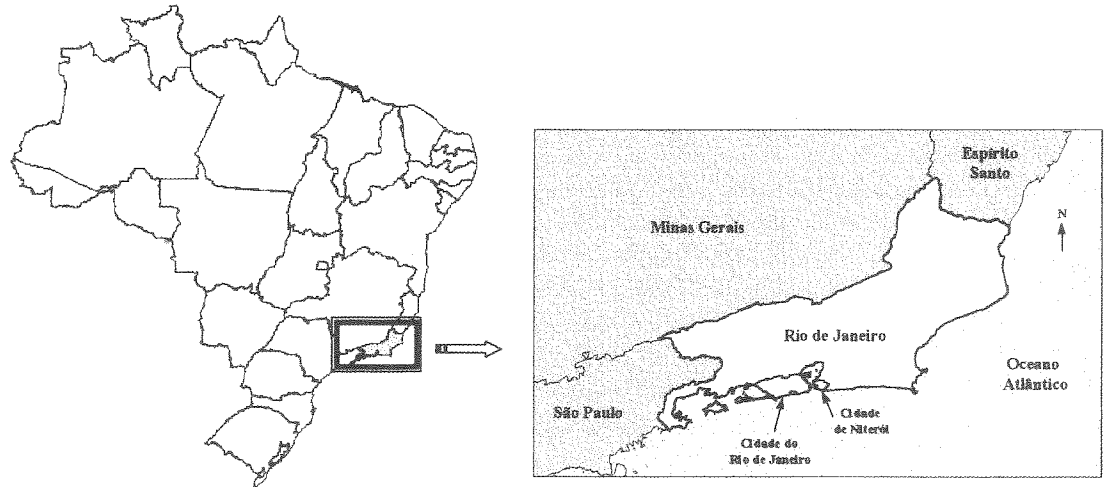


Figura 2 - Localização geográfica do estado do Rio de Janeiro, destacando-se sua capital (Cidade do Rio de Janeiro) e a cidade da UFF (Cidade de Niterói)

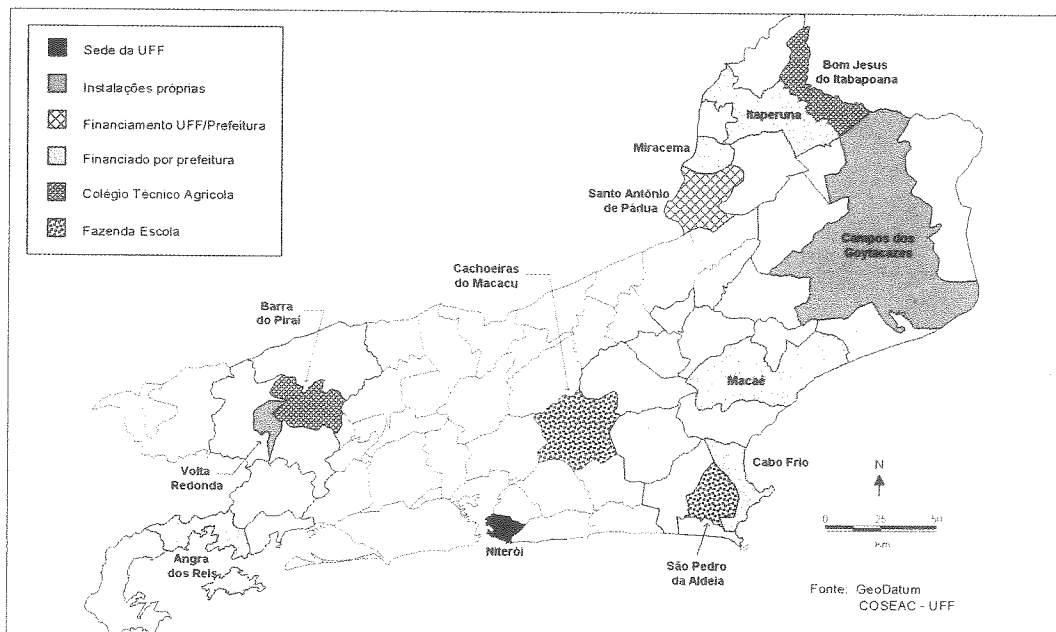


Figura 3 - Actuação da UFF no Estado do Rio de Janeiro

Até 1988 o ingresso nos cursos de graduação das várias universidades do estado do Rio de Janeiro, incluindo-se a UFF, era feito através de um exame único, organizado por uma entidade específica para este fim. Esse exame é denominado vestibular. A partir de 1989, cada universidade, adoptando critérios próprios, passou a realizar seu próprio vestibular.

Devido à interiorização já mencionada, a UFF, notadamente a partir de 1996, realiza a inscrição e as provas de seu vestibular em vários municípios do Estado do Rio de Janeiro, em número cada vez maior, mas escolhidos sem uma análise quantitativa da estrutura já existente e do interesse e utilidade da expansão. A Figura 2 representa a localização geográfica do Estado do Rio de Janeiro. A actuação da UFF e seu vestibular no estado podem ser vistos nas Figuras 3, 4 e 5.

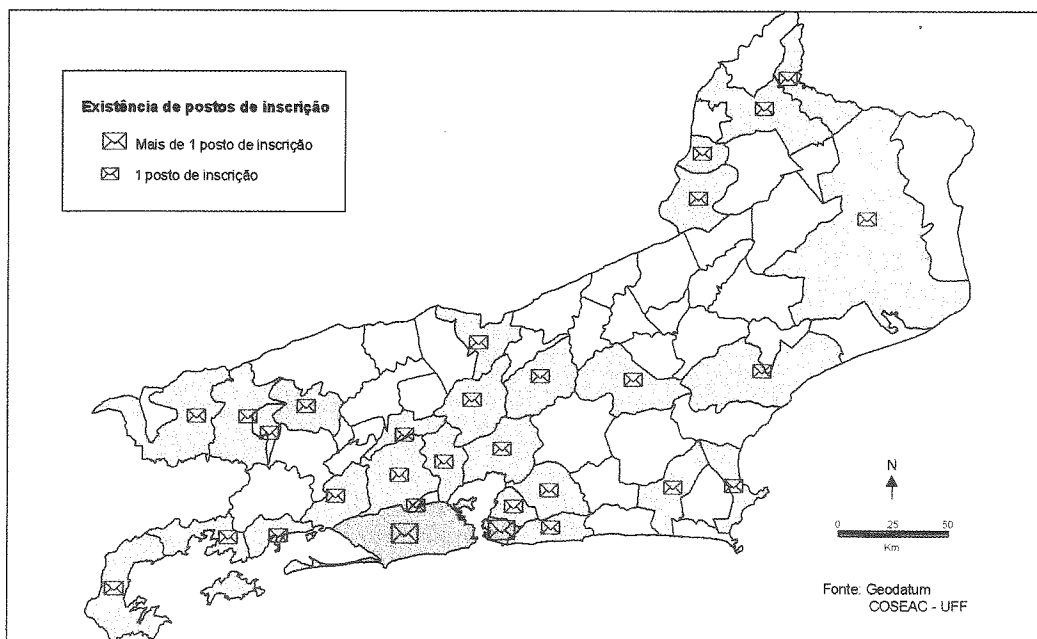


Figura 4 - Mapeamento da existência de postos de inscrição para o vestibular

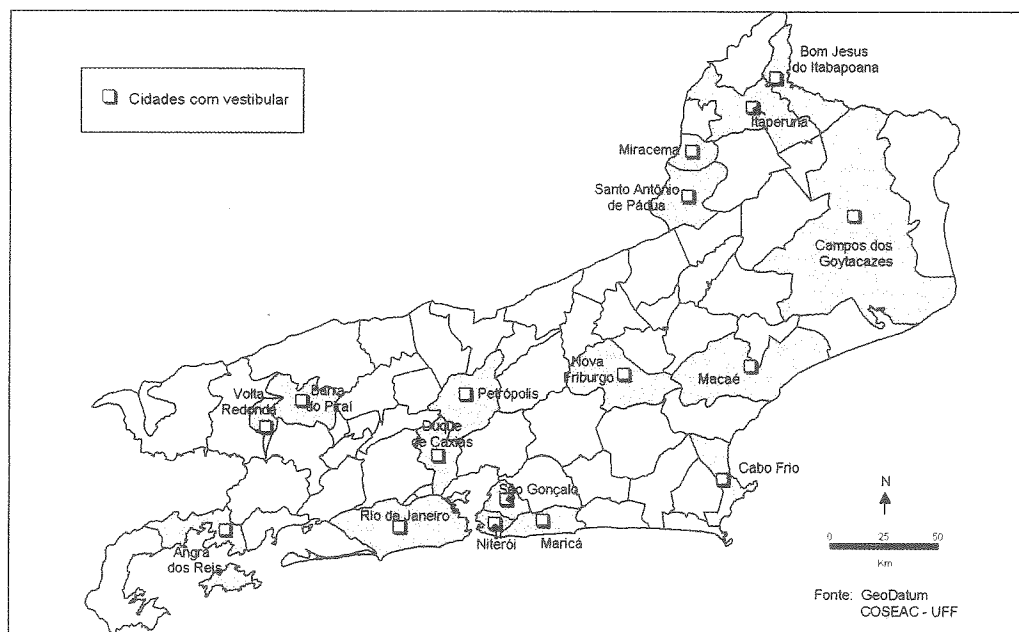


Figura 5 - Identificação dos municípios que são locais de prova do vestibular da UFF

Uma observação deve ser feita. Os mapas apresentados representam a configuração municipal do Estado do Rio de Janeiro para o ano de 1991, quando o mesmo possuía 70

municípios (atualmente o Estado possui 91 municípios). Assim, as instalações da UFF que na Figura 3 aparecem em Barra do Piraí e São Pedro da Aldeia, ficam na realidade em Pinheiral e Iguaba Grande, municípios que ainda não tinham sido criados quando da digitalização desta base gráfica para utilização em SIG.

Neste trabalho pretende-se avaliar a eficiência da UFF e dos municípios em relação ao vestibular. Para esta avaliação foi utilizado o modelo DEA CRS (*constant return to scale*), com dois *outputs*, a saber, número de inscritos e números de aprovados no vestibular, em cada município (Figuras 6 e 7), e dois *inputs*, população de cada município (1993) (Figura 10), e uma parâmetro denominado influência de vagas, ou vagas equivalentes. Este último parâmetro pretende ser uma suavização da variável "vagas oferecidas". Esta variável assume valor nulo na grande maioria dos municípios, não levando em conta a proximidade das cidades.

Com efeito, as vagas oferecidas, por exemplo, para cursos em Niterói influem sobremaneira nesta cidade. Mas têm igualmente uma grande influência nas vizinhas Rio de Janeiro, São Gonçalo e Maricá. Esta influência é já menor nas cidades quase vizinhas de Duque de Caxias, Itaboraí, Itaguaí, etc.. Em cidades a maior distância rodoviária, a influência vai decrescendo assintoticamente, embora seja tão pequena que possa ser considerada constante.

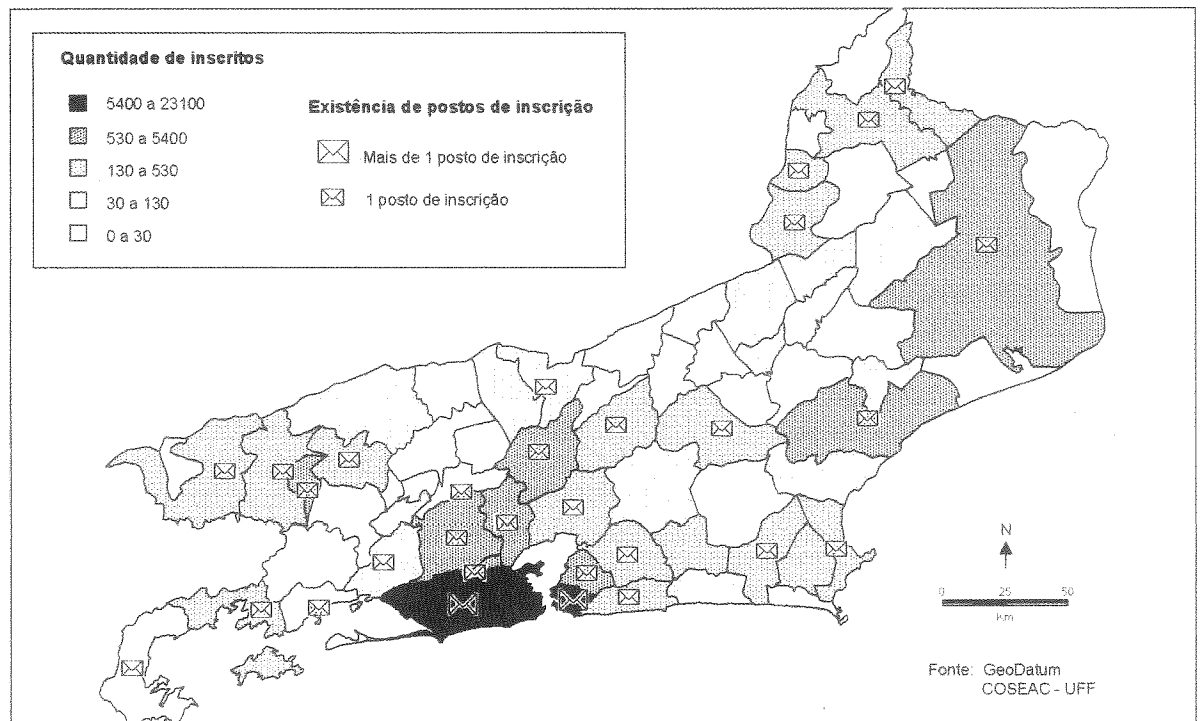


Figura 6 - Número de inscritos no vestibular da UFF de 1999

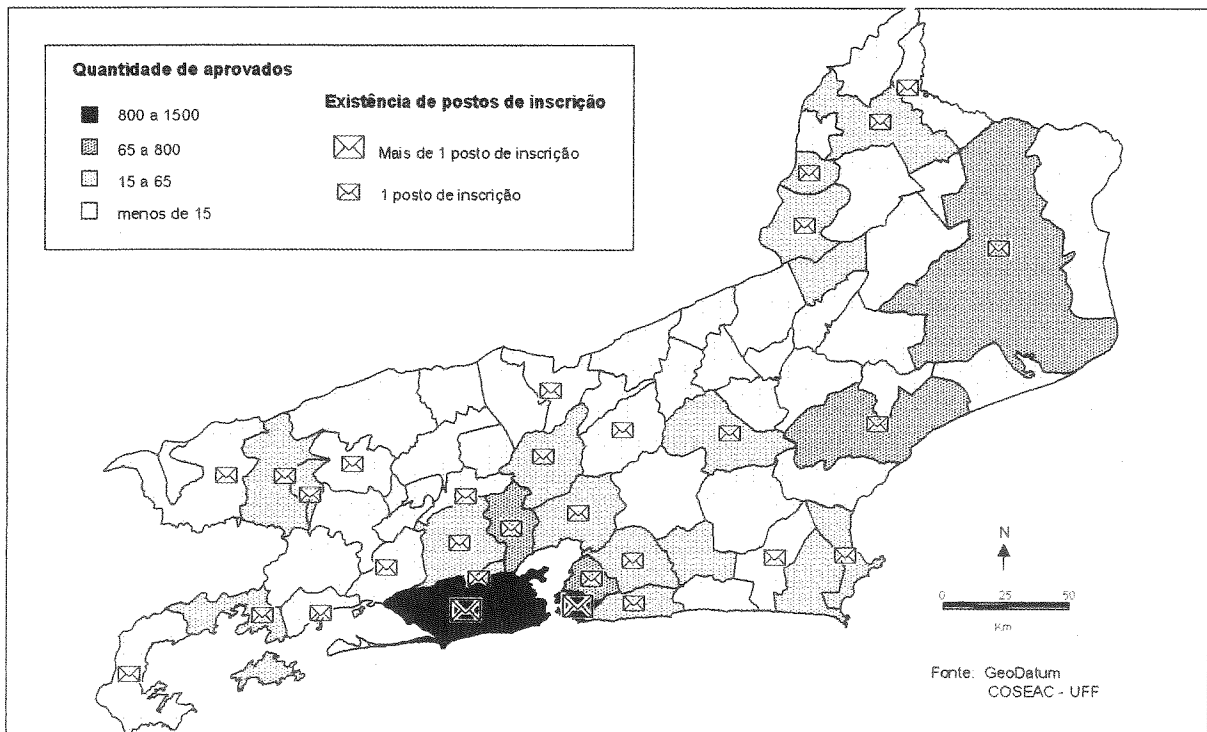


Figura 7 - Número de aprovados no vestibular da UFF de 1999

As vagas equivalentes (ou influência de vagas) são calculadas semi-empiricamente, da seguinte maneira: influência de vagas do município A sobre o município B é igual ao número de vagas que a UFF ofereceu no município A multiplicado por um coeficiente relacionado à ligação rodoviária entre os municípios A e B (Tabela 1, Figuras 6, 7 e 8), a partir de distâncias medidas em um grafo monovalorado (Boaventura Netto, 1996) [10] (Figura 9), onde cada aresta é ligação rodoviária entre dois municípios adjacentes e os vértices representam o centróide de cada município. Este parâmetro é semelhante à estimativa de Kernel (Bailey e Gatrell, 1995) [1]. A influência de vagas no município B é o somatório deste coeficiente para todos os municípios onde a UFF oferece vagas. O uso do *input* "vagas equivalentes" no lugar das vagas tira o carácter pontual do modelo, inserindo em cada DMU algumas características de DMUs próximas, ou seja, as propriedades topológicas são incorporadas ao modelo DEA.

Distância	Factor de multiplicação
0	1,0
1	0,9
2	0,5
> 2	0,1

Tabela 1 - Valores do coeficiente de influência de vagas

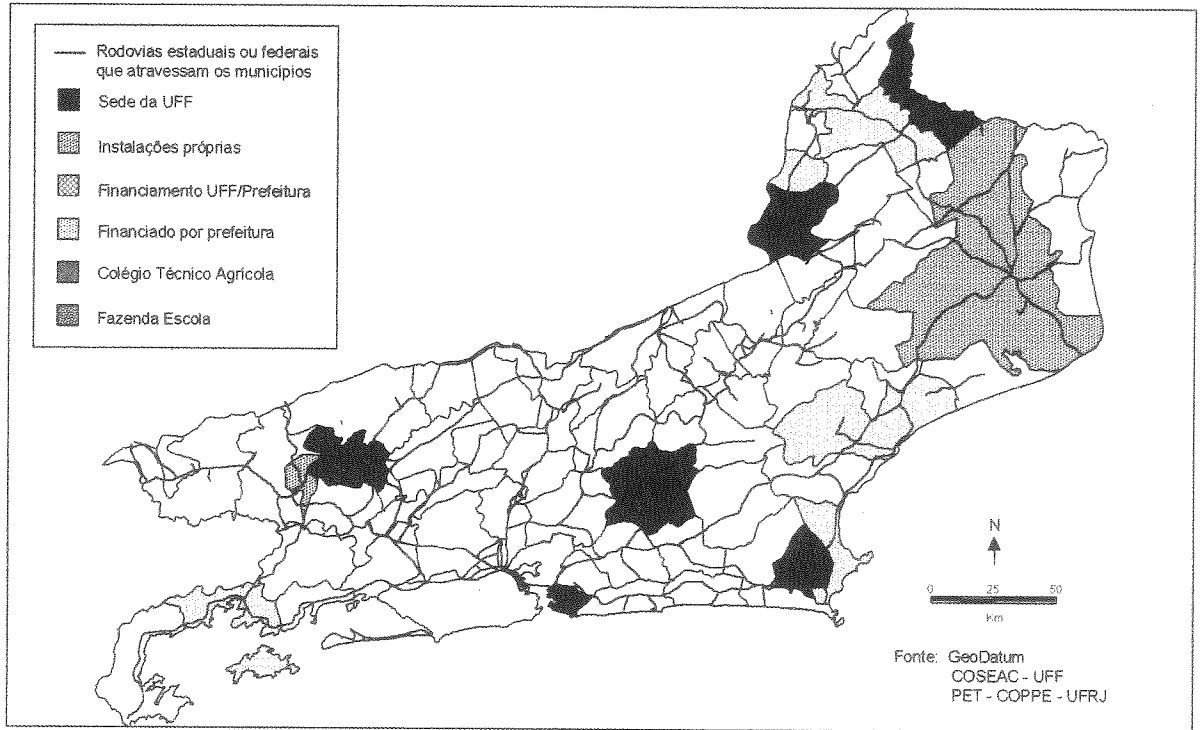


Figura 8 - Malha rodoviária do Estado do Rio de Janeiro

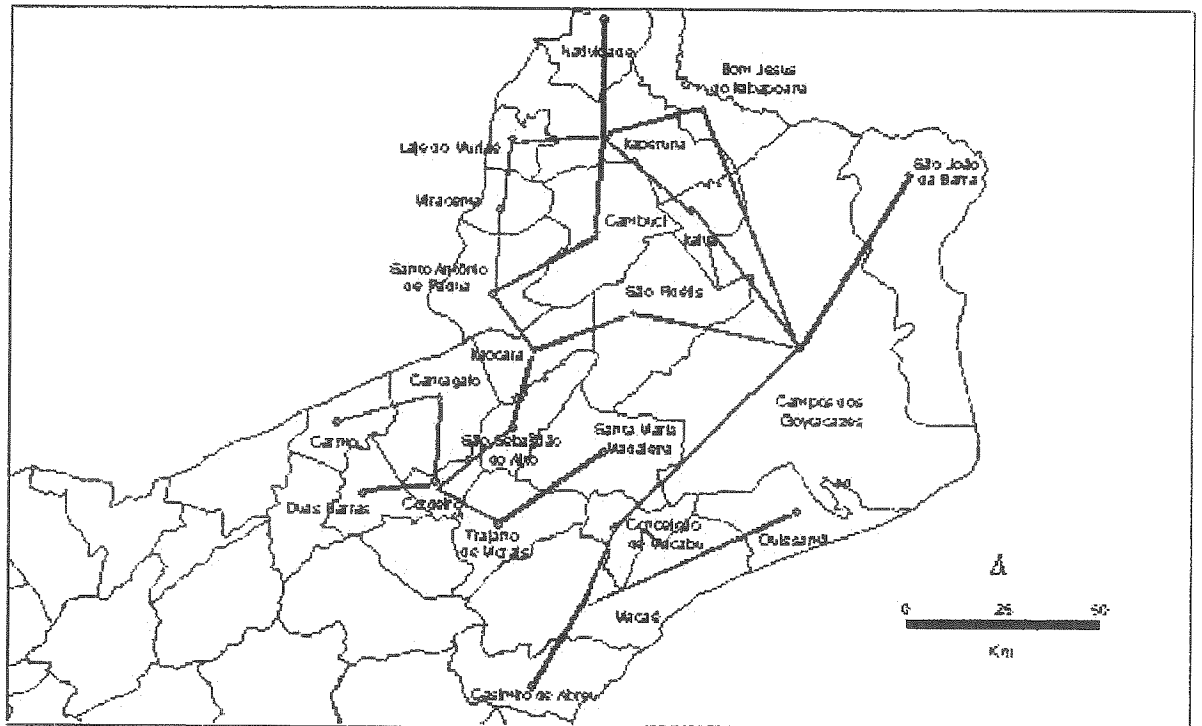


Figura 9 - Sub-grafo da malha rodoviária abrangendo as regiões Norte e Noroeste Fluminenses

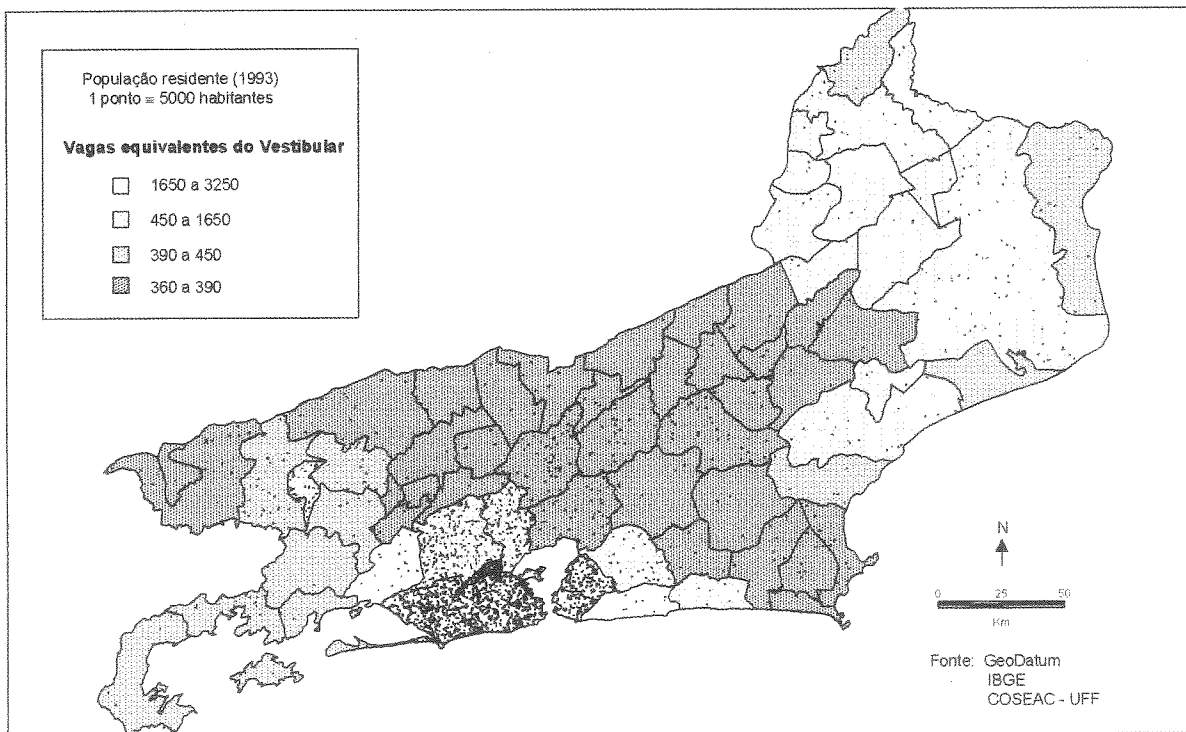


Figura 10 - *Inputs* do modelo DEA - população e influência de vagas

Na implementação do modelo DEA, com a finalidade de se obter uma melhor discriminação das DMUs, a abordagem tradicional, que considera uma única eficiência considerando todos os *outputs*, foi preterida em favor da abordagem proposta por Rangel et al. (1999) [31], na qual é calculada uma eficiência parcial para cada *output*, com uma posterior agregação de resultados. No presente caso, estas eficiências parciais são denominadas eficiência segunda a óptica dos municípios, quando o *output* é número de aprovados, e eficiência segunda a óptica da UFF, quando o *output* é número de inscritos no vestibular. A designação adoptada vem do facto de, para os municípios o importante é conseguir que os seus habitantes ingressem no ensino superior, enquanto para a UFF (pelo menos para a sua comissão de vestibular) o interesse é ter mais inscritos em cada cidade.

Os resultados da implementação deste modelo podem ser vistos nas Figuras 11 e 12. Um resultado não aparente nestes mapas é que os municípios do Rio de Janeiro (onde a UFF não tem nenhuma presença) e de Niterói (sede da UFF) apresentaram eficiência de 100% sob qualquer das duas ópticas adoptadas.

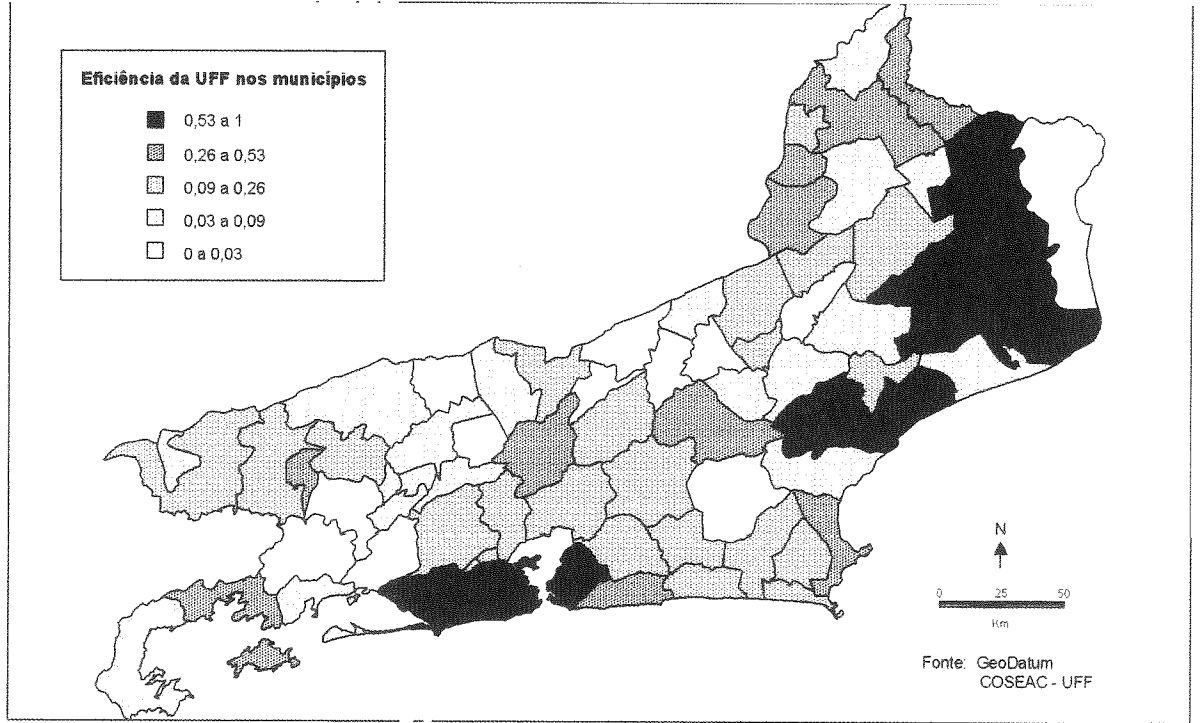


Figura 11 - Eficiência segundo a óptica da UFF

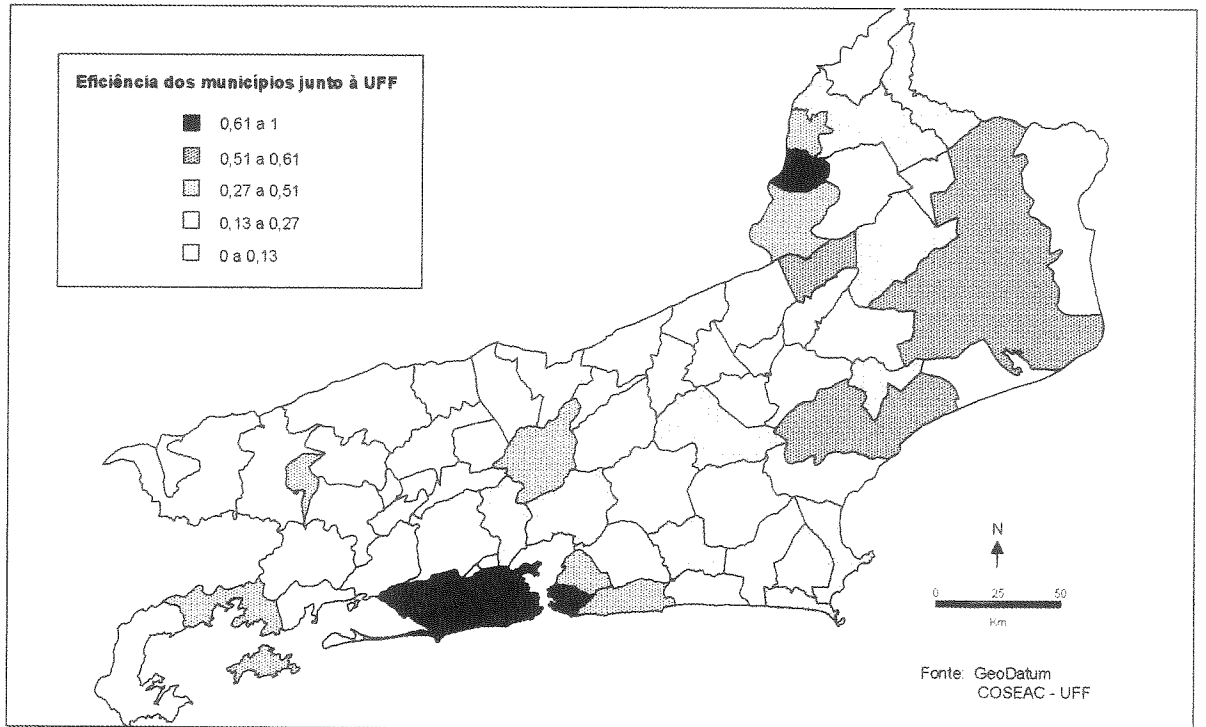


Figura 12 - Eficiência segundo a óptica dos municípios

Após o cálculo das eficiências, os seus valores, a forma de actuação directa da UFF, e a forma de actuação do seu vestibular em cada município foram considerados como critérios de avaliação da presença da instituição em cada cidade. Estes critérios foram agregados, através de

soma ponderada, em um critério síntese (Roy e Bouyssou, 1993 [32], Barba-Romero e Pomerol, 1997 [6]) que foi denominado "Índice de influência", ou "Função de Afinidade" da UFF em cada município. Ao fazer esta agregação deve-se levar em consideração que as eficiências já são uma escala cardinal, enquanto as formas de actuação são dadas em uma escala ordinal que tem que ser transformada em escala cardinal. Esta transformação foi feita usando a metodologia MACBETH (Bana e Costa e Vansnick, 1994 [2], 1995 [3]). As Tabelas 2 e 3 apresentam os resultados finais desta análise, já normalizada no intervalo [0,1]. Já as Figuras 13 e 14 mostram a tela com os resultados obtidos no *software* MACBETH, que levaram à construção destas escalas.

Formas de actuação da UFF	Quantificação normalizada
A1 (sede da UFF)	1,00
A2 (instalações próprias)	0,85
A3 (instalações cedidas*)	0,70
A4 (colégio ou fazenda agrícolas)	0,31
A5 (sem actuação)	0,00

Tabela 2 - Quantificação das formas de actuação da UFF

* instalações cedidas correspondem a Financiamento por Prefeitura e Financiamento UFF/Prefeitura da Figura 2

Formas de actuação do vestibular da UFF	Quantificação normalizada
A1 (local de prova e mais de um posto de inscrição)	1,00
A2 (local de prova e um posto de inscrição)	0,92
A3 (apenas mais de um posto de inscrição)	0,42
A4 (apenas um posto de inscrição)	0,25
A5 (sem local de prova e sem posto de inscrição)	0,00

Tabela 3 - Quantificação das formas de actuação do vestibular da UFF

A função de afinidade (FA) é definida como a soma ponderada dos vários critérios (C1, C2, C3 e C4) $FA_i = V_i.C1_i + X_i.C2_i + W_i.C3_i + Y_i.C4_i$, onde os pesos (V_i , X_i , W_i , Y_i) são dados pelo método MACBETH (Bana e Costa e Vansnick, 1997) [4], após as seguintes considerações hierárquicas: forma de actuação da UFF (critério C1i) é mais importante que a eficiência segundo a óptica do município (critério C2i), que é mais importante que a eficiência segundo a óptica da UFF (critério C3i), que por sua vez, é muito mais importante que a forma de actuação do vestibular (critério C4i). Neste contexto, considera-se que um critério α é mais importante que um critério β quando a alternativa fictícia a_i , ótima no critério α e péssima em todos os outros critérios, é preferível a uma alternativa fictícia a_j , ótima no critério β e péssima em todos os outros critérios.

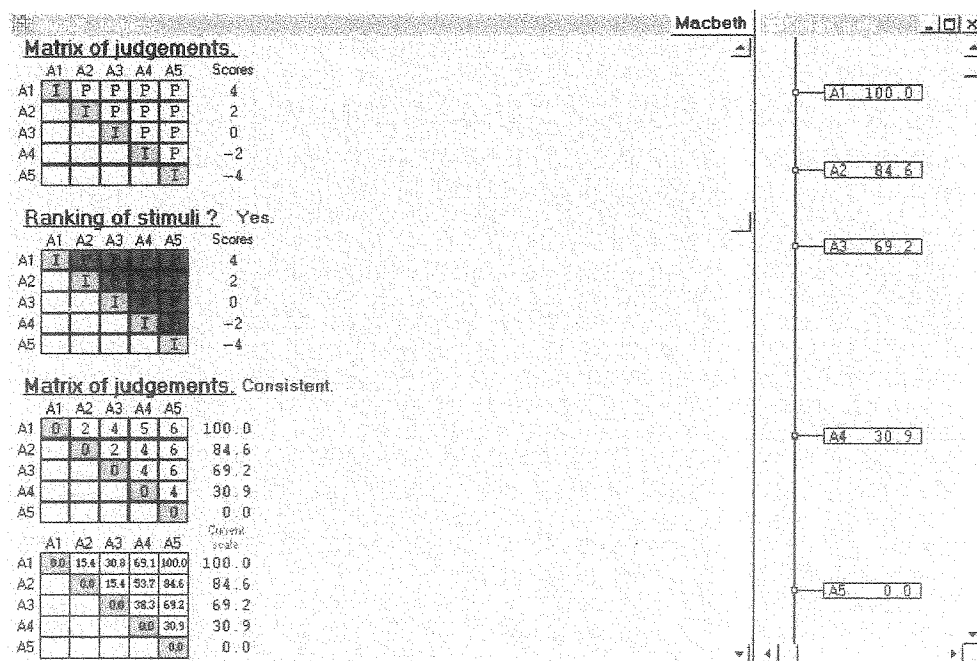


Figura 13 - Tela do MACBETH para construção da escala referente à actuação da UFF

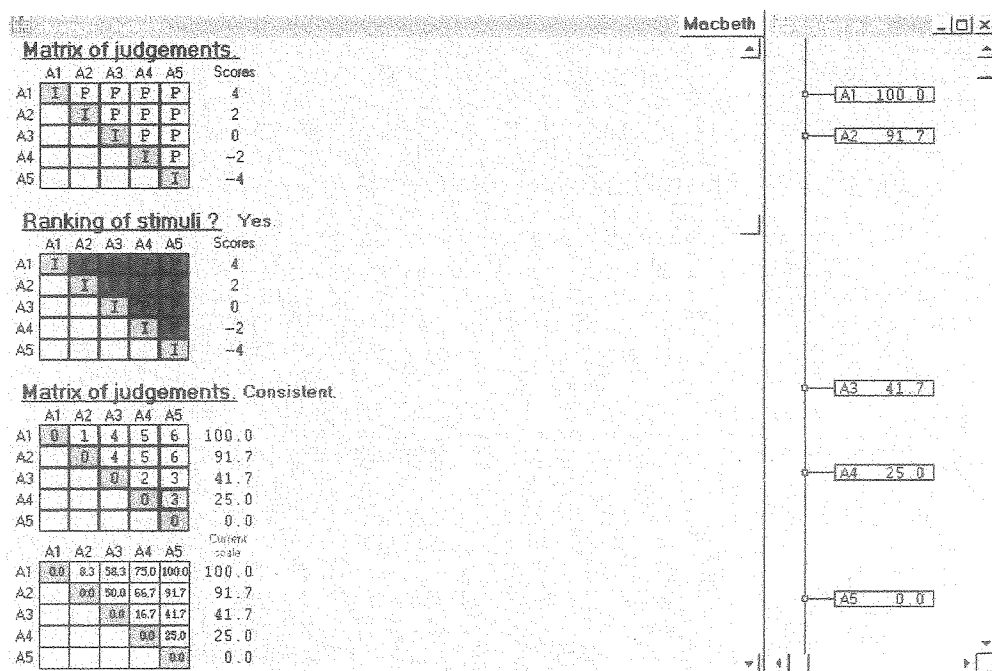


Figura 14 - Tela do MACBETH para construção da escala referente à actuação do vestibular

Devido às características da abordagem MACBETH foi considerado um critério artificial, em relação ao qual todos os outros são mais importantes, com a finalidade de impedir que algum critério possa ter peso igual a zero. Este critério artificial corresponde a uma alternativa que seria péssima em todos os critérios reais.

A Figura 15 mostra os resultados do cálculo dos pesos. As telas apresentadas nas Figuras 13, 14 e 15 não são aquelas fornecidas originalmente pelo MACBETH, mas já incorporam

alterações nas escalas, que consideram as experiências e sensibilidade dos decisores em relação ao problema em questão.

Finalmente, a distribuição espacial da função de afinidade é apresentada na Figura 16, destacando-se que somente o município Niterói obteve valor da função de afinidade igual a 1.

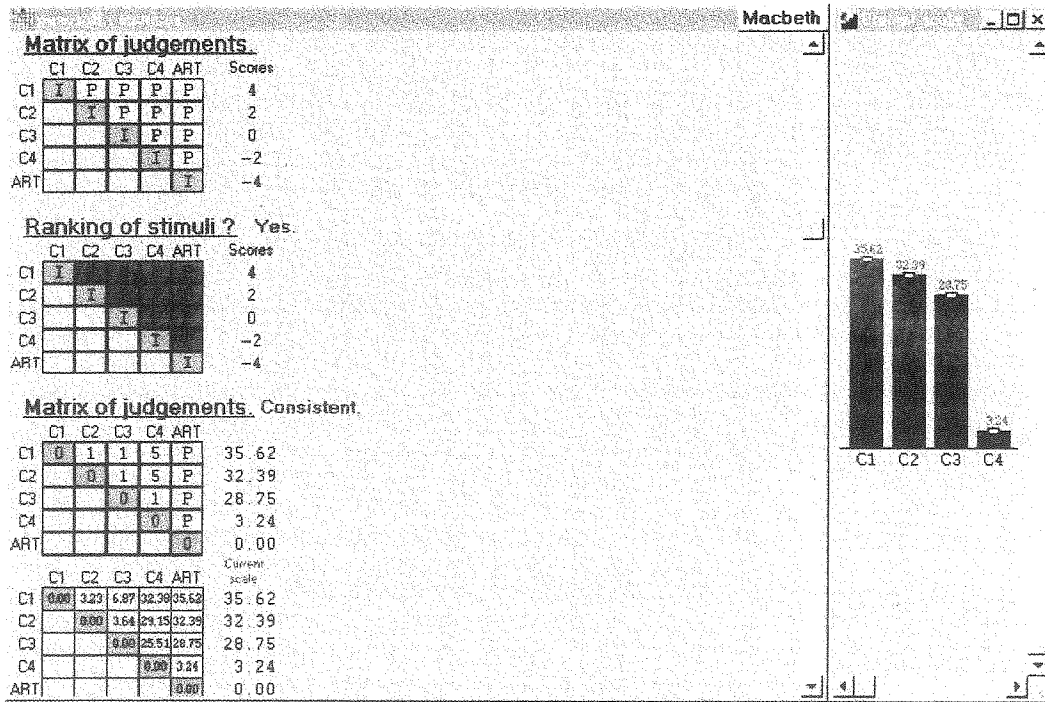


Figura 15 - Tela do MACBETH para cálculo da variação dos pesos da função de afinidade

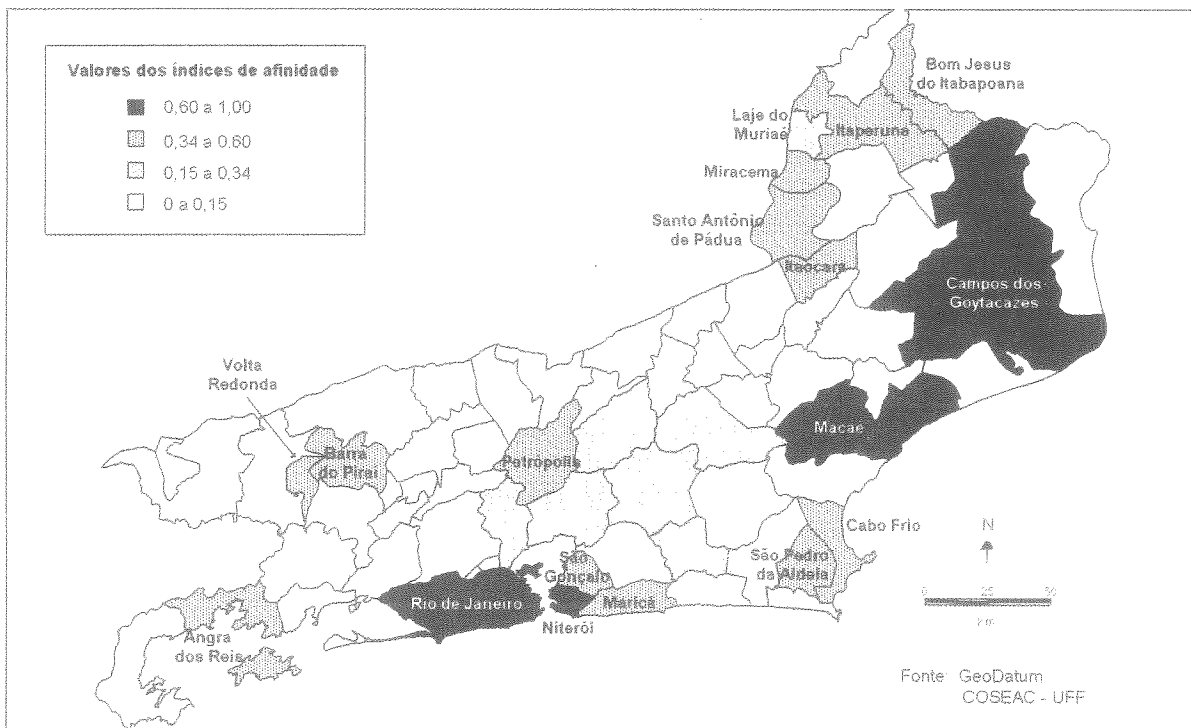


Figura 16 - Afinidade entre os municípios do Estado do Rio de Janeiro e a UFF

6. Conclusões

Como resultados pôde-se observar que os SIG, destacada a sua importante função de visualização, permitem uma boa caracterização da actuação de instituições, como a UFF, geograficamente dispersas, mostrando as regiões do estado e/ou os municípios com maior afinidade com a UFF. A região adjacente à sede, que inclui os municípios de Maricá, Niterói, Rio de Janeiro e São Gonçalo, e a região Norte Noroeste Fluminense aparecem como as de maior afinidade com a instituição analisada.

Além disto, o uso conjunto de modelos, análise multicritério e de SIG, permitirá, por exemplo, o planeamento de algumas características de outros vestibulares, identificando municípios que merecem especial atenção e outros que se mostraram pouco atractivos após a análise. A aceitação destes resultados pelos responsáveis pela tomada de decisão final é bastante melhorada pela facilidade de visualização, vencendo a habitual resistência a métodos quantitativos, por eles não facilmente compreendidos. Tal facto foi verificado em teste prático, quando pessoas em posição de influenciarem nas decisões, ao olharem os mapas, imediatamente viram a necessidade de aumentar a participação da UFF no município de Itaocara, onde a UFF está totalmente ausente, mas com o qual apresentou grande afinidade.

A Figura 16, que mostra a distribuição espacial das afinidades entre os municípios do Estado do Rio de Janeiro e a UFF, é uma excelente ferramenta de estudo. Deve-se, no entanto, evitar o erro de usá-la de forma absoluto e isolada. É preciso igualmente observar como cada município atingiu o seu índice de afinidade:

- na categoria superior (em preto no mapa da Figura 16), encontram-se três municípios com forte actuação da UFF, Niterói, Macaé e Campos dos Goutacazes, e ainda Rio de Janeiro, onde a UFF não tem actuação, mas está em sua esfera de influência, com o maior número de inscritos e aprovados no vestibular;
- na categoria seguinte encontram-se todos os demais municípios onde a UFF tem algum tipo de actuação: Angra dos Reis, Barra do Piraí (Pinheiral), Bom Jesus do Itabapoana, Cabo Frio, Itaperuna, Miracema, Santo António de Pádua, São Pedro da Aldeia (Iguaba Grande) e Volta Redonda. Encontram-se também três municípios onde a UFF não tem actuação directa, mas onde há uma forte presença de seu vestibular: Maricá e São Gonçalo, pela proximidade a Niterói, e Petrópolis, cujo grande índice de afinidade é em parte resultante pelo fato de lá se concentrarem a maioria dos candidatos oriundos do estado de Minas Gerais (ver Figura 2);
- ainda nesta categoria encontra-se o município de Itaocara, onde a UFF não possui actuação directa ou por meio de vestibular. Portanto, a sua inclusão nesta faixa é resultante do grande interesse dos habitantes de Itaocara em ingressar na UFF e de seu sucesso em conseguí-lo. Uma análise mais acurada, mostra que grande parte dos alunos que são aprovados para os cursos sediados em Santo António de Pádua, Miracema e Itaperuna são oriundos de Itaocara. Isto mostra uma necessidade estratégica de expansão da UFF para este município.

A vantagem do uso da função de visualização dos SIG fica bastante evidente quando se comparam os resultados apresentados na forma de mapas e os mesmos resultados apresentados de forma tabular, conforme apresentado na Tabela 4.

O uso integrado de SIG e modelos matemáticos supre os decisores de um ambiente de tomada de decisão que permita que a análise da informação geográfica seja conduzida de maneira flexível. O uso de SIG permite, de forma rápida e interactiva, sobrepor camadas de informação usando símbolos e cores. Assim, consegue-se uma agilidade na comparação e exploração prévia de dados. Este artigo, por se tratar de uma apresentação escrita, não consegue retractar por completo toda esta funcionalidade, que é perdida seja pelo facto da não interactividade, ou pela ausência de cores.

Município	População (1993)	Vagas equivalentes	Eficiência da UFF
Angra dos Reis	90132	400	0,42
Barra do Piraí	80565	412	0,12
Bom Jesus do Itabapoana	30087	484	0,28
Cabo Frio	89441	364	0,26
Campos dos Goytacazes	398169	474	0,56
Duque de Caxias	683451	1646	0,19
Itaguaí	116951	1686	0,02
Itaocara	23238	452	0,17
Itaperuna	80168	473	0,29
Macaé	106784	476	0,53
Maricá	49026	2928	0,42
Miracema	25717	468	0,33
Nilópolis	159197	1646	0,12
Niterói	442786	3248	1,00
Nova Friburgo	174507	364	0,31
Nova Iguaçu	1332173	1646	0,16
Petrópolis	259879	364	0,36
Resende	94307	388	0,12
Rio de Janeiro	5547033	2928	1,00
Santo António de Pádua	40587	470	0,28
São Gonçalo	807774	2928	0,61
Teresópolis	124447	364	0,19
Volta Redonda	226534	488	0,39

Tabela 4 - Alguns resultados na forma tabular

Referências

- [1] Bailey, T.C. and Gatrell, A.C., *Interactive Spatial Data*, Longman Group Limited, London (1995).
- [2] Bana e Costa, C.A. and Vansnick, J.C., MACBETH - An interactive path towards the construction of cardinal value functions, *International Transactions in Operations Research* 1 (1994) 489-500.
- [3] Bana e Costa, C.A. and Vansnick, J.C., A theoretical framework for Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH), in Clímaco J. (Ed.), *Multicriteria Analysis*, Springer Verlag, Berlin (1995) 15-24.
- [4] Bana e Costa, C.A. and Vansnick, J.C., Applications of the MACBETH approach in the framework of an additive aggregation model, *Journal of Multicriteria Decision Analysis* 6 (1997) 107-114.
- [5] Banker, R.D., Charnes, A. and Cooper, W.W., Some models for estimating Technical and Scale inefficiencies in Data Envelopment Analysis, *Management Science* 30 (1984) 1078-1092.
- [6] Barba-Romero, S. and Pomerol, J.C., *Decisiones Multicriterio: Fundamentos Teóricos e Utilización Práctica*, Colección de Economía, Universidad de Alcalá (1997).
- [7] Barcellos, C. and Santos, S.M., Colocando dados no mapa: a escolha da unidade espacial de agregação e integração de bases de dados em saúde e ambiente através do geoprocessamento, *Informe Epidemiológico do SUS Ano V* (1997) 21-29.
- [8] Belton, V. and Vickers, S.P., Demystifying DEA - A Visual Interactive Approach based on Multiple Criteria Analysis, *Journal of Operational Research Society* 44 (1993) 883-896.
- [9] Birkin, M., Clarke, G., Clarke, M. and Wilson, A., *Intelligent GIS: Location Decisions and Strategic Planning*, GeoInformation International, London (1996).
- [10] Boaventura Netto, P.O., *Grafos: Teoria, Modelos, Algoritmos*, Edgard Blücher, São Paulo (1996).
- [11] Boclin, R., *Indicadores de Desempenho: "Novas Estratégias da Educação Superior"*, *Ensaio - Avaliação e Políticas Públicas em Educação* 7 (1999) 299-308.
- [12] Burrough, P.A., *Principles of Geographical Information Systems for Land Resource Assessment*, Oxford University Press (1986).
- [13] Charnes, A., Cooper, W.W., Lewin, A.Y. and Seiford, L.M., *Data Envelopment Analysis: Theory, Methodology, and Application*, Kluwer Academic Publishers, London (1996).
- [14] Charnes, A., Cooper, W.W. and Rhodes, E., Measuring the Efficiency of Decision-Making Units, *European Journal of Operational Research* 2 (1978) 429-444.
- [15] Clímaco, J.C.N., A Investigação Operacional e o preço do bacalhau, *Boletim nº 30 da APDIO*, Julho (1999). [<http://ctpc22.inescn.pt/~apdio/BoletimInf.html>]
- [16] Cook, W.D. and Kress, M., A Data Envelopment Analysis for Aggregating Preference Rankings, *Management Science* 36 (1990) 1302-1310.
- [17] Cook, W.D. and Kress, M., A Multiple Criteria Decision Model with Ordinal Preference Data, *European Journal of Operational Research* 54 (1991) 191-198.
- [18] Cook, W.D., Kress, M. and Seiford, L.M., Priorization Models for Frontier Decision Making Units in DEA, *European Journal of Operational Research* 59 (1992) 319-323.
- [19] COSEAC, 1999, A UFF no estado do Rio de Janeiro. Disponível em <http://www.coseac.uff.br/vest99/cidades.htm>
- [20] Densham, P.J., Visual interactive location analysis, In: Longley, P. Batty, M. (eds.), *Spatial Analysis: Modelling in a GIS Environment*, United Kingdom GeoInformation International (1996).
- [21] Doyle, J.R. and Green, R.H., Data Envelopment Analysis and Multiple Criteria Decision Making, *OMEGA International Journal of Management Science* 21 (1993) 713-715.
- [22] Farinaccio, F., *Data Envelopment Analysis como strumento di Analisi Multicriteria*, Tese de Doutorado, Università di Pisa, Itália (1998).
- [23] Farrell, M.J., The measurement of productive efficiency, *Journal of the Royal Statistical Society, Series A, General*, 120, Part 3 (1957) 253-281.
- [24] Friedman, L. and Sinuany-Stern, Z., Combining Ranking Scales and Selecting Variables in the DEA Contexts: the Case of Industrial Branches, *Computer & Operations Research* 25 (1998) 781-791.
- [25] Gomes, E.G. and Lins, M.P., Integração entre sistemas de informação geográfica e métodos de análise multicritério no apoio à decisão espacial, *Anais do XXXI SBPO - Simpósio Brasileiro de Pesquisa Operacional*, Juiz de Fora, Minas Gerais, Outubro (1999).
- [26] Gonzales-Araya, M.C., Lins, M.P.E. and Gomes, L.F.A.M., A integração entre a análise de envoltória de dados e o apoio multicritério à decisão: uma revisão; *Anais do XXXI SBPO - Simpósio Brasileiro de Pesquisa Operacional*, Juiz de Fora, Minas Gerais, Outubro (1999).
- [27] Jankowski, P., Integrating geographical information systems and multiple criteria decision-making methods, *International Journal of Geographical Informations Systems* 9 (1995) 251-273.
- [28] Jankowski, P. and Richard, L., Integration of GIS-based suitability analysis and multicriteria evaluation in a spatial decision support system for route selection, *Environment and Planning B* 21 (1994) 323-340.
- [29] Lins, M.P.E. and Meza, L.A., *Análise Envoltória de Dados e perspectivas de integração no ambiente de Apoio à Decisão*, Editora da COPPE/UF RJ, Rio de Janeiro (2000).

- [30] Maguire, D.J., An overview and definition of GIS, In: Maguire, D.J., Goodchild, M.F., Rhind, D.W. (eds.), *Geographical Information Systems: Principles and Applications 1* (1991) 9-20.
- [31] Rangel, L.A.D., Lins, M.P.E. and Gomes, L.F.A.M., *Agregação parcial das eficiências das unidades de decisão em análise envoltória de dados*, Anais do XXXI SBPO - Simpósio Brasileiro de Pesquisa Operacional, Juiz de Fora, Minas Gerais, Outubro (1999).
- [32] Roy, B. and Bouyssou, D., *Aide Multicritère à la Décision: Méthode et cas*, Economica, Paris (1993).
- [33] Sinuany-Stern, Z., Mehrez, A. and Barboy, A., *Academic Departments Efficiency in DEA*, *Computers & Operational Research* 21 (1994) 543-556.
- [34] Stewart, T.J., *Relationships between DEA and Multicriteria Decision Analysis*, *Journal of the Operational Research Society* 47 (1996) 654-665.
- [35] Tavares, G. and Antunes, C.H., *Avaliação da Modernização dos Serviços de Telecomunicações nos Países da OCDE usando DEA - A Situação de Portugal*, *Investigação Operacional* 19 (1999) 177-192.

SENSIBILIDADE À PONDERAÇÃO DE CRITÉRIOS NUM PROBLEMA MCDM DE PLANEAMENTO DE RECURSOS HÍDRICOS

Victor G. Sousa

Ana M. Faustino

J. Ferreira Lemos

Departamento de Engenharia Civil
Faculdade de Engenharia da Universidade do Porto
4099 Porto - Portugal

Abstract

Multicriteria Decision Making techniques are here applied to establish the volume of a reservoir to be held by a dam, as well as other facilities needed by the several uses of the water to be delivered.

The water system is simulated as if a dam and all other facilities were there, allowing the construction of basic scenarios as initial solutions to the problem. The criteria are established, the respective weights chosen and utility functions are built for each use of the water and each criterion. An additive weighted aggregation of utilities leads to an evaluation of the performance of any solution.

Searching for an efficient solution with a better performance is achieved by means of a piecewise linear problem with mixed variables, solved as a bilevel problem.

A systematic analysis of these better solutions is done according to the weights assigned to each criterion and indication is provided about the sensitivity of the problem to the weighting policy and about more robust combinations of solutions before weighting subjectivity.

Resumo

As técnicas de MCDM (Multicriteria Decision Making) são aplicadas ao dimensionamento de um reservatório a realizar por uma barragem, bem como dos respectivos órgãos hidráulicos necessários aos diversos usos da água previstos no projecto.

A simulação do sistema hídrico associado à barragem permite a obtenção de cenários-base como soluções iniciais do problema, a partir das quais são designados os critérios, estabelecidos os respectivos pesos e construídas curvas de utilidade de cada uso da água com respeito a cada critério. A agregação ponderada das utilidades conduz a uma avaliação do desempenho de qualquer solução.

A busca de uma solução eficiente com o melhor desempenho faz-se com base num problema segmentado de variáveis mistas que é resolvido com a sua transformação num problema de dois níveis.

É realizada a análise sistematizada destas melhores soluções em função dos pesos atribuídos a cada critério, fornecendo-se indicadores de sensibilidade do problema à ponderação dos critérios escolhida e das combinações de maior robustez perante a subjectividade da escolha dos pesos.

Keywords

multicriteria, efficient solution, piecewise linear program, bilevel program

1. Introdução

O objecto do estudo em que os autores estão envolvidos é o problema geral do planeamento de sistemas de recursos hídricos. Considera-se que o sistema está de algum modo confinado (uma bacia hidrográfica, uma sub-bacia) e que se está perante uma situação de referência sobre a qual é necessário actuar.

A intervenção no sistema incluirá a construção de novas infra-estruturas (barragens, captações em aquíferos, redes de distribuição, estações elevatórias, etc) que permitam uma intensificação e porventura diversificação das utilizações da água (irrigação, abastecimento de água potável, caudais ecológicos, produção hidro-eléctrica, etc). Note-se que o problema trata de forma diversa super-estruturas (como é uma barragem, por exemplo) às quais não é possível cometer, designadamente, os custos associados a cada utilização da água, e estruturas que são executadas para o serviço exclusivo de uma dada utilização.

No problema há também que contar com a estocasticidade dos caudais nas linhas de água e com a incerteza na evolução potencial dos consumos. A circunstância de haver mais do que um tipo de utilizações de água, maioritariamente conflituosas, exige a participação de vários parceiros no processo de decisão, representando os diversos interesses em jogo.

Para apoiar este processo de decisão, optou-se por uma metodologia que tem como base uma análise multicritério, num processo que é inovador no campo dos recursos hídricos, designadamente porque os respectivos princípios são levados até às últimas consequências (não há, como é usual, a remissão de critérios para condições de restrição sobrando um que contribui para a função-objectivo) e porque se propõe uma partição do espaço das soluções que, simplificando o problema, o torna mais consonante com a realidade.

O problema é assim submetido a um conjunto de critérios (incidência ambiental, incidência social, custos, receitas) no sentido de avaliar a utilidade das diferentes utilizações da água e, portanto, das infra-estruturas que directamente lhes devem ser cometidas.

Procura-se, deste modo, encontrar zonas de consenso, que uma simples comparação, m^3 a m^3 , entre as utilizações da água, não permite estabelecer com igual comodidade.

Conforme se depreende da sequência deste artigo, o seu objectivo fundamental centra-se na discussão da importância dos pesos atribuídos a cada critério, que vão intervir na avaliação da utilidade global dos vários cenários de distribuição da água disponibilizada por uma super-estrutura pelas diversas utilizações em cotejo.

Assim, no capítulo 2 descreve-se a metodologia usada na agregação das utilidades dos vários usos da água e indica-se um problema de optimização que conduz a uma solução eficiente em cada patamar de satisfação das necessidades globais, ou seja, para cada uma das barragens que estão a ser analisadas em termos de dimensão.

No capítulo 3 enuncia-se a proposta de realizar uma análise de sensibilidade da ponderação dos critérios e no capítulo 4 desenvolve-se esse estudo, inicialmente no efeito da variação da utilidade global da solução associada à variação privilegiada de um dos pesos, depois, com uma

orientação inversa, procurando determinar a amplitude de variação dos pesos que não afecta a escolha prévia de uma solução eficiente como sendo a de melhor utilidade global.

Consegue-se, desse modo, determinar uma "região de óptimo global" associada a uma solução eficiente em que esta permanece como a solução a escolher. Apresenta-se então um caso de estudo, para três critérios, que ilustra o interesse desta metodologia permitindo a visualização de um polítopo inserido no espaço dos pesos (neste caso com 3 dimensões) correspondente à solução óptima global.

2. O problema Multicritério

O problema inicia-se com a construção de cenários básicos [6], determinados pelas políticas de desenvolvimento definidas para a região e as condicionantes resultantes de documentos regulamentadores, designadamente Planos de Desenvolvimento Regional e Planos Directores Municipais. É em torno dos cenários básicos possíveis que se usa um modelo de simulação das alternativas básicas (ligadas à eleição das infra-estruturas que tornarão a água disponível), que utiliza séries históricas e sintéticas de afluências, entra com níveis de aceitação máximos e mínimos para cada utilização e com modelos de evolução de consumos até ao horizonte do estudo. Este modelo de simulação fornece assim um primeiro conjunto de soluções associadas às diversas hipóteses de infra-estruturas a implementar e de consumos a satisfazer e é também usado para reflectir a estocasticidade e incerteza dos dados, conduzindo à aceitação por todos os parceiros de um limiar de risco.

A abordagem multicritério das alternativas assim geradas começa pela identificação dos critérios e indicação dos respectivos pesos e pela construção de funções $u_{ij}(v_i)$ de utilidade de cada utilização de água i segundo cada critério j [7]. Estas funções são lineares segmentadas, eventualmente com descontinuidades e não côncavas, e o desempenho de cada alternativa de repartição da água a (vector com os volumes de água v_i atribuídos a cada utilização) é avaliado pelo método aditivo com

$$U(a) = \sum_{i=1}^m \sum_{j=1}^{n_c} w_j u_{ij}(v_i), \quad \sum_j w_j = 1, \quad w_j \geq 0 \quad j = 1, \dots, n_c \quad (1)$$

onde w_j representa o peso associado ao critério j , u_{ij} o valor da função utilidade para a utilização i e o critério j , n_c o número de critérios e m o número de utilizações de água.

Este tipo de agregação assume que os critérios de avaliação respeitam as condições de exaustividade e independência.

A busca de uma solução não dominada (eficiente) para cada conjunto de infra-estruturas - que não possam ser associadas a uma única utilização da água - é realizada a partir de uma distribuição equitativa da água pelos diversos consumos (prescreve-se um nível global de satisfação que se mantém para todos os usos) a que corresponde uma solução não eficiente, no interior do espaço das soluções. Para essa busca usa-se um programa linear do tipo

$$\begin{aligned} \max \quad & \sum_{i=1}^m \frac{\partial U}{\partial v_i} \Delta v_i \\ \text{s. a.} \quad & \sum_{i=1}^m \Delta v_i = 0 \end{aligned} \quad (2)$$

que avalia o efeito das variações aos volumes v_i de cada utilização sobre a função de utilidade agregada U , supondo invariável o volume total e portanto a dimensão das infra-estruturas em causa. Este programa, aplicado a funções segmentadas, gera um problema que é resolvido recorrendo a uma transformação num problema de dois níveis [3].

No artigo referido é proposta uma via para usar códigos do problema de dois níveis na resolução de problemas segmentados (piecewise), que foi utilizada no nosso problema.

Para essa transformação foram usadas as equivalências apresentadas em [3], artigo segundo o qual, dados os coeficientes angulares b_i dos segmentos de recta, as abcissas dos pontos de quebra q_i e a intersecção b_0 da recta do primeiro segmento com o eixo dos yy , se tem:

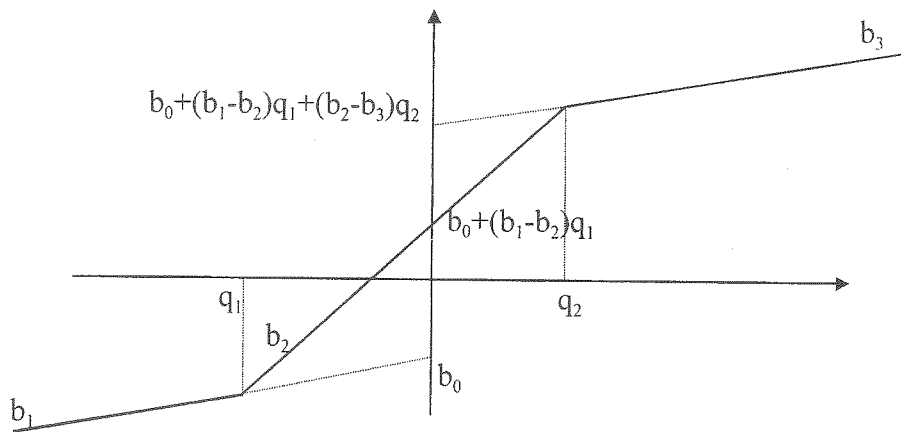
$$f(z) = \begin{cases} b_0 + b_1 z & \Leftarrow z \leq q_1 \\ b_0 + b_1 q_1 + b_2 (z - q_1) & \Leftarrow q_1 \leq z \leq q_2 \\ b_0 + (b_1 - b_2) q_1 + b_2 q_2 + b_3 (z - q_2) & \Leftarrow q_2 \leq z \end{cases}$$

que é equivalente a

$$f(z) = b_0 + b_1 z + (b_2 - b_1) w_1 + (b_3 - b_2) w_2$$

onde

$$w_i = \max\{0, z - q_i\} \quad i = 1, 2$$



Por outro lado, como, segundo o mesmo artigo,

$$\min \left\{ cx+by : Sx \leq s, y = \max \{ A^1x+b^1, A^2x+b^2, \dots, A^kx+b^k \} \right\}$$

é equivalente a

$$\min_{Sx \leq s} \left\{ \begin{array}{l} cx+by \\ y = \arg \min \left\{ \sum_i y_i : y \geq A^1x+b^1, \dots, y \geq A^kx+b^k \right\} \end{array} \right\}$$

O problema anterior pode então ser posto na forma

$$\max f(z) = b_0 + b_1z + (b_2-b_1) y_1 + (b_3-b_2) y_2$$

$$\begin{array}{l} \text{s.a} \quad \min \sum_{k=1}^2 y_k \\ y_1 \geq z - q_1 \\ y_2 \geq z - q_2 \\ y_1 \geq 0 \\ y_2 \geq 0 \end{array}$$

Usando esta transformação, e dado que no problema inicial $z \geq 0$ é dado em percentagem e U é separável relativamente às utilizações, o problema toma a forma

$$\begin{array}{l} \max \sum_i f(z_i) = \sum_i \left(b_{0i} + b_{1i}z_i + \sum_{k=1}^{r_i-1} (b_{k+1,i} - b_{k,i}) y_{ki} \right) \\ \text{s.a} \quad \sum_{i=1}^m v_i z_i = vt \\ \min_y e^T y \\ y_{ki} \geq 0 \quad \text{para } k = 1, \dots, r_i-1; \quad i = 1, \dots, m \\ y_{ki} \geq z_i - q_k^i \end{array} \quad (3)$$

onde vt é o volume total disponível, r_i é o número de ramos da função segmentada associada à utilização i e, se $p = \sum_i r_i$, e é um vector de 1 's em \mathbb{R}^{p-m}

Para este problema temos $z \in \mathbb{R}_+^m$ no primeiro nível, $y \in \mathbb{R}_+^{p-m}$ no segundo nível, 2 restrições de desigualdade no primeiro nível e $(p-m)$ restrições no segundo nível, o que o torna de uma dimensão reduzida face a alternativas conhecidas (por exemplo, a reformulação de problemas segmentados em problemas MIP [5, pp11] e destes em BLP [1]). Para a resolução do problema de dois níveis foi usado o algoritmo apresentado em [4].

Obtém-se assim, para uma região do espaço das soluções associadas a determinadas infra-estruturas, uma solução eficiente que é óptima nessa região, e comparável através do respectivo "score" com soluções eficientes dos outros sub-espços associados a infra-estruturas diferentes ou de outra dimensão, para as quais é também diferente o volume global de água envolvido (ver figura 1). Entenda-se por "score" o valor associado à utilidade global de uma solução que "mede" o respectivo desempenho em termos de satisfação das necessidades de água.

Nesta figura, cada um dos três planos representados indica um lugar geométrico de todas as possíveis soluções de distribuição de água pelas três utilizações, associadas a uma mesma infra-estrutura (barragem, no caso). De baixo para cima, os planos correspondem sucessivamente a satisfações globais de 70%, 80% e 90% dos valores pedidos pelos utilizadores. Na procura da melhor solução eficiente, parte-se de uma solução equitativa (pontos no centro do triângulo) chegando-se à melhor solução (associada a uma ponderação de critérios bem definida), com pontos nas arestas ou vértices dos planos triangulares.

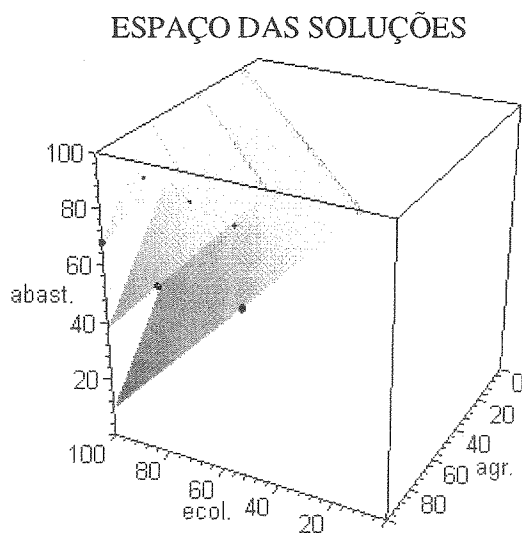


Figura 1

Estas melhores soluções são depois comparadas, uma vez complementadas com a utilidade da super-estrutura correspondente.

3. A importância da Ponderação dos Critérios

A metodologia adoptada para a avaliação de soluções para um problema de planeamento assenta em vários pressupostos que são discutíveis, e por isso se reitera a ideia que uma tal metodologia só tem viabilidade se ancorada numa discussão participada pelas entidades que de algum modo são parte interessada no processo de planeamento.

De facto, uma abordagem multicritério, sendo a que nos parece melhor se adaptar ao problema, não é única. O modelo aditivo tem condições severas para ser aplicado e há sempre uma reserva de insegurança quanto à sua aplicabilidade, designadamente no que se refere à

designação dos critérios que, como se disse, devem ser exaustivos e independentes. A construção das funções de utilidade (ou atractividade) de cada utilização em relação a cada critério está longe de ser um processo pacífico, baseando-se em grande medida na subjectividade dos intervenientes no processo de decisão.

Em todo este processo que antecede a agregação das funções de utilidade e a obtenção do correspondente "score" para cada solução, entende-se que há um número tão grande de variáveis e das suas combinações que qualquer análise de sensibilidade com pretensões a obter conclusões generalizáveis parece-nos fora de causa.

Entretanto, a decisão quanto ao peso de cada critério, um importante passo na metodologia que antecede a agregação, tem grande influência na avaliação das soluções e permite uma abordagem útil e não demasiadamente elaborada [8].

De facto, se a hierarquização dos critérios pode já não ser uma decisão fácil, mais discutível se torna quando se chega ao momento de se atribuir pesos a cada um deles, no pressuposto (necessidade convencional) de que o seu somatório é igual à unidade. A influência da designação deste simples vector de ponderação nos resultados é quase decisiva e pequenas alterações podem conduzir à escolha de uma solução eficiente acentuadamente diferente da que fora escolhida anteriormente.

Esta circunstância tem sido objecto de estudo de vários autores, quer apresentando software de visualização rápida dessas alterações, quer propondo metodologias para a obtenção de consenso entre parceiros na designação desses pesos [2].

O objectivo deste trabalho é o de fazer uma análise de sensibilidade da ponderação dos critérios, avaliando a robustez de soluções eficientes perante variações nos pesos e procurar, tanto quanto possível, e para um problema concreto, repartir o espaço de ponderação por diferentes soluções eficientes.

Este estudo insere-se num espaço de soluções contínuo e tem como fundamento que as funções de utilidade sejam lineares (segmentadas quase sempre). Não se trata de uma restrição incontornável para outros tipos de funções, já que será sempre possível aproximá-las por funções segmentadas lineares.

4. Análise de Sensibilidade

4.1 Variações em torno de um peso

Uma primeira abordagem do problema foi realizada sobre uma dada solução de volumes (v_1, v_2, \dots, v_m) , através de um algoritmo simples que determina o "score" (valor que traduz a utilidade global dessa solução) para um dado vector de pesos. Um dos pesos w_i percorre todos os valores de 0 a 1 (supõe-se que não há outras fronteiras superior e inferior para nenhum dos pesos), enquanto os restantes tomam todas as combinações de valores possíveis que mantêm o somatório dos pesos igual à unidade, determinando-se os respectivos "scores" s .

Desta análise resultaram diagramas do tipo que se apresenta a seguir, que traduzem o facto de que é convexo o conjunto K de todos os pontos (w_j, s) assim obtidos e que os valores máximos e mínimos de s se desenvolvem segundo rectas, como à frente se demonstra

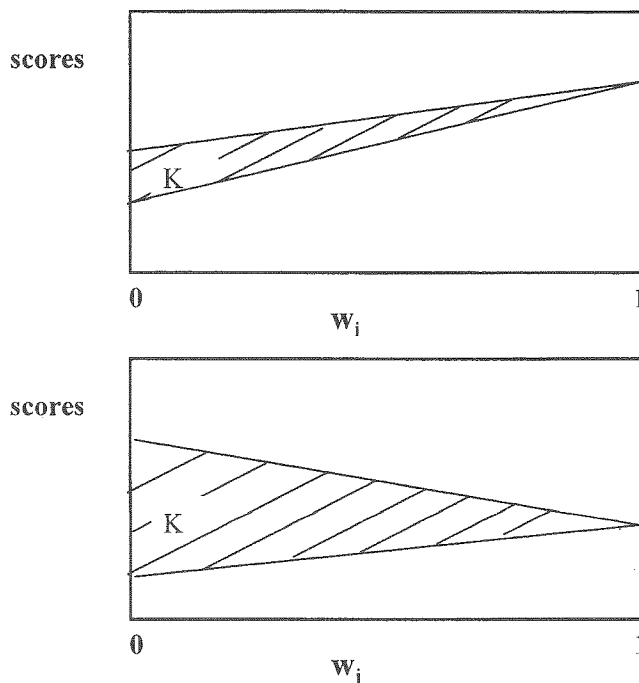


Figura 2

4.1.1 A convexidade de K

Sejam (w_i, s) e (w'_i, s') dois pontos do conjunto K , onde

$$s = \sum_j a_j w_j, \quad \sum_j w_j = 1, \quad w_j \geq 0$$

onde a_j é dado por

$$a_j = \sum_{i=1}^m u_{ij}(v_i)$$

ou seja, a contribuição do critério i para o "score" da solução que envolve os volumes (v_1, v_2, \dots, v_m) .

O conjunto K será então convexo se

$$\theta(w_i, s) + (1-\theta)(w'_i, s') \in K, \quad 0 \leq \theta \leq 1 \quad (4)$$

Demonstração

Aqueles dois pontos de K podem ser escritos na forma

$$(w_i, \sum_j a_j w_j) \text{ e } (w'_i, \sum_j a_j w'_j).$$

Com estes dois pontos e $0 \leq \theta \leq 1$, resulta de (4)

$$\theta(w_i, s) + (1-\theta) (w'_i, s') = (\theta w_i + (1-\theta) w'_i, \sum_j a_j (\theta w_j + (1-\theta) w'_j))$$

e

$$\sum_j (\theta w_j + (1-\theta) w'_j) = \theta \sum_j w_j + (1-\theta) \sum_j w'_j = 1 \text{ com } \theta w_j + (1-\theta) w'_j \geq 0$$

o que mostra que a combinação linear de dois pontos de K também pertence a K .

4.1.2 Limites máximos e mínimos dos scores

Fixando-se o peso w_i no valor \bar{w}_i , o problema de determinação do score máximo e mínimo corresponde a resolver, nas variáveis

$$w_j, j \in \{1, \dots, n_c\} \setminus \{i\}$$

o problema linear

$$\begin{aligned} \max(\min) \quad & s = a_i \bar{w}_i + \sum_{j \neq i} a_j w_j \\ \text{s.a.} \quad & \bar{w}_i + \sum_{j \neq i} w_j = 1 \\ & w_j \geq 0 \end{aligned} \quad (5)$$

O máximo (mínimo) ocorre quando a variável básica w_k está associada ao coeficiente a_k (com $k \neq i$) máximo (mínimo), a_{\max} (a_{\min}), pois neste caso os coeficientes de custo reduzidos satisfazem a condição de optimalidade

$$-a_j + a_k \geq 0. \quad (a_j - a_k \geq 0) \quad \Delta_j$$

tendo a variável básica o valor

$$w_k = 1 - \bar{w}_i$$

Assim, no espaço K , sendo os a_j fixos, a variável básica mantém-se fixa quando w_i varia entre 0 e 1 e tem-se, portanto

$$s_{\max} = a_i w_i + a_{\max} w_k = a_i w_i + a_{\max} (1 - w_i) = a_{\max} + (a_i - a_{\max}) w_i$$

e

$$s_{\min} = a_{\min} + (a_i - a_{\min}) w_i$$

(6)

o que revela que os scores máximos e mínimos estão sobre rectas, de declives, respectivamente $a_i - a_{\max}$ e $a_i - a_{\min}$, que se intersectam no ponto $(1, a_i)$.

Está assim explicado que estes diagramas (figura 2) tenham como característica o facto de serem convexos limitados por duas linhas rectas que se unem num vértice quando o peso em análise toma o valor 1 e se tem portanto apenas uma combinação de pesos. Demais conclusões são do tipo tendencial: um aumento no caso de w_i favorece o score da solução, para w_j não se

pode tirar a mesma conclusão nem a oposta e w_i afigura-se como um peso de maior relevância, mantendo os scores máximo e mínimo entre valores apertados.

Fixando um dos pesos, a solução eficiente correspondente estará numa linha vertical algures entre as linhas de scores máximo e mínimo, sendo que qualquer alteração aos demais pesos que faça subir a solução no sentido da linha superior conservará para essa solução a qualidade de óptima para esse conjunto de pesos.

Por outro lado, um deslocamento horizontal do ponto, que indica a combinação de pesos adoptada para a solução, e que intersecte a linha superior, estará a conduzir a solução para uma situação em que esta deixará de ser óptima para esse novo vector de pesos.

Contudo, como se referiu, todas estas conclusões são do tipo qualitativo, e em alguns casos imprecisas, pelo que se optou por uma abordagem diferente que tivesse em conta a configuração das funções de utilidade para a solução em estudo.

4.2 Variações nas funções de utilidade

Suponha-se conhecida uma solução eficiente (v_1, v_2, \dots, v_m) , obtida para um vector de pesos $(w_1, w_2, \dots, w_{nc})$.

Vamos analisar a sensibilidade do problema linear (2) à alteração dos pesos, colocando-o na forma

$$\begin{aligned} \max \quad & \sum_{i=1}^m (c_i^+ x_i^+ - c_i^- x_i^-) \\ \text{s.a} \quad & \sum_{i=1}^m (x_i^+ - x_i^-) = 0 \\ & 0 \leq x_i^+ \leq u_i \\ & 0 \leq x_i^- \leq l_i \end{aligned} \tag{7}$$

onde

c_i - declives "pesados" à esquerda (-) e à direita (+) para a solução dada das funções de utilidade agregadas, que aqui funcionam como custos;

x_i - variações de v_i à esquerda (-) e à direita (+).

A solução nula deste problema está associada à solução eficiente (v_1, v_2, \dots, v_m) e o que se pretende é determinar os custos para os quais a solução nula continua a ser óptima.

Do método simplex revisto pode escrever-se, em relação à base B , a condição de optimalidade

$$\bar{c}_N^T = -c_N^T + c_B^T B^{-1} N \geq 0$$

onde N é a matriz associada às variáveis não básicas.

Como, neste caso, os elementos B e N são unitários (+1 ou -1), então as condições de optimalidade podem escrever-se na forma

$$\bar{c}_i^+ = -c_i^+ + c_k^* \geq 0 \Leftrightarrow -\bar{c}_i^+ = c_i^+ - c_k^* \leq 0 \quad (8)$$

$$\bar{c}_i^- = c_i^- - c_k^* \geq 0$$

onde c_k^* é o custo associado à variável básica x_k^* . Note-se entretanto que a expressão (8) permite confirmar que, num ponto de máximo, o declive à direita é negativo e à esquerda é positivo.

4.3 Região do óptimo local

Quanto aos custos, ou declives "pesados" das funções de utilidade, são dados pela expressão

$$c_i = \sum_{j=1}^{n_c} w_j d_{ij}$$

onde d_{ij} são os declives, à esquerda e à direita, da função de utilidade i para o critério j nos pontos correspondentes à solução eficiente (v_1, v_2, \dots, v_m) . As condições de optimalidade podem então ser escritas na forma

$$\sum_{j=1}^{n_c} w_j d_{ij}^+ - \sum_{j=1}^{n_c} w_j d_{kj}^* \leq 0$$

$$\sum_{j=1}^{n_c} w_j d_{ij}^- - \sum_{j=1}^{n_c} w_j d_{kj}^* \geq 0 \quad (9)$$

$$\sum_{j=1}^{n_c} w_j = 1, w_j \geq 0$$

Fica assim delimitada uma região de ponderação para a qual a solução eficiente permanece como óptimo local, isto é, desde que a variação de volumes não ultrapasse a intersecção dos intervalos de segmentação onde a solução se inscreve. No entanto, como o problema não é côncavo, nada garante que uma variação dos pesos dentro da região de ponderação atrás definida não possa conduzir a uma outra solução eficiente afastada da inicial.

4.4 Região do óptimo global

A região determinada no ponto anterior é demasiado ampla (ver figura 3) e não contribui para qualquer tipo de partição do espaço de ponderação pelas soluções eficientes. A região surge mais restrita e ligada à presença de uma única solução eficiente se as condições de optimalidade aumentarem a sua exigência aos declives extremos das funções de utilidade em todo o intervalo onde estão definidas.

Sejam assim

\bar{d}_{ij}^+ e \bar{d}_{ij}^- respectivamente os valores algébricos máximo e mínimo dos declives, exteriores ao intervalo em estudo, da função de utilidade segmentada da utilização i e o critério j ,

Verifica-se então que

$$\sum_{j=1}^{n_c} w_j d_{ij}^+ - \sum_{j=1}^{n_c} w_j d_{kj}^* \leq \sum_{j=1}^{n_c} w_j \bar{d}_{ij}^+ - \sum_{j=1}^{n_c} w_j d_{kj}^* \leq 0 \quad (10)$$

$$\sum_{j=1}^{n_c} w_j d_{ij}^- - \sum_{j=1}^{n_c} w_j d_{kj}^* \geq \sum_{j=1}^{n_c} w_j \bar{d}_{ij}^- - \sum_{j=1}^{n_c} w_j d_{kj}^* \geq 0$$

e as novas inequações conduzem à definição de uma nova região de ponderação, mais restrita, onde garantidamente a solução eficiente se comporta como o óptimo global (ver figura 3).

Como é óbvio, o mesmo tipo de região pode ser encontrada para outras soluções eficientes que se comportem como óptimas para outras combinações de pesos, o que poderia indiciar uma repartição da região em sub-regiões para outras tantas soluções eficientes. Tal não é exactamente assim como se verifica do exemplo que foi testado.

4.5 Caso de estudo

Para ilustrar este estudo foi seleccionado um problema simplificado de planeamento de uma barragem (e do volume da respectiva albufeira) com 3 utilizações possíveis da água (abastecimento de água para uma população equivalente de 15 000, irrigação de culturas e descargas no Verão e no Outono para manter o rio com qualidade mínima de água).

Como critérios foram designados 4: custos, benefícios ambientais, benefícios sociais e receitas. Para vector dos pesos foi escolhido

$$(w_1, w_2, w_3, w_4) = (0.4, 0.3, 0.2, 0.1).$$

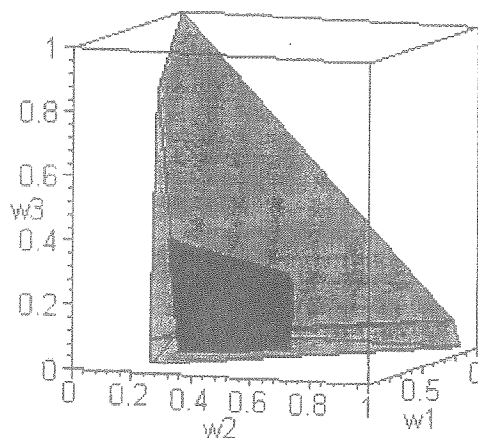


Figura 3

Note-se que o critério dos custos foi subdividido em custos de capital, custos de operação e custos de manutenção, o que implica a adopção de pesos de segunda escala extraídos do peso atribuído ao custos totais. Esta repartição de segunda ordem é no entanto fixa, pelo que o problema só tem que se debater com os quatro pesos para os quatro critérios fundamentais.

Registe-se ainda que a separação entre custos e receitas tem em vista uma mais fácil leitura das soluções encontradas, uma vez que os custos e as receitas, variáveis, são em grande parte distribuídos por diferentes entidades.

Na resolução do problema foram ensaiadas várias hipóteses de albufeira, e portanto de volume total disponível, tendo-se concluído, pela aplicação da metodologia descrita e em particular do problema de programação linear (2) resolvido segundo a forma (3), que a solução mais favorável correspondia a uma barragem que assegurava um volume total de 270 (as unidades foram convencionadas de acordo com uma escala), com uma distribuição entre as utilizações para a solução eficiente de $(v_1, v_2, v_3) = (100, 100, 70)$.

A aplicação das inequações (9) permitiu determinar aquilo que foi designado por "região do óptimo local" através da intersecção de semi-espacos no espaço de ponderação (w_1, w_2, w_3) , o que é suficiente dado que w_4 fica definido pela condição do somatório ser 1 originando o simplex definido pelo plano $w_1 + w_2 + w_3 = 1$ (ver figura 3).

Como foi referido, a região assim definida apenas garante que a solução eficiente seja indiferente aos pesos localmente (ou seja, na intersecção dos intervalos dos segmentos das funções de utilidade correspondentes à solução), pelo que foi possível encontrar dentro dessa região soluções eficientes afastadas para outras combinações de pesos: $(100, 70, 100)$, $(80, 100, 90)$ e $(70, 100, 100)$. Como era de esperar todas estas soluções eficientes se localizam na fronteira do espaço das soluções.

Passou-se então à determinação da "região do óptimo global" utilizando as segundas desigualdades de (10), obtendo-se o polítopo com sombreado mais escuro representado na figura 3.

Este polítopo define uma região convexa em cujo interior a solução $(100, 100, 70)$ permanece como óptima. No entanto, a mesma solução aparece em pontos exteriores anexos ao polítopo, o que significa que não está conseguida a partição rigorosa do espaço de ponderação pelas várias soluções eficientes. Aliás, a determinação dos outros polítopos para as outras soluções eficientes confirmou que eles não são adjacentes a não ser em situações pontuais.

5. Conclusões

Os estudos realizados confirmaram que é possível obter, para qualquer solução eficiente uma região do espaço de ponderação onde ela permanece óptima, dando uma medida da sua robustez face às variações dos pesos dos critérios. Essa região está definida no entanto por defeito, pois não regista todas as combinações de pesos em que a solução se conserva.

Concluiu-se então que, com a metodologia adoptada, permanecem no espaço de ponderação regiões indefinidas que haverá que investigar de modo a obter-se a partição desse espaço pelas várias soluções eficientes.

Os resultados obtidos são no entanto encorajadores quanto à possibilidade de, com esta metodologia, se permitir um melhor conhecimento da influência da ponderação dos critérios, sempre útil num processo de decisão, e de se alargarem espaços de consenso entre as partes

intervenientes nesse processo, vindo aliás ao encontro de trabalhos e ideias como a mesma perspectiva, designadamente com o conceito de "Overall Compromise Criterion" que C. Bana e Costa introduz no seu artigo [2].

Referências

- [1] Audet, C., Hansen, P., Jaumard, B. and Savard, S., Links between the Linear Bilevel and Mixed 0-1 Programming Problems, Les Cahiers du GERARD (1995) (publicação interna).
- [2] Bana e Costa, C.A., A multicriteria decision aid methodology with conflicting situations on the weights, European Journal of Operational Research 26 (1986) 22-34.
- [3] Frangioni, A., On a new class of Bilevel Programming Problems and its use for reformulating Mixed Integer Problems, European Journal of Operational Research 82 (1995) 615-646.
- [4] Júdice, J.J. and Faustino, A.M., A sequential LCP method for bilevel linear programming, Annals of Operation Research 34 (1992) 89-106.
- [5] Nemhauser, G.L. and Wolsey, L.A., Integer and Combinatorial Optimization, John Wiley & Sons, Inc. (1988).
- [6] Stewart, T.J. and Scott, L., A scenario-based framework for multicriteria decision analysis in water resources planning, Water Resources Research 31 (1995) 2835-2843.
- [7] Vincke, P., Gassner, M. and Roy, B., Multicriteria Decision-aid, John Wiley and Sons, Chichester (1992).
- [8] Watkins Jr., D.W. and McKinney, D.C., Finding Robust Solutions to Water Resources Problems, Journal of Water Resources Planning and Management 123 (1997) 49-58.

UTILIZAÇÃO DE SIMULATED ANNEALING EM OPTIMIZAÇÃO DIFUSA

Maria Leonilde Rocha Varela

Univ. do Minho- Escola de Engenharia
Dept. Produção e Sistemas
Campus de Azurém
4800-058 Guimarães - Portugal

Rita Almeida Ribeiro

Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia
Dept. Informática
2825-114 Monte de Caparica - Portugal

Abstract

The Simulated Annealing algorithm (SA) is an adequate tool for solving fuzzy optimisation problems, through the selection of the best solution among a finite number of possible solutions. It is a particularly attractive technique to solve fuzzy optimisation problems because it allows finding close to optimal solutions, without big computational effort, which in a fuzzy environment is usually good-enough. In this context, we present the results for a set of problems, selected with the purpose of testing the SA algorithm performance. The problems tested were formulated following a complete fuzzification method proposed by Ribeiro and Moura-Pires (1999). These examples show the flexibility and adaptability of the fuzzy method as well as the SA algorithm, for the resolution of fuzzy linear optimisation problems.

Resumo

O algoritmo *Simulated Annealing* (SA) é uma ferramenta adequada à resolução de problemas de optimização linear, de natureza difusa, através da selecção da melhor solução de entre um número finito de soluções possíveis. Trata-se de uma técnica particularmente atractiva para resolver problemas de optimização difusa pois permite encontrar soluções próximas da óptima, à custa de um esforço computacional geralmente pouco exagerado. Neste contexto, apresentam-se os resultados para um conjunto de casos seleccionados, com a finalidade de testar o algoritmo de SA implementado. Uma vez que se pretende testar problemas de optimização difusos, usou-se um método de formulação proposto por Ribeiro e Moura-Pires (1999). Estes casos pretendem mostrar a flexibilidade e adaptabilidade do método difuso utilizado e do algoritmo de resolução SA, em problemas optimização linear difusos.

Keywords

Mathematical Programming, Fuzzy Optimisation, Fuzzy Sets, Simulated Annealing.

1. Introdução

Um Problema de Otimização Difuso pode ser entendido como um problema de otimização clássico transformado de modo a incluir alguma forma de "fuzificação" (do Inglês *fuzzification*), isto é, de flexibilização de modo a permitir uma tomada de decisão mais alargada e abrangente e, conseqüentemente, mais flexível e adequada em determinados cenários.

Em geral, os problemas de otimização visam a maximização ou a minimização de um objectivo simples (único) ou composto (múltiplo) enquanto satisfazem as restrições do problema, que representam as condições do modelo. O principal objectivo da abordagem difusa utilizada neste tipo de problemas é o de encontrar a solução ou a alternativa de decisão que satisfaz "melhor" uma determinada situação específica, que ocorre num ambiente difuso [9].

Subjacente a este trabalho está o desenvolvimento de uma aplicação informática que implementa um algoritmo de resolução do tipo *Simulated Annealing* (SA) pois é uma ferramenta adequada à resolução de problemas de otimização linear difusa, através da selecção da melhor solução de entre um número finito de soluções possíveis. Trata-se de uma técnica bastante atractiva para resolver problemas de otimização difusa pois permite encontrar soluções próximas da óptima, à custa de um esforço computacional geralmente não muito grande e é relativamente simples de implementar e manipular. No contexto deste trabalho um dos objectivos principais consiste portanto em mostrar a utilidade ou conveniência desta ferramenta para a resolução de problemas de otimização linear difusos e a principal medida de avaliação do seu desempenho consiste no valor que se obtém para a função objectivo, perante uma determinada situação de compromisso, subjacente a uma certa necessidade de violação do nível de satisfação das condições que caracterizam o problema em causa. Além disso também se introduziu na aplicação a possibilidade de proceder a uma análise do tempo de resposta do algoritmo SA em função do tipo de problema "fuzificado" (do Inglês *fuzzified*) considerado e far-se-á uma análise comparativa destes valores, face aos resultados obtidos para os diversos tipos de problemas considerados, expressos em termos de valor da função objectivo e do correspondente nível de satisfação das condições do problema, delimitado por um determinado valor de *threshold* imposto no problema.

Adicionalmente, foi introduzida a possibilidade de o utilizador seleccionar uma "semente", para a geração dos números aleatórios, permitindo repetir experiências, a fim de possibilitar uma análise comparativa dos resultados obtidos para diferentes execuções do programa, com determinadas variações na formulação do problema. Além disso, também se inclui uma ligeira adaptação do algoritmo original, de modo a permitir a resolução de problemas de programação inteira.

O objectivo último da implementação deste algoritmo consiste não só em fornecer um resultado para problemas de otimização difusos, mas também servir como um meio de apoio à tomada de decisão, de modo a facultar um conjunto de soluções alternativas ao decisor, em tempo útil, indicando-lhe as conseqüências, em termos de vantagens e inconvenientes

(compromissos), subjacentes a cada decisão, avaliados, essencialmente, pelo valor obtido para a F.O. e pelo nível de satisfação das condições do problema (*threshold*) e, deste modo, apoiá-lo na escolha da solução mais adequada a cada caso particular.

Neste trabalho pretende-se, também, ilustrar a flexibilidade de um método de otimização difusa proposto por Ribeiro e Moura-Pires [8,9] na formulação de diversos tipos de problemas de otimização linear difusos, que variam no tipo de "fuzificação" considerada e/ou nos desvios permitidos nas condições do problema e/ou no nível de satisfação (*threshold*) e/ou na dimensão do problema, além de outros parâmetros de controlo do algoritmo, que também podem variar, como se irá explicar mais adiante. O objectivo principal consiste em permitir apoiar a tomada de decisão na resolução deste tipo de problemas. Em particular, pretende-se mostrar que a principal vantagem da utilização deste método difuso reside na liberdade que o utilizador dispõe de flexibilizar qualquer problema específico de otimização. Sendo assim, a formulação difusa do problema pode variar desde um modelo simples em que, por exemplo, apenas se admitem variações nos parâmetros das restrições, até um modelo difuso global do problema, em que são permitidos desvios nos parâmetros e nos coeficientes das restrições e no valor e nos coeficientes da função objectivo.

A fim de expor estes assuntos achou-se conveniente estruturar este artigo do seguinte modo. No ponto 2 faz-se uma breve introdução à Tomada de Decisão Difusa, incluindo uma descrição de funções de pertença comuns, o modelo *Maxmin* de Belman e Zadeh [2], o modelo de otimização difusa de Zimmermann [11] e o método de "fuzificação" completo proposto por Ribeiro e Moura-Pires [8]. No ponto 3 descreve-se o algoritmo *Simulated Annealing (SA)* e apresentam-se os detalhes de implementação do algoritmo. No ponto 4 apresentam-se as formulações *crisp* e "fuzificadas" de um conjunto de casos que irão ser analisados (cobrindo um total de 84 casos de problemas difusos). Esta secção inclui a "fuzificação" do problema através da flexibilização que resulta da combinação de diversos tipos de "fuzificação", em termos de parâmetros das restrições, dos coeficientes da função objectivo e das restrições e do valor da função objectivo (meta). O ponto 5 apresenta os resultados para os casos considerados, resolvidos através do algoritmo de *SA* implementado e efectua uma análise destes, para as diferentes formulações consideradas. Finalmente, no ponto 6 extraem-se algumas conclusões e apresentam-se algumas intenções em termos de trabalho futuro.

2. Tomada de Decisão Difusa

Em geral poderá dizer-se que a Tomada de Decisão Difusa integra dois ramos fundamentais que são o "Multi-atributo Difuso" e a "Otimização Difusa" [6,11]. Neste trabalho aborda-se a problemática da Otimização Difusa quando tanto os coeficientes como os recursos e metas são difusos [8].

A tomada de decisão difusa recorre à Teoria de Conjuntos Difusos, desenvolvida por Zadeh [10]. Bundy [4] define a Teoria de Conjuntos Difusos como uma extensão à teoria de conjuntos convencional em que o grau de pertença para um elemento num conjunto toma um valor algures

no intervalo $[0,1]$, em vez de apenas 0 ou 1. Esta teoria foi desenvolvida no intuito de tornar tratável a complexidade subjacente a descrições de processos subjectivos ou mal entendidos.

Segundo Bellman [2] a tomada de decisão num ambiente difuso é considerada como um processo de decisão em que as metas a atingir e/ou as restrições, mas não necessariamente o sistema sob controlo, são de natureza difusa. Isto significa que as metas e/ou as restrições constituem classes de alternativas cujas fronteiras não estão definidas com precisão.

As metas e as restrições podem assim ser definidas de um modo adequado na forma de conjuntos difusos, no espaço de alternativas. Um conjunto difuso é uma classe de objectos em que não existe uma fronteira bem definida entre os objectos que pertencem à classe e os que não pertencem. Uma definição mais precisa poderá ser como se segue [2]:

Seja $X = \{x\}$ uma classe de objectos (pontos) representados genericamente por x . Então um conjunto difuso A em X é um conjunto de pares ordenados

$$A = \{(x, \mu_A(x))\}, x \in X$$

onde $\mu_A(x)$ é designado por grau de pertença de x em A , e $\mu_A: X \rightarrow M$ é a função de X para um espaço M , designado de espaço de pertença. Geralmente assume-se que M pertence ao intervalo $[0, 1]$, em que 0 e 1 representam, respectivamente, os graus ou níveis de pertença da lógica Booleana clássica. Sendo assim, assume-se que um conjunto difuso A pode ser definido com precisão, associando a cada objecto x um valor entre 0 e 1, que representa o seu grau de pertença em A .

2.1 Otimização difusa

Os problemas tradicionais de otimização consistem em maximizar ou minimizar uma determinada função objectivo, sujeita a um conjunto de restrições, que exprimem, por exemplo, a limitação de recursos. O objectivo consiste então em maximizar ou minimizar a função objectivo, satisfazendo as restrições do problema.

$$\text{Max } f(x) \text{ sujeita a } x \in X$$

O objectivo principal da otimização difusa é o de encontrar a "melhor" solução (alternativa de decisão) na presença de informação incompleta, isto é, informação imprecisa e/ou a existência de limites vagos na informação. Existem muitas formas de imprecisão, quando se manipulam ou abordam problemas de otimização difusa. Nomeadamente, coeficientes das variáveis que não se conhecem com precisão (por exemplo, "tempos de produção de cerca de uma hora, para a montagem de uma peça") ao nível de satisfação de restrições, que têm limites imprecisos (por exemplo, "o tempo total de produção disponível deverá rondar as 100 horas"). Um desafio proposto actualmente é o de permitir a construção de modelos que permitem utilizar e interpretar uma linguagem vaga e imprecisa, que possam ser traduzidos em métodos quantitativos difusos [9].

Um problema de otimização linear clássico pode ser formalizado como,

$$\begin{aligned} \max \quad & Z = C'x \\ & Ax \leq B \\ & x \geq 0 \end{aligned}$$

A versão "fuzificada" deste problema é geralmente formalizada como,

$$\begin{aligned} \max/\min \quad & \tilde{Z} = \tilde{C}x \\ & \tilde{A}x \{ \geq, \leq, = \} \tilde{B} \\ & x \geq 0 \end{aligned}$$

onde \tilde{Z} representa uma meta difusa, \tilde{C} é o vector de custos difusos, \tilde{A} é a matriz que contém os coeficientes difusos do(s) objectivo(s) e das restrições e \tilde{B} é o correspondente vector dos limites (parâmetros) do(s) objectivo(s) e dos recursos. Por vezes usa-se como notação um til em cima do sinal da equação/inequação.

As primeiras propostas de "fuzificação" do problema de otimização apenas consideravam a "fuzificação" dos recursos (restrições) e de metas para os objectivos, como veremos nas próximas secções. Será no entanto de notar que existem vários métodos de "fuzificação" propostos na literatura, tanto para os recursos e metas, como para os coeficientes (ver detalhes em [5]). Neste artigo apenas focamos, em detalhe, o método de "fuzificação" completo proposto por Ribeiro e Moura-Pires [8], pois trata-se de um abordagem flexível e adaptável a problemas lineares e não lineares, conforme descrito na secção 2.4.

Um tipo de função de pertença, bastante utilizadas para representar os desvios aceites para os objectivos (metas), restrições e/ou coeficientes difusos são as funções triangulares. Com estas funções podemos representar facilmente qualquer tipo de "fuzificação" do problema de otimização difuso. Considerando um desvio p_i para flexibilizar as restrições e/ou para as metas de objectivos, os três tipos de funções mais comuns para representar equações de menor ou igual, maior ou igual e apenas igual, são:

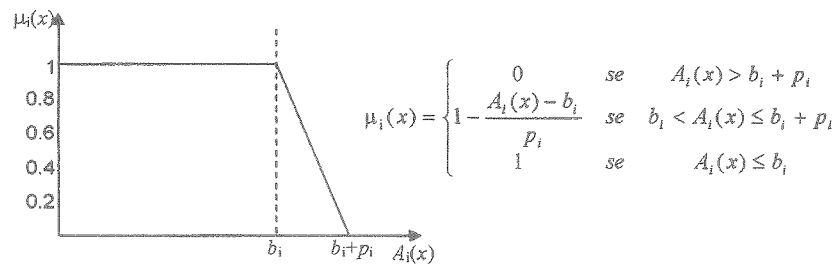


Figura 2.1 - Função de pertinência para restrições do tipo «menor ou igual».

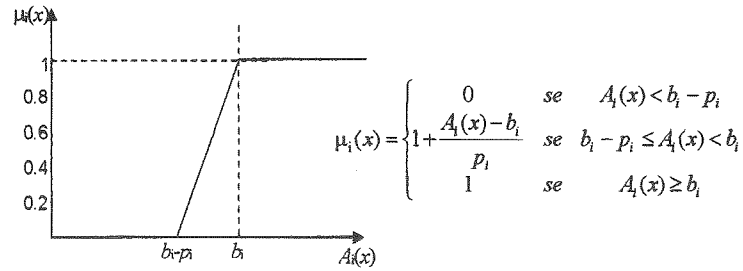


Figura 2.2 - Função de pertinência para restrições do tipo «maior ou igual».

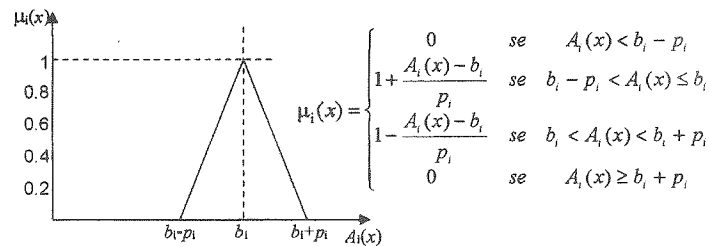


Figura 2.3 - Função de pertinência para restrições do tipo «igual».

A função representada na Figura 2.3 também pode ser usada para representar coeficientes difusos, como por exemplo "o custo por unidade x é de cerca de 200 escudos". Ou seja, a preferência do decisor é de 200 escudos mas são permitidos desvios à esquerda e à direita desta preferência. Por vezes esta representação triangular também serve para representar números difusos [8].

Note-se que o método de Zimmerman (secção 2.3) usa as funções de maior ou igual ou menor ou igual, apresentadas nas Figuras 2.1 e 2.2, para definir tanto as metas dos objectivos como as restrições de menor ou maior, respectivamente. É com base nestas funções que a transformação matemática proposta por Zimmerman [11] é possível e mantém a linearidade no problema transformado.

Quando se utilizam algoritmos independentes do problema para a sua resolução, como por exemplo o SA, qualquer outra função pode ser utilizada, nomeadamente uma função do tipo sinusoidal, ou trapezoidal, uma vez que não existem problemas de linearidade [9].

2.2 Modelo simétrico

O primeiro método, *Maxmin*, para resolver problemas de decisão em ambiente difuso foi proposto por Bellman e Zadeh [2]. Estes autores consideram que a decisão sobre cada alternativa resulta da intersecção dos objectivos e das restrições (min) e que a melhor decisão é dada pela união das decisões sobre cada alternativa (max).

Uma decisão difusa pode então ser vista como a intersecção de metas e restrições o que significa tratar-se dum modelo simétrico. Uma decisão difusa de maximização é definida como um ponto no espaço de alternativas, no qual a função de pertença ou de membro de uma decisão difusa atinge o seu valor máximo [2]. Este modelo pode ser formalizado pela expressão matemática seguinte:

$$\mu_{D^*}(x) = \max \min_{i=1}^{m+k} \mu_i(x), \forall x \in X$$

onde $\mu_{D^*}(x)$ é a decisão "óptima".

Grande parte da tomada de decisão ocorre num ambiente em que os objectivos, as restrições e as consequências de possíveis acções não são conhecidos com precisão. Para lidar quantitativamente com imprecisão torna-se por vezes necessário recorrer a ferramentas providenciadas pela teoria dos conjuntos difusos, de controlo e de informação [2].

Voltando ao conceito de decisão, observa-se que, intuitivamente, uma decisão é basicamente uma escolha ou um conjunto de escolhas alternativas efectuadas a partir de alternativas disponíveis. Esta ideia pode ser formalizada através da definição seguinte [2]: assumindo que existem metas difusas M e restrições difusas R, num espaço de alternativas X, então, M e R combinam-se para formar um conjunto difuso resultante da intersecção de M e R. Simbolicamente,

$$D = M \cap R$$

e, correspondentemente, $\mu_D = \mu_M \wedge \mu_R$. O "óptimo" é obtido pela escolha do melhor D para cada alternativa X (max de X).

No modelo de Bellman e Zadeh [2], não existe, portanto, diferença entre objectivos e restrições, no modelo do problema, assim como não existe diferença entre objectivos simples e múltiplos (compostos). Com o modelo simétrico, um problema de programação matemática torna-se num problema de satisfação de restrições, onde uma decisão é uma confluência de restrições propriamente ditas e objectivo(s).

2.3 Modelo simétrico para programação linear

A primeira proposta de programação linear difusa deve-se a Zimmermann em 1976 [11]. Este autor considera que existe flexibilidade para a violação das restrições e objectivos e baseia a sua proposta no modelo simétrico de Bellman e Zadeh (isto é, não existe diferença entre restrições e objectivos).

Considerando a formalização clássica do problema de optimização, conforme definido na secção 2.1,

$$\begin{aligned} \max \quad & Z = C'x \\ & Ax \leq b \\ & x \geq 0 \end{aligned}$$

e usando o modelo simétrico [2] podemos formalizar o problema como:

$$R(x) \leq \tilde{B} \text{ onde } \begin{pmatrix} -C \\ A \end{pmatrix} = R \text{ e } \begin{pmatrix} -Z \\ b \end{pmatrix} = B$$

$$e \quad D^*(x) = \max \min \left(1 - \left(\frac{R_i - B_i}{p_i} \right) \right) \text{ para } x \geq 0$$

Considerando as funções de pertença (Figuras 2.1 e 2.2) para as tolerâncias p_i , definidas à priori pelo decisor e que representam os níveis de violação do objectivo e das restrições, Zimmermann [11] propõe a seguinte transformação linear,

$$\begin{aligned} \text{Max} \quad & \lambda \\ \text{s.a.} \quad & \\ & \lambda p_i + R_i(x) \leq p_i + B_i \\ & x \geq 0, i = 1, \dots, m+1 \end{aligned}$$

em que λ representa a medida de satisfação de restrições difusas e onde p_i são os desvios dos valores *crisp*¹.

Note-se que, com esta transformação, o problema pode ser resolvido com um algoritmo simples como o *Simplex*, o que representa uma vantagem deste método. No entanto, este modelo não é compatível com a "fuzificação" de coeficientes tanto na F.O. como nas restrições e esta é a razão de ser necessário recorrer a outros métodos.

2.4 Método de "fuzificação"/flexibilização completa

Em geral a "fuzificação" completa do modelo de programação linear, de optimização difusa, inclui seis formas de imprecisão, como se apresenta a seguir [9]:

Caso 1: Imprecisão nos limites das restrições. Isto implica a "fuzificação" dos limites das desigualdades (por exemplo, "o tempo total para pintar um apartamento deverá ser consideravelmente inferior a 100 horas").

Caso 2: É imposta uma meta difusa para a função objectivo. Isto implica essencialmente a "fuzificação" da função de utilidade (por exemplo, "o custo total para o projecto deverá ser mantido consideravelmente abaixo de cem mil euros").

Caso 3: Imprecisão composta: que implica a combinação de imprecisões dos tipos anteriores.

Caso 4: Os coeficientes das variáveis das restrições não são conhecidos com precisão. Isto significa que os coeficientes são números difusos (por exemplo, "o custo por hora de produção de uma camisa x , é cerca de 5 euros").

¹ Entenda-se por valor *crisp* todo aquele valor normal, que está associado a uma a formalização do problema de optimização, isto é, um valor não "fuzificado".

Caso 5: Os coeficientes das variáveis da função objectivo não são conhecidos com precisão. Isto significa que esses coeficientes são números difusos, tal como em 4 (por exemplo, "o preço de venda do produto x é cerca de 10 euros por item").

Caso 6: Todas as combinações possíveis de incerteza.

Em resumo, o problema de optimização difuso com coeficientes difusos e parâmetros difusos pode ser expresso matematicamente por:

$$\begin{aligned} \max/\min \quad Z &= \tilde{C}x \\ \tilde{A}x \{ \succeq, \preceq, \equiv \} B \\ x &\geq 0 \end{aligned}$$

onde \tilde{C} é o vector de custos difusos, \tilde{A} é a matriz que contém os coeficientes difusos do(s) objectivo(s) e das restrições e \succeq, \preceq, \equiv representa o sinal correspondente à "fuzificação" dos limites (parâmetros) do(s) objectivo(s) e/ ou dos recursos (cuja informação, alternativamente, se pode representar por \tilde{B}).

Vários autores propuseram formalizações para resolver o caso da "fuzificação" completa, como por exemplo Carlsson e Korhonen [citados em 5] e Ribeiro e Moura-Pires [8]. Neste trabalho apenas discutimos em detalhe o método de Ribeiro & Moura-Pires pois é flexível e facilmente adaptável a problemas com objectivo único ou múltiplo, assim como para problemas lineares e não-lineares.

Método de "fuzificação" completa (flexível) proposto por Ribeiro & Moura-Pires

Ribeiro e Moura-Pires propuseram uma "extensão" aos métodos de resolução de problemas de optimização difusos existentes, por forma a incluir coeficientes difusos tanto na função objectivo como nas restrições. Todo o processo é designado de "abordagem flexível" para problemas de optimização difusos [7,9] e a sua formalização é:

$$\begin{aligned} \max_x \quad Z &= \sum_{k=1}^K \tilde{e}_k \cdot x_k \\ \sum_{k=1}^K \tilde{a}_{ik} \cdot x_k &= \tilde{b}_i, \quad i = 1, \dots, N \\ x &\geq 0 \end{aligned}$$

que é convertida no seguinte sistema de equações não-lineares:

$$\begin{aligned} \max_x \quad Z &= \sum_{k=1}^K w_k \cdot x_k \\ \max_{y, w} \quad &(\mu_{a_{ik}}, \mu_{b_i}, \mu_{c_k}) \\ \sum_{k=1}^K y_{i,k} \cdot x_k &\{ =, \leq, \geq \} \tilde{b}_i \end{aligned}$$

$$y_{i,k} = \tilde{a}_{i,k}$$

$$w_k = \tilde{c}_k$$

onde $k = 1, \dots, K$ e $i = 1, \dots, N$ e $x, y, w \geq 0$.

Note-se que todos os parâmetros difusos são apresentados com um til por cima e que as funções de pertença usadas são idênticas à representação constante nas Figuras 2.1, 2.2 e 2.3, isto é, são usadas funções triangulares. Note-se que quaisquer outras funções de pertença poderiam ser utilizadas.

Neste processo de flexibilização c_k e a_{ik} difusos tornam-se variáveis do processo. Poderá mesmo acontecer que um destes coeficientes seja dependente de outro c_k ou a_{ik} . Estas dependências podem ser facilmente incorporadas no modelo expressando os relacionamentos existentes. Note-se que com este novo método apenas as variáveis independentes são variáveis do problema de otimização.

O método de resolução proposto nesta abordagem assenta em 2 passos fundamentais:

- 1º) Encontrar o valor de x , y e w , variáveis independentes que maximizam o μ (Niu) agregado das restrições.
- 2º) Encontrar o valor óptimo de Z que satisfaz todas as restrições com um α -cut² = α_{\max} .

Este método permite manipular qualquer tipo de problema de otimização difuso linear ou não linear. Outra vantagem é a possibilidade de resolver problemas *crisp* que não têm solução, característica que é facultada através do seu funcionamento num ambiente difuso e pela inclusão de imprecisão quando os coeficientes ou os parâmetros não são conhecidos com precisão.

Em conclusão poderá dizer-se que este modelo de "fuzificação" aumenta a possibilidade de estabelecimento de compromissos entre satisfação de restrições e o valor para a meta a atingir, ao dispor do decisor e, deste modo, permite uma tomada de decisão mais flexível.

3. Resolução de Problemas de Otimização

A análise da complexidade indica se é possível resolver um determinado problema de otimização e se é possível encontrar uma solução óptima, num intervalo de tempo superiormente limitado por um polinómio em função do comprimento de entrada (dados) do problema, ou seja, em tempo polinomial. Um algoritmo de complexidade polinomial é um algoritmo cuja função de complexidade temporal é $O(p(k))$, onde p é um polinómio qualquer e k é o comprimento de entrada de uma instância do problema. Cada algoritmo cuja função de complexidade temporal não pode ser limitada deste modo será designado algoritmo de tempo exponencial.

A classe de problemas P consiste em todos os problemas de decisão que poderão ser resolvidos através da máquina determinística de Turing, um modelo abstracto de computação,

² Um α -cut é um valor de limiar que define um limite mínimo para o nível de satisfação das restrições do problema considerado.

em tempo limitado superiormente por um polinómio do comprimento de entrada [3]. Sabe-se que $P \subseteq NP$. A classe de problemas NP, de problemas de decisão, consiste em todos os problemas de decisão que poderão ser resolvidos em tempo polinomial por uma máquina não determinística de Turing. Esta classe de problemas inclui duas subclasses de problemas que são os problemas do tipo *NP-Hard* e *NP-Complete* [3].

De entre a enorme variedade de algoritmos de optimização é possível identificar um conjunto de procedimentos que podem ser agrupados em três categorias fundamentais [3]:

- Procedimentos matemáticos, exactos ou de aproximação;
- Procedimentos baseados em Simulação;
- Procedimentos heurísticos diversos.

Os procedimentos matemáticos exactos correspondem a algoritmos de resolução exactos, que têm por objectivo encontrar a solução óptima. Três classes fundamentais destes algoritmos são os Algoritmos Exactos de Programação Linear, de Programação Dinâmica e os Algoritmos de "Ramificação e Limite" exactos ou puros (*Branch and Bound - B&B*). Estes algoritmos, sendo exaustivos, não são adequados à resolução de problemas de grandes dimensões ou "difíceis". Sendo assim, existem, alternativamente, os algoritmos ou procedimentos heurísticos, que são algoritmos mais apropriados para a resolução de problemas "difíceis". Embora estes métodos de resolução não garantam a obtenção de soluções óptimas, são normalmente mais expeditos e permitem, geralmente, obter soluções boas ou pelo menos aceitáveis, podendo mesmo chegar à solução óptima. Grande parte destes algoritmos são relativamente recentes e têm sido cada vez mais explorados, como é o caso dos Algoritmos de Pesquisa na Vizinhança ou de Pesquisa Local Extendidos (*Extended Neighborhood Search*), que incluem *Simulated Annealing*, Algoritmos de Pesquisa Tabu assim como Algoritmos Genéticos. Existem ainda Algoritmos de Pesquisa em Feixe (*Beam Search Methods*), Algoritmos Aproximados de Programação Dinâmica e outros Algoritmos Aproximados de Ramificação e Limite, baseados em técnicas *B&B*, entre outros, nomeadamente, um conjunto de *Bottleneck Methods* [3].

Muitos dos problemas de optimização são do tipo *NP-Complete*. Geralmente acredita-se que os problemas deste tipo não são resolúveis em tempo polinomial. Consequentemente, existe muito interesse em usar algoritmos heurísticos ou de aproximação, que permitem encontrar uma boa solução, em tempo de execução razoável ou aceitável, o que motivou a realização deste estudo. Este trabalho centra-se num dos algoritmos pertencentes à classe dos algoritmos de pesquisa local ou de pesquisa na vizinhança, mais precisamente, no *Simulated Annealing*. Existem vários esquemas de pesquisa, mas todos têm a mesma característica de possuir uma função de vizinhança subjacente, que é utilizada para guiar a pesquisa até uma boa solução.

O tratamento das vizinhanças depende do problema considerado e encontrar funções de vizinhança eficientes, que conduzem a óptimos locais de elevada qualidade pode ser visto como um dos desafios colocados à pesquisa local em geral. Até agora não existem ainda regras gerais

e cada caso terá de ser resolvido em particular. Estes problemas de otimização podem ainda ser problemas de natureza difusa, como é o caso dos problemas que aqui se pretendem analisar.

3.1 Algoritmo de *Simulated Annealing*

Um algoritmo de *Simulated Annealing* (SA) pertence a uma classe de algoritmos de pesquisa local que também são conhecidos sob a designação genérica de "algoritmos de limiar" (do Inglês, *Threshold Algorithms*). Estes algoritmos desempenham um papel importante na pesquisa local, por duas razões; em primeiro lugar, porque têm sido bem sucedidos, quando aplicados a um vasto conjunto de problemas práticos, o que lhes atribuiu uma considerável afirmação entre os seus utilizadores. Em segundo lugar, alguns desses algoritmos, tal como é o caso do SA, têm uma componente estocástica, que facilita uma análise teórica da sua convergência assintótica e isto tornou-os bastante populares entre os matemáticos [1].

O algoritmo de *Simulated Annealing* foi originalmente proposto por Kirkpatrick e outros, em 1983 e por Cerny, em 1988, por analogia com o processo de fundição (*annealing*) de um sólido (citado em [1]). O objectivo de um algoritmo desta natureza é encontrar a melhor solução de entre um número finito de soluções possíveis. A técnica de SA é uma técnica particularmente atractiva pois permite encontrar soluções próximas da óptima, à custa de um esforço computacional geralmente pouco exagerado. É de notar que neste tipo de algoritmo, não é possível saber se a melhor solução encontrada, é o óptimo global. Esta característica restringe a sua utilização a casos em que um bom óptimo local é aceitável [3].

Em procedimentos do tipo SA, as sequências das soluções não tendem linearmente para um óptimo local, como acontece noutras técnicas de pesquisa local. Sendo assim, verifica-se que as soluções traçam um percurso ou trajecto variável, através de um conjunto S, de soluções possíveis e este percurso tende a ser guiado numa direcção "favorável".

Regra geral, poderá dizer-se que o SA consiste num procedimento fiável para usar em situações em que o conhecimento é escasso ou se aparenta difícil de aplicar algoritmicamente. Mesmo para dar soluções a problemas complexos, esta técnica é relativamente fácil de implementar e normalmente executa um procedimento do tipo *hill-climbing* com múltiplos recomeços. Tipicamente, esta técnica gera um caminho "Markoviano", em que o sucessor de um ponto actual é escolhido estocasticamente, com uma probabilidade que depende apenas do ponto actual e não da história prévia da busca. A propriedade "Markoviana" é uma bênção misturada, ou seja, permite resultados heurísticos, que são bastante bons em muitos casos, mas torna a análise teórica do algoritmo difícil. O algoritmo de SA é claramente desapropriado para a resolução óptima de problemas de optimização.

Em resumo, o algoritmo de *Simulated Annealing* (SA) é um algoritmo estocástico com uma analogia "física" de "fusão" do sistema a ser optimizado (como um sólido) usando "temperaturas elevadas" e posteriormente prosseguindo através de uma redução progressiva e suave da temperatura, até que o sistema "cristalize" e não ocorram mais alterações. O processo de fusão pode ser visto como a busca pela melhor solução. Extensões a este tipo de algoritmo

permitem também resolver problemas de otimização difusa pois não são dependentes da linearidade da função objectivo ou das restrições [9]. Sendo assim, poderão referir-se as seguintes vantagens. Ao usar uma implementação deste algoritmo em que os limiares (α -cuts) podem ser definidos, o decisor poderá definir o grau de satisfação (μ_{A_k} e/ou $\mu_{a_{ij}}$) para o parâmetro de cada restrição e/ou de cada coeficiente (da(s) função(ões) objectivo e/ou da(s) restrição(ões)) do problema. Acresce que esta técnica de resolução permite tratar problemas de otimização difusos, que podem ser abordados, de um modo controlado, através da consideração da possibilidade de existência de desvios (violações) em relação aos valores normais (*crisp*). Isto é, o algoritmo *SA* permite lidar com qualquer um dos tipos de "fuzificação" referidos no ponto 2.4.

3.2 Implementação do *Simulated Annealing*

Como referido, o algoritmo *SA* é um algoritmo usado para problemas de otimização onde a função objectivo (F.O.) corresponde à energia dos estados de um sólido. O algoritmo *SA* requer a definição de uma estrutura de vizinhança, assim como os parâmetros para a programação do arrefecimento. Um parâmetro de temperatura permite distinguir entre alterações profundas ou ligeiras na função objectivo. Alterações drásticas ocorrem a elevadas temperaturas e modificações pequenas a temperaturas baixas. Trata-se de um processo evolucionário, que se move em pequenos passos, de um estado para outro e existe o problema de ficar preso num óptimo local, o que constitui uma consequência deste tipo de algoritmo [9]. De acordo com [9], os quatro requisitos básicos para o utilizar o algoritmo *SA* em problemas de otimização combinatorial são:

- descrição concisa do problema;
- geração aleatória das alterações de uma configuração para outra;
- uma F.O. contendo a função de utilidade dos compromissos;
- uma definição do estado inicial, do número de iterações a serem executadas para cada temperatura e o seu esquema de *annealing*.

O pseudo código do algoritmo de *SA* utilizado (baseado na implementação apresentada em [9]) é o que se apresenta a seguir:

- Inicializar o contador de mudança de temperatura $t := 0$;
- Seleccionar a temperatura inicial $T > 0$;
- Seleccionar *threshold* > 0 ;
- Inicializar a melhor solução $Z_m :=$ solução inicial *crisp*;
- Inicializar o Gerador de números aleatórios *Seed* ≥ 0 ;
- Gerar Coef. das restrições *fuzzy* na vizinhança dos coef. *crisp*;
- Para $k:=1$ até N_r fazer $\mu(k) :=$ valor de pertença da restrição $A_k(x)$;
- $Niu1 =$ agregação($\mu(1), \dots, \mu(N_r)$);
- Repetir
 - Colocar contador de repetição $n := 0$;

```

Repetir
  Gerar estado y na vizinhança de x;
  Gerar Coef. das restrições fuzzy na vizinhança dos coef. crisp;
  k := 1;
  Repetir
     $\mu(k) :=$  valor de pertença da restrição  $A_k(x)$ ;
    k++;
  até k > Nr ou  $\mu(k-1) \leq \text{threshold}$ ;
  se  $\mu(k-1) \geq \text{threshold}$  então
    Niu2 = agregação( $\mu(1)$ , ,  $\mu(Nr)$ );
     $\delta :=$  Niu2 - Niu1;
    se  $\delta > 0$  então x:= y; Niu1:=Niu2;
      senão se  $\text{random}(0,1) < \exp(\delta/T)$ 
        então x:=y; Niu1:=Niu2;
    se  $Z_m$  melhor que  $Z_{fuzzy}$  então
      Guardar melhor óptimo intermédio
  n := n + 1;
  até n = N(t) ;
  t := t + 1;
  T := T(t);
  Até se verificar o critério de paragem;
fim;

```

A variável k é um contador para o número de restrições. É usada para calcular o grau de pertença de cada restrição $\mu(k)$. Niu representa a agregação (intersecção) dos graus de pertença das restrições difusas e representa o menor nível de satisfação das restrições (ou seja, o melhor possível obtido). Qualquer operador de agregação poderia ser utilizado, mas neste trabalho foi utilizado o operador *t-norm min*. A variável δ exprime a diferença entre o valor prévio do Niu e o novo valor obtido (Niu corrente) e é utilizada para escolher o máximo Niu, ou seja, o melhor nível de satisfação das restrições. Esta variável representa a diferença de energia entre o estado anterior e o novo estado. A função de probabilidade de movimentação, para um estado de melhor energia, (o objectivo) é dada pela exponencial de δ dividido pelo parâmetro de controlo T . Quanto menor a temperatura T , menos provável será a ocorrência de alguma alteração. O $N(t)$ representa o número de soluções vizinhas geradas e, conseqüentemente, o número de soluções possíveis. O $T(t)$ é a função de decréscimo da temperatura.

Foram ainda efectuadas as seguintes escolhas para a implementação SA, baseadas no trabalho apresentado em [14]. Na implementação do algoritmo SA tornou-se necessário especificar os seguintes parâmetros: como gerar um estado y , vizinho de x ; que função de

agregação (Niu) usar; o número de vizinhos gerados, $N(t)$; a função de decréscimo da temperatura, $T(t)$ e finalmente o critério de paragem.

A escolha de como gerar um estado y , vizinho de x , é efectuada através da definição de um novo estado, que é um ponto aleatório y , em que a distância desse novo ponto ao ponto actual z é aleatória e inferior a $\rho(t)$, definida por:

$$\rho(t) = \begin{cases} 1 & \text{se } t < 100 \\ 2 & \text{se } 100 \leq t < 150 \\ 3 & \text{se } 150 \leq t < 250 \\ 5 & \text{se } 250 \leq t < 350 \\ 15 & \text{se } t \geq 350 \end{cases}$$

Como referido a função de agregação é a intersecção de todos os valores de pertença das restrições e o operador usado para a intersecção é o *t-norm min*. A intersecção (dada pelo Niu) representa o "e" lógico, para indicar que todas as restrições têm de ser satisfeitas com um certo nível (min). O objectivo torna-se então o da maximização da minimização dos desvios das restrições.

Os pontos vizinhos são gerados através do uso de coordenadas polares, que garantem a aleatoriedade necessária à selecção dos novos pontos a considerar. Para n variáveis de decisão, tem-se que para Δx ,

$$\begin{aligned} \Delta x_1 &= \rho \times \cos \theta_1 & \theta_1 &\in [0, 2\pi] \\ \Delta x_2 &= \rho \times \sin \theta_1 \times \cos \theta_2 & \theta_2 &\in [0, \pi] \\ &\dots & & \\ \Delta x_{k-1} &= \rho \times \sin \theta_1 \times \dots \times \sin \theta_{k-2} \times \cos \theta_{k-1} & \theta_{k-1} &\in [0, \pi] \\ \Delta x_k &= \rho \times \sin \theta_1 \times \dots \times \sin \theta_{k-2} \times \sin \theta_{k-1} \times \sin \theta_k & \theta_k &\in [0, \pi] \end{aligned}$$

O novo ponto, na vizinhança do anterior é então dado por:

$$\begin{aligned} y &= x + \Delta x \\ (y_1, \dots, y_n) &= (x_1, \dots, x_n) + (\Delta x_1, \dots, \Delta x_n) \end{aligned}$$

Deve notar-se que, após a geração de um novo ponto na vizinhança é necessário testar se esse ponto (vector de valores, y) satisfaz as restrições do problema difuso e além disso, no caso de problemas do tipo inteiro, haverá a necessidade de proceder a um arredondamento dos componentes do vector y (ou a selecção ou adaptação das funções de pertença para estes casos), donde que:

$$y = x + \text{int}(\Delta x + 0.5)$$

A selecção do número de vizinhos a gerar $N(t)$, segue a seguinte heurística:

$$N(t) = \begin{cases} 200 & \text{se } t < 400 \\ 250 & \text{se } t \geq 400 \end{cases}$$

Esta heurística tem em consideração o facto de que mais sucessores deverão ser gerados quando a temperatura T diminui, para ter mais opções para testar. A função de temperatura $T(t)$ é uma função que usa um factor de decréscimo de 0.99. O critério de paragem do algoritmo é atingido quando a temperatura é inferior ou igual a 0.0001.

4. Problemas de Otimização Linear Testados

4.1 Formulação dos problemas lineares *crisp*

Neste trabalho foram considerados dois problemas do tipo *crisp*, como problemas de base para a formulação de problemas "fuzificados". O primeiro problema ("Crisp 1") considera 10 restrições, desde a restrição 11 à restrição 20, e o segundo problema ("Crisp 2") contempla a totalidade das restrições (22), que se apresentam na Tabela 4.1.

Restrição	x_1	x_2	x_3	x_4	x_5	x_6	x_7	sinal	parâmetro
1	0.1	0.13	0.11	0.13	0.13	0.1	0.11	\leq	500
2	0.77	1	0.91	1.11	1.25	0.83	0.77	\leq	3000
3	0.91	0.83	0.91	0.91	0.91	0.83	0.77	\leq	3600
4	0.08	0.09	0.09	0.08	0.1	0.1	0.09	\leq	600
5	1	1.67	1.43	1.25	1.11	1.25	1	\leq	6000
6	0	2.5	2.5	3.33	0	0	0	\leq	3600
7	0	0.17	0.25	0.2	0	0	0	\leq	500
8	0	0	0	0	2.86	0.2	0.25	\leq	600
9	0.33	0.5	0.5	0.56	0.67	0.33	0.63	\leq	1500
10	0.2	0.2	0.25	0.33	0.25	0.29	0.29	\leq	1800
11	0.25	0	0.5	0	0.33	0	0	\leq	600
12	0.25	0.29	0.29	0.25	0.2	0.25	0.33	\leq	1800
13	0	0	1	0	0	0	0	\leq	50
14	0	0	0	1	0	0	0	\leq	60
15	0	0	0	0	0	1	0	\leq	35
16	0	0	0	0	0	0	1	\leq	45
17	1	1	1	1	1	1	1	\leq	150
18	1	0	0	0	0	0	0	\geq	10
19	0	1	0	0	0	0	0	\geq	20
20	0	0	1	0	0	0	0	\geq	20
21	0	0	0	1	0	0	0	\geq	20
22	0	0	0	0	1	0	0	\geq	10

$$x_j \geq 0 \quad (j = 1, 2, \dots, 7)$$

Tabela 4.1 - Restrições dos problemas *crisp*.

A função objectivo escolhida para teste é a seguinte.

$$\text{Max } 200 x_1 + 300 x_2 + 400 x_3 + 500 x_4 + 600 x_5 + 600 x_6 + 650 x_7$$

Ao problema "Crisp 1" corresponde um valor de F.O. de 78250 e ao problema "Crisp 2" um valor de 76250, para os seguintes valores das variáveis de decisão:

Problema	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Crisp1	10	20	20	0	55	0	45
Crisp2	10	20	20	20	10	25	45

Tabela 4.2 - Valores das variáveis de decisão para os problemas *crisp*.

4.2 "Fuzificação" do problema

Nos problemas de optimização linear foram considerados os seguintes 7 casos de flexibilização:

- Caso 1. "Fuzificar" os coeficientes da função objectivo e o valor da função objectivo F.O. (meta);
- Caso 2. "Fuzificar" valor da função objectivo e o valor das restrições (parâmetros).
- Caso 3. "Fuzificar" os coeficientes e o valor da F.O. e o valor das restrições.
- Caso 4. "Fuzificar" o valor da F.O. e os coeficientes das restrições.
- Caso 5. "Fuzificar" os coeficientes e o valor da F.O. e os coeficientes das restrições.
- Caso 6. "Fuzificar" o valor da F.O. e os coeficientes e o valor das restrições.
- Caso 7. "Fuzificar" os coeficientes e o valor da F.O. e os coeficientes e o valor das restrições.

Estas formas de flexibilização foram então aplicadas a cada uma das formulações *crisp* de base, referidas no ponto 4.1, perfazendo 14 problemas. Cada um destes problemas foi ainda submetido a uma análise de flexibilização com desvios permitidos de 10 e 15%, dando origem a 28 casos distintos. Estes 28 casos, por sua vez, foram ainda testados sob 3 cenários distintos, em termos de valor de limiar imposto no nível de satisfação das condições destes problemas (*threshold*), correspondentes a 30, 60 e 90%. Sendo assim, no total, foram testadas 84 situações distintas, com a finalidade de proceder a uma análise suficientemente detalhada e diversificada dos resultados obtidos pelo algoritmo de SA para os diferentes cenários considerados.

De acordo com o formalismo apresentado no ponto 2.4, quando num problema surgem coeficientes flexibilizados, tanto na F.O como nas restrições e/ou se considera a "fuzificação" de parâmetros das restrições (e/ou do valor da F.O - meta) estas flexibilizações serão denotadas, na correspondente formulação, através de um til, colocado por cima do respectivo coeficiente e/ou sinal da inequação, por exemplo, $0.\tilde{3}3$ e $\tilde{\leq}$.

5. Resultados dos Problemas de Otimização

Em geral o processo para resolver um problema de otimização difuso consiste nos seguintes passos:

1. Dependendo do tipo de problema, formalizá-lo na forma de problema de programação linear ou na forma de um problema multi-objectivo, ou mesmo como um problema de programação não-linear.
2. Se se pretender "fuzificar" os objectivos, definir a meta para esses objectivos.
3. Definir a função de pertença para representar a "fuzificação" de qualquer restrição. (Ex: triangular, sinusoidal, trapezoidal, ou outra).
4. Definir limiares para o grau de aceitação de desvios (violações) na satisfação das restrições ou condições.
5. Definir o operador de agregação para combinar as restrições (e metas) como por exemplo uma "t-norma".
6. Resolver o problema com um algoritmo do tipo SA.

Note-se que existe a possibilidade de correr o algoritmo SA para diferentes valores de limiar (α -cut ou *threshold*) e o objectivo consiste em permitir uma comparação entre diferentes resultados obtidos, para um mesmo problema, a fim de analisar se é possível convergir para um determinado ponto ou zona de pontos óptimos ou para tentar extrair as conclusões que se apresentem mais promissoras para a resolução de um determinado problema ou, simplesmente, para ter uma noção de que existem diferentes alternativas de decisão, às quais correspondem determinadas situações de compromisso, em termos de optimização do valor da F.O. (Z) e do nível de satisfação global das condições do problema (Niu).

A Figura 5.1 permite uma visualização das janelas de diálogo que servem de interface para a introdução dos dados e para a apresentação dos resultados dos diversos problemas, respectivamente.

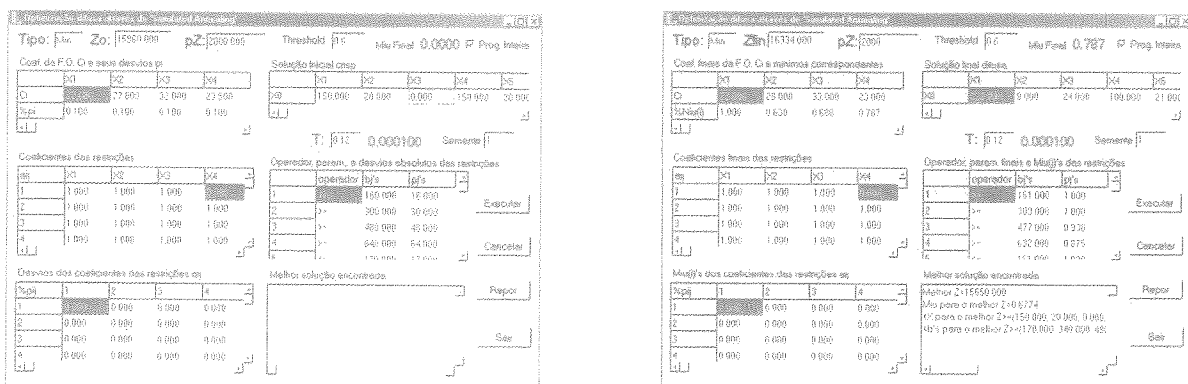


Figura 5.1 - Janelas de diálogo da aplicação, para a introdução dos dados (a) e para a apresentação de resultados (b).

É ainda importante lembrar que os exemplos que se apresentam a seguir foram executados para valores de limiar (*threshold*) de 0.3, 0.6 e 0.9, para os diferentes casos de "fuzificação" abordados.

Relativamente aos casos de "fuzificação" considerados estes foram divididos em dois grupos fundamentais de problemas, A e B, que são os problemas em que se consideram, respectivamente, desvios ou violações de 10%, relativamente aos valores *crisp* (Tabela 5.1) e desvios de 15% (Tabela 5.2).

Também se achou conveniente utilizar uma "semente", no algoritmo, para a geração de números aleatórios, para permitir a geração dos mesmos valores, em diferentes execuções do programa, para um mesmo exemplo, a fim de ser possível comparar os diferentes resultados obtidos, face a pequenas variações nas condições de execução de cada problema. Nestes problemas utilizou-se uma semente de valor 1.

Além disso também se implementou um relógio, que permite avaliar o tempo de resposta (em segundos) aos vários problemas e que estão expressos pela variável "Tempo", nas tabelas abaixo.

As tabelas seguintes apresentam ainda o valor da F.O. (Z), para cada problema considerado e o correspondente nível de satisfação das condições do problema (Niu). Esta variável contempla desvios no valor das restrições, no valor da F.O. e nos correspondentes coeficientes e combinações destas situações, de acordo com os casos de "fuzificação" referidos. Será ainda de acrescentar que Niu representa o nível mínimo de violação das restrições, ou seja, a melhor satisfação possível encontrada para o valor da F.O.

As designações dos problemas contêm informação relativa ao tipo de caso de "fuzificação" de que se trata (casos 1 a 7) (Ex.: Caso1.yz, para um problema da classe dos problemas "Caso 1"), informação relativa ao problema *crisp* em que se baseia (Ex.: Casox.1Z), que corresponde ao problema "Crisp 1") e ainda informação relativa à classe de problemas em que se insere, relativamente aos desvios ou violações permitidas (Ex.: Casox.yA, para um problema pertencente à classe A, caracterizado por desvios de 10%).

Os resultados para os problemas da classe A foram os seguintes:

Desvios10%	Threshold = 0.3			Threshold = 0.6			Threshold = 0.9		
	Z	Niu	Tempo	Z	Niu	Tempo	Z	Niu	Tempo
Caso1.1A	82223.584	0.4059	297	80321.289	0.6050	648	78606.265	0.9059	2116
Caso1.2A	79775.478	0.3458	990	77901.232	0.6232	1402	76575.161	0.9676	2524
Caso2.1A	82935.475	0.3003	59	80623.626	0.6037	51	78796.001	0.9083	42
Caso2.2A	81520.987	0.3019	135	79040.269	0.6044	129	76889.279	0.9109	126
Caso3.1A	88864.288	0.3366	93	82535.396	0.6064	140	79276.940	0.9010	681
Caso3.2A	85833.658	0.3237	142	80753.384	0.6058	154	77348.281	0.9005	169
Caso4.1A	79496.303	0.3310	934	79748.248	0.6332	2686	78645.183	0.9013	33

Caso4.2A*	76163.606*	0.3385	193	76055.217*	0.6037	372	76104.835*	0.9005	2287
Caso5.1A	85689.700	0.3238	116	81218.748	0.6311	314	78967.500	0.9014	329
Caso5.2A	78994.391	0.3065	137	77604.922	0.6169	273	76433.001	0.9001	1267
Caso6.1A	82995.997	0.3240	182	81379.872	0.6000	240	78907.454	0.9001	850
Caso6.2A	82411.910	0.3001	210	79290.528	0.6012	255	77000.635	0.9016	631
Caso7.1A	91032.668	0.3127	179	84524.696	0.6031	277	79200.534	0.9004	2643
Caso7.2A	85582.212	0.3100	226	80973.784	0.6028	293	77079.792	0.9041	796

Tabela 5.1 - Resultados para os problemas da classe A.

Os resultados para os problemas da classe B foram os seguintes:

Desvios10%	Threshold = 0.3			Threshold = 0.6			Threshold = 0.9		
	Z	Niu	Tempo	Z	Niu	Tempo	Z	Niu	Tempo
Caso1.1B	84037.685	0.8449	161	81499.400	0.6270	95	78795.698	0.9738	1978
Caso1.2B	80953.229	0.3051	192	77999.003	0.6044	181	76709.386	0.9109	1671
Caso2.1B	86308.608	0.3215	114	82371.776	0.6046	160	78951.227	0.9034	373
Caso2.2B	85024.635	0.3404	135	80704.633	0.6200	129	77367.162	0.9003	124
Caso3.1B	89596.860	0.3077	83	86290.221	0.6000	108	79704.568	0.9003	422
Caso3.2B	88283.349	0.3397	139	83891.885	0.6025	147	77864.135	0.9008	153
Caso4.1B	83591.246	0.3238	86	80539.869	0.6050	83	78909.421	0.9024	2627
Caso4.2B*	76104.068*	0.3304	83	76044.617*	0.6120	130	76045.359*	0.9023	1178
Caso5.1B	88934.092	0.3185	75	82575.194	0.6022	377	79091.023	0.9031	3435
Caso5.2B	81857.123	0.3044	97	78794.372	0.6126	265	76377.843	0.9012	955
Caso6.1B	83719.261	0.3011	146	81379.492	0.6001	335	78972.459	0.9006	1121
Caso6.2B	86197.369	0.3022	187	79298.176	0.6002	218	77012.363	0.9000	524
Caso7.1B	89047.650	0.3035	157	84771.080	0.6084	452	79815.201	0.9000	1898
Caso7.2B	90734.469	0.3036	203	83884.405	0.6109	243	77649.240	0.9002	615

Tabela 5.2 - Resultados para os problemas da classe B.

Os problemas "Caso4.2A" e "Caso4.2B", assinalados com '*', nas Tabelas 5.1 e 5.2, correspondem aos únicos exemplos de problemas, de entre os que foram analisados, em que não se verifica uma melhoria no valor da F.O. (Z), como se pode observar nos resultados tabelados.

5.1 Análise de resultados

As experiências, correspondentes aos 7 casos de "fuzificação" apresentados no ponto 4.2, foram executadas partindo de uma temperatura inicial de 0.12. Os diferentes tipos de violações ou desvios permitidos foram de 10% nos problemas do tipo A e de 15 % nos problemas do tipo B, isto tanto nos valores dos parâmetros das restrições (e da F.O.) como nos coeficientes das

restrições e da F.O. Estes desvios serviram para elaborar as funções de pertinência, de acordo com o respectivo sinal da (in)equação e utilizando as funções triangulares descritas na secção 2.1.

A seguir, nos pontos 5.1.1 e 5.1.2, mostram-se alguns gráficos, para uma melhor visualização e análise dos resultados obtidos, anteriormente apresentados.

5.1.1 Resultados do valor da F.O versus o nível de satisfação global das condições do problema

A Figura 5.2 apresenta os resultados obtidos para os problemas das classes A e B, para um threshold de 30%.

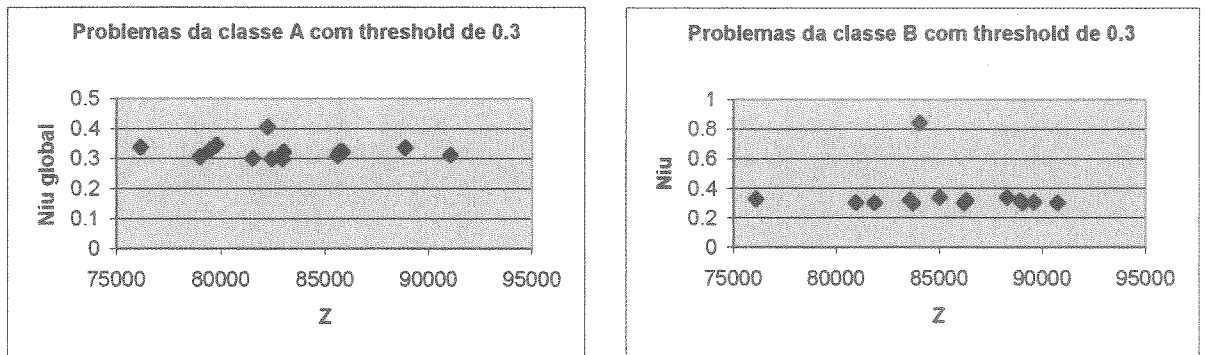


Figura 5.2 - Valores de Z versus Niu global para um threshold de 30%.

Na classe de problemas A, para um threshold de 30%, o problema que corresponde a um melhor valor da F.O. é o "Caso7.1A" (Z=91032.668), para um nível de satisfação dos coeficientes da F.O. e das restrições e do valor das restrições de 31.27% (Niu=0.3127). Na classe de problemas B, para um mesmo threshold, o melhor Z (90734.469) verificou-se para um Niu de 30.36%, no "Caso7.2B".

A Figura 5.3 apresenta os resultados obtidos para os problemas das classes A e B, para um threshold de 60%.

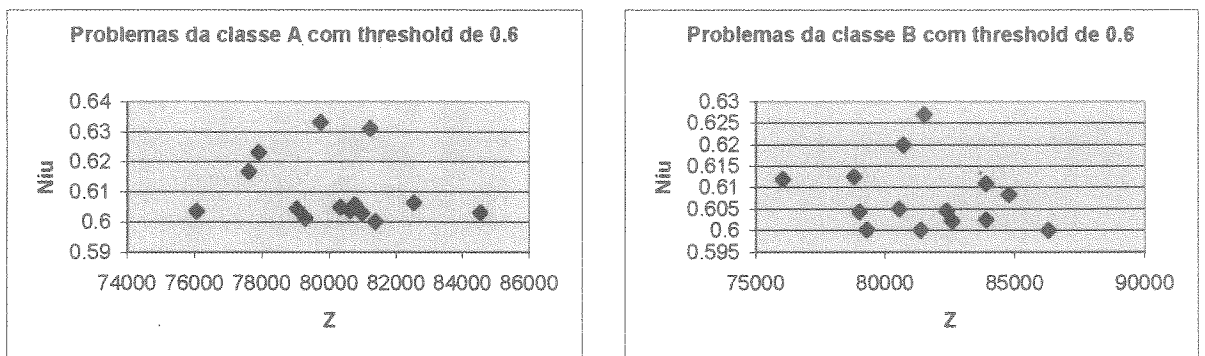


Figura 5.3 - Valores de Z versus Niu global para um threshold de 60%.

No conjunto dos problemas com *threshold* de 60%, a melhor ocorrência, da classe A, foi também o "Caso7.1A" ($Z=84524.696$) e da classe B o "Caso3.1B" ($Z=86290.221$).

A Figura 5.4 apresenta os resultados obtidos para os problemas das classes A e B, para um *threshold* de 90%.

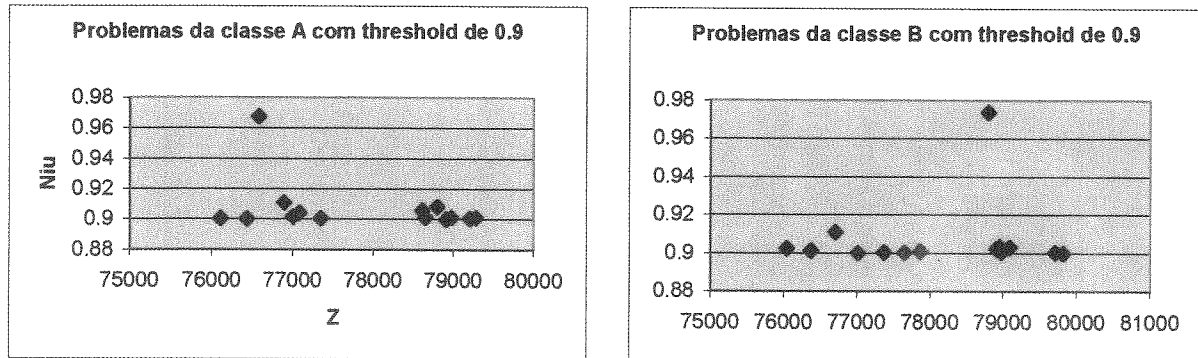


Figura 5.4 - Valores de Z versus Niu global para um *threshold* de 90%.

Relativamente aos problemas com *threshold* de 90% a melhor ocorrência, em termos de valor da F.O., verificou-se para o "Caso3.1A", na classe A e para o "Caso7.1B", na classe B.

5.1.2 Resultados do valor da F.O versus tempo de resolução

A seguir apresenta-se uma outra análise aos tipos de problemas considerados, desta vez com o objectivo de relacionar a variável Z (valor da F.O.) com o tempo de resposta do algoritmo a esses problemas.

A Figura 5.5 apresenta os resultados obtidos para os problemas das classes A e B, para um *threshold* de 30%.

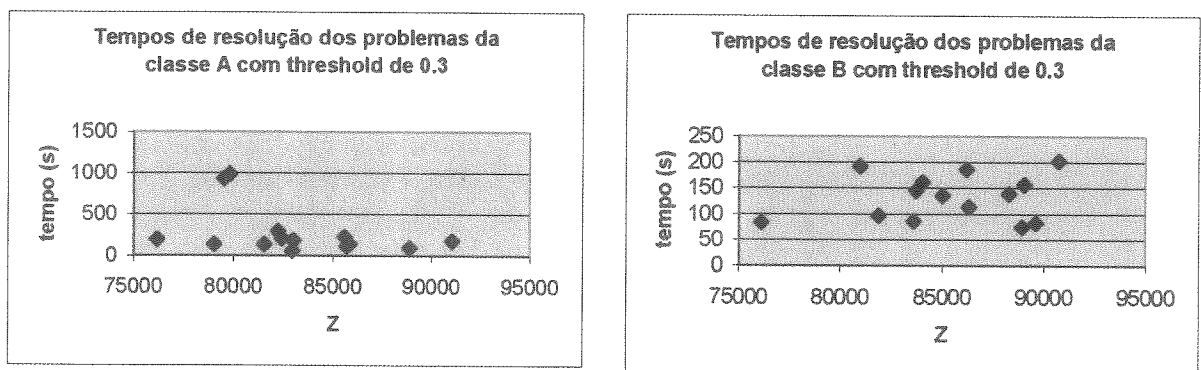


Figura 5.5 - Valores de Z versus tempos de resolução para um *threshold* de 30%.

Para um *threshold* de 30% o melhor tempo de resposta, para a classe de problemas A verificou-se para o "Caso2.1A", que foi de 59 segundos, ao qual corresponde um Z de 82935.474. Na classe B o melhor tempo registado corresponde ao problema "Caso5.1B" ($Z=88934.092$) e foi de 75 segundos.

A Figura 5.6 apresenta os resultados obtidos para os problemas das classes A e B, para um *threshold* de 60%.

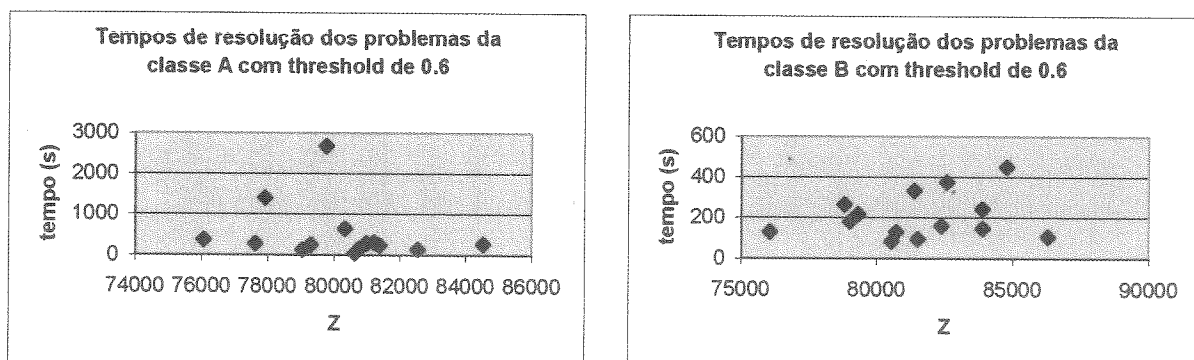


Figura 5.6 - Valores de Z *versus* tempos de resolução para um *threshold* de 60%.

Relativamente aos problemas com *threshold* de 60% o melhor tempo foi de 51 segundos, que ocorreu para o "Caso2.1A", nos problemas da classe A e de 83 segundos para o "Caso4.1B", para a classe B.

A Figura 5.7 apresenta os resultados obtidos para os problemas das classes A e B, para um *threshold* de 90%.

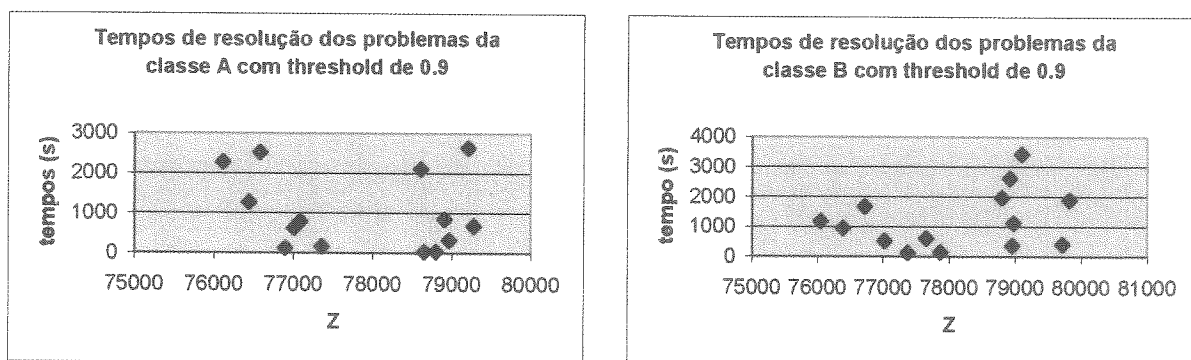


Figura 5.7 - Valores de Z *versus* tempos de resolução para um *threshold* de 90%.

Por último, no que se refere aos problemas com *threshold* de 90% o melhor tempo ocorreu para o "Caso4.1A", para a classe A, no valor de 33 segundos ($Z=78645.183$) e para a classe B foi de 124 segundos, correspondente ao "Caso2.2B", com um Z de 77367.162.

Fazendo uma análise geral final aos resultados obtidos pode constatar-se que, a um melhor valor para Z, isto é, valor da função objectivo, está geralmente associado um pior valor em termos de satisfação global das condições do problema *crisp* correspondente (Niu global). Também se pode verificar que todos os problemas, excepto o "Caso4.2A" e o "Caso4.2B", permitem uma satisfação plena do valor da F.O. (meta), o que indica que o valor da F.O. satisfaz a correspondente restrição. Além disso, ainda se pode concluir que o melhor resultado, em termos de valor de Z, nestas experiências, foi obtido para o "Caso7.1A", correspondente a uma "fuzificação" total (dos coeficientes e do valor da F.O. e dos coeficientes e do valor das restrições), obtendo-se um valor de 91032.668 para Z, associado a um valor de satisfação global de 0.3127 (muito próximo do limite mínimo aceitável, isto é, do valor limite imposto pelo *threshold* de 0.3, neste caso), portanto muito próximo da máxima violação total permitida nesta classe de problemas. Em contrapartida, o pior valor, em termos de Z, foi de 76104.068, para uma flexibilização do valor da F.O. e dos coeficientes das restrições (ligeiramente inferior ao correspondente valor do problema *crisp*, que é de 76250).

Poderá então concluir-se que, a melhor decisão a tomar, em cada cenário alternativo, perante os resultados obtidos, será uma decisão "flexível", na medida em que existem diferentes alternativas de decisão e nenhuma domina completamente outra. Sendo assim, a decisão final dependerá sempre de um compromisso entre tentar violar o menos possível, ou não violar nada, como se verifica para as soluções *crisp* ("Crisp 1" e "Crisp 2") ou violar o mais admissível, de modo a obter o "melhor" resultado desejado, em termos de valor da F.O., ou então optar por uma situação intermédia, em que se tenta chegar a uma situação de compromisso entre o valor obtido para a F.O. (Z) e o nível de satisfação global das "restrições" do problema (Niu). Como explicado anteriormente, o Niu é um valor agregado e consequentemente representa a melhor satisfação obtida para as restrições para o respectivo valor da F.O.

No que se refere à eficiência ou tempo de resposta do algoritmo SA implementado aos problemas poderá dizer-se que este apresenta comportamentos bastante diversos, ou seja, existem problemas para os quais ele consegue fornecer rapidamente uma resposta (apenas algumas dezenas de segundos), mesmo para problemas relativamente complexos, nomeadamente para os problemas pertencentes às classes de "fuzificação" incluídas no Caso 3 ("fuzificar" os coeficientes e o valor da F.O. e o valor das restrições) e no Caso 5 ("fuzificar" os coeficientes e o valor da F.O. e os coeficientes das restrições). Contudo, outras vezes verifica-se um tempo de resposta maior, que pode atingir algumas (poucas) dezenas de minutos, por vezes mesmo para problemas não muito complexos, nomeadamente, problemas pertencentes ao caso de "fuzificação" dos coeficientes e do valor da F.O. (Caso 1). Tal situação poderá justificar-se pela ocorrência de uma situação de convergência lenta, devido à entrada numa determinada zona do espaço de pesquisa, não muito favorável, para um determinado problema, que conduz a um processo moroso de resolução deste, principalmente na fase final, aquando do refinamento da qualidade da solução. Sendo assim, constatou-se que, a resolução de problemas mais complexos, isto é, que envolvem uma maior variedade de formas de

"fuzificação" (e conseqüentemente uma maior quantidade de condições a testar) e/ou problemas de maiores dimensões (nomeadamente, um maior número de restrições - problemas que derivam do problema de base "Crisp 2") e/ou problemas em que o grau de exigência relativamente ao nível de satisfação das condições deste é maior (por exemplo, os casos em que o *threshold* é de 90%) não requerem, forçosamente, um maior tempo de resolução. Por exemplo, o "Caso7.2B", para um *threshold* de 90% requereu um tempo de 615 segundos (10.25 minutos), muito inferior ao tempo requerido, por exemplo, para a resolução do "Caso1.1A", para um mesmo valor de *threshold*, que foi de 2116 segundos (aproximadamente 35.27 minutos).

Em termos deste factor de avaliação do desempenho do algoritmo SA implementado constatou-se que, para este conjunto de problemas testado, o melhor tempo que se obteve foi de 33 segundos, que se registou para o "Caso4.1A", a que correspondeu um valor de Z de 78645.183 (que representa uma melhoria de cerca de 0.5050% em relação ao correspondente problema crisp - "Crisp1") e um Niu de 90.13%, relativamente à "fuzificação" do valor da F.O. e dos coeficientes das restrições. Em contrapartida, o pior tempo verificou-se para o "Caso5.1B", no valor de 3435 segundos (57.25 minutos), para um valor da F.O. de 79091.023 (uma melhoria na ordem de 1.075%) e um nível de satisfação global das condições do problema de 90.31%, relativo à "fuzificação" dos coeficientes da F.O. e das restrições e ao valor da F.O..

6. Conclusões

Este trabalho mostra que o algoritmo SA é uma boa técnica de resolução para tratar problemas de optimização difusos, podendo as suas vantagens resumir-se do seguinte modo: simplicidade e fácil implementação; independência em relação ao problema a tratar (pode resolver problemas lineares e não lineares); não requer transformações matemáticas do problema a tratar para o resolver (como é o caso do método de Zimmerman). Note-se que na implementação do algoritmo SA foi, adicionalmente, introduzida a possibilidade, por parte do utilizador, de escolher uma semente para a geração dos valores aleatórios, através de um processo de selecção pseudo aleatório desses valores, subjacente a tal geração. A principal desvantagem do algoritmo SA é a necessidade de definir os estados iniciais que satisfazem as restrições, para cada variável. Outra desvantagem é a escolha de uma temperatura (T) adequada, porque pode implicar uma pesquisa maior e conseqüentemente mais tempo de computação.

Além dos parâmetros Z (valor da F.O.) e Niu (nível de satisfação global das condições do problema) achou-se também conveniente avaliar o tempo de resposta do algoritmo aos problemas, sendo este um *output* adicional fornecido pela aplicação desenvolvida. Relativamente a este aspecto será conveniente referir que o tempo de resposta aos problemas não apresentou um comportamento linear e claramente previsível, mas antes um comportamento que dependia, fortemente, do tipo de convergência verificado para cada problema. Sendo assim, verificaram-se situações em que, apesar de uma relativa complexidade do problema

tratado o algoritmo fornecia uma resposta rápida ao problema (ou mesmo quase imediata) e vice versa. Daqui se pode concluir que o tempo de resposta do algoritmo SA embora também dependa, em geral, do tipo de problema considerado (por exemplo, do caso de "fuzificação" e da dimensão do problema) depende essencialmente do tipo de convergência verificado, isto é, da zona do espaço de pesquisa para onde um determinado problema é canalizado, em cada execução do algoritmo (o que também é influenciado pela escolha da semente) e do nível de satisfação das condições do problema (*threshold*). Sendo assim, enquanto que o primeiro factor é, de certa forma, mais imprevisível e, desta forma, conduz a um problema mais dificilmente contornável, o segundo factor apresenta um crescimento na razão directa do tempo de resposta do algoritmo, ou seja, o tempo de resposta aumenta com o aumento do valor de *threshold*, dentro de cada classe de problemas, como também seria de esperar, pois aumenta o nível de exigência dos resultados obtidos, em termos de satisfação das condições do problema.

O tipo de convergência verificado pelo algoritmo implementado e o correspondente desempenho, em termos de tempo de resposta aos problemas é, de certa forma, um aspecto característico dos métodos de pesquisa não determinísticos, como é o caso do método de *Simulated Annealing*. Uma forma de tentar melhorar esta situação será através da introdução de determinadas heurísticas, adequadas a um determinado tipo de problemas, enquadrados num determinado contexto específico, o que conduzirá a um processo de adaptação mais rigorosa do algoritmo às suas necessidades de utilização concretas, ditadas numa determinada situação particular.

O método de "fuzificação" proposto por Ribeiro & Moura-Pires foi implementado com o algoritmo de resolução SA e ilustrado com um conjunto de exemplos, que mostram quão flexível e adaptável o método é na resolução de problemas de optimização difusos. Como se mostrou, a principal vantagem do uso deste método consiste na liberdade de que o decisor dispõem para escolher qualquer modelo que este considere apropriado para um problema específico.

A fim de permitir tratar também problemas do tipo inteiro, tornou-se necessário incluir a opção de proceder a um arredondamento dos valores (componentes) do vector de variáveis de decisão geradas.

Através da análise de resultados efectuada verificou-se que os melhores resultados, em termos de valor da F.O. foram de $Z = 91032.668$, para a classe de problemas A e $Z = 90734.469$, para a classe B e foram obtidos com a "fuzificação" total (Caso 7), isto é, "fuzificação" dos coeficientes e valor da F.O. e dos coeficientes e do valor das restrições, mas à custa de uma violação ou insatisfação das condições (*threshold* de 30%). Deve também notar-se que, em geral, se o decisor pretender atingir um melhor nível de satisfação das condições, isto é, dos parâmetros ou valores das restrições e/ou do valor da F.O. e/ou da satisfação dos valores dos coeficientes da função objectivo e/ou das restrições, isso irá conduzir a um pior resultado em termos do resultado esperado para a função objectivo e vice versa.

Em termos de trabalho futuro seria interessante estender este estudo, através de uma comparação dos resultados aqui obtidos com os resultados para outros valores para os parâmetros de controlo do algoritmo, nomeadamente a temperatura inicial e eventualmente tentar também ajustar melhor a forma de variação (diminuição) da temperatura, face à natureza dos problemas específicos que se pretendem resolver (nomeadamente, a dimensão do problema e os casos de "fuzificação" que se pretendem abordar), adicionalmente à averiguação da possibilidade de introdução de heurísticas, para facilitar o processo de resolução dos problemas a tratar.

Um outro aspecto interessante seria avaliar o comportamento do algoritmo na resolução de problemas não lineares.

Além disso, também seria interessante tentar ajustar, mais rigorosamente, este algoritmo para o caso das variáveis do tipo inteiro, através da adaptação das funções de pertença para valores discretos, de modo a tornar o algoritmo mais rigoroso e adequado à resolução deste tipo de problemas.

Referências

- [1] Aarts E., Lenstra J. K., Local Search in Combinatorial Optimization, John Wiley & Sons, Inc. (1997).
- [2] Bellman R. E., Zadeh L. A., Decision-Making in a Fuzzy Environment, Management Science 17 (1970).
- [3] Blazewics J, Ecker K.H., Pesch E., Schmidt G., Weglarz J., Scheduling Computer and Manufacturing Processes, Springer-Verlag (1996).
- [4] Bundy A., Artificial Intelligence Techniques, Springer-Verlag (1997).
- [5] Lai Y. J., Hwang C. L., Fuzzy Mathematical Programming - Methods and Applications, Springer-Verlag (1992).
- [6] Lai Y. J., Hwang C. L., Fuzzy Multiple Objective Decision Making - Methods and Applications, Springer-Verlag (1994).
- [7] Moura-Pires F., Ribeiro R. A., A New Risk Function for Fuzzy Linear Programming, Proceedings of the World Automation Congress, Alaska (WAC'98) 138 (1998) 1-10.
- [8] Ribeiro R. A., Pires F. M., Pires J.M., Solving Fuzzy Optimisation Problems: Flexible Approaches using Simulated Annealing, Proceedings of the World Automation Congress (WAC'96), Montpellier-France (1996).
- [9] Ribeiro R. A., Pires F. M., Fuzzy Linear Programming via Simulated Annealing, Kybernetika 35 (1999) 57-67.
- [10] Zadeh, L. A., Fuzzy Sets, Information and Control 8 (1965) 338-353.
- [11] Zimmermann H. J., Fuzzy Set Theory and its applications, Kluwer Academic Publishers (1991).

NOTA DO EDITOR-PRINCIPAL

Um lamentável lapso ocorreu na impressão do artigo “An algorithm for calculating K shortest paths with a maximum number of arcs”, de T. Gomes, L. Martins e J. Craveirinha, do número 1 do volume 21 e impediu a publicação da sua versão original. O Editor-Principal da Investigação Operacional assume por inteiro a responsabilidade do sucedido e apresenta as maiores desculpas aos autores do artigo e aos leitores da revista. A versão original do artigo é apresentada nas páginas seguintes deste número.

NOTE FROM THE EDITOR-IN-CHIEF

In the number 1 of volume 21, the article “An algorithm for calculating K shortest paths with a maximum number of arcs”, by T. Gomes, L. Martins and J. Craveirinha has been printed out with some errors. The Editor-in-Chief of Investigação Operacional apologizes to the authors of the article and the readers of the journal for such inconvenience. The original version of the article is included in the next pages of this number.

AN ALGORITHM FOR CALCULATING K SHORTEST PATHS WITH A MAXIMUM NUMBER OF ARCS

Teresa Gomes*

Lúcia Martins

José Craveirinha

Departamento de Engenharia Electrotécnica/INESC-Coimbra
Pólo II da Universidade de Coimbra
Pinhal de Marrocos
3030-290 Coimbra
Portugal
teresa@dee.uc.pt
lucia@dee.uc.pt
jcraiv@dee.uc.py

Abstract

In multiexchange telecommunication networks the problem of calculating shortest paths involves, in general, constraints on the maximum number of arcs in any path. In highly connected networks the number of available paths between pairs of nodes is very high and if no additional constraints are introduced, shortest paths (for a given objective function) may occur having a significant number of arcs.

In this paper we present an efficient algorithm for calculating the k shortest loopless paths with a maximum number of arcs per path, derived from the MPS algorithm [5,6] for calculating the k shortest loopless paths.

The proposed algorithm will be compared with the original MPS algorithm when both procedures are used for enumerating the k shortest loopless paths with a maximum number of arcs, in order to show the significant efficiency improvement obtained from the algorithm. The comparison will be made in terms of the total number of generated paths in each algorithm.

Keywords

Routing, shortest paths enumeration

1. Introduction

The calculation of the K -shortest paths in a network has many applications in science and engineering, namely as a solution to a sub-problem (see ex. [3, 2]). An interesting application in telecommunications as sub-procedure in a multi-objective routing model was recently proposed [1]. In a telecommunication multiexchange network routing context, the determination of optimal/sub-optimal loopless paths (according to predefined criteria) is normally constrained by the maximum number of arcs which may constitute any path. It must

* Work partially supported by FCT, project PRAXIS/P/EEI/13219/1998, Um estudo sobre encaminhamento dinâmico multi-objectivo e dependente do estado em redes multi-serviço.

be noted that in highly connected networks the number of available paths between pairs of nodes is very high and if no additional constraints are introduced, shortest paths (for a given objective function) may occur having a significant number of arcs.

A first (simplistic) approach to the K-shortest loopless path problem with a maximum (D) number of arcs would be to generate the sequence of the shortest paths, using a very efficient algorithm, such as the one proposed by Martins et al. [6] and then to select a subset of K paths by ignoring those which have more than D arcs. Note that to the best of our knowledge this algorithm is so far the most efficient one available in the literature. As an alternative an algorithm is proposed in this paper for solving this K shortest path constrained problem, which is based on the algorithm for loopless paths in [6] which is a variant of the algorithm in [5]. It will be shown that the proposed algorithm is much more efficient than the former (simplistic) approach since it prevents, by construction, the enumeration of paths with more than D arcs, therefore reducing the number of generated paths which would be disregarded later on.

This paper is organised as follows. In section 2, beyond reviewing the definitions and notation used in [5] some new definitions will be introduced. The foundations of the algorithm will be presented in section 3. The application of the algorithm to various network examples and its comparison with the simple method of rejecting the paths with more than D arcs in the MPS algorithm, will be performed in section 4. Finally some conclusion on this work and the envisaged application of the algorithm in a bi-objective routing model for telecommunication networks, will be briefly mentioned.

2. Definitions and Notation

This section begins by reviewing the basic definitions and notation used in [5]. Let (N,A) denote the graph defining a network¹ topology, where $N = \{v_1, \dots, v_n\}$ (or $N = \{1, \dots, n\}$ to simplify) is the finite set of nodes (or vertices) and $A = \{a_1, \dots, a_m\}$ is the finite set of arcs or edges, such that each arc is a pair of nodes. If all the pairs are ordered (unordered) the graph is said to be ordered (unordered). Hereafter and unless something explicit is said, the graph is supposed to be ordered. A path p from s to t ($s, t \in N$) is defined by an alternating sequence of nodes and arcs, $\langle s = v_1', a_1', v_2', \dots, a_{r-1}', v_r' = t \rangle$, where $a_k' \in A$ for any $k = 1, \dots, r-1$ and $v_k' \in N$ for any $k = 1, \dots, r$; $a_k' = (v_k', v_{k+1}')$ for any $k = 1, \dots, r-1$. The value (or cost) $c(p)$ of a path p is obtained by associating a cost $c_{ij} \in \mathbb{R}$ with each arc (i,j) :

$$c(p) = \sum_{(i,j) \in p} c_{ij} \quad (1)$$

A path from s to t in (N,A) is a loopless path iff all its nodes are different. Let p_{ij} be a path from i to j ; the concatenation of paths p_{ij} and p_{jl} is the path, $p_{ij} \diamond p_{jl}$, from i to l , which coincides with p_{ij} from i to j and with p_{jl} from j to l .

¹ The term network will be reserved to designate a broader entity which includes the graph (N,A) as a fundamental descriptive structure - e.g. a telecommunication multiexchange network.

Let $P^K = \{p_1, \dots, p_K\}$ be the set of the K shortest (unconstrained) loopless paths between a given pair of nodes (s and t). An algorithm enabling to construct \mathcal{T}_K , a pseudo-tree of the K-shortest paths (from s to t), algorithm T1, is given in [5]. When a path p_k , $k < K$, is added to \mathcal{T}_{k-1} , creating \mathcal{T}_k , the deviation node of p_k , v_k , is determined as the node whose tail is the first arc where $\mathcal{T}_{k-1} \neq \mathcal{T}_k$ differ, according to algorithm T1. In the context of the MPS algorithm the node v_k of path p, from which a new candidate path is generated is the deviation node of that new path. By definition s is the deviation node of p_1 . The sub-path from v_k to t in p_k is the deviation path of p_k , $p_{v_k t}^k$. Note that although nodes are repeated in \mathcal{T}_K , they are all considered as being different in the pseudo-tree. Let \mathcal{T}_t designate a tree where there is a unique path from any node i to t (tree rooted at t) and $\pi_i(\mathcal{T}_t)$ denote the cost of the path p_{it} in \mathcal{T}_t ; the reduced cost $\bar{c}_{i,j}$ of arc $(i,j) \in A$ associated with \mathcal{T}_t is $\bar{c}_{i,j} = \pi_j(\mathcal{T}_t) - \pi_i(\mathcal{T}_t) + c_{ij}$. So all arcs in \mathcal{T}_t have a null reduced cost. The advantage of using reduced costs was first noted by Eppstein [3] and is shown, by theorems 8 and 9 in [5] and by theorem 2.1 in [6], in the context of the MPS algorithm, to lead to less arithmetic operations and path generation simplification.

Next additional notation and definitions used in the proposed algorithm will be introduced. Let $|p|$ denote the number of arcs in path p, from s to t, or path length. The notation ${}^D p$ will designate a path such that ${}^D p = p$ if $|p| \leq D$; otherwise ${}^D p$ is the sub-path with origin s and D arcs. If $D = 0$ then ${}^D p$ is the empty set ϕ . Let p be a path which contains nodes v_i and v_j . The distance from v_i to v_j in p will be given by the number of arcs in p, from v_i to v_j ; if v_i, v_j , are extreme nodes of an arc then the distance is 1; the distance of a node to itself is zero. Let p be a path from s to t, which contains v_i ; the depth of v_i in p is given by the distance from s to v_i in p and will be designated by $d_p(v_i)$. Let \mathcal{T}_t^* be the tree of the shortest paths from all nodes to t. The height of node v_i in \mathcal{T}_t^* is given by the distance from v_i to t; $p_{v_i t}^*$ represents the shortest path from v_i to t in \mathcal{T}_t^* . Let the set of arcs A of (N,A) be written in terms of $A(v)$, the set of arcs the tail node of which is $v \in N = \{1, \dots, n\}$, i.e. $A = A(1) \cup A(2) \cup \dots \cup A(n)$ such that $A(i) \cap A(j) = \phi$ for any $i \neq j$ ($i, j \in N$). Let $a'_k \in A(\xi)$ and $a'_l \in A(\theta)$. Assuming that $\xi \neq \theta$ an order relation ' $<$ ' is defined for the arcs such that $a'_k < a'_l$ iff $\xi < \theta$. Moreover if $\xi = \theta$ then $a'_k < a'_l$ if $\bar{c}(a'_k) \leq \bar{c}(a'_l)$. This means that the set A is sorted in such a way that for any two arcs $(k,j), (i,l) \in A$, $(k,j) < (i,l)$ if $k < i$ or $(k = i$ and $\bar{c}_{kj} \leq \bar{c}_{il})$. The resulting set $A = \{a_1, \dots, a_m\}$ is said to be in the sorted forward star form, which leads to the path generation simplification.

3. The algorithm

Let one consider that only shortest paths with length less than or equal to D, be acceptable. Then the K-shortest path constrained problem which we are seeking to solve for two given nodes s,t is equivalent to determine a set $P_D^K = \{p_1, p_2, \dots, p_K\} \subset P \equiv P_{st}$ where P_{st} is the set of paths which can be defined from s to t in (N,A) such that:

1. $|p_k| \leq D, k \in \{1, \dots, K\}$
2. $c(p_k) \leq c(p_{k+1}), k \in \{1, \dots, K\}$
3. $c(p_k) \leq c(p)$ for any $p : |p| \leq D$ and $p \in P - P_D^K$

The proposed algorithm has the additional (desirable) property that p_k is identified just before p_{k+1} for any $k \in \{1, \dots, K-1\}$.

Let \mathcal{T}_k designate the pseudo-tree of the k -shortest loopless paths from s to t , with a maximum of D arcs and X the set of the candidate paths to the following shortest path in the context of the algorithm MPS in [5]. The justifications underlying the proposed variant to the algorithm MPS [5] in the version in [6] are now presented.

Let p_u be the shortest path in X , and v its deviation node. From p_u , paths with deviation node at depth at most $D-1$, will be obtained from p_u only if their deviation node $v_i, v_i \neq t$ and $v_i \in p_{vt}^u$ are such that $|p_{sv_i}^u| < D-1$ or, $|p_{sv_i}^u| = D-1$ and $p_{sv_i}^u \diamond \langle v_i, (v_i, t), t \rangle \neq p_u$, where $p_{sv_i}^u$ represents the sub-path of p_u from s to v_i . Therefore paths of length greater than D will be placed in X if their deviation node is at a depth less than D . The efficiency of the proposed algorithm stems precisely from not generating paths of length greater than D which are of no interest, because, when selected, they would be discarded and any path obtained from them would also certainly be discarded too.

The MPS algorithm begins by obtaining the shortest path tree from all nodes to t . In telecommunication networks, most measures of interest are non negative, therefore the Dijkstra algorithm can be used, for it will be considered that $c(a) \geq 0, a \in A$; in fact, any suitable labelling algorithm could be used for this purpose [4]. The inverse Dijkstra algorithm, used as a starting point of the algorithm, will be adapted to associate each node with its height in the shortest paths tree from all nodes to t . This is easily obtained by giving to t the height 0 and establishing that whenever the successor of a node v_i is changed, being for example v_j , then the height associated with v_i is also changed and becomes the height of v_j plus 1.

A final note on the used notation. Given a path p (from s to t) and a node v in p , p_{sv} represents the sub-path of p from s to v and p_{vt} represents the sub-path of p from v to t . Also let p_{vt}^* represent the shortest path from v to t in \mathcal{T}_t^* .

Next the algorithm for generating the K -shortest paths of length less or equal to D , is formalised.

Algorithm KD

1. Input: the representation of the graph (N, A) and arc costs c_{ij} .
2. By using Dijkstra inverse algorithm (including the node heights calculations), obtain the shortest paths from every node to t , hence constructing the shortest tree rooted at t , \mathcal{T}_t^* .
3. Calculate the reduced cost \bar{c}_{ij} for every $(i, j) \in A$.

4. Rearrange the arcs of (N, A) in the sorted forward star form (for the computed \bar{c}_{ij} , leading to $A = \{a_1, \dots, a_m\}$ such that for any $h \in \{1, \dots, m-1\}$, $\bar{c}_{a_h} \leq \bar{c}_{a_{h+1}}$ if $v = \text{tail}(a_h) = \text{tail}(a_{h+1})$).
5. $p \leftarrow$ shortest path from s to t ($p \in \mathcal{T}_t^*$).
6. $k \leftarrow 0$.
7. $X \leftarrow \{p\}$ (X is the set of paths that are candidates to shortest path).
8. While $(k < K)$ and $(X \neq \emptyset)$ do
 - (a) $p \leftarrow$ shortest path in X
 - (b) $X \leftarrow X - \{p\}$
 - (c) If $|p| \leq D$ and p has no loops then
 - i. $k \leftarrow k + 1$
 - ii. $p_k \leftarrow p$ (the k -th path which has at most D arcs, was found)
 EndIf
 - (d) Let v be the deviation node of p
 - (e) If $(|p| = D)$ and $(|p_{sv}| < D - 1)$
 then $w \leftarrow D - 2$ (try to generate new paths)
 else If $(|p| \neq D)$ and $(|p_{sv}| < D)$
 then $w \leftarrow \min(|p| - 1, D - 1)$ (try to generate new paths)
 else $w \leftarrow -1$ (no path will be generated from p)
 EndIf EndIf
 - (f) If $w \neq -1$ then (possibly a path will be placed in X)
 - i. $v_i \leftarrow v$
 - ii. Repeat
 - A. $a_h \leftarrow$ the arc of p the tail of which is v_i
 - B. $p_{sv_i} \leftarrow$ sub-path of p from s to v_i
 - C. While $(v_i$ is the tail of $a_{h+1})$ and $(a_{h+1}$ forms a loop with $p_{sv_i})$ do
 $h \leftarrow h + 1$
 EndWhile
 - D. If v_i is the tail a_{h+1} then
 - $v_j \leftarrow$ head of a_{h+1}
 - $X \leftarrow X \cup \{p_{sv_i} \diamond \langle v_i, a_{h+1}, v_j \rangle \diamond p_{v_j t}^*\}$
 EndIf
 - E. If $d_p(v_i) = w$
 then $w \leftarrow -1$ (no more paths will be obtained from p)
 else $v_i \leftarrow$ following node in ${}^w p$ (truncated path p after w arcs)
 EndIf
 Until $(p_{sv_i}$ has a loop) or $(w = -1)$
 EndIf $(w \neq -1)$

EndWhile

4. Analysis of Results

Having in mind to compare the algorithm KD proposed in section 3, (which will be designated hereafter by "KD") with results from the algorithm MPS [6] (hereafter designated by "MPS") for the purpose of generating the K-shortest loopless paths with a maximum of D arcs, numerous computational experiments were carried out. Strongly meshed networks and fully meshed networks (corresponding to strongly or fully connected graphs) were considered², with 20, 50, 200 and 1000 nodes. The comparison measure was the total number of paths generated by each algorithm until the K-shortest path with at most D arcs was obtained. To facilitate the comparison, values were obtained by averaging the number of generated paths corresponding to solve the problem for pairs of origin-destination nodes in a given network and at least in two networks obtained from different random number sequences, which determine the values of the arc cost (in a given interval) and also the network arcs in the case of strongly meshed networks.

A major factor which influences the relative performance of MPS is the range of variation of arcs costs. For this reason results were presented for two ranges of variation: a moderate variation, from 1 to 10 in 20 node networks up to 1 to 40 in 1000 node networks and a second range of larger relative variations of 1 to 50 in 20 node networks and 1 to 150 for the remaining networks. In MPS the higher is the cost variation, the higher is the average number of arcs of the shortest paths, hence the poorer is the relative performance of MPS, as illustrated in figures 1, 2 and 3.

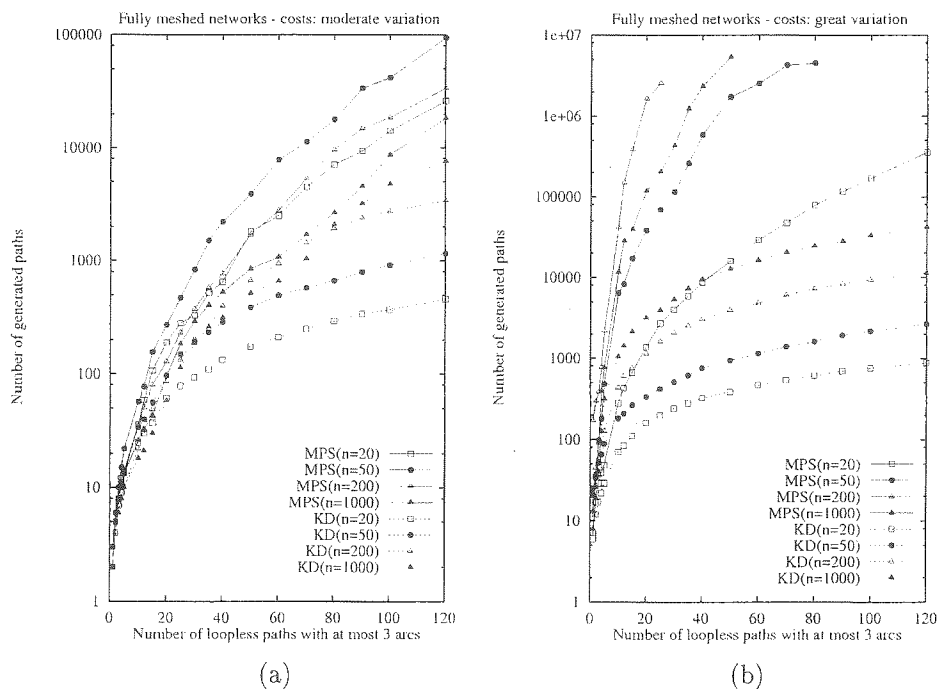


Figure 1 - Comparison of the algorithms for $D = 3$; (a) $n = 20, 50, c_{ij} \in [1, 10]$; $n = 200, c_{ij} \in [1, 20]$ and $n = 1000, c_{ij} \in [1, 40]$ (b) $n = 20, c_{ij} \in [1, 50]$ and $n = 50, 200, 1000, c_{ij} \in [1, 150]$

² The used program for network generation was kindly borrowed by José Luís Santos.

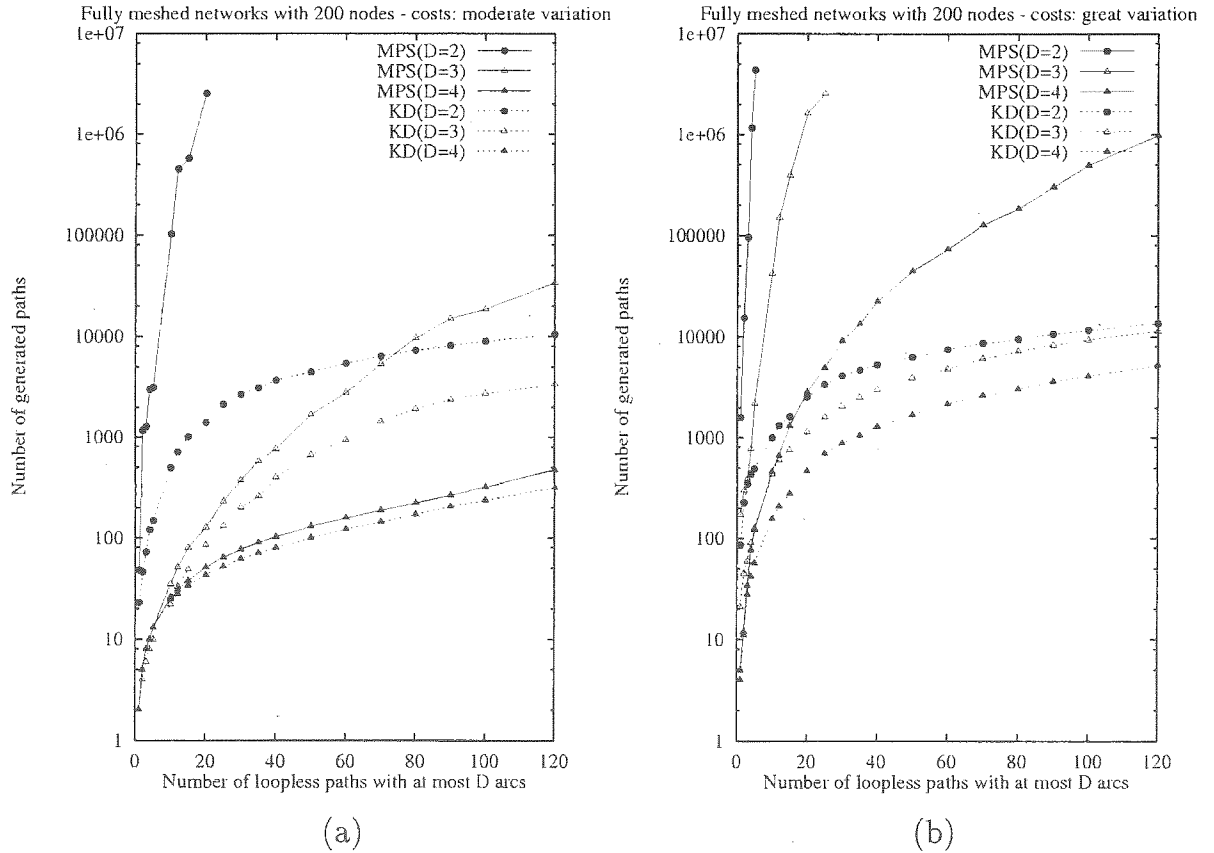


Figure 2 - Comparison of the algorithms for fully meshed networks with 200 nodes and $D = 2, 3, 4$; (a) arc costs $\in [1, 20]$ (b) arc costs $\in [1, 150]$

Some of the most significant results are condensed in figures 1, 2 and 3. The presented curves represent averages over the total number of paths generated for obtaining up to 120 shortest paths (whenever they exist) between two pairs of nodes with a maximum of $D = 2, 3$ or 4 arcs. Note that these are the values of D (specially $D = 2, 3$) which are of interest to applications in route generation algorithms in multiexchange networks. It should be noted that for $D = 2$ the use of any of the algorithms is not necessarily advantageous. In fact in many cases it would be more efficient simply to generate all the paths with two arcs and order them by decreasing cost. The considered number of generated paths in the number of paths inserted in the set X of candidate paths in the course of algorithm execution. To prevent excessive times in some cases (due to swapping) any of the algorithms is stopped when the 10⁷-th is generated even if the last shortest path with up to D arcs hasn't been obtained. This limiting situation occurred some times with the MPS algorithm, but never with the KD algorithm. This has to do with the fact that KD uses much less memory than MPS because the pseudo-tree of the K -shortest paths generated by KD is much smaller. This is obviously a significant advantage of the KD algorithm.

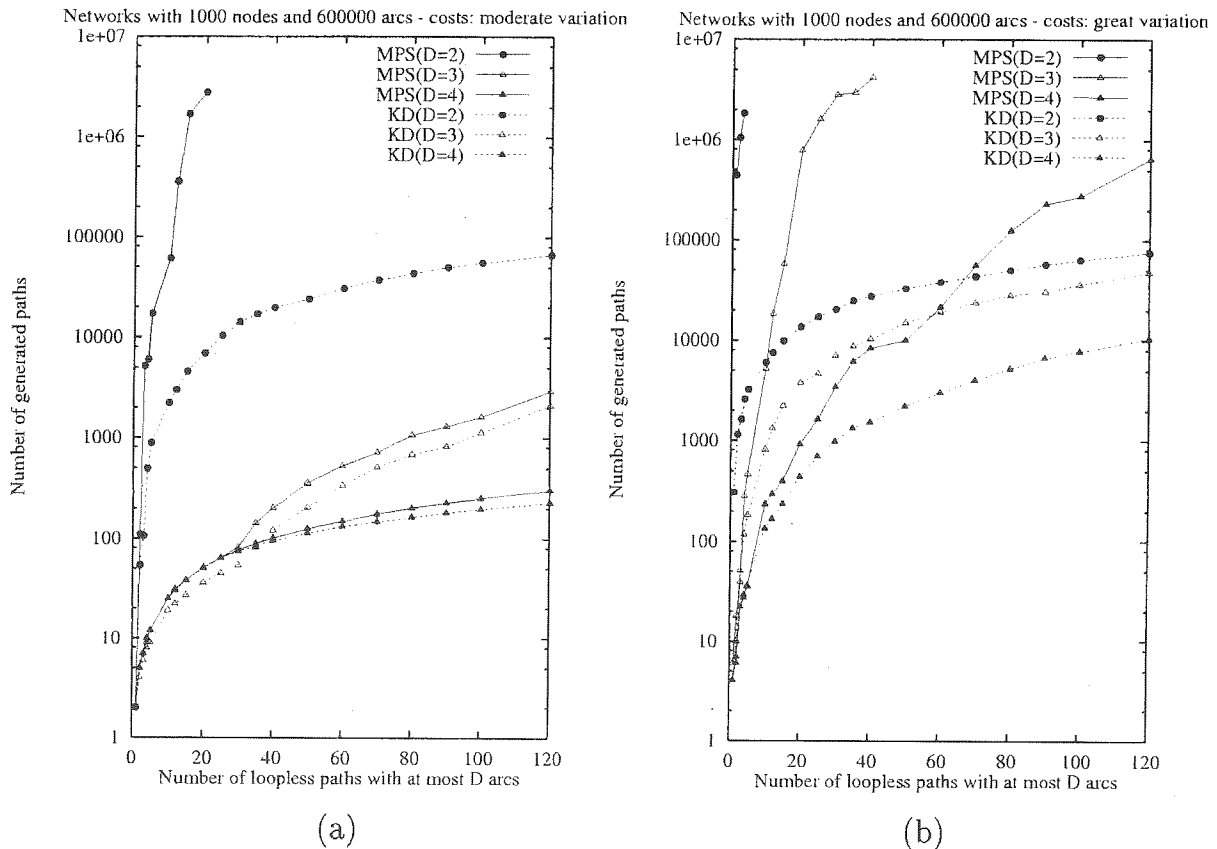


Figure 3 - Comparison of the algorithms for non-fully meshed networks with 1000 nodes and 600000 arcs, $D = 2, 3, 4$; (a) arc costs $\in [1, 20]$ (b) arc costs $\in [1, 150]$

It is apparent from the results the quite significant advantage of KD with respect to MPS in most cases. This highly improved efficiency of KD is expectable having in mind that KD does not generate unnecessary paths (that is with more than D arcs - when such paths are of no potential interest) while it makes the most of the inherent efficiency of MPS. Other advantage of KD stems from the fact that it manages to identify the phase in which there are no more paths with the desired number of arcs. This is particularly important in non fully meshed networks where one seeks to obtain paths with 2 or 3 arcs, a common situation in telecommunication networks.

In general the total number of paths generated by MPS is closer to the equivalent number of KD, when the maximum number of arcs is higher and the relative variation of the arc costs is smaller. This stems from the fact that, in these cases, almost all the first shortest paths generated by MPS have a number of arcs less than or equal to D .

It was also verified that the relative performance of the algorithms did not depend so much on the network dimensions that on the cost variations. This may be appreciated from figures 1(a) and 1(b) where the results are presented for several fully meshed networks and for two ranges of cost variations, with $D = 3$.

5. Applications and Conclusions

In some applications of the K shortest path problem an important constraint is the maximum number of arcs in any chosen path. In particular, for multiexchange telecommunication networks, the determination of optimal or suboptimal (according to some criteria) paths which may be used for routing calls between pairs of nodes (associated with exchanges in terms of network representation) imposes, for technical reasons, that only paths with up to a certain number of arcs, may be considered. This has to do with the necessity of simplifying control and signalling functions and prevent undesirable situations from the point of view of the traffic flow distributions. Note also that multiexchange telecommunication networks are usually either fully meshed or strongly meshed and technology enables the utilisation of sophisticated routing mechanisms. Such mechanisms usually require frequent recalculation (for different network conditions e.g. related to traffic loads or channel occupations) of shortest paths for each pair of origin-destination nodes. An example of this type of application environment for this constrained K -shortest path problem is a multi-objective routing model proposed in [1]. The basis of this model is to consider two objective functions (to be minimised) in the routing model such as cost and blocking probability or cost and delay. This model involves the selection of non-dominated solutions (paths) of a bi-objective problem taking into account some preference criteria defined on the objective function space. The algorithm used for tackling this problem uses repeatedly the K -shortest path algorithm [6] and the authors are developing a new variant of the model enabling the choice of alternative paths when for example the first selected path is blocked (ie. at least one arc is unavailable), a technique that is known as alternative routing, also considering as explicit constraint the maximum number D of arcs per path ($D = 2, 3, 4$).

An algorithm was therefore proposed in this paper for solving the K -shortest path problem with a constraint on the number of arcs per path, which is based on the MPS algorithm for loopless paths in [6] (which is a variant of the algorithm in [5]). It was shown that the proposed algorithm (algorithm KD) is much more efficient in most cases than the simplistic approach of ignoring the paths selected by MPS with more than D arcs. This was confirmed by extensive computational results some of which were discussed in the paper. This highly improved efficiency of KD is expectable having in mind that KD does not generate unnecessary (that is with more than D arcs - when such paths are of no potential interest) while it makes the most of the inherent efficiency of MPS. Other significant advantage of KD is the fact that it uses much less memory than the MPS algorithm, because the pseudo-tree of the shortest paths generated by KD is much smaller. Also KD , unlike MPS, in non-fully meshed networks manages to identify the phase in which there are no more paths with the desired number of arcs. This is particularly important in non fully meshed networks where one seeks to obtain paths with 2 or 3 arcs, a common situation in telecommunication networks.

The experimental results also suggested that in general the total number of paths generated by the two algorithms is closer when the maximum number of arcs is higher and the relative variation of the arc is smaller. This stems from the fact that, in these cases, almost all the first shortest paths generated by MPS have a number of arcs less than or equal to D .

Acknowledgements

The authors are very grateful to Professor Ernesto Martins, Marta Pascoal and José Luís Santos for their explanations concerning the MPS algorithm and its implementation.

References

- [1] Antunes, C.Hengeller, Craveirinha, J., Clímaco, J. and Barrico, C., A multiple objective routing algorithm for integrated communication networks, In *Teletraffic Engineering in a Competitive World - ITC 16*, Elsevier (1999) 1291-1300.
- [2] Clímaco, J.C.N. and Martins, E.Q.V., A bicriterion shortest path algorithm, *European Journal of Operational Research* 11 (1982) 399-404.
- [3] Eppstein, D., Finding the k shortest paths, *SIAM Journal on Computing* 28 (1998) 652-673.
- [4] Martins, E., Pascoal, M., Rasteiro, M. and Santos, J., The optimal path problem, *Investigação Operacional* 19 (1999) 43-60.
- [5] Martins, E., Pascoal, M. and Santos, J., Deviation algorithms for ranking shortest paths, *International Journal of Foundations of Computer Science* 10 (1999) 247-263.
<http://www.mat.uc.pt/~marta/research.html>.
- [6] Martins, E., Pascoal, M. and Santos, J., An algorithm for ranking loopless paths, Research Report 99/007, CISUC (1999). <http://www.mat.uc.pt/~marta/Publicacoes/mps2.ps.gz>.

INSTRUÇÕES AOS AUTORES

Os autores que desejam submeter um artigo à Investigação Operacional devem enviar três cópias desse trabalho para:

Prof. Joaquim J. Júdice
Departamento de Matemática
Faculdade de Ciências e Tecnologia
Universidade de Coimbra
Apartado 3008
3001-454 Coimbra, Portugal

Os artigos devem ser escritos em Português ou Inglês. A primeira página deve conter a seguinte informação:

- Título do artigo
- Autor(es) e instituição(ões) a que pertence(em)
- Abstract (em inglês)
- Resumo
- Keywords (em inglês)
- Título abreviado

As figuras devem aparecer em separado de modo a poderem ser reduzidas e fotocopiadas. As referências devem ser numeradas consecutivamente e aparecer por ordem alfabética de acordo com os seguintes formatos:

Artigos: autor(es), título, título e número da revista (livro com indicação dos editores), ano, páginas.

Livros: autor(es), título, editorial, local de edição, ano.

ÍNDICE

J. Antunes Ferreira, L. Valadares Tavares, J. Silva Coelho, A general generator of project networks in terms of their morphological features	107
M. E. Cunha, J. A. Assis Lopes, A. P. B. Póvoa, Modelo de Simulação para a gestão de um sistema multi-escalão de artigos reparáveis	123
M. M. L. Marquês, J. O. Soares, M. C. V. Coutinho, Desigualdades regionais em Portugal Continental: Uma análise estatística multivariada	139
T. S. Fernandes, M. C. Mourão, A. Pimenta, A eficiência técnica das estações de correio portuguesas: Uma aplicação da metodologia DEA — Data Envelopment Analysis	153
J. C. Mello, E. G. Gomes, M. E. Lins, L. M. Vieira, Um caso de estudo de integração SIG-DEA-MCDA: A influência de uma instituição de ensino superior em vários municípios do estado do Rio de Janeiro	171
V. G. Sousa, A. M. Faustino, J. F. Lemos, Sensibilidade à ponderação de critérios num problema MCDM de planeamento de recursos hídricos	191
M. L. R. Varela, R. A. Ribeiro, Utilização de simulated annealing em optimização difusa	205
Nota do Editor-Principal	233
T. Gomes, L. Martins, J. Craveirinha, An algorithm for calculating K shortest paths with a maximum number of arcs	235

