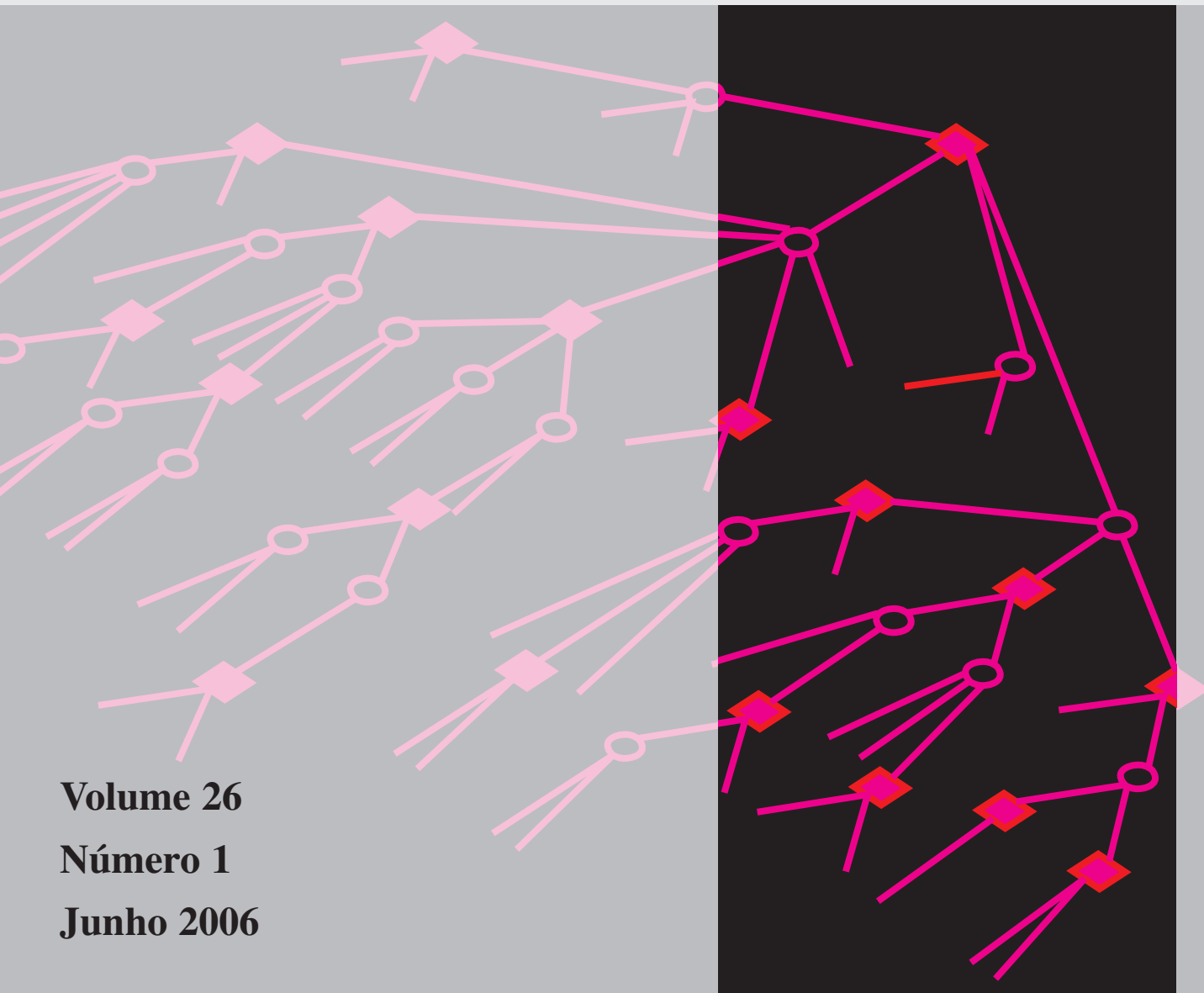


# *Investigação Operacional*



**Volume 26**  
**Número 1**  
**Junho 2006**

## **INVESTIGAÇÃO OPERACIONAL**

Propriedade:

APDIO

Associação Portuguesa  
de Investigação Operacional

Patrocinadores

**Fundação Calouste Gulbenkian**

The logo for FCT (Fundação para a Ciência e a Tecnologia) consists of the letters 'FCT' in a bold, green, sans-serif font.

**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

420 Ex.

2006/06

# INVESTIGAÇÃO OPERACIONAL

Volume 26 — nº 1 — Junho 2006

Publicação Semestral

Editor Principal: José F. Oliveira  
Universidade do Porto

## Comissão Editorial

M. Teresa Almeida  
Inst. Sup. Econ. e Gestão

J. Rodrigues Dias  
Univ. de Évora

N. Maculan  
Univ. Fed., Rio Janeiro

C. Henggeler Antunes  
Univ. de Coimbra

Laureano Escudero  
IBM, Espanha

Rui Oliveira  
Inst. Superior Técnico

Marcos Arenales  
Univ. de São Paulo

Edite Fernandes  
Univ. do Minho

J. Pinto Paixão  
Univ. de Lisboa

Jaime Barceló  
Univ. de Barcelona

J. Soeiro Ferreira  
Univ. do Porto

M. Vaz Pato  
Inst. Sup. Econ. e Gestão

Eberhard E. Bischoff  
Univ. of Wales, Swansea

J. Fernando Gonçalves  
Univ. do Porto

Mauricio G. Resende  
AT&T Labs Research

C. Bana e Costa  
Inst. Superior Técnico

Luís Gouveia  
Univ. de Lisboa

A. Guimarães Rodrigues  
Univ. do Minho

M. Eugénia Captivo  
Univ. de Lisboa

Rui C. Guimarães  
Univ. do Porto

António J. L. Rodrigues  
Univ. de Lisboa

Domingos M. Cardoso  
Univ. de Aveiro

Joaquim J. Júdice  
Univ. de Coimbra

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 J. Luz  
Inst. Polit. Setúbal

L. Valadares Tavares  
Inst. Superior Técnico

João P. Costa  
Univ. de Coimbra

Virgílio P. Machado  
Univ. Nova de Lisboa

B. Calafate Vasconcelos  
Univ. do Porto

Ruy Costa  
Univ. Nova de Lisboa

Manuel Matos  
Univ. do Porto

L. Nunes Vicente  
Univ. de Coimbra

Victor V. Vidal  
Techn. Univ. of Denmark



# Reflexões sobre os *rankings* do Secundário

Manuel Matos<sup>†</sup>  
Carla Teixeira Lopes<sup>‡</sup>  
Sérgio Nunes<sup>§</sup>  
Isabel Venâncio<sup>¥</sup>

<sup>†</sup> FEUP & INESC Porto  
mam@fe.up.pt

<sup>‡</sup> ESTSP – IPP  
carla.lopes@acm.org

<sup>§</sup> FEUP  
sergio.nunes@fe.up.pt

<sup>¥</sup> EB2/3 Soares dos Reis  
isabel.venancio@gmail.com

---

## Abstract

The publication of rankings about high schools, started in 2001, had a significant impact at several levels, initiating a series of comments and speculations in main stream media and influencing the decisions (or at least the aspirations) of many people. Taking advantage of the availability of five years of data, and the possibility of comparing the rankings, our data analysis has tried to verify a posteriori if the implicit hypothesis in the enthusiastic adoption of rankings is reasonable. Using other words, are the rankings an accurate measure of the quality, or of the effectiveness, of schools? The main conclusion points to the necessity of not overrating the rankings based on examinations, nevertheless considering them as a useful piece of information, for school's self-assessment and for comparing comparable things. On the other side, being evident the social interest for these kind of orderings, it is worth investing on the identification of quality characterization factors that could be used to complement the information, always preferring understandable, objective and accessible aspects.

## Resumo

A publicação de *rankings* das escolas secundárias, iniciada em 2001, teve um impacto apreciável a vários níveis, suscitando uma série de comentários e especulações nos diversos meios de comunicação social e influindo realmente nas decisões (ou pelo menos aspirações) de muitas pessoas. Beneficiando da disponibilidade de cinco anos de dados, e da possibilidade de comparar os *rankings* respectivos, o nosso exercício de análise de dados procurou verificar a posteriori até que ponto a hipótese implícita na utilização entusiástica dos *rankings* é razoável. Ou seja, serão os *rankings* um indicador fiável da qualidade, ou da eficácia, das escolas? A conclusão geral do exercício aponta para a necessidade de não sobrevalorizar os *rankings* baseados em exames, sem deixar de

os considerar uma peça de informação útil, para as escolas se situarem e para se compararem coisas comparáveis. Por outro lado, sendo evidente o interesse social por este tipo de ordenações, valerá a pena investir na identificação dos factores de caracterização da qualidade que poderão ser utilizados para complementar a informação dos exames, com o cuidado de privilegiar aspectos compreensíveis, objectivos e acessíveis.

**Palavras Chave:** *Rankings*, Ensino secundário, Análise de dados, Decisão

**Título:** Considerations about the High School Rankings

---

## 1 Introdução

A publicação de *rankings* das escolas secundárias, iniciada em 2001, teve um impacto apreciável a vários níveis, suscitando uma série de comentários e especulações nos diversos meios de comunicação social e influenciando realmente nas decisões (ou pelo menos aspirações) de muitas pessoas.

Para além da óbvia atracção que qualquer *ranking* exerce junto das entidades que nele ficam mais bem classificadas, constata-se que poucos ficam indiferentes à apresentação de uma ordenação clara, baseada em factos (as classificações dos exames) e aparentemente despida de subjectividades e ambiguidades. Ainda por cima, a história do processo inclui recusas de disponibilização dos dados brutos (classificações dos exames) por parte de sucessivos Ministérios da Educação, o que terá aumentado a adesão afectiva ao primeiro exercício concreto de utilização desses dados, e depois aos que se lhe seguiram.

A esse propósito, e antes de prosseguir, querem os autores deste artigo deixar claro que, na sua opinião, dados deste tipo devem estar disponíveis para o público, cabendo aos interessados dar-lhes boa utilização e denunciar conclusões erradas ou abusivas. O problema, aliás, situar-se-á mais na escassez de dados complementares aos das classificações dos exames, impedindo estudos mais completos do que a mera organização de *rankings* de resultados de exames.

Voltando ao que nos ocupa, verifica-se que os *rankings*, que tecnicamente são apenas ordenações das escolas pela média das classificações nos exames do 12º ano dos alunos que levam a exame, acabam por ser na prática considerados indicadores da qualidade ou da eficácia das escolas, com todas as consequências daí decorrentes, em que os *rankings* funcionam como auxiliares de decisão para alunos, pais, escolas e ministério. As limitações deste tipo de exercício para caracterizar e ajudar a melhorar entidades e processos tão complexos ficam patentes quando se tomam em consideração trabalhos académicos e estudos realizados ao longo de várias décadas sobre eficácia e qualidade nas escolas (Venâncio et al., 2003), mas os *rankings* existem e não podem ser ignorados.

O nosso propósito aqui é bastante diferente daquilo que temos visto nas discussões relacionadas com este tópico. Beneficiando da disponibilidade de cinco anos de dados, e da possibilidade de comparar os *rankings* respectivos, o nosso exercício de análise de dados procurou verificar *a posteriori* até que ponto a hipótese implícita na utilização entusiástica dos *rankings* é razoável. Ou seja, serão os *rankings* um indicador fiável da qualidade, ou da eficácia, das escolas?

A análise desta questão trouxe outras, também abordadas neste artigo, a começar na identificação de uma inconsistência sistemática, inerente à utilização de indicadores de

média pesada (mas agravada pelas grandes variações das médias das diferentes disciplinas) e terminando na comparação público-privado que tem sido um dos cavalos de batalha da corrente de opinião que se apoia nos *rankings* para exigir o cheque-ensino. Pelo meio, analisa-se ainda a importância e evolução da diferença entre classificações internas e do exame e discute-se a eventual influência do número de provas nos resultados obtidos por uma escola.

Em todo o artigo procurou-se sobretudo analisar os dados e identificar situações claramente decorrentes dessa análise, deixando para outros especulações ou explicações mais profundas, que estão fora das competências principais dos autores, mas são certamente desejáveis.

Convém neste ponto explicar que se utilizou, como base de discussão, os *rankings* Público/Universidade Católica, sem preocupação de os comparar com aqueles publicados em outros jornais, apesar de tudo não tão diferentes, quando a perspectiva é aquela em que nos encontramos.

A estrutura do artigo a partir deste ponto é a seguinte: começamos por um pequeno enquadramento científico, histórico e metodológico (Secção 2), abordando em seguida a questão da inconsistência e da influência nesta das médias das disciplinas (Secção 3), antes de iniciar a discussão do valor dos *rankings* para auxílio à decisão (Secção 4). A análise dos restantes aspectos relevantes decorre na Secção 5, a que se seguem as conclusões e a lista de referências.

## 2 Enquadramento

### 2.1 Eficácia e qualidade

A perspectiva que se poderia dizer “industrialista” da eficácia traduz-se na observação das relações entre recursos e resultados. Nesta perspectiva, uma escola eficaz é aquela que optimiza os recursos disponíveis, em termos dos resultados escolares (académicos e não académicos) alcançados. Em consequência, mesmo que se aceite que o sucesso é medido unicamente pelos resultados nos exames, qualquer análise (ou *ranking*) que não tome em conta os recursos utilizados, incluindo a situação inicial dos alunos, será sempre deficiente.

Numa outra perspectiva, o conceito de eficácia das escolas associa-se à capacidade de diminuir as desigualdades entre alunos e de oferecer mais oportunidades de aprendizagem de forma diferenciada e conforme as necessidades de cada um, quer se trate de alunos com dificuldades, quer se trate de alunos excepcionais. A escola eficaz promove o êxito educacional dos alunos e reduz as diferenças existentes entre eles, procurando conduzir o maior número ao mais alto nível das suas capacidades. Aqui, o próprio conceito de resultado é muito mais complexo do que as classificações nos exames.

Pammela Sammons (cf. Clímaco et al., 1992), define “eficácia” como “o impacto da educação sobre a sociedade ou da acção educativa sobre uma geração ou grupo de alunos”. Os “efeitos da escola” referem-se aos factores da escola que podem ser apontados como correlacionados ou associados ao sucesso dos alunos.

O conceito de “qualidade da escola” aparece frequentemente referido na literatura educacional, em textos oficiais ou no debate político, não sendo, contudo, descrito de forma explícita - raramente se apresentam os critérios de observação que fundamentam os juízos de valor. Acredita-se que o nível de qualidade reflecte a dimensão político-social

da democracia e tem subjacente a igualdade de oportunidades, a necessidade de modernizar os equipamentos educativos, a distribuição de recursos, a orientação escolar, os programas de formação dos agentes educativos, a avaliação dos alunos, a orientação pedagógica e as relações da educação com os modelos de desenvolvimento económico e o mercado de emprego.

As escolas eficazes e as escolas de qualidade estão muito ligadas ao conceito de desenvolvimento dos alunos, que se mede pelos resultados cognitivos, académicos e não académicos, como as expectativas positivas, as atitudes face à escola e à aprendizagem, a sociabilidade e a capacidade de trabalhar em grupo, o espírito de iniciativa, a capacidade de tomar decisões e aquisição de valores relacionados com o espírito de cidadania, de liberdade e de respeito pela diferença.

A hipervalorização dos *rankings* conduz, inevitavelmente, à simplificação do conceito de sucesso (= resultados dos exames). Implícito, subjaz o “efeito escola” como único factor de sucesso. Mas nada é tão simples. Curiosamente, completam-se, em 2006, 40 anos do célebre relatório Coleman (Coleman et al, 1966), segundo o qual o grau de sucesso dos alunos resultaria fundamentalmente da sua origem social. Analisando de que forma as variáveis sociais, culturais e familiares interferiam no sucesso dos alunos, verificou-se que as variáveis directamente ligadas aos alunos (estatuto sócio-económico, raça, etc) explicavam de 65% a 85% do seu sucesso escolar. Estas conclusões minimizavam qualquer estratégia de intervenção ao nível da escola, apontando para reformas que pretendiam compensar os “handicaps sócio-culturais”.

No entanto, havia indícios de que a estrutura da escola, a forma de liderança e o estilo das práticas docentes poderiam afectar o rendimento dos alunos. Analisando as características organizacionais e contextuais dessas escolas de bom desempenho, esperava-se identificar as condições favorecedoras da eficácia. O estudo mais conhecido, intitulado 15.000 Horas de Escola (Rutter et al., 1979) conduziu a resultados que foram, ao mesmo tempo, confirmados por um grande número de autores e violentamente questionados por outros. De facto, nesses estudos, o efeito da escola é estimado através de análises multivariadas complexas, que, quase sempre, resultam em correlações fracas e instáveis e a um grau de sofisticação metodológica que aumenta o risco de artificialidade.

E, desde há 40 anos, investigações teórico-práticas associadas à eficácia e melhoria da eficácia da escola tentam encontrar resposta para algumas questões-chave para a qualidade da escola:

- Que factores fazem com que algumas escolas alcancem os seus objectivos melhor do que outras?
- Que fazer para que uma escola melhore?
- Que “efeito escola” terá relação directa com o sucesso do processo ensino-aprendizagem?
- Pode a escola potenciar o “efeito escola” no sentido da melhoria real da eficácia, tendo em atenção que cada aluno é reflexo de inúmeras interacções, ao longo do seu percurso escolar?

Naturalmente, tudo isto está muito longe da mera ordenação das escolas por médias globais de resultados em exames, exercício que não se destina a procurar resposta a nenhuma destas perguntas.



## 2.2 História

O acesso público aos resultados dos exames nacionais do 12º ano tornou-se possível em Agosto de 2001, com a publicação dos dados em bruto pelo Ministério da Educação (ME). Esta decisão surge após a tomada de posse do ministro Júlio Pedrosa e a deliberação da Comissão de Acesso aos Documentos Administrativos (CADA). Este parecer resultou de uma queixa apresentada pelo jornal Público, tendo a CADA deliberado em Julho de 2001 que “o acesso aos documentos administrativos de carácter não nominativo constitui um direito fundamental de todos os cidadãos” (CADA, 2001).

Em 2001, vários órgãos de comunicação social publicaram análises realizadas com base nestes dados. O jornal Público, em colaboração com a Universidade Católica Portuguesa (UCP), publicou um suplemento com a manchete “A LISTA” (este título viria a ser alvo de críticas por parte do provedor do jornal (Fidalgo, 2001)). Desde esta data que o jornal Público, em colaboração com a UCP, publica anualmente um *ranking* que tem por base os mesmos critérios que aquele publicado em 2001.

Em 2002, com um novo ministro à frente da pasta, o ME produz e publica, em colaboração com a Universidade Nova de Lisboa, um *ranking* que incorpora indicadores sócio-económicos. Esta análise suscita várias críticas e leva à publicação de um livro de apreciação crítica sobre o tema em 2004 (Correia e tal., 2004). O próprio ministro David Justino reconheceu, na época, que o *ranking* produzido pelo ME teve aspectos negativos que lhe retiraram a credibilidade (Chitas et al., 2002). Nos anos seguintes, o ME abandona a publicação desta análise, mantendo apenas a publicação dos dados em bruto.

## 2.3 Metodologia

As bases de dados relativas aos anos entre 2001 e 2005, e disponíveis nas páginas web da Direcção-Geral de Inovação e Desenvolvimento Curricular do ME (DGIDC, 2006), foram descarregadas durante o mês de Agosto de 2006. As bases de dados são fornecidas em ficheiros individuais por ano, no formato específico do programa Microsoft Access. A análise dos dados foi realizada usando o Microsoft Access para a extracção dos dados e o programa SPSS para a análise estatística. Para algumas tarefas específicas, nomeadamente para identificar exemplos que satisfaçam determinados critérios, foi usada uma combinação de PHP e MySQL em ambiente UNIX.

Numa primeira fase, procurámos reproduzir os resultados obtidos pelo Público/UCP e, usando os critérios divulgados no jornal, produzimos os *rankings* para cada um dos 5 anos. Assim, seleccionámos apenas o conjunto dos exames realizados na primeira fase, as 7 disciplinas com mais provas (Biologia, Física, História, Português B, Psicologia, Química e Matemática) mais Português A (por ser obrigatória no agrupamento de Humanidades) e as escolas onde foi realizado pelo menos 1 exame numa destas 8 disciplinas. Com esta selecção, a seriação das escolas foi feita com base na média dos exames realizados no conjunto das 8 disciplinas.

A primeira constatação tem a ver com o facto dos *rankings* reconstruídos não apresentarem sempre ordenações iguais às aquelas apresentadas pelo jornal. Uma análise mais detalhada permitiu verificar que o *ranking* apresentado pelo jornal não usa sempre o mesmo critério, havendo pequenas mudanças que têm impacto na ordenação final. Por exemplo, enquanto que em 2001 o *ranking* do Público/UCP incluiu todas as provas realizadas por alunos internos, em 2003 foram incluídas apenas as provas dos alunos internos realizadas para aprovação, sendo excluídas aquelas realizadas para melhoria. Para além destas diferenças ao nível dos critérios usados, detectámos também diferenças

ao nível dos dados brutos. Por exemplo, em várias escolas, o número de provas definido nas bases de dados de 2003 é diferente dos valores publicados pelo jornal. Tendo em conta estas diferenças, optámos por usar os dados disponíveis em Agosto de 2006 e ter por base o critério, na forma como foi originalmente enunciado em 2001 pelo Público/UCP – *todas as provas da primeira fase realizadas por alunos internos a uma das 8 disciplinas*. Ao longo deste texto, todas as análises e comparações realizadas são feitas com base neste critério e nos *rankings* do Público/UCP reconstruídos por nós.

### 3 Inconsistências e diferenças de sucesso entre disciplinas

#### 3.1 A questão do sucesso nas diferentes disciplinas

Quando se constrói um indicador baseado em médias globais por escola, está a esquecer-se que os exames não estão aferidos, ou seja, nada garante que meçam de forma equivalente a desempenho dos alunos nas diferentes disciplinas. Mais ainda, nada garante que meçam sequer de forma equivalente o desempenho dos alunos *na mesma disciplina*, em diferentes anos lectivos. A primeira circunstância gera desvantagens injustificadas para escolas com mais exames realizados nas disciplinas que, incidentalmente, têm médias nacionais mais baixas (e, naturalmente, o efeito contrário em outras escolas), enquanto que a segunda faz com que comentários sobre a evolução entre anos sucessivos perca significado. Surpreendente é a ausência de reflexão sobre estas variações de média (inter e intra disciplinas), no sentido de melhorar o equilíbrio do grau de dificuldade dos exames e das suas indicações de correcção (por vezes incompreensivelmente draconianas) e de introduzir mecanismos correctores disponíveis há muitas décadas.

A diferença entre as médias das disciplinas em 2005 (Figura 1) mostra esse desequilíbrio (que afecta negativamente as escolas com mais exames a Matemática), mas é mais reveladora a análise dos resultados ao longo dos anos, mostrados na Figura 2.

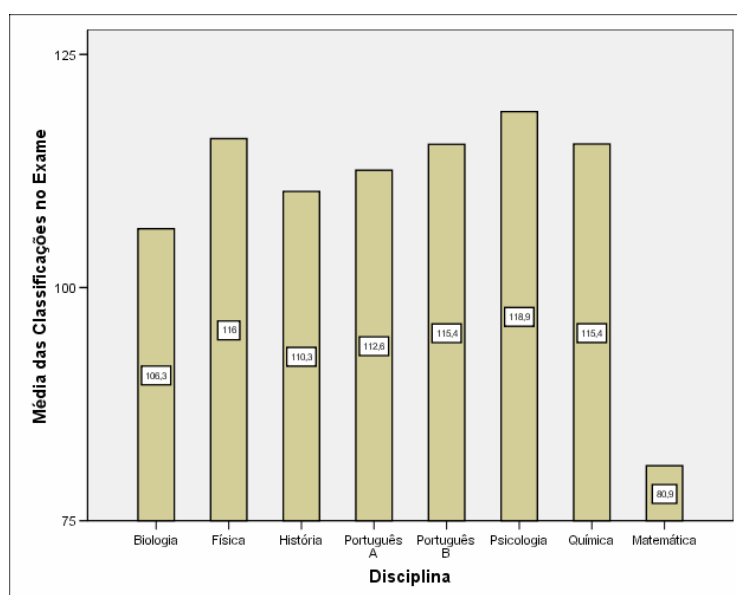


Figura 1 – Diferentes médias nacionais nas disciplinas (2005)

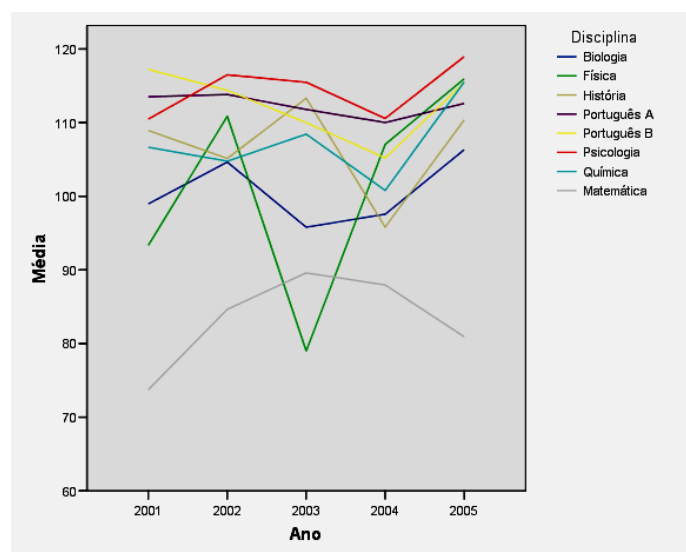


Figura 2 - Variação das médias nacionais (2001-05)

Na verdade, qualquer teoria que pretenda tomar os resultados dos exames pelo seu valor *facial* teria que conseguir explicar como é que em Física, a globalidade das escolas piorou tanto de 2002 para 2003, para logo a seguir voltar aos níveis anteriores (2004) e melhorar ainda mais em 2005, sabendo-se que caíram de novo em 2006 para níveis negativos pouco superiores aos de 2003. Claro que se poderia dizer que estas variações se devem exclusivamente aos alunos (note-se que esta hipótese invalidaria os *rankings* como instrumento de avaliação das escolas), mas ninguém acreditaria em diferenças tão grandes entre gerações tão próximas, nas mesmas escolas.

Para se entender que a diferença da distribuição de resultados entre disciplinas tem influência num *ranking* agregado como o que se discute, deve ter-se em conta que é muito desigual o número de disciplinas diferentes com que cada escola trabalha (Figura 3).

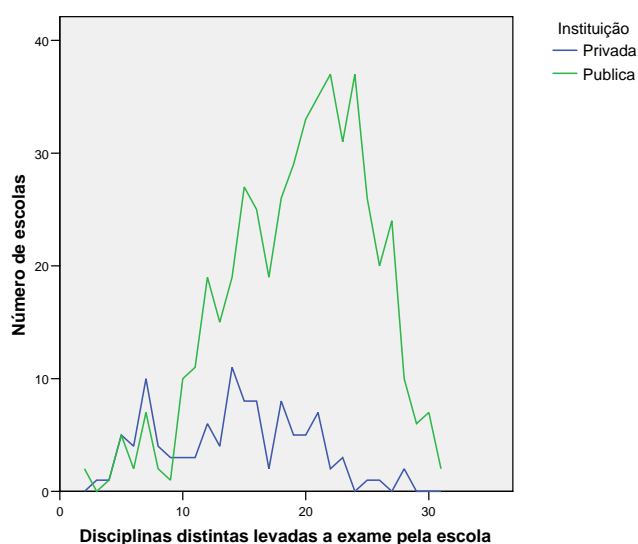


Figura 3 - Estatística da carteira de disciplinas das escolas

De qualquer modo, a influência deste aspecto no *ranking* Público/UCP, não sendo radical ao ponto de o invalidar, pode afectar de forma significativa e injusta a posição de

algumas escolas, para além de contribuir artificialmente para a instabilidade do *ranking* em anos sucessivos, assunto a que voltaremos na Secção 4.

### 3.2 Inconsistências na comparação binária

Um ponto importante é que não é apenas pelo efeito das médias de disciplinas diferentes que surgem inconsistências – o facto do indicador global ser uma média pesada, só por si, provoca situações anómalas quando se comparam escolas duas a duas, sendo possível a uma escola mostrar superioridade em todos os *rankings* parciais (por disciplina), ao mesmo tempo que fica menos bem colocada no *ranking* global.

A Tabela 1 (*ranking* Público/UCP de 2005, reconstruído) mostra duas dessas situações. No primeiro exemplo, o Instituto de S. Tiago realizou exames a seis disciplinas, obtendo a média 79,9 que a colocou na posição 584. Ora a Escola Pré-Universitária Autónoma obteve melhores resultados em *todas as seis disciplinas*, ficando no entanto em lugar inferior do *ranking* (590), por influência do maior número de provas realizadas a Matemática, Física e Química, disciplinas com médias nacionais baixas.

Tabela 1 – Casos notórios de inconsistência

Escola	Bio.	Fis.	Hist.	Port. A	Port. B	Psic.	Quí.	Mat.	Média	Posição
Instituto de S. Tiago – Cooperativa de Ensino	77,11	95,00	-	-	94,09	100,67	86,77	40,15	<b>79,90</b>	584
Escola Pré-Universitária Autónoma	91,07	101,67	-	-	120,50	117,09	88,50	44,43	<b>77,50</b>	590
Escola Secundária Soares dos Reis	-	103,15	-	-	120,90	124,20	-	79,19	<b>115,30</b>	139
Colégio Dr. Luís Pereira da Costa	110,41	113,70	123,67	99,50	127,93	135,89	97,00	87,35	<b>113,5</b>	164

O segundo exemplo é mais característico das situações deste tipo (423 ao todo) no *ranking* Público/UCP de 2005. Trata-se de uma escola (Secundária Soares do Reis) que apresentou alunos apenas a algumas disciplinas e que no *ranking* fica à frente de uma outra escola (Colégio Dr. Luís Pereira da Costa) que em todas essas disciplinas apresenta melhor média, mas que é prejudicada pelos restantes exames realizados pelos seus alunos.

Este tipo de inconsistência é tão forte que não se compreende como nunca foi mencionada a propósito do *ranking* Público/UCP (ou de qualquer outro *ranking* português que conheçamos), sobretudo quando variações relevantes da média de cada disciplina em anos sucessivos também se tornaram evidentes. De qualquer modo, saliente-se que uma parte da inconsistência é inultrapassável, uma vez que resulta do cálculo da média pesada, ou seja, é sempre possível uma escola ter melhores médias do que outra em todas as disciplinas consideradas individualmente e ficar em posição inferior no *ranking*. Por isso, o indicador MCS que se propõe na secção seguinte mitiga este problema (247 casos, contra os já citados 423), ao anular o efeito das médias nacionais, mas nunca poderia eliminar a inconsistência inerente a qualquer sistema baseado em médias pesadas.

### 3.3 Proposta de um indicador alternativo

Uma parte do problema discutido nas secções anteriores decorre, como se disse, das diferenças, por vezes significativas, entre as distribuições das classificações das diferentes

disciplinas consideradas no *ranking* Público/UCP. Na verdade, o indicador que suporta a construção do *ranking* não distingue disciplinas, limitando-se a calcular a média das classificações de todos os exames da escola (embora seja por vezes apresentado como a média ponderada das médias das classificações em cada disciplina, o que vai dar ao mesmo):

$$pub(E) = \frac{\sum_{x \in E} class(x)}{n_e} = \frac{\sum_{d \in D} n_d M_d(E)}{\sum_{d \in D} n_d}$$

onde  $E$  se refere à escola,  $x$  é uma prova,  $D$  é o conjunto das disciplinas,  $M_d$  é a média na escola da disciplina  $d$ ,  $n_d$  é o número de provas à disciplina  $d$  e  $n_e$  é o número total de provas na escola  $E$ .

O novo indicador que se propõe (e que baptizámos de MCS) corrige as classificações brutas, tomando em conta a média nacional da disciplina ( $MN_d$ ):

$$mcs(E) = \frac{\sum_{x \in E} \frac{class(x)}{MN_d(x)}}{n_e} = \frac{\sum_{d \in D} \left( \frac{1}{MN_d} \sum_{x \in E \cap D} class(x) \right)}{\sum_{d \in D} n_d} = \frac{\sum_{d \in D} n_d \frac{M_d(E)}{MN_d}}{\sum_{d \in D} n_d}$$

Como é fácil de ver, o valor deste indicador para uma escola que, em todas as disciplinas, obtivesse as médias nacionais, seria exactamente 1. Naturalmente, escolas melhores do que a média nacional na maior parte das provas terão um indicador maior do que 1, independentemente das disciplinas em causa. Passa também a ser possível comparar anos sucessivos, sem influência da variação de dificuldade dos exames respectivos.

Podemos agora regressar aos exemplos da Tabela 1, aplicando o indicador MCS e o *ranking* correspondente. A Tabela 2 permite verificar que o indicador MCS está mais de acordo, nos exemplos estudados, com a percepção holística da posição relativa entre as Escolas (para maior facilidade de análise, repetiram-se os dados anteriores).

Tabela 2 – Aplicação do indicador MCS

Escola	Bio.	Fís.	Hist.	Port. A	Port. B	Psic.	Quí.	Mat.	<b>Média</b>	Posição
Instituto de S. Tiago – Cooperativa de Ensino	77,11	95,00	-	-	94,09	100,67	86,77	40,15	<b>79,90</b>	584
(MCS)	0,73	0,82	-	-	0,82	0,85	0,75	0,50	<b>0,73</b>	586
Escola Pré-Universitária Autónoma	91,07	101,67	-	-	120,50	117,09	88,50	44,43	<b>77,50</b>	590
(MCS)	0,86	0,88	-	-	1,04	0,98	0,77	0,55	<b>0,75</b>	577
Escola Secundária Soares dos Reis	-	103,15	-	-	120,90	124,20	-	79,19	<b>115,30</b>	139
(MCS)	-	0,89	-	-	1,05	1,04	-	0,98	<b>1,03</b>	207
Colégio Dr. Luís Pereira da Costa	110,41	113,70	123,67	99,50	127,93	135,89	97,00	87,35	<b>113,5</b>	164
(MCS)	1,04	0,98	1,12	0,88	1,11	1,14	0,84	1,08	<b>1,05</b>	173

De qualquer modo, convirá salientar, para terminar este tópico, que a vantagem que vemos no novo indicador e *ranking* MCS se resume a ter uma melhor representação cardinal e ordinal das escolas em relação aos resultados obtidos pelos alunos que

propuseram a exame, que é o que está em discussão quando se fala de *rankings* como o do Público/UCP. As secções seguintes deste artigo são bastante elucidativas sobre a limitação destes indicadores, e dos *rankings* deles resultantes, para outros fins.

Quanto à inconsistência resultante do uso da média pesada (em qualquer dos *rankings*): é uma questão que deixamos em aberto, dada a natureza do nosso estudo, mas é certamente merecedora de posterior reflexão.

## 4 Utilidade do *ranking* para tomar decisões

### 4.1 Colocação da questão

Apesar do indicador que lhes está na base se referir, como se disse, apenas ao desempenho médio nos exames, este tipo de *rankings* tem sido interpretado, e por vezes mesmo apresentado, como *ranking* das escolas, num sentido muito mais global do que seria adequado, tendo em conta a informação em que se baseia. Há uma percepção social de que as escolas mais bem classificadas serão as melhores, de que um aluno que nelas seja integrado terá melhores classificações nos exames que vier a fazer do que se for integrado numa escola com número de ordem mais baixo, de que se “devia fazer qualquer coisa” em relação às escolas menos bem classificadas. Significativamente, as escolas mais bem classificadas são convidadas a explicar o “segredo do seu sucesso”, ficando implícita uma relação forte entre a organização da escola e os resultados dos alunos que ela propõe a exame.

Ou seja, os *rankings* têm sido apresentados, ou pelo menos interpretados, como valiosos auxiliares de decisão:

- a) Para os pais escolherem a “melhor” escola para os filhos, na eventualidade de lhes ser permitida ou possível essa opção;
- b) Para o Ministério da Educação actuar junto das escolas “piores” ou “menos boas”, responsabilizando órgãos de gestão e professores (pelo menos) pelas suas baixas posições no *ranking*;
- c) Para as próprias escolas melhorarem a qualidade.

Repare-se que nesta visão se está implicitamente a considerar que o lugar da escola no *ranking* permite:

- a) Prever a futura eficácia da escola, no que respeita aos resultados dos alunos – isto corresponde, em termos de previsão, a considerar um modelo de persistência (a melhor estimativa dos resultados futuros são os resultados actuais);
- b) Avaliar a qualidade actual da escola – o que neste caso corresponde a identificar qualidade com resultados nos exames, esperando-se alguma estabilidade do indicador em anos sucessivos.

Na ausência de informação adicional, não é possível discutir seriamente até que ponto estas ideias são sustentáveis, embora o material que incluímos na Secção 2 e algum bom-senso nos levem a pensar que a origem social dos alunos, o seu nível anterior e outros factores poderão influenciar bastante os resultados, e que a análise da qualidade também teria mais que ver com o valor acrescentado e outros indicadores associados ao processo do que meramente com resultados de exames.

O nosso exercício é, portanto, de outro tipo, beneficiando da existência de um histórico de 5 anos de resultados publicamente acessíveis, a partir dos quais foram construídos os *rankings* correspondentes. Torna-se assim possível verificar até que ponto se confirmam

as qualidades de persistência e estabilidade que se indicaram atrás como inerentes à utilização dos *rankings* como auxiliares de decisão.

A Figura 4 (*Ranking Público/UCP*) dá desde logo uma ideia algo contrária a essa confirmação, embora algumas escolas se mantenham no grupo durante os 5 anos, e a Figura 5 (*Ranking MCS*) vai no mesmo sentido.

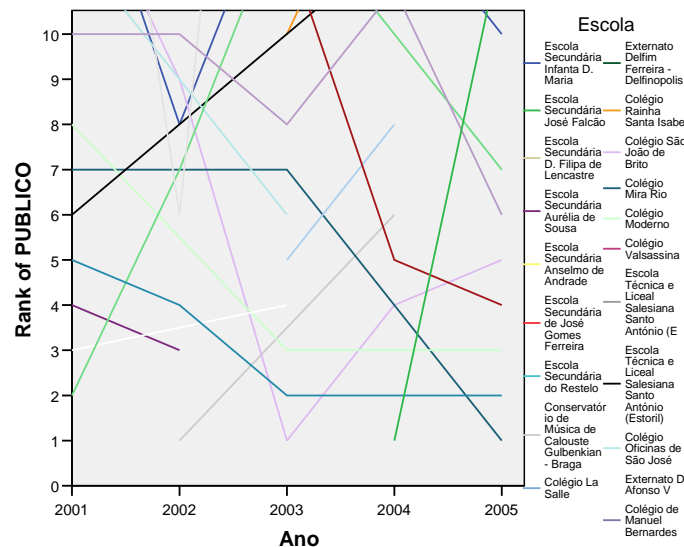


Figura 4 – Variações nas 10 primeiras posições do *ranking* Público/UCP

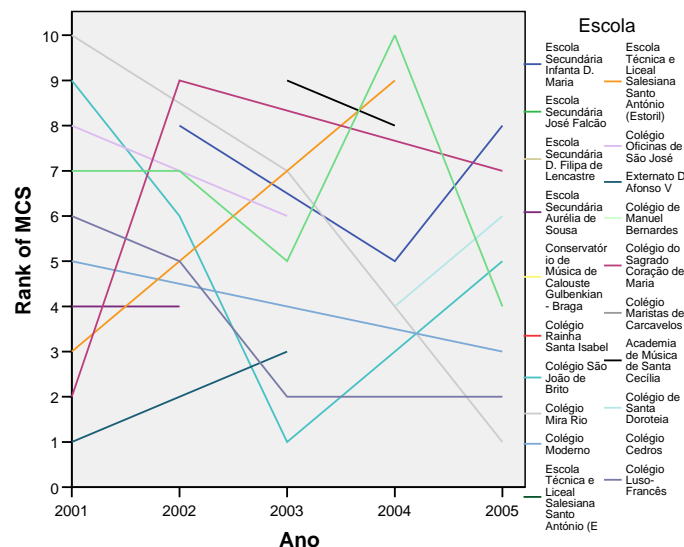


Figura 5 - Variações nas 10 primeiras posições do *ranking* MCS

## 4.2 Variação absoluta de lugar no período de análise

Entretanto, a natureza temporal do exercício (*rankings* de anos sucessivos) e a preocupação em avaliar a capacidade preditiva levou à construção de uma medida da variabilidade ao longo dos anos, correspondendo simplesmente à soma das 4 variações absolutas de lugar que a escola experimentou nos 5 anos cujos resultados estão disponíveis. A aplicação desse indicador levou às conclusões patentes na Tabela 3, onde

se salienta a média de quase 380 variações (ou seja quase 95 em cada transição de um ano para o seguinte) no *ranking* Público/UCP, com valores extremos de 1800 variações (450 por ano!). Na mesma tabela pode ver-se que os valores para o *ranking* MCS são mais moderados, como seria de esperar por se ter anulado o efeito da variação do grau de dificuldade dos exames, mas ainda elevados.

Tabela 3 – Variações absolutas de lugar (2001-05) para todas as escolas

Variables	Mean	Minimum	Maximum	Sum	Percentiles		
					25	50	75
RMCS_DIST	349,12	4	1766	202141	198,00	326,00	459,00
RPUBLICO_DIST	377,54	3	1806	218593	223,00	356,00	497,00

Repare-se, entretanto, que mesmo em zonas tipicamente mais estáveis dos *rankings* (as posições superiores) se verificam grandes variações médias, como mostra a Tabela 4 em relação às primeiras 100 escolas do *ranking* quinquenal. Mais uma vez, o *ranking* MCS apresenta variações menores.

Tabela 4 - Variações absolutas de lugar (2001-05) para as primeiras 100 escolas.

	Mean	Minimum	Maximum	Sum	Percentiles		
					25	50	75
RMCS_DIST	154,50	4	642	14214	65,50	119,50	206,75
RPUBLICO_DIST	161,33	3	698	14842	71,25	127,50	226,00

Finalmente, a Figura 6 é bastante elucidativa das piores situações encontradas nos dados, correspondentes a variações totais (pela ordem da legenda) de 1806, 998, 1495, 1036 e 1221 lugares no *ranking* Público/UCP e 1766, 787, 1029, 883 e 861 no *ranking* MCS.

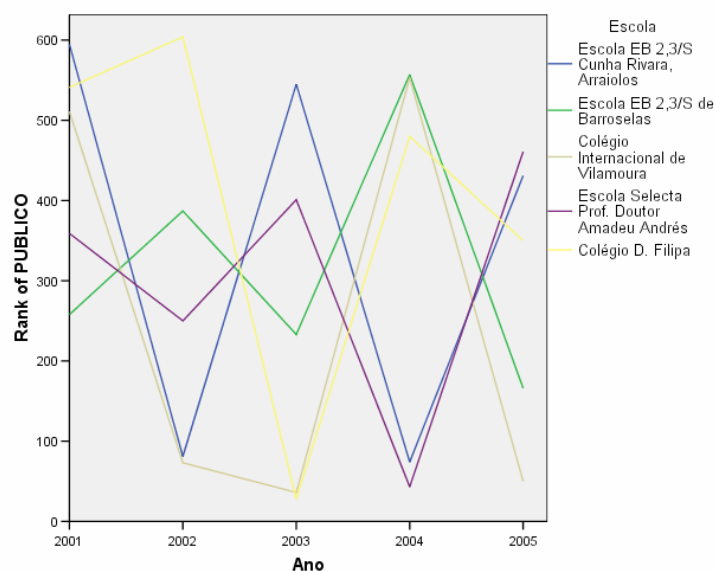


Figura 6 – Escolas com maiores variações totais absolutas de lugar



### 4.3 Permanência no topo

Analizou-se também a estabilidade da permanência nos lugares cimeiros do *ranking*, com os resultados mostrados na Figura 7 – ao fim de 4 anos, apenas cerca de 40% das escolas do “Top 100” de 2001 se mantinham nessa lista.

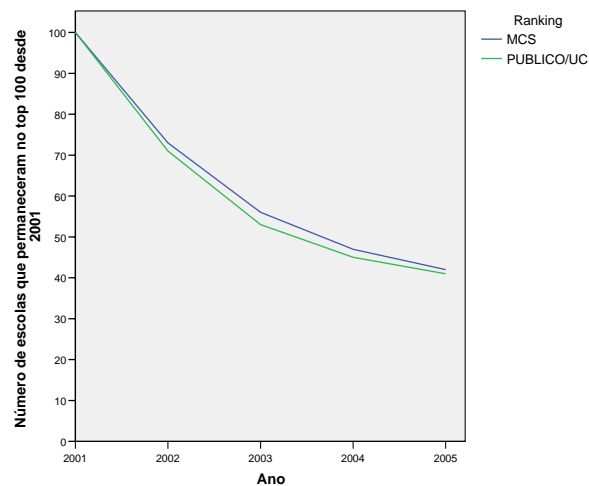


Figura 7 – Despromoções do top 100

Uma análise complementar é a que se mostra na Figura 8, onde em abcissa se pode escolher um “top” de qualquer dimensão, lendo depois em ordenada o número de escolas que já pertenceram a esse *top* pelo menos um vez no período 2001-05. A figura do lado direito é uma ampliação da zona mais proeminente, deixando ver que quase 200 escolas já pertenceram ao grupo das 100 primeiras.

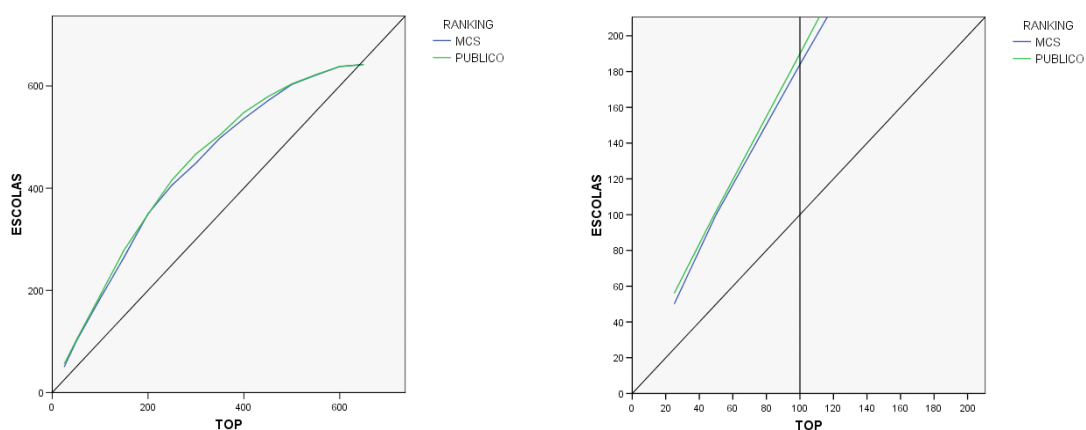


Figura 8 – Número de escolas que já pertenceram ao top

### 4.4 Intervalo de Confiança para a Média

Um aspecto que tem sido muito discutido é o do diferente número de exames que as escolas apresentam (entre a dezena e o milhar), seja pela possível influência nos resultados, seja em relação ao significado estatístico dos valores apresentados nos *rankings*.

Sendo a questão da influência nos resultados discutida mais adiante neste artigo (Secção 5), ficamos agora pela discussão do significado estatístico, assunto sempre tratado um pouco superficialmente na imprensa. Reportando-nos ao Público de 2005, a questão tem sido abordada da seguinte forma: quando uma escola tem menos de 10 ou 15 exames (de acordo com as disciplinas), considera-se que a média não tem significado estatístico e a escola não aparece no *ranking* parcial da disciplina respectiva. No entanto, esses resultados são incorporados nos cálculos para a média global, conduzindo ao *ranking* geral. Não é conhecido o detalhe do raciocínio que levou ao estabelecimento dos limites indicados de 10 e 15 exames.

Esta questão é bastante interessante, porque em rigor não estamos a trabalhar com amostras, mas com a totalidade da população que fez exame sob os auspícios daquela escola. A média calculada é a média, e não uma estimativa da média. No entanto, há realmente um ponto de vista, implícito na interpretação que se quer fazer dos *rankings* como auxiliares de decisão, que justifica de certo modo esta perspectiva. Na verdade, as médias que se calculam num certo ano são implicitamente consideradas como sendo obtidas sobre amostras duma população de resultados que inclui os resultados futuros, ou seja, o modelo de previsão por persistência que já mencionámos.

Admitindo essa interpretação um pouco peculiar, já fará sentido analisar a confiança associada a cada uma das médias, tendo em conta a dimensão da “amostra”. A rejeição nos *rankings* parciais não tem qualquer impacto real, pelo que não seguimos esse caminho, mas é fácil determinar os intervalos de confiança em torno da média. Realce-se que se pode assumir a normalidade da média das classificações dos exames na maioria das escolas, seja pela aplicação do Teorema do Limite Central, seja pela normalidade da classificação do exame na escola. Nas pequenas excepções a esta situação (como é o caso da escola nº 1 da Figura 9), o intervalo de confiança para a média poderá ser inferior ao indicado.

A Figura 9 mostra os intervalos de confiança para as primeiras 20 escolas do ranking Público/UCP de 2005, sendo notória a grande amplitude de alguns dos intervalos, o que mais uma vez revela a fragilidade associada a este tipo de exercícios, se deles se pretender retirar indicações mais substanciais do que as que eles na verdade fornecem. Por exemplo, o valor inferior do intervalo da escola nº 18 é melhor que o valor inferior da escola nº 8. Noutra direcção, repare-se que os intervalos das escolas nº 1 e nº 19 não são disjuntos. Em conclusão, também nesta perspectiva a capacidade preditiva dos *rankings* fica um pouco posta em causa.

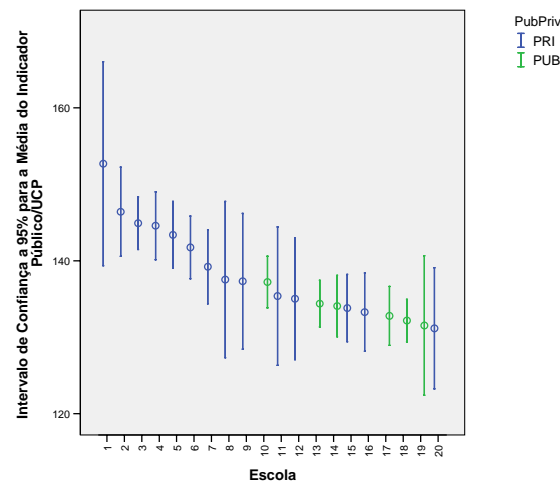


Figura 9 – Intervalos de confiança das médias (95%) em 2005

## 5 Análise de outros aspectos

### 5.1 Diferenças entre a classificação interna e a do exame

Um ponto que tem levantado alguma polémica é a diferença entre as classificações nos exames (CE) e as classificações internas finais (CIF). A discussão é mais complexa do que parece à primeira vista, mas podemos começar por uma constatação, em face dos resultados mostrados na Figura 10: toda a zona à esquerda dos 10 valores contribui para o défice, uma vez que os alunos chegaram aqui com notas positivas.

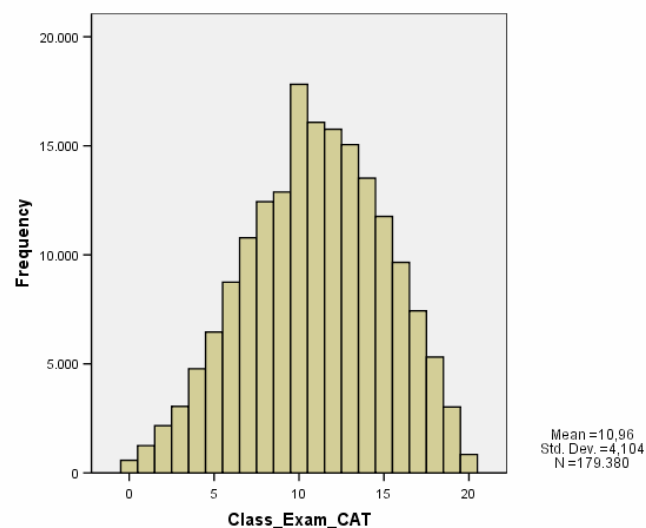


Figura 10 – Distribuição dos resultados dos exames (2005)

A partir daí, não há meio de saber se o sistema de ensino propicia, em média, classificações internas superiores às que os alunos deveriam ter, ou se são os exames nacionais que são demasiado difíceis ou são corrigidos de forma demasiado severa. Para

além disso, há a questão de saber se existem, ou não, escolas com desvios sistematicamente superiores aos médios, configurando situações que deveriam ser investigadas (no sentido policial do termo). No entanto, mesmo descontando esses eventuais casos anómalos, este aspecto deveria ser monitorizado, pois a variância dos desvios é bastante grande, provocando distorções injustas nas classificações de acesso ao ensino superior.

A Figura 11 mostra a experiência acumulada no período 2001-05, separada por disciplinas para evitar o mascaramento devido à diferença entre médias. Verifica-se que, durante este período, o déficit médio a Matemática ultrapassa os 4 valores, enquanto que, no outro extremo, Português A e B têm um déficit médio menos notório, inferior a 2 valores. A separação de escolas públicas e privadas permite ver também que o déficit médio tem sido maior nestas últimas, sobretudo nas disciplinas de Física e Português A.

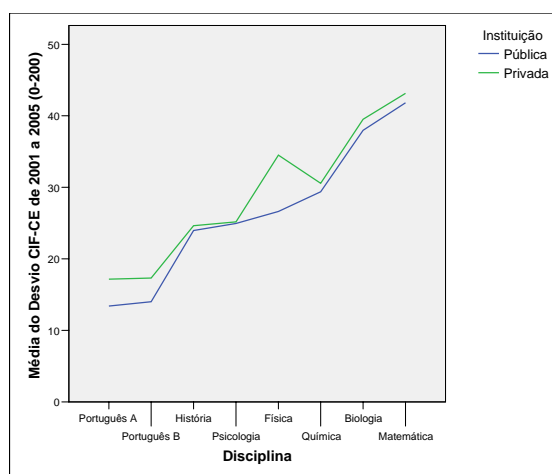


Figura 11 – Déficit classificativo médio por disciplina (2001-05)

No entanto, se se atentar na Figura 12, vê-se que esta tendência se tem desvanecido, ou seja:

- Inicialmente os alunos das escolas privadas tiveram, em média, classificações internas superiores às dos seus colegas equivalentes (ou seja, com a mesma classificação no exame) do ensino público em todas as disciplinas (em 2001, mais de 1 valor em média a Física, quase 0,9 valores a Português A), ficando assim numa posição mais vantajosa para o acesso ao Ensino Superior;
- A evolução até 2005 foi no sentido da convergência entre os dois subsistemas, em parte porque os défices de cada um dos subsistemas em geral diminuíram (embora não em todas as disciplinas), mas sobretudo porque no subsistema privado houve um decréscimo mais acentuado da generalidade dos défices.

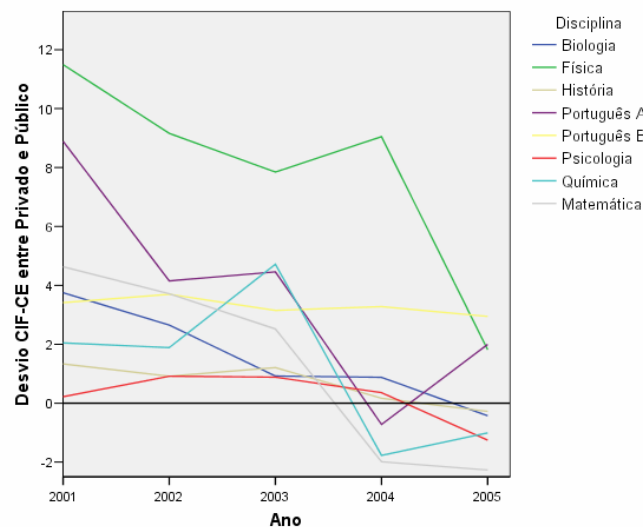


Figura 12 – Evolução do défice classificativo (Público vs Privado)

## 5.2 Influência do número de alunos levados a exame

Na discussão sobre as limitações dos *rankings* é muitas vezes mencionada a eventual influência do número de alunos no desempenho global das escolas, normalmente com a ideia de que escolas com muitos alunos levados a exame ficariam prejudicadas. Analisada essa questão, a Figura 13 (conclusões semelhantes em 2001-04) mostra que não há evidência desse tipo de influência: se por um lado os melhores resultados aparecem sobretudo em escolas com poucas provas, o mesmo se passa com os piores resultados. A principal conclusão é então a de uma maior dispersão de resultados nas escolas com poucas provas, assunto cujo aprofundamento está para além dos objectivos deste artigo.

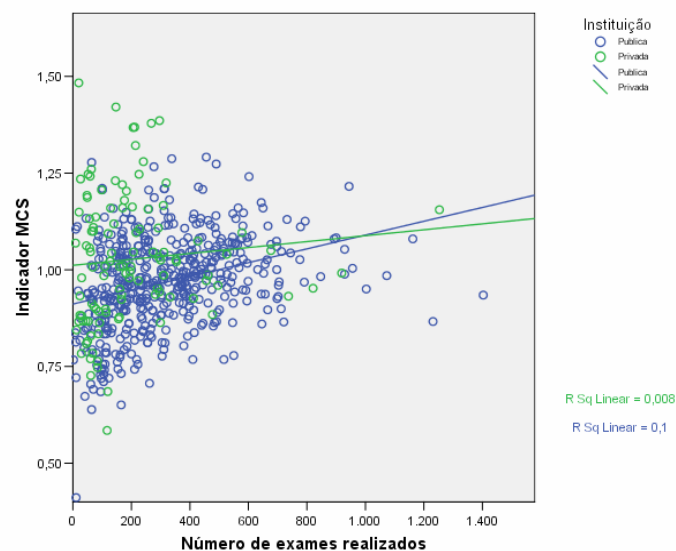


Figura 13 – Influência do número de exames no desempenho (2005)

Saliente-se que as rectas de regressão incluídas na figura são meramente indicativas, dada a proximidade dos coeficientes de determinação ao seu limite inferior (0). Estes valores dizem-nos que o número de provas não explica praticamente nenhuma variação no indicador MCS.

### 5.3 Comparação dos subsistemas público e privado

Embora constituindo um aspecto algo marginal do *ranking*, até porque ao aglomerar subsistemas se torna mais notória a ausência, seja de variáveis explicativas de índole social, seja do grau de retenção (alunos não propostos a exame), a comparação entre ensino público e privado tem sido muito salientada nos meios de comunicação, muitas vezes como suporte para as propostas do cheque-ensino, outras por mera satisfação ideológica, e sempre como veículo de publicidade para as escolas privadas mais bem classificadas, em relação às outras escolas privadas.

Trata-se de matéria que merece bastante reflexão, e onde muitas comparações se revelam injustas, mas, seguindo a perspectiva deste trabalho, vamos sobretudo analisar os dados disponíveis e tirar as conclusões que deles directamente emanam. Assim, podemos começar por verificar, na Figura 14, o seguinte:

- As médias por disciplina em cada subsistema têm sido sempre bastante semelhantes;
- Ao longo dos anos que levam os *rankings*, a média das classificações no ensino privado tem vindo a melhorar de forma sustentada, tendo por referência a média dos resultados das escolas públicas.

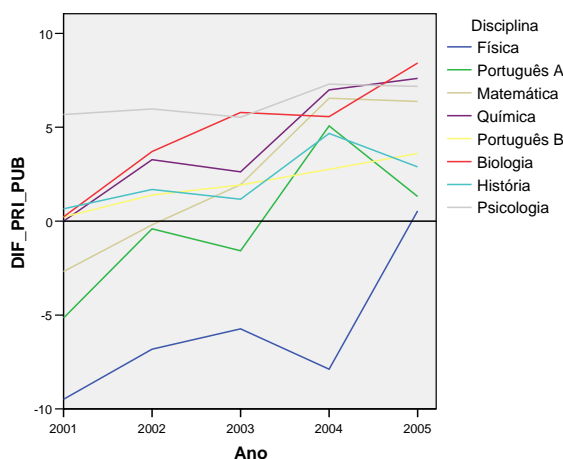


Figura 14 – Diferenças de média entre os sistemas Público e Privado (2001-05)

A evolução mencionada atrás é talvez a conclusão mais interessante, por ser independente de considerações sociais ou outras variáveis que distinguem, presumivelmente, os dois subsistemas. Na verdade, parece que as escolas privadas se têm adaptado melhor do que as públicas a um ambiente onde os *rankings* são bastante valorizados. Seria importante, neste ponto, saber se isso foi conseguido à custa de um aumento do grau de retenção ou se se trata efectivamente de um maior ajuste à obtenção de boas classificações nos exames. Haveria também que discutir, evidentemente, se é isso que realmente se pretende das escolas, mas sabe-se que a publicação de *rankings* induz esse comportamento.

Um ponto de análise menos vulgar é aquele que se apresenta na Figura 15, correspondendo ao seguinte exercício: começando com as escolas mais bem classificadas no *ranking*, foram-se acumulando, de forma separada para os subsistemas público e privado, o número de provas efectuadas, como se todas tivessem tido a média da escola. Isto corresponde a uma das utilizações especulativas do *ranking*: a média representa a escola, portanto é como se todos os seus exames tivessem esse valor, ou, de outra forma,

se os alunos pudessem escolher, inscrever-se-iam nas escolas de maior média, até às respectivas capacidades.

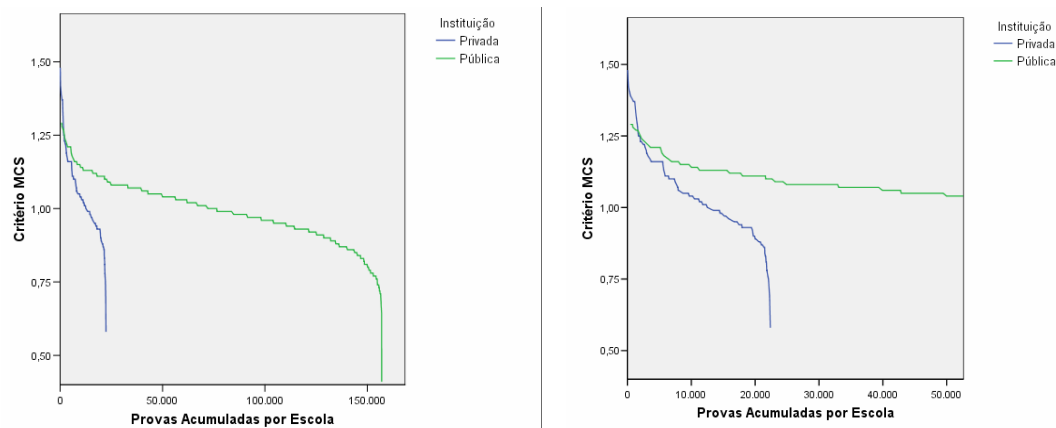


Figura 15 – Colocação de alunos nas escolas segundo os *rankings* público e privado

As curvas (a da direita é uma ampliação) são bem elucidativas da situação. As escolas privadas do topo do *ranking* (até à posição 14 do *ranking* global) e as escolas públicas (até à posição 18 do *ranking* global) acolheriam 1630 provas em cada subsistema, o que corresponde a cerca de 800 alunos. É este o ponto de cruzamento das duas curvas, e estamos a falar de escolas no mínimo 27% acima da média nacional.

A partir deste ponto, as escolas públicas disponíveis estão sistematicamente mais bem colocadas no *ranking* global. A prova nº 5000 no ensino público corresponde a uma escola que está na posição 31, enquanto que a prova nº 5000 no ensino privado corresponde a uma escola que está na posição 48. Se passarmos à prova nº 10000, as posições são 54 no ensino público e 189 no ensino privado.

Esta constatação, que tem a ver com a capacidade muito diferente dos dois subsistemas, não só em número de escolas, mas também na dimensão média das escolas, mostra como é irrealista qualquer comparação dos dois subsistemas em termos da satisfação das necessidades de formação ao nível secundário, mesmo que não se inclua na discussão o grau de retenção, o enquadramento social das escolas ou os custos de formação por aluno.

## 6 Conclusões

O trabalho de análise efectuado pelos autores, do qual este artigo relata as principais conclusões, permitiu assentar algumas ideias relativamente à construção e utilização de *rankings* baseados unicamente nas classificações dos alunos nos exames.

Como preâmbulo técnico, mostrou-se que há uma inconsistência inerente à forma como os *rankings* são construídos (médias pesadas), que é agravada pela utilização dos valores brutos das classificações, devido à grande variação, inter e intra disciplinas, do grau de dificuldade dos exames. As situações anómalas podem atingir o extremo de uma escola, com média em todas as disciplinas em que se apresentou mais baixa do que uma outra, acabar por ficar à frente desta no *ranking*. Para mitigar esta situação, propôs-se o novo indicador MCS, que harmoniza os resultados recorrendo às médias nacionais por disciplina em cada ano. No entanto, a inconsistência permanece.

Em relação ao ponto chave da utilidade deste tipo de *ranking* para auxílio à tomada de decisões, constatou-se uma excessiva variabilidade da posição das escolas ao longo dos anos (quase 100 posições em média, por ano), que pode atingir extremos de variação de 450 posições por ano. Além disso, em muitos casos, o intervalo de confiança da média é bastante alargado, verificando-se importantes sobreposições, o que fragiliza a própria ordenação com base nas médias. Os *rankings* têm, assim, um fraco valor preditivo e não parecem constituir um indicador fiável da qualidade das escolas.

Em relação a outros aspectos, não há evidência de que as escolas que propõem mais alunos a exame (em termos absolutos) tenham melhores ou piores resultados, e confirmou-se a existência de um défice classificativo (CIF-CE) médio importante, que ultrapassa os 4 valores em Matemática, é historicamente maior nas escolas privadas do que nas públicas mas mostra tendência a equilibrar-se entre os dois subsistemas.

Na comparação entre os subsistemas público e privado, com base unicamente nos dados disponíveis, a principal conclusão é a da grande diferença de capacidade, mesmo nas posições mais elevadas do *ranking*, que faz com que, preenchendo escolas pelo *ranking*, o aluno nº 1000 colocado no ensino público estaria com certeza numa escola com melhor posição no *ranking* do que o aluno nº 1000 colocado no ensino privado. É também de mencionar, de qualquer forma, uma tendência, em todas as disciplinas, para a média geral dos exames de alunos de escolas privadas subir em relação à das escolas públicas, sendo superior em todos os casos no ano de 2005.

A conclusão geral do exercício aponta para a necessidade de não sobrevalorizar os *rankings* baseados em exames, sem deixar de os considerar uma peça de informação útil, para as escolas se situarem e para se compararem coisas comparáveis (p.ex. as escolas privadas de Lisboa, excluindo as que estão ligadas a instituições de solidariedade social, as escolas públicas urbanas nas grandes cidades, etc). Por outro lado, sendo evidente o interesse do público por este tipo de ordenações, valerá a pena investir na identificação dos factores de caracterização da qualidade que poderão ser utilizados para complementar a informação dos exames, com o cuidado de privilegiar aspectos compreensíveis, objectivos e acessíveis. Depois, há que garantir que eles estão mesmo acessíveis, e trabalhar!

## Nota

As opiniões dos autores são aqui expressas a título individual, não envolvendo as instituições educativas a que estão ligados, nenhuma das quais apresenta alunos a exames do 12º ano.

## 7 Referências

- Venâncio, Isabel e Otero, Agustin (2003), *Eficácia e Qualidade na Escola*, Ed. Asa: Porto
- Clímaco, M<sup>a</sup> do Carmo (1992), *Monitorização e práticas da avaliação das escolas*, Lisboa: GEP.
- Coleman, J.S. et al (1966), *Equality of educational opportunity*, Washington U.S.: Government Printing Office.
- Rutter, M., Maugham, B., Mortimore, P., Ouston, J., Smith, A. (1979), *Fifteen thousand hours: Secondary schools and their effects on children*, Cambridge, MA: Harvard University Press.
- Comissão de Acesso aos Documentos Administrativos – Parecer nº 126/2001 [Em linha]. Lisboa, 2001. [Consultado em 29 de Outubro de 2006]. Disponível em URL: <http://www.cada.pt/paginas/parecer01/parec126.htm>.



Fidalgo, Joaquim – Enfim “a Lista”. E Agora?.... Jornal Público. Lisboa. ISSN 0872-1556. 2 de Setembro de 2001.

Correia, F., Santiago, R. A., Tavares, O., Pimenta, C. – Um olhar sobre os *rankings*. 1ª ed Coimbra/Matosinhos : CIPES/FUP, 2004. ISBN 972-98848-8-9.

Chitas, P., Pena, P. – Entrevista com David Justino. Revista Visão. Lisboa. Nº 505 (2002).

Direcção-Geral de Inovação e de Desenvolvimento Curricular [Em linha]. 2006. [Consultado em 29 de Outubro de 2006]. Disponível em URL: [www.dgidec.min-edu.pt](http://www.dgidec.min-edu.pt).



# Applying Data Warehousing Technology to Support Planning and Control on Mass Transit Companies

João Mendes Moreira \*

Jorge Freire de Sousa \*

\* INEGI – Instituto de Engenharia Mecânica e Gestão Industrial  
Rua Dr. Roberto Frias s/n; 4200-465 Porto – Portugal  
FEUP – Faculdade de Engenharia da Universidade do Porto  
Tel.: +351-225081639, Fax: +351-225081797  
{jmoreira, jfsousa}@fe.up.pt

---

## Abstract

This paper describes the methodology used in the construction of a Data Warehousing project to support planning and control on mass transit companies. The data sources used were the ones from the GIST98/EUROBUS system - a computer application for supporting operational planning in public transport companies. Firstly, the main Data Warehousing concepts and definitions are introduced. After the problem description and a brief analysis of different Data Warehousing methodologies, the analysis and development phases of the Data Warehousing project are presented. Finally, in the conclusions, the main advantages to implement such a system in the situation described are emphasized.

**Keywords:** Data Warehousing, OLAP, Mass Transit, Transportation Management

---

## 1 Introduction

When computation began to be used massively in the sixties/seventies, the data moved from paper files to device drivers but, in the main, the role that data had in the companies did not change so much. The manager of a company made use of the same information no matter if it was on paper or in a device driver. Along time, information and time won an increasing importance, and the pressure to get more and faster information also increased. In the nineties, Data Warehousing technology was an answer to that tendency that is still going on, particularly with the possibility

to use the data to extract hidden information. It is the era of the data mining and knowledge discovery.

In this paper the construction of a Planning Indicators Board (PIB) is described. The PIB is a top level oriented module of the GIST98/EUROBUS system, a Decision Support System (DSS) for mass transit companies. The proposal is to collect and to show the planning and control information to support the decision-making process using Data Warehousing technology.

Firstly, the main Data Warehousing concepts and definitions are introduced, followed by the description of the GIST98/EUROBUS system and the presentation of a Data Warehousing development methodology. The PIB's construction is described by using the structure of that methodology.

## **2 Data Warehousing concepts and definitions**

Traditionally, the operational databases are:

- Data oriented: the database design depends on data itself and not on the way the data is read by end-users applications. Roughly we can say that the paradigm is: "The less is the space needed to store the data, the better is the database design".
- Not integrated: the data is spread across several databases not necessarily consistent.
- Volatile: the data can be changed.
- Time invariant: often, it is not possible to know the same information at different moments in time.

Databases with these characteristics are not adequate to support management's decisions because queries are time consuming, information is not integrated, and it is not guaranteed that the information is accurate and stored along time. The answer to these problems is the Data Warehouse: "a subject oriented, integrated, non-volatile and time variant collection of data in support of management's decisions" [6]. This concept became quite popular on Information Technology in the nineties, because, on one hand, the information won an increasing strategic importance in the companies and, on the other hand, hardware became cheaper and technologically much better.

The definition of Data Warehouse is data-oriented and does not include all the processes connected with the Data Warehouse technology. In order to get a process-oriented definition, the term Data Warehousing became more popular: "Warehousing refers to a set of processes or an architecture that merges related data from many operational systems to provide an integrated view of data that can span multiple business divisions" [16].

Connected with the Data Warehouse concept appears in 1993 the term OLAP (OnLine Analytical Processing) that is “a category of software technology that enables analysts, managers and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user (. . .)” [8].

## **2.1 Data Warehousing reference architecture**

As it can be seen in Figure 1, the Data Warehousing architecture can be divided, typically, in three big components:

- The sources: operational databases and/or data files.
- The Data Warehouse and all the software needed to load, administer and manage it: the Data Warehouse administration and management software is like a Data Base Management System but with more capacity to manage huge amount of data and to data reading operations. The data acquisition is the software responsible by the data migration from the sources to the Data Warehouse.
- The decision support tools: the frontend & report tools – they are easy to use tools that allow enduser to see and manipulate data in a useroriented way; the OLAP tools – they allow to organise and manipulate data in a multidimensional way; and vertical and Data Mining solutions – they are able to find, in a semi-automatic way, patterns, associations, changes and disturbances, using statistical methods and a big variety of algorithms.

Data Marts can be seen as small Data Warehouses. If the Data Marts load data directly from operational systems they are said to be independent. If they load data from the Data Warehouse, they are defined as dependent. The main difference between a Data Warehouse and an independent Data Mart is that the Data Mart is departmentally or functionally structured, while the Data Warehouse is organisationally structured.

The Data Warehouses and the Data Marts can be stored using the relational model (ROLAP), the multi-dimensional model (MOLAP) or a hybrid model that uses both relational and multi-dimensional models (HOLAP). The relational model [3] is the most popular of all database models. The multi-dimensional model [15] can be seen as a set of multi-dimensional arrays. For each array, the number of results (the NULL is a possible result) is the Cartesian product of the several possible arguments. A multi-dimensional array has the name of Multi-Dimensional Structure or Cube.

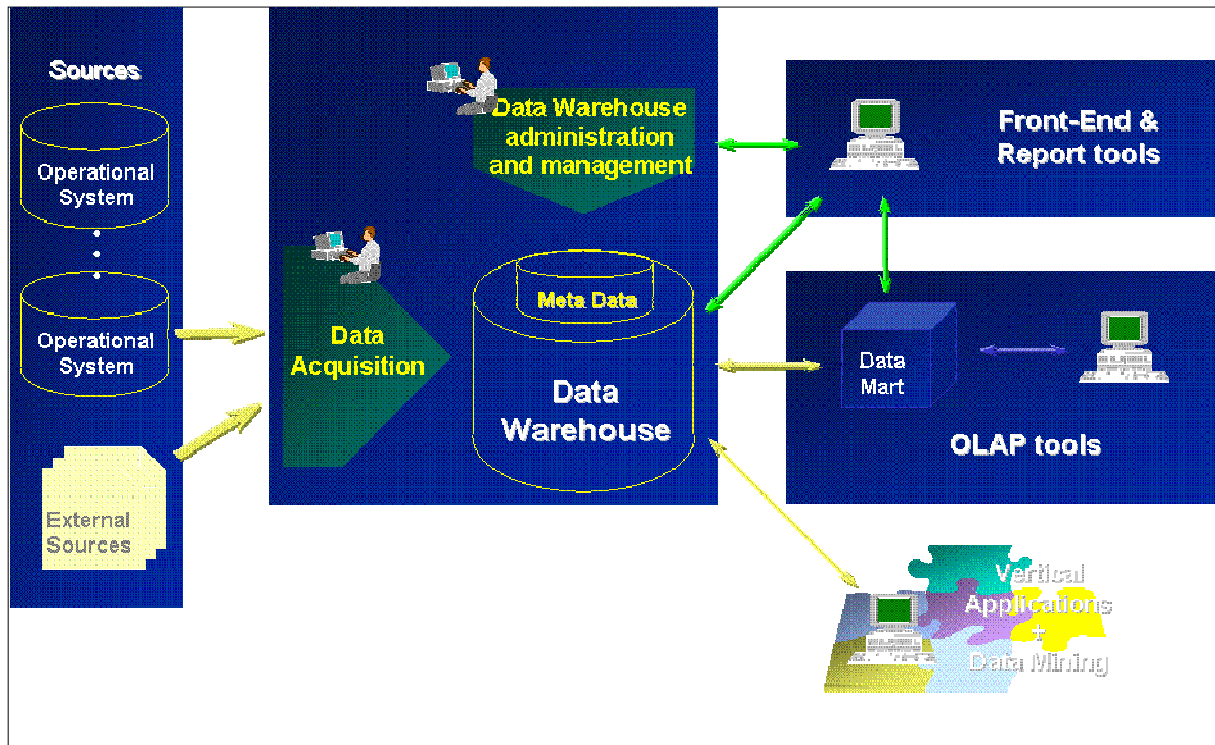


Figure 1: Data Warehousing reference architecture (translated from [12])

### 3 The GIST system

The GIST system [2], [13] is a computer application for supporting operational planning in public transport companies. It was developed as a decision support system that aims to help mass transit companies to improve the operation of critical resources, such as vehicles, drivers and planning staff. The system is also an important software tool to support tactical and strategic management studies regarding companies operations. A consortium of 5 leading Portuguese transport companies (CARRIS, STCP, Horários do Funchal, Empresa Barraqueiro and Vimeca) and 2 R&D institutes (INEGI-FEUP and ICAT-FCUL) is responsible for the GIST system.

The GIST system was successfully installed in those companies in 1996. In 2001, this application was developed under the name of GIST98/EUROBUS. This evolution represents improvements both to the GIST functions and extensions of its functionality. Nowadays, it is being used by eight Portuguese companies that operate, daily, about 2800 vehicles, corresponding broadly to 21% of the road public transport market in Portugal, including Madeira and Azores.

#### 3.1 The GIST98/EUROBUS system

The GIST98/EUROBUS system (Figure 2) contains the following modules:

- Network Module, allowing the definition of the transportation network;
- Gist-Line Module, the route information module;
- Trip and Vehicle Scheduling Module, allowing the trip timetable definition and the vehicle scheduling information management and optimisation;
- Crew Scheduling Module, the crew scheduling information management and optimisation module;
- Crew Management and Rostering Module, which defines the daily tasks to every employee and where various optimisation algorithms are applied to the rostering rules.
- User Information Module, an user-oriented module producing information to the users of public transports;
- Performance Indicators Board Module, a top level oriented module providing planning and control indicators to support the decision making process.

The first four modules are upgrades of the GIST modules. The Crew Scheduling Module, the Crew Management and Rostering Module, the User Information Module and the Performance Indicators Board Module belong to the EUROBUS project that has been financially supported by a public institution named 'Agência de Inovação'. All these modules form the GIST98/EUROBUS system.

### **3.2 The Performance Indicators Board Module (PIB)**

The PIB Module filters the information derived from the other modules so that the toplevel managers can access, in an easy way, the relevant information for the decision making process. The main problem in structuring the PIB Module is to identify those indicators that are significant to top-level managers. All other information is included in the corresponding module. As the GIST98/EUROBUS is a decision support system to the operational planning, the information that it can provide it is essentially the planning and control information.

## **4 The methodology**

There are several Data Warehousing methodologies [1], [16]. In the following, instead of describing them, it will be highlighted what is common to those methodologies (Figure 3), namely the Hadden-Kelly method [5], the methodology from NCR Corporation [4] and the methodology from DataWing Consulting Services, LLC [17]. In fact, it is possible to say that these methodologies have three main phases, even if in some of them, those phases are splitted in more than three. The phases are analysis, development and implementation. The analysis phase includes all the tasks related to the process definition, the requirement analysis and the software and hardware architecture design. The development phase corresponds to the database

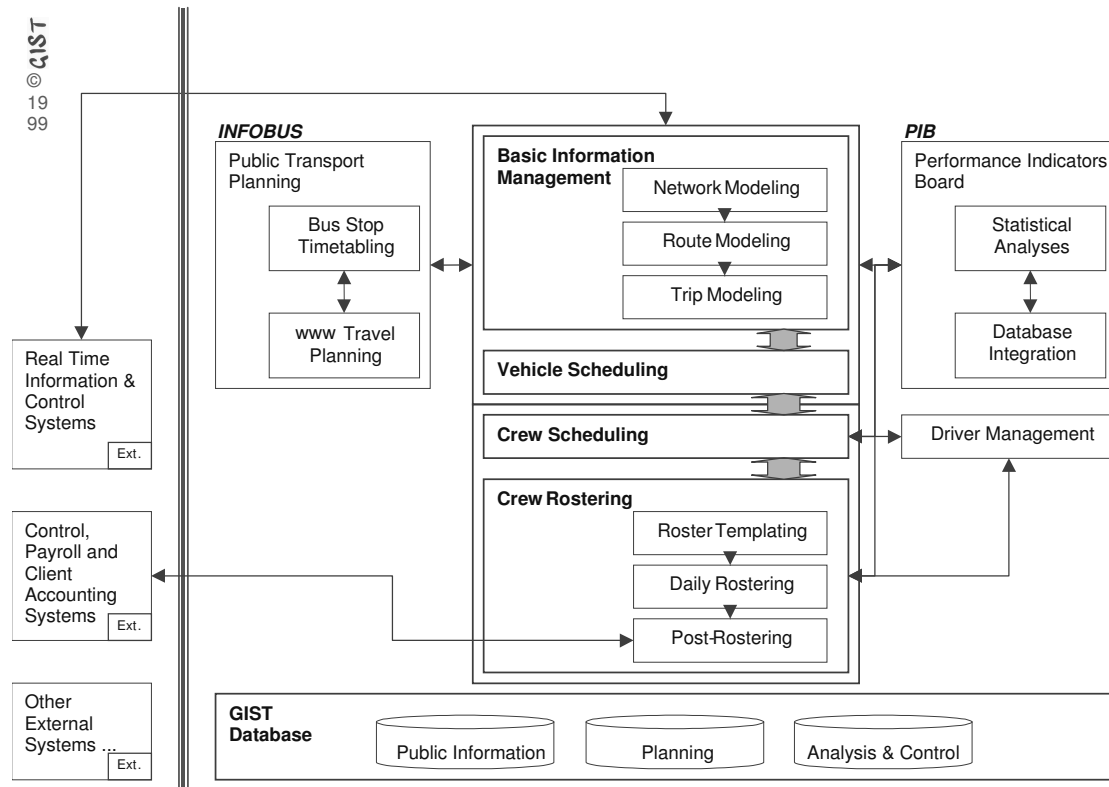


Figure 2: The GIST98/EUROBUS system architecture model

design and development, as well as the transformation and integration programs development and the Cube's definition. The last phase is the system implementation and configuration that is done in the company.

Another common aspect to the main Data Warehousing methodologies is their iterative nature. An iteration, that is every set of three phases, has between three to six months of duration, allowing a good expectation management. So, it can be a never stop process.

#### 4.1 The methodology used to build the PIB module

In the methodology used to build the PIB module (Figure 4) it is clear the sequential structure of the iterations. The only exception is in the analysis phase. In fact, the study of the indicators that the company needs and the study of the data sources are done almost in parallel.



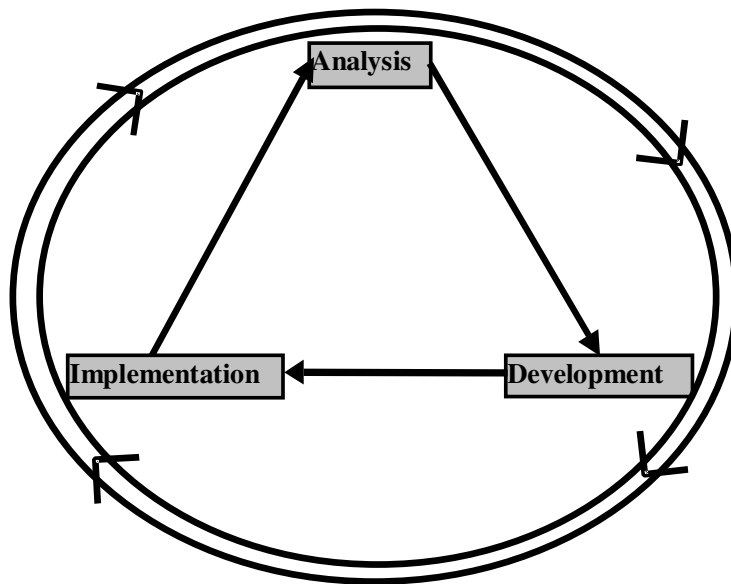


Figure 3: Data Warehousing reference methodology

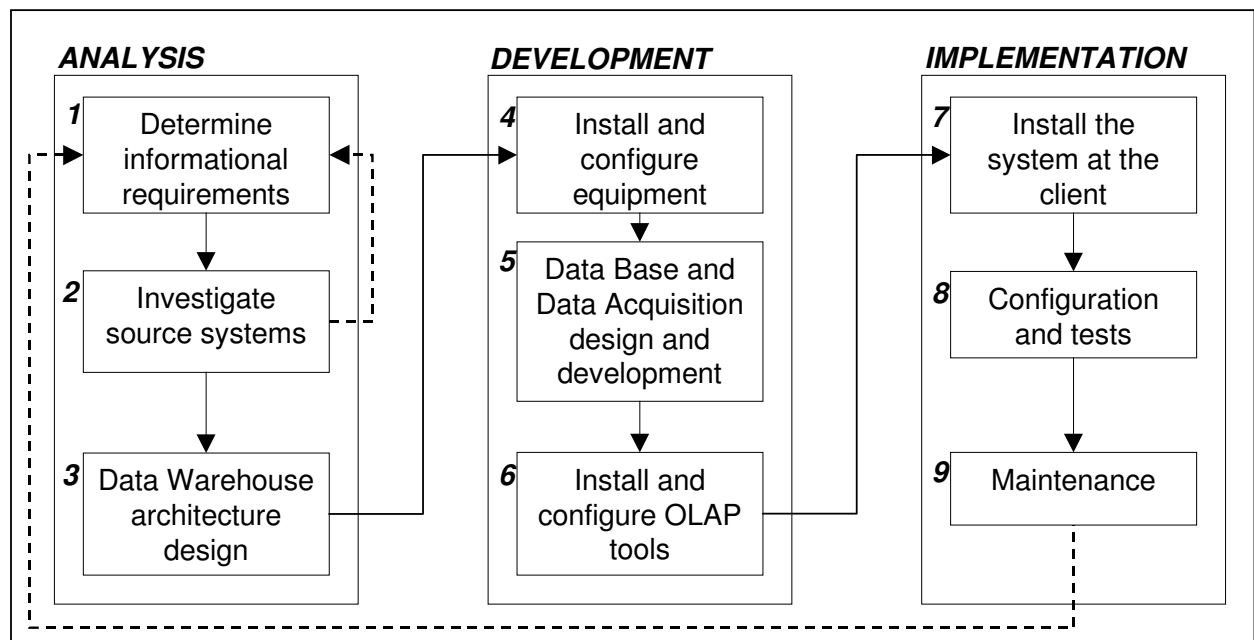


Figure 4: The methodology used to build the PIB module

## **5 The PIB module: Analysis, Development and Implementation**

By definition, a Data Warehouse integrates all the data of the company that is relevant to support management's decisions. Since the data used on PIB module is exclusively the one that comes from one or more GIST98/EUROBUS systems, we cannot say that the PIB's database is a Data Warehouse. In fact, it is more correct to define the PIB's database as an independent Data Mart. Independent because the data sources are operational systems (one or more GIST98/EUROBUS databases) and Data Mart because it is functionally structured (it uses the data relevant to the planning and control activities).

In this section it is described the analysis and development phases of the first iteration.

### **5.1 The analysis phase**

The analysis phase has, typically, a duration of three months.

#### **5.1.1 Informational requirements**

The first questions we tried to answer were: "Which are the big groups of indicators that managers need?" and "Which indicators must be included in each group?"

The results presented in this text are derived from the study accomplished at the STCP Company (the Oporto Public Transport Authority).

##### **"Which are the big groups of indicators that managers need?"**

Taking attention to the information that GIST98 contains and to the big areas where top-level managers need planning and control information, we have defined two main groups of performance indicators: service indicators and crew indicators. The first group gives information about the level of service that the company is offering, which is especially important to the operations management department, while the crew indicators give information relevant to the human resource department.

##### **"Which indicators must be included in each group?"**

The first thing to do was to collect information from the companies. It was a long, interactive process between the analysts and the companies' staff. The service reports produced monthly by the companies were another major source of information.

Based on the information collected it was possible to notice that each potential indicator has no more than 4 parameters. This means that it is possible to identify a specific indicator by answering to 4 questions: "Which is the indicator denomination?", "Which entity does the indicator refers to?", "Which is the aggregation level

used?” and “Which aggregation function does it represent (in other words, is it an average, a maximum or any other kind of function)?” As an example: if we want to know the average distance of trips per route, the indicator is the distance, the entity is the trip, the aggregation level is the route, and the aggregation function is the average. Using this methodology it was possible to identify the indicators group, the entities group and the aggregation function group. The aggregation level was more complex to define. In fact, using the example above, the trips can be aggregated by route, by line, by day type, by route and by day type at the same time, etc., i.e., they can be aggregated by one entity or by a combination of entities. The first thing we observed was that when an indicator can be aggregated by one entity it can also be aggregated by more generic entities. In the example, if routes can aggregate trips, lines can also aggregate trips (notice that a line is a set of routes). The step forward was to define the different dimensions to be used. Each dimension refers to a sequence of entities. The order by which they appear on table 1 is related with the degree of detail, i.e., the first entity of each dimension is the most generic one and the last one is the most specific. In other words, the entity referred at the first column of each dimension is a set of the entities referred at the next one, and so on.

In tables 1, 2 and 3 the results obtained by this methodological approach are presented (the aggregation functions were ignored just for the sake of simplicity).

Table 1: Dimensions

Dimensions	Levels			
Dates	Year	Quarter	Month	Day
Network Dates	Network Date	Route		
Timetable Dates	Timetable Date			
Companies	Company			
Year Seasons	Year Season			
Depot	Depot			
Free Days Groups	Free Days Groups			
Periods of the day	Period of the day			
Network	Line			
Day Types	Day Type			
Route Types	Route Type			
Network Types	Network Type	Crew		
Situation Types	Situation Type			
Trip Types	Trip Type			
Crew	Category			
Vehicle	Logic Vehicle	Stretch		

With these tables we can easily identify four groups of indicators: the route extension and the number of routes in the first group of indicators, the number of logic vehicles in the second group, all the others indicators from table 2, in the third group, and the indicators presented in table 3, in the fourth group. The identification of these groups of indicators derives from the indicator's entity and from the set of dimensions used. Each group will be implemented using the Cube concept. Once the cubes are defined tables 4 and 5 are used in substitution of tables 2 and 3 because they present the same information in a more concise and practical way.

Table 2: Service Indicators

Service Indicators	Dimensions												
	Companies	Network Types	Network	Network Dates	Timetable Dates	Year Seasons	Depots	Periods of the Day	Dates	Route Types	Day Types	Trip Types	Vehicle
Route Extension	x	x	x	x						x			
Number of Routes	x	x	x	x						x			
Number of Trips	x	x	x		x	x	x		x		x	x	x
Trip Distance	x	x	x		x	x	x		x		x	x	x
Trip Driving Time	x	x	x		x	x	x		x		x	x	x
Trip Driving Time with Support Time	x	x	x		x	x	x		x		x	x	x
Trip Commercial Speed	x	x	x		x	x	x		x		x	x	x
Trip Exploration Speed	x	x	x		x	x	x		x		x	x	x
Number of Logic Vehicles	x	x	x		x	x	x	x	x		x	x	x

Table 3: Crew Indicators

	Dimensions						
Crew Indicators	Companies	Depots	Situation Types	Free Days Groups	Network Types	Crew Crew	Dates Dates
Number of Drivers	x	x	x	x	x	x	x
Number of Planned Situations	x	x	x	x	x	x	x
Number of Real Situations	x	x	x	x	x	x	x
Driver Planned Actual Duration	x	x	x	x	x	x	x
Driver Planned Extra Duration	x	x	x	x	x	x	
Driver Planned Night Duration	x	x	x	x	x	x	x
Driver Real Actual Duration	x	x	x	x	x	x	x
Driver Real Extra Duration	x	x	x	x	x	x	x
Driver Real Night Duration	x	x	x	x	x	x	x

Table 4: Dimensions by Cube

Cubes	Commercial Network	Vehicle's Services	Vehicle's Services along the day	Crew
Dimensions				
Dates	x	x	x	
Network Dates	x			
Timetable Dates		x	x	
Companies	x	x	x	x
Year Seasons		x	x	
Depots		x	x	x
Free Days Groups				x
Periods of the Day			x	
Network	x	x	x	
Day Types		x	x	
Route Types	x			
Net Types	x	x	x	x
Situation Types				x
Trip Types		x	x	
Crew				x
Vehicle		x	x	

Table 5: Indicators by Cube

Commercial Network	Vehicle's Services	Vehicle's Services along the day	Crew
Route Extension	Number of Trips	Number of Logic Vehicles	Number of Drivers
Number of Routes	Trip Distance		Number of Planned Situations
	Trip Driving Time		Number of Real Situations
	Trip Driving Time with Support Time		Driver Planned Actual Duration
	Trip Commercial Speed		Driver Planned Extra Duration
	Trip Exploration Speed		Driver Planned Night Duration
			Driver Real Actual Duration
			Driver Real Extra Duration
			Driver Real Night Duration

### 5.1.2 Source systems

In order to know how to obtain the identified indicators from the available data, maps that relate the sources (modules) with the dimensions and indicators have been done (table 6).

The “New” column exists because there is some information that is created during the data acquisition process. An example is the company owner of each source. That information is implicit to each source but it must be explicit in the Data Warehouse.

In order to obtain that information with more detail, tables have been created that describe, for each level of every dimension and for each indicator of every Cube, which are the tables from the sources that are needed to obtain them. Table 7 shows this information for the different levels of each dimension.

### 5.1.3 Data Warehouse architecture design

To define the Data Warehouse architecture the first step is to preview the database dimension. Secondly, it is necessary to choose the software tools that will be evaluated. Once the evaluation is done, the architecture is established.

To preview the database dimension it was assumed the use of the star schema [11] since it is the most space consuming database design used for these proposals. Since the PIB module is part of the GIST98/EUROBUS system, the scalability problem is a minor one. The formula used to preview the database space was:

$$fmidx \times fmdiv \times (ech \times ndim + eind \times nind) \times nregmax \times fmgra \quad (1)$$

Table 6: Indicators by Cube

From To	New	Network	Gist-Lines	Trip & Vehicle Scheduling	Crew Management & Rostering
Dates	x				x
Network Dates	x				
Timetable Dates				x	
Companies	x				
Year Seasons		x			
Depots					x
Free Days Groups					x
Periods of the Day			x		
Network			x		
Day Types	x				
Route Types		x			
Net Types	x				
Situation Types					x
Trip Types				x	
Crew					x
Vehicle				x	
Commercial Network Indicators		x	x		
Vehicle's Services Indicators		x	x	x	
Vehicle's Services along the day Indicators		x	x	x	
Crew Indicators					x

Table 7: Data sources for each level of each dimension

Dimension →Level	Data source
Dates	Data acquisition process and <i>Shifts_schedules; Daily_rosters</i>
Network Dates →Network Date	Data acquisition process
Timetable Dates →Timetable Date	Data acquisition process
Companies →Company	Direct introduction
Year Sessions →Year Session	<i>Year_seasons</i>
Depots →Depot	<i>Depots</i>
Free Days Groups →Free Days group	<i>Freeday_groups</i>
Periods of the Day →Period of the Day	<i>Periods</i>
Network →Line	<i>Lines</i>
Network →Route	<i>Paths e Lines.paths</i>
Day Types →Day Type	<i>Day.types</i>
Route Types →Route Type	Direct introduction
Net Types →Net Type	Direct introduction
Situation Types →Situation Type	<i>Situation_types</i>
Trip Types →Trip Type	Direct introduction
Crew →Category	<i>Categories</i>
Crew →Crew	<i>Crew.members</i>
Vehicle →Logic Vehicle	<i>Shifts_schedules, Shifts e Trips</i>
Vehicle →Stretch	<i>Shifts_schedules, Shifts, Trips, Trips.nodes, Glines.nodes.types e Nodes.types</i>

where:

*fmidx* – a factor to preview the space used by the indexes: 2.

*fmdiv* – a factor to preview the space used by metadata, or other no counted space: 1,5.

*ech* – size of each primary key field: 6 Bytes.

*ndim* – number of independent dimensions: marked with **X** on table 8.

*eind* – size of each indicator field: 4 Bytes.

*nind* – number of the Cube's indicator: see table 5.

*nregmax* – number of records (it was used the data from CARRIS - the biggest company): it is the Cartesian product of the number of members of the independent dimension's lowest levels (marked with **X** on table 8).

*fmggra* – is a granularity factor. It is the average of records' number needed for each value of *nregmax*. It is assumed that the data detail level is already defined. It is a critical decision once it is a compromise between the data space and the information detail level. For example, to the vehicle's services Cube, it is important to save the trips information because, even if indicators are not aggregated by trip, this information allows, in the future, the calculation of other indicators with a low cost. The average of trips by vehicle's stretches is around 15. This is the *fmggra* to the vehicle's services Cube. To the other Cubes the *fmggra* will be 1.

The values of *fmidx* and *fmdiv* will be studied using a more reliable sample.

Looking to table 8 it can be seen that around 80% of the database dimension is needed to obtain the indicator 'Number of logic vehicles'. It is possible to get this indicator by calculus once the beginning and the end of each stretch are stored. With this change, the database dimension for a year will be around 1,9 Giga Bytes. For three years – the time considered as necessary by the companies to keep the information – it will be around 5,7 Giga Bytes.

The software tools chosen to be evaluated were: *MicroStrategy 7* (a reference between the ones that use the ROLAP architecture); *Express Server* from *Oracle Corporation* (a reference between the ones that use the MOLAP architecture); *iTM1* from *Applix, Inc* (a very versatile tool); and *SQL Server7/OLAP Services* from *Oracle Corporation* (a tool with a very competitive ratio price/quality).

The weights used for each criteria were the ones that seem to be more adequate for this particular case (see table 9). The points were based in some tests and mainly in [10]. *MicroStrategy7* was classified with 8 in the 'Database management system robustness' despite it has not an own Database management system. That classification is the one attributed to *Oracle8 Enterprise Edition* once it is the Database management system used by all the other modules of the GIST98/EUROBUS system and it would be naturally the one to use with the *MicroStrategy 7* option.

Figure 5 presents the chosen architecture. An important feature of *SQL server 7' OLAP services* is the possibility to choose between ROLAP, MOLAP or HOLAP architecture. The *DTS – Data Transformation Services*, the *SQL Server 7*, and the



Table 8: Prevision study of the database dimension for a year

Dimensions	Number	Commercial Network	Vehicle's Services	Vehicle's Services along the day	Crew
Dates	365		<b>X</b>	<b>X</b>	<b>X</b>
Network Dates	10	<b>X</b>			
Timetable Dates	10		x	x	
Companies	1	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>
Year Seasons	3		x	x	
Depots	6		x	x	x
Free Days Groups	10				x
Periods of the Day	96			<b>X</b>	
Network	350	<b>X</b>	x	x	
Day Types	3		x	x	
Route Types	2	x			
Net Types	1	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>
Situation Types	100				x
Trip Types	2		x	x	
Crew	2500				<b>X</b>
Vehicle	2000		<b>X</b>	<b>X</b>	
Number of records		3.500	10.950.000	70.080.000	912.500
Space size (MB)		0,32	1.503,75	6.817,02	146,20

Table 9: OLAP tools evaluation

	Weight	MicroStrategy 7	Express	iTM1	SQL Server / OLAP Services
Capacity	10	10	8	6	9
Database structural simplicity	50	4	5	9	8
Database structural flexibility and calculation functionality	50	7	9	7	9,5
Data access functionality	70	7	10	7	9
Technical support and documentation	50	7	9	7	7
Database management system robustness	70	n.a. (8)	8	8	10
Decision support tools	70	6	8	6	9
Price	90	6	2	5	10
	4600	<b>3010</b>	<b>3230</b>	<b>3130</b>	<b>4175</b>
<b>Classification (%)</b>		<b>65,43%</b>	<b>70,22%</b>	<b>68,04%</b>	<b>90,76%</b>

n.a. - not applicable because it has not a own database management system

Pontos from 1 to 10

OLAP services are all part of the *SQL Server / OLAP Services* option. The *Knosys ProClarity 2.0* is an easy to use front-end tool as the quite known *Microsoft Excel 2000*.

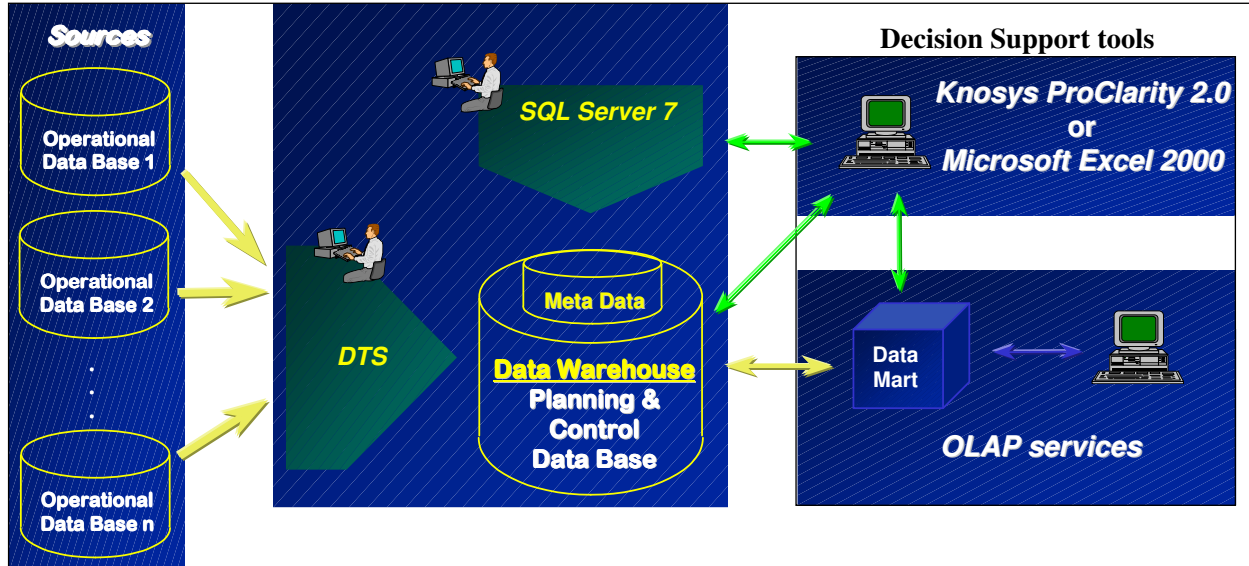


Figure 5: PIB architecture

## 5.2 The development phase

The development phase has, typically, a duration of three months.

### 5.2.1 Install and configure equipment

It is a technical work without any special difficulty.

### 5.2.2 Data Base and data acquisition design and development

In the development phase of a Data Warehousing system there are three main steps to consider [14]:

- The database design: As shown in table 3 and figure 6, the data is stored according to the indicators and dimensions required. The data is pre-processed before the user can access it; so, the concern with efficiency when designing the database has impact, especially on the processing time. It has no visible impact on the efficiency as seen by the end-user. The efficiency for the end-user is mainly determined by the architecture used. The ROLAP architecture is more scalable (manages more data) but is slower while MOLAP architecture is less scalable but faster. The PIB module uses the HOLAP architecture. Figure

6 shows the database design for the crew indicators. As it can be seen, it is directly obtained from tables 1 and 3. The database design was done according to the dimensional model [7].

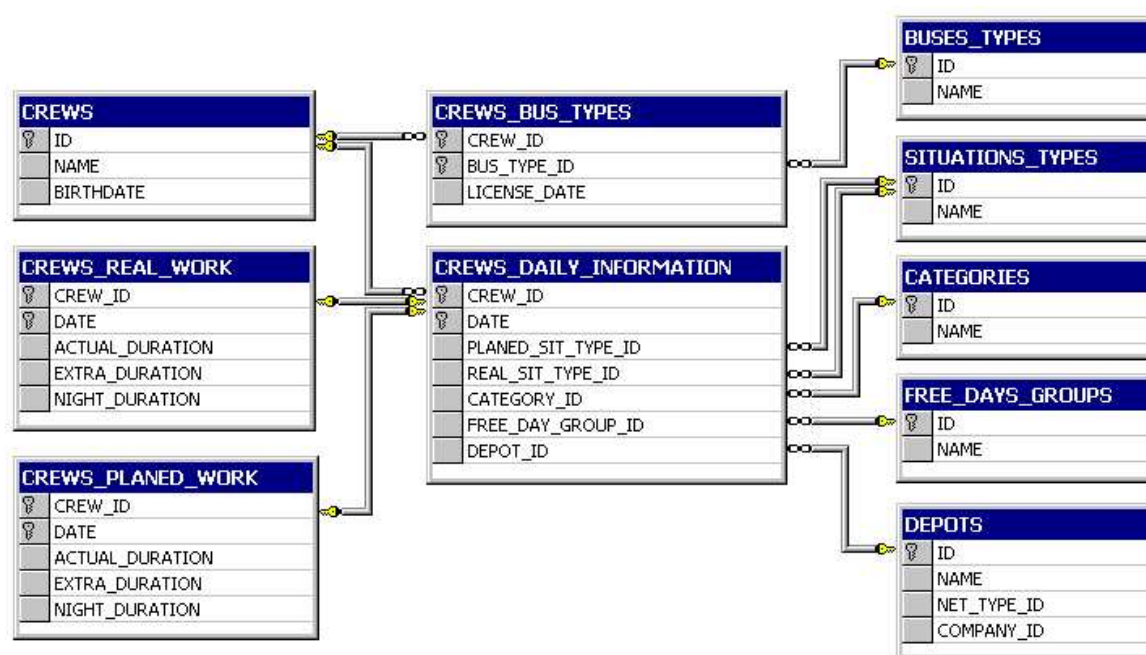


Figure 6: database design for the crew indicators

- The transformation and integration programs [16]: The development of all these programs takes typically around 80% of all the development effort. It includes the extraction, transformation, transport and load processes. Previously, it is important to analyse the source data and clean it. This is a very important issue because without reliable data it is not possible to guarantee the results. The source of information can be multiple: operational databases, other Data Warehouses/Marts or external data. In PIB's case, the source data can be one or more GIST98/EUROBUS databases. The execution of these programs is done every night assuring that the Data Warehouse is always up to date. This step was done using the *DTS – Data Transformation Services*.
- Cubes definition [15]: A cube is a structure that can be seen as a huge table (fact table) where the primary key is the set of all dimensions' identifiers and the attributes are the indicators. Taking attention to tables 1, 4 and 5, it is possible to define the cubes easily. The first thing to do is to define the dimensions as in table 1. The second step is to implement the cubes defining the indicators and selecting the related dimensions. This step was done using the *OLAP Services*.

### 5.2.3 Install and configure OLAP tools

Finally, how can the information be useful to the decision making process? There are several non-expensive analysis and reporting tools that allow the user to do all kind of analysis, just executing basic commands with the mouse. In figures 7 and 8 some examples using the Knosys – ProClarity 2.0 are presented. They show two specific types of analysis. Figure 7 shows the decomposition tree with the information about the distance planned to the day 28-05-1999 with different levels of detail: by net type; by logic vehicle; by stretch; by travel's type and by route. Figure 8 shows the relationship between the planned distance and the number of trips by logic vehicle for the day 28-05-1999.

Using this tool it is possible to do much more types of analysis such as, to compare indicators for the same month of consecutive years, or for consecutive months, to see line, column, pie, bar or other chart types, to create other indicators using formulae, etc. All this may be done using the aggregation level chosen by the user. To aggregate or detail data, it is enough to drill up or down.

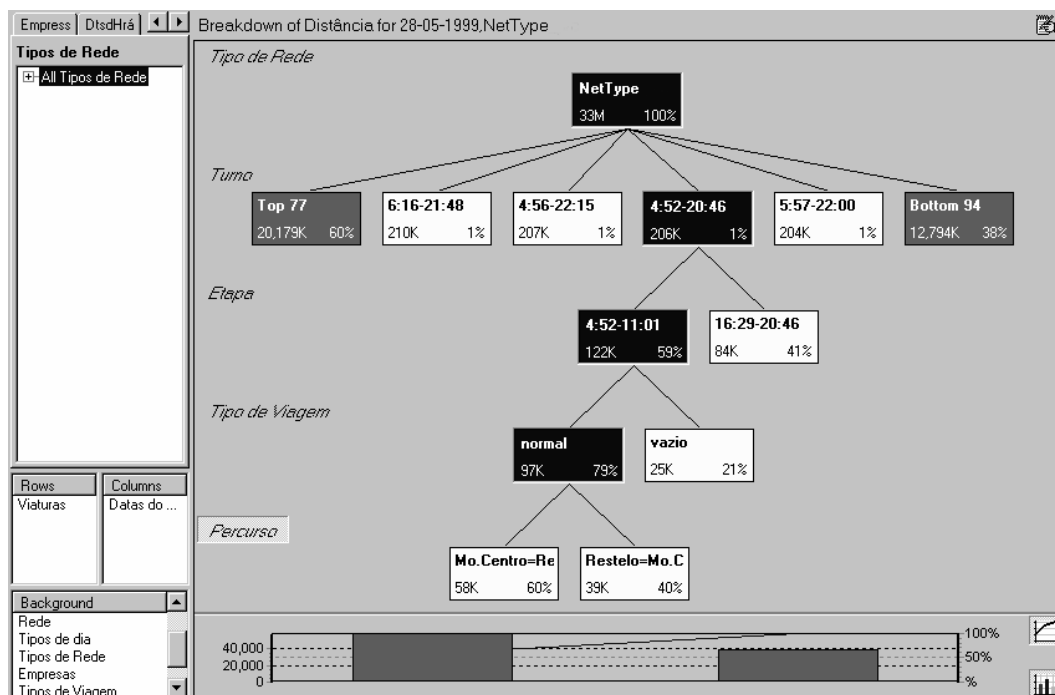


Figure 7: Decomposition Tree

### 5.3 Some considerations about the implementation phase

The implementation phase has a set of tasks that are done once and other tasks that must be done periodically. In the first group are:

- To install and configure hardware and software at the client: like the install and configure equipment task, it is a technical work without any special difficulty.

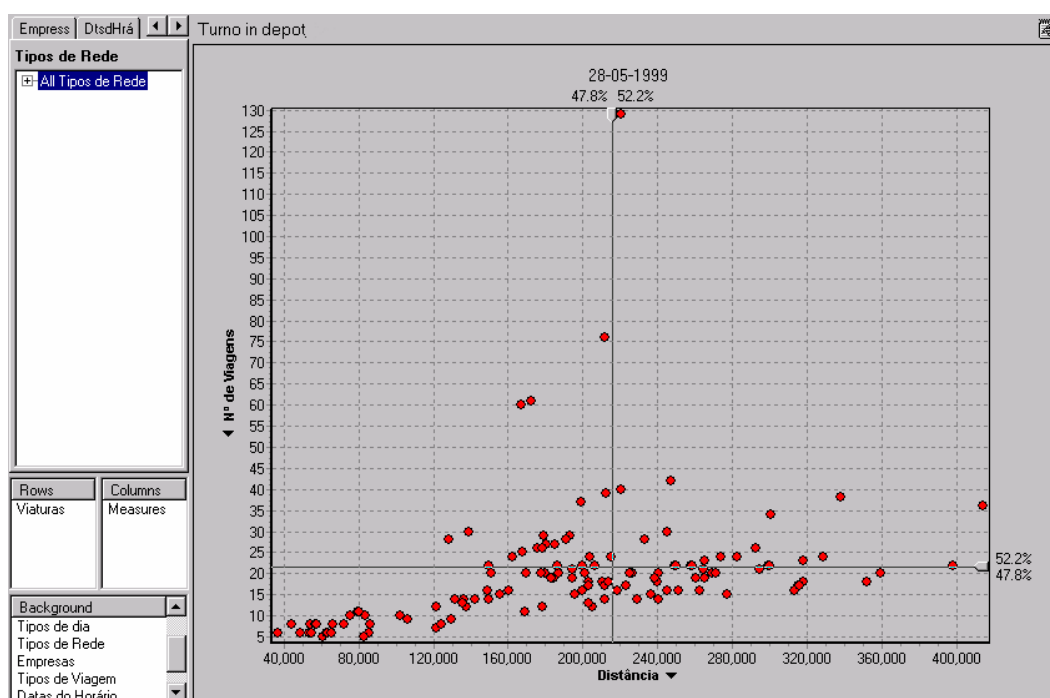


Figure 8: Perspective chart

- To create database, users and access policy: it is a typical task for a database manager.
- To verify and configure OLAP Services: when the OLAP Services will be tested with real data.
- To create a backup and data acquisition policy: it is a very important issue where the acquisition and backup periodicities are defined. The data acquisition policy depends on the aggregation level previously chosen. The backup policy is a basic task for every database system.

In the second group are:

- To verify log files: it is a regular task to do at least after every acquisition procedure since it is, usually, the main source of errors.
- To verify indicators evolution along time to detect eventual data errors: it is important to control the indicators evolution because there are several possible errors that are difficult to detect by software.

## 6 Conclusions

There are three main reasons to implement a Data Warehousing system [9]:

- There is the perception that the information exists but it is not so useful as it could be;
  - Each department has its own language and communication is difficult;
  - The reports production is expensive and inefficient.

To transform data in information is the main goal of a Data Warehousing system. Traditionally, when a manager wants information he asks it to the information department in order to get a report. If the manager wants additional information he needs to wait longer. A Data Warehousing system lets the manager free to pick up the information he wants, with the type of display he prefers, just using the mouse. This flexibility leads to a higher concentration on information instead on data, giving more time to analyse information. The report production is also reduced since time is not spent doing queries to the database. Another important feature is that all the information is centralised, with everybody using the same indicators with the same definitions, which makes communication easier. But the main advantage of the Data Warehousing systems is to turn data into a competitive advantage because it can be used to analyse relationships between variables, to analyse trends, in a word, to put the information at the service of the decision making process.

## 7 References

- [1] Sid Adelman, Joe Oates, *Data Warehouse Project Management*, Data Management Review, May 1998
- [2] João Falcão e Cunha, Jorge Pinho de Sousa, Teresa Galvão, José Luis Borges., *A Decision Support System for the Operational Planning at Mass Transport Companies*, presented at the 6<sup>th</sup> International Workshop on Computer-Aided Scheduling of Public Transport, July 1993, Lisbon, Portugal.
- [3] C. J. Date, *An Introduction to Database Systems (6th Edition)*, Addison-Wesley, 1995
- [4] Stephen R. Gardner, *Building the Data Warehouse*, Communications of the ACM, September 1998
- [5] Hadden & Company, <http://www.hadden-kelly.com>, 1999
- [6] Inmon W.H., *Building the Data Warehouse*, Wiley, 1996, pp.33
- [7] Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite, *The Data Warehouse Lifecycle Toolkit*, Wiley, 1998, pp. 137-314
- [8] OLAP Council, <http://www.olapcouncil.org/research/glossaryly.htm>, 1997
- [9] John Onder, Todd Nash, *Building a business-driven Data Warehouse*, Data Management Review, October 1998
- [10] Nigel Pendse, Richard Creeth, *The OLAP Report*, <http://www.olapreport.com>, red in October 2000
- [11] Neil Raden, *Data Modeling – Star Schema 101*, <http://www.archerdecision.com>, 1995/96
- [12] João Ranito, seminar presentation at Faculdade de Engenharia da Universidade do Porto, 1999.

- [13] Jorge Pinho de Sousa, Jorge Freire de Sousa and Rui Campos Guimarães, *Un système informatique d'aide à la génération d'horaires de bus et de chauffeurs*, in: *Gestion de l'économie et de l'entreprise - l'approche quantitative*, Editions CORE, Série Balises, De Boeck, Brussels, 1988, pp. 477-492 (in French).
- [14] The Data Warehousing Institute, *1998 Data Warehousing Buyer's Guide*, <http://www.dwinsitute.com/buyersguide/>, 1998
- [15] Erik Thomsen, *OLAP Solutions*, Wiley, 1997, pp. 229-264
- [16] Karen Watterson, *Attention, Data-Mart shoppers*, BYTE Magazine - July 1997
- [17] J. D. Welch, <http://www.datawing.com/BOOKS.F.HTM>, 1998





## **Carta $\bar{X}$ com Amostras de Tamanho Variável: Um Novo Procedimento Dinâmico**

Paulo Infante \*

J. Rodrigues Dias \*

\* Departamento de Matemática  
Universidade de Évora  
{pinfante, jrd}@uevora.pt

---

### **Abstract**

In the last years a new class of control charts which parameters change during production was developed. In this paper, we present a new procedure that consists in a very simple and attractive method to vary the samples size. In this procedure, which is based in the idea presented by Rodrigues Dias (1999) to obtain different sampling intervals, the size of the following sample depends on the information in the last instant of inspection, assuming different values, in a distinct form that it happens when we use the technique usually known by VSS, which considers only two values.

Some statistical properties are studied using, in particular, Markov chains. Using a control chart for the sample mean, we present and analyse, in a critical way, some results obtained for different changes of the quality. A comparative study with some previously developed procedures is also presented.

### **Resumo**

Nos últimos anos foi desenvolvida uma nova classe de cartas de controlo cujos parâmetros variam durante a produção. Neste trabalho apresenta-se um novo procedimento que consiste num método muito simples e atractivo de variar o tamanho das amostras. Neste procedimento, que tem por base a ideia apresentada por Rodrigues Dias (1999) para obter intervalos de amostragem variáveis, o tamanho da amostra seguinte depende da informação obtida no instante anterior de inspecção, podendo assumir vários valores, de uma forma diferente do que acontece com a técnica usualmente conhecida por VSS, que considera apenas dois valores.

Recorrendo, em particular, às cadeias de Markov para estudar algumas propriedades, e utilizando uma carta de controlo para a média, apresentam-se e analisam-se, de uma forma crítica, alguns resultados obtidos para diferentes alterações da característica da qualidade. Efectua-se um estudo comparativo com outros procedimentos apresentados na literatura.

**Keywords:** Statistical quality control,  $\bar{X}$  control chart, variable sample sizes.

**Title:**  $\bar{X}$  chart with variable sample sizes: a new adaptive method.

---

## 1 Introdução

As cartas de controlo do tipo Shewhart, introduzidas por volta de 1930, continuam a ser uma das ferramentas mais utilizadas em controlo de qualidade. Na sua concepção e utilização é necessário ter em conta, para além da recolha de subgrupos racionais, três aspectos fundamentais: os instantes de amostragem, os tamanhos das amostras e os limites de controlo. Nas cartas Shewhart clássicas, os valores destas grandezas são fixos durante todo o processo de controlo. Em particular, usando uma carta de controlo para a média, utilizam-se usualmente os limites “3-sigma”, as amostras têm um tamanho que, embora dependa do processo, está habitualmente compreendido entre 4 e 9 e são recolhidas periodicamente para análise. Estas cartas são consideradas estáticas.

Convém, no entanto, referir que os seus parâmetros, apesar de não variarem durante o processo de controlo não são necessariamente constantes. Por exemplo, os instantes de inspecção podem não ser constantes, embora sejam determinados antes do início do controlo do processo. Neste contexto pode ver-se uma nova metodologia, apresentada em Rodrigues Dias (2002), na qual os instantes de amostragem são definidos com base na taxa cumulativa de risco do sistema. Esta metodologia é comparada com outras em Infante (2004) e em Rodrigues Dias e Infante (em preparação).

Nos últimos anos foi desenvolvida uma nova classe de cartas de controlo, designadas por cartas dinâmicas ou cartas adaptáveis. Nestas cartas, pelo menos um dos seus parâmetros varia durante o processo de controlo baseado nos valores da estatística amostral que fornecem informação actualizada acerca do estado do processo produtivo. A flexibilidade deste tipo de carta de controlo poderá resultar numa maior eficácia desta em detectar alterações da qualidade. Foram estes supostos benefícios que motivaram o aparecimento de um grande número de publicações a partir de Reynolds *et al.* (1988). Neste artigo os autores consideram uma carta de controlo para a média com intervalos de amostragem variáveis. O procedimento apresentado, usualmente conhecido por VSI (Variable Sampling Intervals), consiste em utilizar um intervalo de amostragem pequeno se a média amostral estiver próxima dos limites de controlo e um intervalo de amostragem grande se a média amostral estiver próxima da linha central.

A ideia de variar o tamanho das amostras numa carta de controlo para a média é intuitivamente a mesma. Em Prabhu *et al.* (1993) e Costa (1994) analisa-se a eficácia de cartas de controlo para a média com duas possíveis dimensões amostrais, introduzindo o procedimento conhecido por VSS (Variable Sample Sizes). Neste, utiliza-se um tamanho de amostra pequeno quando a média amostral cai junto aos limites de controlo, numa região designada de aviso, e uma amostra de maior

dimensão quando a média amostral cai na região central. Estas cartas de controlo são concebidas de modo a que o tamanho médio das amostras seja o mesmo que o tamanho constante das amostras de uma carta de controlo clássica. Zimmer *et al.* (1998) consideram uma terceira dimensão amostral, verificando que daí resulta apenas uma modesta melhoria na eficácia da carta.

Daudin (1992) propõe uma carta de controlo para a média com amostragem dupla. Neste procedimento, conhecido por DS (Double Sampling), extraem-se 2 amostras de tamanhos diferentes, sendo a segunda amostra apenas analisada se a primeira amostra não é suficiente para decidir se o processo está sob controlo. No caso de ser necessária uma segunda amostra, a decisão baseia-se na informação conjunta das duas amostras.

Outras ideias foram introduzidas e têm sido estudadas as propriedades estatísticas das cartas associadas. Prabhu *et al.* (1994) apresentam um procedimento que permite variar simultaneamente o tamanho dos intervalos de amostragem e a dimensão das amostras. Zimmer *et al.* (2000) introduziram uma quarta dimensão amostral numa carta de controlo com tamanho de amostras variáveis e analisaram diversas variações de um terceiro estado de uma carta de controlo com intervalos de amostragem e tamanhos de amostra variáveis. Costa (1999) analisa uma carta de controlo para a média com todos os três parâmetros variáveis. Um levantamento bibliográfico de artigos relacionados com esta abordagem dinâmica, que foram publicados até 1997, pode ser visto em Tagaras (1998). Em particular, refira-se que em Rodrigues Dias (1999 a, b) é apresentada e analisada uma nova metodologia, simples e interessante, recorrendo à função densidade de probabilidade da variável normal reduzida, para obter diferentes intervalos de amostragem. Esta metodologia foi estudada, em termos das suas propriedades estatísticas, e comparada com outros métodos adaptativos em Infante (2004) e em Infante e Rodrigues Dias (em preparação) e foi analisada, em termos da sua robustez perante desvios à normalidade, em Infante e Rodrigues Dias (2003). Finalmente, em Infante (2004) e em Infante e Rodrigues Dias (2004) apresenta-se e analisa-se um novo esquema de amostragem que combina a metodologia de intervalos predefinidos proposta em Rodrigues Dias (2002) com o método adaptativo de amostragem apresentado em Rodrigues Dias (1999a, b), o qual revela excelentes potencialidades.

Neste trabalho, começa-se por introduzir um novo procedimento dinâmico para definir o tamanho das amostras, estudando-se algumas das suas propriedades estatísticas. Depois, efectua-se um estudo comparativo com alguns outros métodos, em termos do número médio de amostras e do número médio de itens inspecionados. Analisa-se, finalmente, a sensibilidade deste método quando o tamanho das amostras é limitado superiormente. Para além das considerações finais, as conclusões vão sendo apresentadas ao longo do texto. Dada a dimensão atingida pelo artigo com as questões anteriores, não se apresenta qualquer exemplo numérico de aplicação, com base na optimização de uma dada função objectivo. Do mesmo modo, outras comparações com outras abordagens, apesar de importantes, foram omitidas (como é o caso das Somas Acumuladas CUSUM e das Médias Móveis Exponencialmente Amortecidas EWMA).

Em síntese, parece-nos que este estudo é revelador do bom desempenho deste

novo procedimento quando se pretendem detectar pequenas e moderadas alterações da média (definidas adiante).

## 2 Novo Procedimento Dinâmico para Definir o Tamanho das Amostras

Vamos admitir que a produção, estando sob controlo estatístico, segue uma distribuição normal com média  $\mu_0$  e desvio padrão  $\sigma_0$ . Por outro lado, admitimos que num determinado instante, como resultado da presença de uma causa assinalável, a média do processo se altera para  $\mu_1 = \mu_0 \pm \lambda\sigma_0$ ,  $\lambda = |\mu_1 - \mu_0|/\sigma_0$ .

Neste trabalho, considera-se que uma alteração da média do processo é pequena quando  $\lambda$  é menor ou igual que 1, é moderada quando é maior que 1 e menor ou igual que 2 e é grande nos outros casos.

Utilizando uma carta de controlo para a média, consideramos que o processo está fora de controlo quando a média amostral  $\bar{X}$  cair fora dos limites de controlo, dados por  $\mu_0 \pm L\sigma_0/\sqrt{n}$ .

Na técnica conhecida por VSS a escolha do tamanho das amostras é difícil e não teoricamente determinada. Este novo método proposto neste artigo, daqui em diante designado por método RDN, é muito simples, rápido e intuitivo, e tem por base a mesma ideia apresentada em Rodrigues Dias (1999a, b) para obter intervalos de amostragem diferentes.

Seja  $\bar{x}_i$  o valor médio da  $i$ -ésima amostra (dentro dos limites de controlo) e  $n_i^*$  a sua dimensão. O procedimento por nós proposto sugere que o tamanho da amostra seguinte seja dado por

$$n_{i+1}^* = \text{Int} \left[ \Theta \left( \frac{k}{\phi(u_i)} \right) \right] \quad (1)$$

em que :

$$u_i = \frac{\bar{x}_i - \mu_0}{\sigma_0/\sqrt{n_i^*}} \quad (2)$$

$$\bar{x}_0 = \mu_0, -L \leq u_i \leq L \quad (3)$$

$$n_1^* = \text{Int} \left[ \Theta \left( k\sqrt{2\pi} \right) \right] \quad (4)$$

$\phi(u)$  representa a função densidade de probabilidade da distribuição normal padronizada,

$L$  é o coeficiente dos limites de controlo,

$\text{Int}(x)$  representa o maior inteiro que não excede  $x$ ,

$\Theta(x)$  é uma função que se considere ser adequada e

$k$  uma constante conveniente.

Note-se, por um lado, que o valor da constante  $k$  dependerá, em particular, do custo de amostragem. Note-se, por outro lado, que para valores de  $|u_i| > L$  se está numa situação fora dos limites de controlo, podendo corresponder a um falso alarme (caso em que se pode convencionar, por exemplo, que o tamanho da amostra seguinte é igual à maior dimensão amostral possível).

Neste artigo, consideramos para  $\Theta(x)$  as funções  $\ln(x)$  e  $x^{1/2}$ , e calculamos o valor de  $k$  por forma a podermos comparar diferentes métodos. É claro que outras funções poderão e deverão ser consideradas em trabalhos futuros.

De acordo com este método, em função do valor da média da amostra de ordem  $i$ , facilmente se obtém o tamanho da amostra seguinte. Assim, quanto mais próxima estiver a média amostral da média inicial (linha central), menor será o tamanho das amostras e quanto mais se aproximar dos limites de controlo maior será o tamanho das amostras. Tal como no método VSS, no procedimento proposto o tamanho da amostra seguinte depende da informação obtida no instante actual de inspecção. Contudo, no método RDN é considerado um maior número de dimensões amostrais, sendo obtidas de uma forma perfeitamente definida.

### 3 Estudo de Algumas Propriedades Estatísticas

As propriedades de uma carta de controlo com um período fixo de inspecção são usualmente determinadas pelo número de amostras  $N_A$  e pelo número de itens  $N_I$  que são inspeccionados desde o instante em que o sistema falha até um ponto cair fora dos limites de controlo. Recorrendo às cadeias de Markov vamos procurar determinar o número médio de amostras e o número médio de itens. Em Prabhu *et al.* (1993) e Costa (1994), este tipo de abordagem é utilizada para determinar o número médio de amostras até detectar uma alteração para uma carta de controlo de médias usando o método VSS.

Começamos por dividir a região entre os limites de controlo em  $r$  sub-regiões  $R_1, R_2, \dots, R_r$ , tantas quanto o número possível de dimensões amostrais utilizando o método RDN. O tamanho da amostra de ordem  $i+1$  será igual a  $n_j$ , se a média da amostra de ordem  $i$ ,  $\bar{X}_i$ , pertencer à região  $R_j$ ,  $j=1, 2, \dots, r$ . Deste modo, em cada instante de inspecção, um de  $r$  estados transientes é atingido. O estado absorvente é atingido sempre que a média de uma amostra sai fora dos limites de controlo, estando o processo fora de controlo. A matriz de transição é dada por

$$P = \begin{bmatrix} p_{1,1}^\lambda & p_{1,2}^\lambda & \cdots & p_{1,r+1}^\lambda \\ p_{2,1}^\lambda & p_{2,2}^\lambda & \cdots & p_{2,r+1}^\lambda \\ \vdots & \vdots & \ddots & \vdots \\ p_{r+1,1}^\lambda & p_{r+1,2}^\lambda & \cdots & p_{r+1,r+1}^\lambda \end{bmatrix} \quad (5)$$

onde

$$P_{i,j}^\lambda = P(\bar{X}_i \in R_j | n_i, \lambda) \quad (6)$$

representa a probabilidade de passar do estado  $i$  para o estado  $j$ , isto é, a probabilidade de a próxima amostra ser de tamanho  $n_j$  sabendo que a actual tem dimensão  $n_i$  quando a média do processo se alterou  $\lambda$  desvios padrões.

Sendo  $N$  a variável aleatória que representa o tamanho de uma amostra analisada num determinado instante e  $\bar{X}$  a média dessa amostra, podemos escrever

$$N = n_j \Leftrightarrow n_j \leq \Theta \left( \frac{k}{\phi(u)} \right) < n_j + 1, j = 1, 2, \dots, r \quad (7)$$

pelo que, sendo  $U$  a variável normal padronizada,

$$N = n_j \Leftrightarrow U \in R_j^* \Leftrightarrow \bar{X} \in R_j, j = 1, 2, \dots, r \quad (8)$$

As sub-regiões  $R_j$ , bem como as probabilidades  $p_{i,j}^\lambda$  serão determinadas nos pontos seguintes para as duas funções  $\Theta(x)$  consideradas.

Pelas propriedades elementares das cadeias de Markov tem-se

$$E(N_A) = \vec{b}^T (I - Q_\lambda)^{-1} \vec{1} \quad (9)$$

$$E(N_I) = \vec{b}^T (I - Q_\lambda)^{-1} \vec{s} \quad (10)$$

onde  $I$  representa a matriz identidade de ordem  $r$ ,  $Q_\lambda$  representa a matriz de probabilidades de transição da qual foram retirados os elementos associados com o estado absorvente,  $\vec{1}$  é um vector coluna unitário  $r \times 1$ ,  $\vec{s}^T = (n_1, n_2, \dots, n_r)$  representa o vector dos tamanhos de amostras, onde  $n_1$  é a menor dimensão amostral obtida de acordo com o método, e onde  $\vec{b}^T = (b_1, b_2, \dots, b_r)$  representa o vector das probabilidades iniciais (com o processo sob controlo), usualmente dadas por

$$b_j = \frac{p_{jj}^0}{p_{j1}^0 + p_{j2}^0 + \dots + p_{jr}^0}, j=1, 2, \dots, r \quad (11)$$

### 3.1 Caso em que $\Theta(x)=\ln(x)$

Quando se considera para  $\Theta(x)$  o logaritmo neperiano tem-se:

$$R_1^* = \left[ -\sqrt{2(n_1 + 1 - \ln(k\sqrt{2\pi}))}, \sqrt{2(n_1 + 1 - \ln(k\sqrt{2\pi}))} \right] \quad (12)$$

$$R_j^* = \left[ -\sqrt{2(n_j + 1 - \ln(k\sqrt{2\pi}))}, -\sqrt{2(n_j - \ln(k\sqrt{2\pi}))} \right] \cup \left[ \sqrt{2(n_j - \ln(k\sqrt{2\pi}))}, \sqrt{2(n_j + 1 - \ln(k\sqrt{2\pi}))} \right] \quad j = 2, 3, \dots, r-1 \quad (13)$$

$$R_r^* = \left[ -L, -\sqrt{2(n_r - \ln(k\sqrt{2\pi}))} \right] \cup \left[ \sqrt{2(n_r - \ln(k\sqrt{2\pi}))}, L \right] \quad (14)$$

Admitindo que o tamanho da amostra extraída num determinado instante de inspecção é igual a  $n_i$ ,  $i=1, 2, \dots, r$ , e que essa amostra tem média  $\bar{X}$ , realizando algumas simplificações algébricas obtemos

$$R_1 = \left[ \mu_0 - \sigma_0 \sqrt{\frac{2(n_1 + 1 - \ln(k\sqrt{2\pi}))}{n_i}}, \mu_0 + \sigma_0 \sqrt{\frac{2(n_1 + 1 - \ln(k\sqrt{2\pi}))}{n_i}} \right] \quad (15)$$

$$R_j = \left[ \mu_0 - \sigma_0 \sqrt{\frac{2(n_j+1-\ln(k\sqrt{2\pi}))}{n_i}}, \mu_0 - \sigma_0 \sqrt{\frac{2(n_j-\ln(k\sqrt{2\pi}))}{n_i}} \right] \cup \left[ \mu_0 + \sigma_0 \sqrt{\frac{2(n_j-\ln(k\sqrt{2\pi}))}{n_i}}, \mu_0 + \sigma_0 \sqrt{\frac{2(n_j+1-\ln(k\sqrt{2\pi}))}{n_i}} \right] \quad j = 2, 3, \dots, r-1 \quad (16)$$

$$R_r = \left[ \mu_0 - \sigma_0 \frac{L}{\sqrt{n_i}}, \mu_0 - \sigma_0 \sqrt{\frac{2(n_r-\ln(k\sqrt{2\pi}))}{n_i}} \right] \cup \left[ \mu_0 + \sigma_0 \sqrt{\frac{2(n_r-\ln(k\sqrt{2\pi}))}{n_i}}, \mu_0 + \sigma_0 \frac{L}{\sqrt{n_i}} \right] \quad (17)$$

Podemos então facilmente obter as probabilidades de transição entre estados (i=1, 2, ..., r):

$$P_{i,1}^\lambda = P(\bar{X} \in R_1 | n_i, \lambda) = \Phi\left(\sqrt{2(n_1+1-\ln(k\sqrt{2\pi}))} - \lambda\sqrt{n_i}\right) - \Phi\left(-\sqrt{2(n_1+1-\ln(k\sqrt{2\pi}))} - \lambda\sqrt{n_i}\right) \quad (18)$$

$$P_{i,j}^\lambda = P(\bar{X} \in R_j | n_i, \lambda) = \Phi\left(-\sqrt{2(n_j-\ln(k\sqrt{2\pi}))} - \lambda\sqrt{n_i}\right) - \Phi\left(-\sqrt{2(n_j+1-\ln(k\sqrt{2\pi}))} - \lambda\sqrt{n_i}\right) + \Phi\left(\sqrt{2(n_j+1-\ln(k\sqrt{2\pi}))} - \lambda\sqrt{n_i}\right) - \Phi\left(\sqrt{2(n_j-\ln(k\sqrt{2\pi}))} - \lambda\sqrt{n_i}\right) \quad j = 2, 3, \dots, r-1 \quad (19)$$

$$P_{i,r}^\lambda = P(\bar{X} \in R_r | n_i, \lambda) = \Phi\left(-\sqrt{2(n_r-\ln(k\sqrt{2\pi}))} - \lambda\sqrt{n_i}\right) - \Phi(-L - \lambda\sqrt{n_i}) + \Phi(L - \lambda\sqrt{n_i}) - \Phi\left(\sqrt{2(n_r-\ln(k\sqrt{2\pi}))} - \lambda\sqrt{n_i}\right) \quad (20)$$

Note-se que as probabilidades de transição não dependem do tamanho da amostra no estado actual quando o processo se encontra sob controlo ( $\lambda=0$ ), o que seria de esperar, pois a probabilidade de um ponto sair fora dos limites sem ter ocorrido qualquer alteração (falso alarme) não depende do tamanho da amostra utilizado.

Representando por  $\alpha$  a probabilidade de ocorrer um falso alarme,

$$\alpha = 2(1 - \Phi(L)) \quad (21)$$

as probabilidades de se utilizar uma determinada dimensão amostral, estando o processo sob controlo, são dadas por:

$$P(N = n_1) = \frac{2}{1-\alpha} \left[ \Phi\left(\sqrt{2(n_1+1-\ln(k\sqrt{2\pi}))}\right) - 0.5 \right] \quad (22)$$

$$P(N = n_j) = \frac{2}{1-\alpha} \left[ \Phi\left(\sqrt{2(n_j+1-\ln(k\sqrt{2\pi}))}\right) - \Phi\left(\sqrt{2(n_j-\ln(k\sqrt{2\pi}))}\right) \right], \quad j = 2, \dots, r-1 \quad (23)$$

$$P(N = n_r) = \frac{2}{1 - \alpha} \left[ \Phi(L) - \Phi \left( \sqrt{2 \left( n_r - \ln(k\sqrt{2\pi}) \right)} \right) \right] \quad (24)$$

Desta forma, o tamanho médio das amostras utilizadas sob controlo é obtido pela fórmula

$$E(N) = \sum_{j=1}^r n_j P(N = n_j) \quad (25)$$

Para se obter o valor da constante k por forma a que  $E(N)=n$ , sendo n o tamanho da amostra utilizado no procedimento clássico, podemos utilizar para valor inicial de k o obtido através da aproximação

$$k \approx \exp \left[ n + 0.45 + \frac{L \exp(-L^2/2)}{\sqrt{2\pi}} - \frac{1 - \alpha}{2} - \frac{L^2 \cdot \alpha}{2} \right] / \sqrt{2\pi} \quad (26)$$

que se obtém, mediante algum tratamento algébrico, atendendo a que

$$E \left[ \text{Int} \left( \ln \left( \frac{k}{\phi(U)} \right) \right) \right] \approx \ln(k\sqrt{2\pi}) + \frac{E(U^2)}{2} - 0.45 \quad (27)$$

No caso em que  $n=5$  e se utilizam os usuais limites “3-sigma” tem-se para valor inicial  $k=56.46$ , obtendo-se  $n_1=4$  e  $n_r=9$ , pelo que através de (25) se obtém  $k=57.30$ . Repare-se que os tamanhos possíveis das amostras são bastante sugestivos, sendo valores usuais num controlo por variáveis. Na Tabela 1 apresentamos os valores da constante k e os tamanhos máximo  $n_r$  e mínimo  $n_1$  associados a valores de n entre 2 e 9.

Tabela 1: Valores de k,  $n_1$  e  $n_r$ , para diferentes valores de n.

<b>n</b>	<b>k</b>	<b><math>n_1</math></b>	<b><math>n_r</math></b>
2	2,853	1	6
3	7,755	2	7
4	21,080	3	8
5	57,302	4	9
6	155,762	5	10
7	423,406	6	11
8	1150,936	7	12
9	3128,569	8	13

### 3.2 Caso em que $\Theta(\mathbf{x})=\sqrt{x}$

Quando se considera para  $\Theta(x)$  a raiz quadrada tem-se:

$$R_1^* = \left[ -2 \sqrt{\ln \left( \frac{n_1 + 1}{k^{1/2} (2\pi)^{1/4}} \right)}, 2 \sqrt{\ln \left( \frac{n_1 + 1}{k^{1/2} (2\pi)^{1/4}} \right)} \right] \quad (28)$$



$$R_j^* = \left[ -2\sqrt{\ln\left(\frac{n_j+1}{k^{1/2}(2\pi)^{1/4}}\right)}, -2\sqrt{\ln\left(\frac{n_j}{k^{1/2}(2\pi)^{1/4}}\right)} \right] \cup \left[ 2\sqrt{\ln\left(\frac{n_j}{k^{1/2}(2\pi)^{1/4}}\right)}, 2\sqrt{\ln\left(\frac{n_j+1}{k^{1/2}(2\pi)^{1/4}}\right)} \right] \quad j = 2, 3, \dots, r-1 \quad (29)$$

$$R_r^* = \left[ -L, -2\sqrt{\ln\left(\frac{n_r}{k^{1/2}(2\pi)^{1/4}}\right)} \right] \cup \left[ 2\sqrt{\ln\left(\frac{n_r}{k^{1/2}(2\pi)^{1/4}}\right)}, L \right] \quad (30)$$

De forma análoga ao que se fez para o caso anterior, considerando que a amostra analisada num determinado instante tem dimensão  $n_j$  ( $j=1, 2, \dots, r$ ) e média  $\bar{X}$ , encontrando as sub-regiões  $R_j$  e depois de algum tratamento algébrico podemos obter as probabilidades de transição entre estados ( $i=1, 2, \dots, r$ ):

$$P_{i,1}^\lambda = P(\bar{X} \in R_1 | n_i, \lambda) = \Phi\left(2\sqrt{\ln\left(\frac{n_1+1}{k^{1/2}(2\pi)^{1/4}}\right)} - \lambda\sqrt{n_i}\right) - \Phi\left(-2\sqrt{\ln\left(\frac{n_1+1}{k^{1/2}(2\pi)^{1/4}}\right)} - \lambda\sqrt{n_i}\right) \quad (31)$$

$$P_{i,j}^\lambda = P(\bar{X} \in R_j | n_i, \lambda) = \Phi\left(-2\sqrt{\ln\left(\frac{n_j}{k^{1/2}(2\pi)^{1/4}}\right)} - \lambda\sqrt{n_i}\right) - \Phi\left(-2\sqrt{\ln\left(\frac{n_j+1}{k^{1/2}(2\pi)^{1/4}}\right)} - \lambda\sqrt{n_i}\right) + \Phi\left(2\sqrt{\ln\left(\frac{n_j+1}{k^{1/2}(2\pi)^{1/4}}\right)} - \lambda\sqrt{n_i}\right) - \Phi\left(2\sqrt{\ln\left(\frac{n_j}{k^{1/2}(2\pi)^{1/4}}\right)} - \lambda\sqrt{n_i}\right) \quad j = 2, 3, \dots, r-1 \quad (32)$$

$$P_{i,r}^\lambda = P(\bar{X} \in R_r | n_i, \lambda) = \Phi\left(-2\sqrt{\ln\left(\frac{n_r}{k^{1/2}(2\pi)^{1/4}}\right)} - \lambda\sqrt{n_i}\right) - \Phi(-L - \lambda\sqrt{n_i}) + \Phi(L - \lambda\sqrt{n_i}) - \Phi\left(2\sqrt{\ln\left(\frac{n_r}{k^{1/2}(2\pi)^{1/4}}\right)} - \lambda\sqrt{n_i}\right) \quad (33)$$

Tal como no caso anterior, pode observar-se que as probabilidades de transição não dependem do tamanho da amostra num dado instante quando o processo se encontra sob controlo.

As probabilidades de se utilizar uma determinada dimensão amostral, estando o processo sob controlo, são, para este caso, dadas por:

$$P(N = n_1) = \frac{2}{1-\alpha} \left[ \Phi\left(2\sqrt{\ln\left(\frac{n_1+1}{k^{1/2}(2\pi)^{1/4}}\right)}\right) - 0.5 \right] \quad (34)$$

$$P(N = n_j) = \frac{2}{1-\alpha} \left[ \Phi\left(2\sqrt{\ln\left(\frac{n_j+1}{k^{1/2}(2\pi)^{1/4}}\right)}\right) - \Phi\left(-2\sqrt{\ln\left(\frac{n_j}{k^{1/2}(2\pi)^{1/4}}\right)}\right) \right] \quad j = 2, 3, \dots, r-1 \quad (35)$$

$$P(N = n_r) = \frac{2}{1-\alpha} \left[ \Phi(L) - \Phi\left(2\sqrt{\ln\left(\frac{n_r}{k^{1/2}(2\pi)^{1/4}}\right)}\right) \right] \quad (36)$$

Tal como no caso anterior, obtemos o valor da constante  $k$  por forma a que  $E(N)=n$ , sendo  $n$  o tamanho da amostra utilizado no procedimento estático. Neste caso,

torna-se mais complicado encontrar uma fórmula de aproximação para obter um valor inicial de  $k$ , atendendo a que a esperança matemática de  $\text{Int}(k/\phi(u))$  apresenta uma grande variabilidade. Consequentemente, procedemos a uma discretização dos valores de  $U$  entre  $-L$  e  $L$ , obtendo os correspondentes valores de  $N$  e as probabilidades associadas. Pode assim obter-se um valor de  $k$  mais fiável. Note-se que este procedimento pode também ser usado no caso anterior.

No caso em que  $n=5$  e se utilizam os usuais limites “3-sigma” tem-se  $k=6.309$ , obtendo-se  $n_1=3$  e  $n_r=37$ . Repare-se que os tamanhos possíveis das amostras não são tão sugestivos quanto os do caso anterior quando comparados com os tamanhos usuais das cartas Shewhart para variáveis. Contudo, podem ser valores razoáveis num procedimento em que varia apenas o tamanho da amostra, apresentando resultados muito bons quando a média sofre pequenas alterações, conforme veremos no ponto seguinte. Por outro lado, pode sempre fixar-se para maior tamanho das amostras um valor inferior, podendo os resultados globais não sofrer grandes alterações. Mais adiante analisaremos a sensibilidade deste método quando se impõe um limite superior para o tamanho das amostras. Mais concretamente, veremos qual a diferença em se considerar, neste caso concreto,  $n_r=25$ ,  $n_r=20$  e  $n_r=15$  em vez de  $n_r=37$ .

Na Tabela 2 apresentamos os valores da constante  $k$  e os tamanhos máximo  $n_r$  e mínimo  $n_1$  associados a valores de  $n$  entre 2 e 9.

Tabela 2: Valores de  $k$ ,  $n_1$  e  $n_r$ , para diferentes valores de  $n$ .

$N_0$	$k$	$n_1$	$n_r$
2	1,405	1	17
3	2,730	2	24
4	4,217	3	30
5	6,309	3	37
6	9,158	4	45
7	12,086	5	52
8	15,118	6	58
9	19,189	6	65

Tabela 2- Valores de  $k$ ,  $n_1$  e  $n_r$ , para diferentes valores de  $n$ .

## 4 Comparação com Outros Procedimentos

Neste ponto vamos comparar este novo procedimento usando uma carta de controlo para a média, com o método clássico de Shewhart, bem como com os procedimentos dinâmicos VSS e DS.

No caso da carta de médias clássica, com o tamanho das amostras constante e igual a  $n$ , as expressões para o número médio de amostras desde a falha até detecção  $E(N_A)$  e para o número médio de itens inspeccionados desde a falha até detecção  $E(N_I)$ , são facilmente obtidas, pois, neste caso,  $N_a$  segue uma distribuição

geométrica. Tem-se

$$E(N_A) = 1/p \quad (37)$$

$$E(N_I) = n/p \quad (38)$$

onde  $p$  é a probabilidade da média de uma amostra estar fora dos limites de controlo, sendo dada por:

$$p = 1 - \Phi(L - \lambda\sqrt{n}) + \Phi(-L - \lambda\sqrt{n}) \quad (39)$$

Em Prabhu *et al.* (1993) e em Costa (1994) são analisadas as propriedades e a performance do método VSS com dois tamanhos possíveis de amostras, que denotaremos por  $n_1$  e  $n_2$ , com  $n_1 < n < n_2$ . A região entre os limites de controlo é dividida em duas sub-regiões

$$R_1^* = ]-W, W] \quad (40)$$

$$R_2^* = ]-L, -W] \cup [W, L[ \quad (41)$$

onde  $0 < W < L$ . Repare-se que a sub-região  $R_1^*$  corresponde a valores centrais, enquanto  $R_2^*$  corresponde a valores mais próximos dos limites de controlo. Caso a média reduzida de uma amostra pertença a  $R_1^*$  então a próxima amostra terá tamanho  $n_1$  e caso pertença a  $R_2^*$  a próxima amostra terá tamanho  $n_2$ . O coeficiente  $W$  é obtido, para um dado valor de  $L$ , de modo a que o tamanho médio das amostras quando o processo está sob controlo seja igual a um valor específico  $n$ . Desta forma, tem-se

$$W = \Phi^{-1} \left[ \frac{2\Phi(L)(n_2 - n) + n - n_1}{2(n_2 - n_1)} \right] \quad (42)$$

Costa (1994) obteve, recorrendo às cadeias de Markov, expressões para  $E(N_A)$  e para  $E(N_I)$ , sendo dadas por

$$E(N_A) = p_1 \left[ \frac{1 - p_{22} + p_{12}}{(1 - p_{11})(1 - p_{22}) - p_{12}p_{21}} \right] + (1 - p_1) \left[ \frac{1 - p_{11} + p_{21}}{(1 - p_{11})(1 - p_{22}) - p_{12}p_{21}} \right] \quad (43)$$

$$E(N_I) = n_1 \left[ \frac{p_1(1 - p_{22}) + (1 - p_1)p_{21}}{(1 - p_{11})(1 - p_{22}) - p_{12}p_{21}} \right] + n_2 \left[ \frac{p_1p_{12} + (1 - p_1)(1 - p_{11})}{(1 - p_{11})(1 - p_{22}) - p_{12}p_{21}} \right] \quad (44)$$

onde

$$p_1 = \frac{2\Phi(W) - 1}{2\Phi(L) - 1} \quad (45)$$

$$p_{i1} = \Phi(W - \lambda\sqrt{n_i}) - \Phi(-W - \lambda\sqrt{n_i}), i = 1, 2 \quad (46)$$

$$p_{i2} = \Phi(-W - \lambda\sqrt{n_i}) - \Phi(-L - \lambda\sqrt{n_i}) + \Phi(L - \lambda\sqrt{n_i}) - \Phi(W - \lambda\sqrt{n_i}), i = 1, 2 \quad (47)$$

Para podermos comparar cartas de controlo com procedimentos diferentes, consideram-se nas mesmas condições numa situação de controlo, isto é, com o mesmo número médio de falsos alarmes, mesmo número médio de amostras analisadas e mesmo número médio de itens inspeccionados. Se o valor do coeficiente dos limites de controlo  $L$  e o intervalo de amostragem forem iguais em ambas as cartas, podemos afirmar que o número médio de falsos alarmes e o número médio de amostras é igual. Sendo igual o tamanho médio das amostras, então em ambas as cartas inspeccionam-se o mesmo número médio de itens quando o processo está sob controlo. Neste trabalho consideramos  $L=3$  (limites “3-sigma”) nos procedimentos clássico, VSS e RDN. Por outro lado, obtivemos  $W$  (no método VSS) e  $k$  (no método

RDN) de modo a que o tamanho médio de amostras fosse igual ao tamanho constante,  $n$ , das amostras do procedimento clássico.

Daudin (1992), por analogia com os planos de amostragem dupla, propôs uma carta de controlo para a média com amostragem dupla (DS), onde duas amostras de tamanhos  $n_1$  e  $n_2$  são retiradas periodicamente do processo produtivo, mas a segunda amostra (de maior dimensão) apenas é analisada quando a primeira não é suficiente para decidir sob o estado do processo. Este procedimento utiliza um coeficiente  $W$  para os limites de aviso e dois coeficientes  $L_1$  e  $L_2$  para os limites de controlo.

Designando por  $U_1$  a média padronizada da primeira amostra, isto é,

$$U_1 = \frac{(\bar{X}_1 - \mu_0) \sqrt{n_1}}{\sigma_0} \quad (48)$$

e por  $U_2$  a média padronizada da média global das duas amostras, dada por

$$U_2 = \frac{\sqrt{n_1 + n_2}}{\sigma_0} \left( \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2} - \mu_0 \right) \quad (49)$$

tem-se o seguinte conjunto de regras:

1. Caso  $|U_1| \leq W$ , conclui-se que o processo está sob controlo e não se inspecciona a segunda amostra;
2. Caso  $|U_1| > L_1$ , conclui-se que ocorreu uma alteração;
3. Caso  $W \leq |U_1| < L_1$ , então a segunda amostra é imediatamente analisada:
  - (a) Caso  $|U_2| \leq L_2$ , conclui-se que o processo está sob controlo;
  - (b) Caso  $|U_2| > L_2$ , conclui-se que ocorreu uma alteração.

No referido artigo são apresentadas expressões para o tamanho médio das amostras e para o número médio de amostras inspeccionadas, embora neste último caso seja necessário recorrer a integração numérica. As referidas expressões são dadas por:

$$E(N) = n_1 + n_2 [\Phi(L_1 + \lambda\sqrt{n_1}) - \Phi(W + \lambda\sqrt{n_1}) + \Phi(-W + \lambda\sqrt{n_1}) - \Phi(-L_1 + \lambda\sqrt{n_1})] \quad (50)$$

$$E(N_A) = 1/p \quad (51)$$

com

$$p = 1 - \Phi(W + \lambda\sqrt{n_1}) + \Phi(-W + \lambda\sqrt{n_1}) - \int_{z \in I} \left\{ \Phi \left[ \frac{1}{\sqrt{n_2}} (\sqrt{n_1 + n_2} L_2 + \lambda(n_1 + n_2) - \sqrt{n_1} z) \right] \phi(z) \right\} dz + \int_{z \in I} \left\{ \left[ \frac{1}{\sqrt{n_2}} (-\sqrt{n_1 + n_2} L_2 + \lambda(n_1 + n_2) - \sqrt{n_1} z) \right] \phi(z) \right\} dz \quad (52)$$

onde

$$I = [-L_1 + \lambda\sqrt{n_1}, -W + \lambda\sqrt{n_1}] \cup [W + \lambda\sqrt{n_1}, L_1 + \lambda\sqrt{n_1}] \quad (53)$$

A expressão para o número médio de itens inspeccionados obtém-se pela identidade de Wald

$$E(N_I) = E(N_A) E(N) \quad (54)$$

Para efeitos de comparação, nas mesmas condições, sob controlo, admitindo um intervalo de amostragem igual ao dos outros procedimentos, os parâmetros  $W$ ,  $L_1$ ,  $L_2$ ,  $n_1$  e  $n_2$  devem ser escolhidos por forma a que o número médio de falsos alarmes e número médio de itens inspeccionados sejam iguais para todos os procedimentos. No referido artigo, são apresentadas diferentes combinações de valores para estes parâmetros. Para o caso em que  $n_1=3$ ,  $n_2=6$ ,  $W=1$ ,  $L_1=3.51$  e  $L_2=3$ , obtém-se  $E(N)=4.9$ , tendo a carta o mesmo número médio de falsos alarmes que uma carta usual com limites “3-sigma” e aproximadamente o mesmo número médio de itens inspeccionados que o procedimento clássico com  $n=5$ .

#### 4.1 Apresentação e Análise de Resultados

Considere-se  $E(N_A)^*$  o número médio de amostras analisadas desde a falha até que um alerta seja emitido pela carta de controlo no caso de se utilizar o método estático (clássico), e seja  $E(N_A)^{**}$  o número médio de amostras analisadas desde a falha até que um alerta seja emitido pela carta de controlo, associado a um dado esquema de amostragem dinâmico. Para compararmos os dois procedimentos, em termos dos respectivos valores de  $E(N_A)$ , podemos considerar

$$Q_1 = \frac{E(N_A)^* - E(N_A)^{**}}{E(N_A)^*} \times 100\% \quad (55)$$

podendo, então, dizer que  $Q_1$  representa a redução relativa, em %, do número médio de amostras analisadas quando se utiliza um determinado procedimento dinâmico em vez do método clássico.

Analogamente, considere-se  $E(N_I)^*$  o número médio de itens inspeccionados no caso de se utilizar o método clássico e seja  $E(N_I)^{**}$  o número médio de itens inspeccionados associado a um dado esquema de amostragem dinâmico. Assim, para compararmos os dois métodos, em termos dos respectivos valores de  $E(N_I)$ , podemos considerar

$$Q_2 = \frac{E(N_I)^* - E(N_I)^{**}}{E(N_I)^*} \times 100\% \quad (56)$$

e podemos dizer que  $Q_2$  representa a redução relativa, em %, do número médio de itens inspeccionados quando se utiliza um determinado procedimento dinâmico em vez do método clássico.

Nas Tabelas 3 e 4 apresentam-se alguns resultados obtidos para  $Q_1$  e  $Q_2$  considerando diferentes alterações da média da qualidade, associadas a diferentes valores de  $\lambda$ , usando as duas versões do novo método, bem como os métodos VSS e DS. No caso do método VSS considerámos 2 combinações possíveis de tamanhos amostrais. Assim, no esquema designado por VSS(a) tem-se  $n_1=2$  e  $n_2=25$  e no esquema designado por VSS(b) tem-se  $n_1=3$  e  $n_2=15$ . Por um lado, estamos a usar valores habitualmente considerados na literatura para este esquema de amostragem (que nos parecem razoáveis na prática) e, por outro lado, valores que permitem um

bom desempenho do método para diferentes alterações da média. Embora seja possível obter para cada alteração considerada uma combinação que minimize o número médio de amostras até detecção, como se pode ver em Prabhu *et al.* (1993) e em Costa (1994), a sua aplicação prática por vezes não parece viável. Refira-se também que a carta pode ter um bom desempenho para uma dada alteração com um dado par de valores  $(n_1, n_2)$ , mas ter um mau desempenho com o mesmo par de valores para outra alteração. Refira-se, por fim, que no método DS, utilizamos os valores dos parâmetros referidos no ponto anterior.

Da observação destas tabelas podemos tecer algumas considerações:

Tabela 3: Valores de  $Q_1$  para diferentes valores de  $\lambda$ , no caso em que  $n=5$  e  $L=3$ .

$\lambda$	<b>RDN</b> $\Theta(\mathbf{x})=\ln(\mathbf{x})$	<b>RDN</b> $\Theta(\mathbf{x})=\mathbf{x}^{1/2}$	<b>VSS(a)</b> $(n_1, n_2)=(2, 25)$ <b>W=1.5032</b>	<b>VSS (b)</b> $(n_1, n_2)=(3, 15)$ <b>W=1.3757</b>	<b>DS</b>
0,2	1,1	6,30	20,0	9,4	28,1
0,4	7,7	39,6	66,1	45,2	45,9
0,6	18,3	61,9	72,6	63,6	53,1
0,8	26,0	60,2	60,0	60,9	54,6
1,0	26,4	47,6	38,1	47,0	51,5
1,2	20,4	30,5	11,3	26,9	44,5
1,4	12,0	13,7	-14,6	5,8	34,5
1,6	4,9	1,3	-33,7	-10,3	23,3
1,8	0,8	-4,7	-43,2	-18,1	13,3
2,0	-0,6	-5,8	-43,6	-18,3	6,3
2,5	-0,3	-1,9	-25,7	-7,1	0,4
3,0	0,0	-0,2	-9,3	-1,2	0,0

1. O novo procedimento dinâmico conduz, em termos globais, a resultados bastante interessantes quanto à redução do número médio de amostras inspeccionadas até detectar uma alteração pequena da qualidade, e também para alguns casos de uma alteração moderada, relativamente ao método clássico, e, consequentemente, uma redução do tempo médio de mau funcionamento. São também interessantes os resultados relativos à redução do número médio de itens inspeccionados quando a alteração é pequena.
2. De um modo geral, para alterações pequenas e também em alguns casos de alterações moderadas, verifica-se que a utilização da raiz quadrada conduz a uma maior eficácia da carta do que aquela que se obtém usando o logaritmo. Repare-se que, para além desta maior eficácia, se consegue também uma redução do número médio de itens inspeccionados quando a alteração é pequena. Para alterações grandes e em certos casos de alterações moderadas, o método clássico tem um desempenho relativamente melhor, sendo também melhor em termos do número médio de itens inspeccionados quando  $\lambda > 1$ .
3. Em termos globais, poder-se-á concluir que o método DS é, entre os que foram analisados, o mais eficiente, pois para além de revelar, no que concerne ao

número médio de amostras, um melhor desempenho na detecção de alterações da média, também é comparativamente melhor quanto ao número médio de itens a inspeccionar. No entanto, em certos casos, como se pode verificar nos quadros em análise, o método DS é pior que outros. Em particular, o método RDN é mais rápido na detecção das alterações  $\lambda=0.6$  e  $\lambda=0.8$  quando se utiliza a raiz quadrada.

4. Os esquemas VSS são claramente menos eficientes que o novo método com qualquer uma das funções  $\Theta(x)$  consideradas quando  $\lambda > 1.2$ , o mesmo acontecendo no método RDN no caso da raiz quadrada para valores de  $\lambda$  a partir de .8. Para pequenas alterações da média, os esquemas VSS são preferíveis.

Tabela 4: Valores de  $Q_2$  para diferentes valores de  $\lambda$ , no caso em que  $n=5$  e  $L=3$ .

$\lambda$	<b>RDN</b> $\Theta(x)=\ln(x)$	<b>RDN</b> $\Theta(x)=x^{1/2}$	<b>VSS(a)</b> $(n_1, n_2)=(2, 25)$ <b>W=1.5032</b>	<b>VSS (b)</b> $(n_1, n_2)=(3, 15)$ <b>W=1.3757</b>	<b>DS</b>
0,2	-0,6	-0,8	0,0	0,2	27,1
0,4	1,6	18,6	5,4	21,9	40,1
0,6	6,9	33,7	40,5	35,9	42,1
0,8	10,8	24,2	47,0	31,0	37,7
1,0	9,6	-0,2	26,9	11,0	28,2
1,2	3,9	-30,0	-18,4	-20,1	14,5
1,4	-2,5	-55,3	-83,5	-53,1	-0,9
1,6	-6,4	-66,6	-150,0	-74,6	-13,4
1,8	-6,7	-61,4	-198,6	-77,6	-18,2
2,0	-4,8	-45,7	-217,7	-65,5	-13,7
2,5	-0,8	-10,8	-208,0	-22,1	15,7
3,0	0,0	-1,2	-121,9	-3,5	34,5

A terminar, refira-se que quer em Prabhu *et al.* (1993), quer em Costa (1994) se questiona se a vantagem do método DS, em diferentes situações, compensa as dificuldades de administração. Daudin (1992) refere que a questão prática reside em saber se a melhoria na eficiência compensa o problema e o custo da sua administração. Prabhu *et al.* (1993) apontam mesmo um conjunto de vantagens do método VSS sobre o método DS.

Por outro lado, o facto da 2<sup>a</sup> amostra ser contígua à primeira pode ser uma dificuldade acrescida de implementação. De facto, apenas em alguns processos poderá ser possível a implementação do método DS, processos em que seja desprezável o tempo de recolha e análise de cada amostra. Além disso, quando se concebe uma carta de controlo a ser utilizada com o método DS existe uma grande variedade de alternativas na escolha dos parâmetros que necessitam ser especificados. Nos esquemas VSS e RDN os limites de controlo são fixos, o que pensamos corresponder a uma maior simplicidade. Daudin (1992) refere que o facto do número de amostras a inspeccionar não ser fixo pode ser uma dificuldade de administração, bem como o facto do processo de decisão ser complexo.

#### 4.2 Caso em que a maior dimensão amostral é predefinida

Como se verificou nos valores apresentados na Tabela 2, este novo procedimento dinâmico, na sua versão “raiz quadrada” pode implicar a utilização de tamanhos de amostras elevados tendo como referência os valores usuais do método clássico. Ora, pode no intervalo de tempo entre duas amostras consecutivas ser fisicamente impossível ou administrativamente muito difícil recolher e analisar amostras de tal dimensão. Neste caso, ainda é possível a utilização do método limitando à partida o tamanho da maior amostra.

Sendo, no método RDN,  $\Theta(x)=\sqrt{x}$ , para analisar em que medida é que os resultados anteriores podem ser afectados, considerámos, em vez de  $n_r=37$ , as situações em que é apenas possível utilizar  $n_r^*=25$ ,  $n_r^*=20$  e  $n_r^*=15$ . Os valores da constante  $k$  passam a ser iguais a 6.325, 6.346 e 6.377, respectivamente, em vez do valor 6.309 obtido no caso em que  $n_r=37$ , de modo a que se continue a ter  $E(N)=n=5$ .

Representemos, então, por  $E(N_a)_1$  e por  $E(N_I)_1$ , respectivamente, o número médio de amostras analisadas e o número médio de itens inspeccionados, adoptando o método RDN, quando se limita a maior dimensão amostral; analogamente, no caso em que tal limitação não é imposta, consideremos  $E(N_a)_0$  e  $E(N_I)_0$ . Sendo assim, as respectivas variações relativas, em percentagem, podem ser dadas pelas grandezas

$$Q_3 = \frac{E(N_a)_0 - E(N_a)_1}{E(N_a)_0} \cdot 100\% \quad (57)$$

$$Q_4 = \frac{E(N_I)_0 - E(N_I)_1}{E(N_I)_0} \cdot 100\% \quad (58)$$

Na Tabela 5 e na Tabela 6 (onde  $n_r^*$  representa o limite superior imposto à dimensão amostral), apresentam-se os valores de  $Q_3$  e de  $Q_4$  para diferentes valores de  $\lambda$  associados a diferentes alterações no nível médio da qualidade. Da observação destas Tabelas, podemos tirar as seguintes conclusões:

Tabela 5: Valores de  $Q_3$  para diferentes valores de  $\lambda$ , com  $L=3$  e  $n=5$ .

$\lambda$	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.5	3.0
$n_r^* = 25$	0,5	3,9	3,6	0,7	-0,1	-0,2	-0,2	-0,2	-0,2	-0,2	-0,1	0,0
$n_r^* = 20$	1,1	9,1	10,4	3,0	0,2	-0,4	-0,5	-0,6	-0,6	-0,6	-0,3	0,0
$n_r^* = 15$	2,0	18,2	26,0	10,9	2,5	-0,3	-1,1	-1,3	-1,4	-1,3	-0,6	-0,1

1. No que concerne ao número médio de amostras, a eficácia da carta de controlo para a média mantém-se sensivelmente igual quando se limita a maior dimensão amostral para alterações de magnitude  $\lambda > 1$ . No caso em que  $\lambda < 1$ , verificam-se algumas reduções relevantes no número médio de amostras analisadas, embora no caso em que  $n_r=25$  (que corresponde a uma redução da maior dimensão amostral superior a 30%), estas reduções sejam inferiores a 4%, o que é bem elucidativo da pequena sensibilidade do método quando se limita a dimensão da maior amostra. Apenas no caso em que  $n_r=20$ , quando



$\lambda=0.6$ , e no caso em que  $n_r=15$ , quando  $0.4 \leq \lambda \leq 0.8$ , se obtiveram reduções superiores a 10%.

2. Quando se limita o tamanho da maior amostra, o número médio de itens inspeccionados sofre um aumento, por vezes bastante considerável, para alterações da média de magnitude  $\lambda \geq 0.8$ . Este aumento acentua-se com a diminuição do tamanho da maior amostra. Para alterações menores da média registam-se reduções do número médio de itens inspeccionados, embora apenas em três das situações consideradas sejam superiores a 5%.
3. Da conjugação das duas alíneas anteriores, podemos concluir que o método RDN é globalmente robusto quando se limita o tamanho da maior amostra, podendo mesmo aumentar a sua eficiência.

Tabela 6: Valores de  $Q_4$  para diferentes valores de  $\lambda$ , com  $L=3$  e  $n=5$ .

$\lambda$	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.5	3.0
$n_r^* = 25$	0,2	2,4	1,3	-2,5	-4,1	-4,6	-4,8	-4,8	-4,7	-4,1	-1,6	-0,3
$n_r^* = 20$	0,3	5,2	4,3	-3,8	-8,2	-9,7	-10,2	-10,3	-9,8	-8,5	-3,2	-0,5
$n_r^* = 15$	0,4	9,6	11,2	-3,0	-12,9	-17,1	-18,4	-18,5	-17,4	-14,9	-5,5	-0,8

## 5 Considerações Finais

Neste trabalho apresentámos um novo procedimento dinâmico para definir o tamanho das amostras em controlo de qualidade. Foram sugeridas duas versões deste procedimento, uma que conduz a dimensões amostrais usuais e outra que conduz a analisar algumas amostras de maior a dimensão. Este procedimento é de implementação muito simples, não requerendo a adição de novos limites à carta Shewhart clássica nem qualquer formação especial por parte do operador. As amostras são retiradas periodicamente, ficando perfeitamente definido o tamanho da amostra seguinte a partir do valor médio da amostra actual.

Com base numa abordagem markoviana, foi possível estabelecer algumas propriedades estatísticas que permitiram a comparação com outros métodos dinâmicos. Os resultados obtidos são bastante interessantes quando se pretendem detectar alterações pequenas e moderadas da média. Por outro lado, podemos concluir que o método é robusto quando se limita o tamanho da maior amostra a inspeccionar, pois tal não afecta de uma forma muito significativa o seu desempenho. Pensamos que este método oferece uma alternativa em determinadas situações, tomando em linha de conta a comparação de desempenhos estatísticos com outros métodos e os requisitos necessários a cada um.

Refira-se que o caso em que se considerou a função logaritmo, pelos tamanhos de amostra obtidos, poderá ser perfeitamente adaptável a um esquema combinado com intervalos variáveis. Por outro lado, pensamos que será possível estender a

ideia a outras cartas de controlo, bem como obter o valor da constante  $k$  de modo a minimizar um custo total. Finalmente, refira-se que outras funções  $\Theta(x)$  poderão e deverão ser consideradas na tentativa de aumentar o desempenho estatístico-económico do método.

### **Agradecimento:**

Os Autores agradecem os úteis comentários feitos por um Referee, que contribuíram para a melhoria do artigo e, em particular, para uma melhor explicitação e compreensão de algumas conclusões.

## **6 Referências**

- Costa, A.,  $\bar{X}$  Chart with Variable Sample Size, *Journal of Quality Technology*, Vol. 26, 1994, pp. 155-163.
- Costa, A.,  $\bar{X}$  Chart with Variable Sample Size and Sampling Intervals, *Journal of Quality Technology*, Vol. 29, 1997, pp. 197-204.
- Costa, A.,  $\bar{X}$  Charts with Variable Parameters, *Journal of Quality Technology*, Vol. 31, 1999, pp. 408-416.
- Daudin, J. J., Double Sampling  $\bar{X}$  Charts, *Journal of Quality Technology*, Vol. 24, 1992, pp. 78-87.
- Infante, P. - *Métodos de Amostragem em Controlo de Qualidade*, Tese de Doutoramento, Universidade de Évora, 2004.
- Infante, P.; Rodrigues Dias, J., Robustez de um Novo Método Dinâmico de Amostragem em Controlo de Qualidade, em *Literacia e Estatística*, editado por P. Brito, A. Figueiredo, F. Sousa, P. Teles e F. Rosado, 2003, pp. 345-360.
- Infante, P.; Rodrigues Dias, J. - Esquema Combinado de Amostragem em Controlo de Qualidade com Intervalos Predefinidos Adaptáveis, em *Estatística com Acaso e Necessidade*, editado por P. Rodrigues, E. Rebelo e F. Rosado, 2004, pp. 335-347.
- Infante, P.; Rodrigues Dias, J., A Comparative Study Between a New Sampling Method and Others Adaptive Sampling Schemes, em preparação.
- Prabhu, S. S.; Runger, G. C.; Keats, J. B., An Adaptive Sample Size  $\bar{X}$ , *International Journal of Production Research*, Vol. 31, 1993, pp. 2895-2909.
- Prabhu, S. S.; Montgomery, D. C.; Runger, G. C., A Combined Adaptive Sample Size and Sampling Interval  $\bar{X}$  Control Scheme, *Journal of Quality Technology*, Vol. 27, 1994, pp. 74-83.
- Reynolds, M. R. Jr.; Amin, R. W. ; Arnold, J. C. ; Nachlas, Joel A.,  $\bar{X}$  Charts with Variable Sampling Intervals, *Technometrics*, Vol. 30, 1988, pp. 181-192.
- Rodrigues Dias, J., Analysis of a New Method to Obtain Different Sampling Intervals in Statistical Quality Control. *Proceedings of the IV Congreso Galego de Estatística e Investigación de Operacións*, Universidade de Santiago de Compostela, 1999a, pp. 155-158.
- Rodrigues Dias, J., A New Method to Obtain Different Sampling Intervals in Statistical Quality Control, 1999b, Universidade de Évora, 18 pp.
- Rodrigues Dias, J., Amostragem em Controlo de Qualidade com Intervalos Diferentes Predefinidos: Uma Nova Abordagem, 2002, *Joclad 2002*, Lisboa.

Rodrigues Dias, J.; Infante, P., Control Charts with Predetermined Sampling Intervals, em preparação.

Tagaras, G., A Survey of Recent Developments in the Design of Adaptive Control Charts, *Journal of Quality Technology*, Vol. 30, 1998, pp. 212-231.

Zimmer, L. S.; Montgomery, D.C.; Runger, G. C., A three-state adaptive sample size  $\bar{X}$  control chart, *International Journal of Production Research*, Vol. 36, 1998, pp. 733-743.

Zimmer, L. S.; Montgomery, D.C.; Runger, G. C., Guidelines for the application of adaptive control charting schemes, *International Journal of Production Research*, Vol. 38, 2000, pp. 1977-1992.



## **Fronteira DEA Difusa na Avaliação de Eficiência em Agricultura**

Eliane Gonçalves Gomes \*

João Carlos Correia Baptista Soares de Mello †

João Alfredo de Carvalho Mangabeira ‡

\* Empresa Brasileira de Pesquisa Agropecuária (Embrapa) – SGE  
Parque Estação Biológica, W3 Norte final, Asa Norte,  
70770-901, Brasília, DF, Brasil  
eliane.gomes@embrapa.br

† Universidade Federal Fluminense – Departamento de Engenharia de Produção  
Rua Passo da Pátria 156, São Domingos,  
24210-240, Niterói, RJ, Brasil  
jcsmello@producao.uff.br

‡ Embrapa Monitoramento por Satélite  
Av. Dr. Júlio Soares de Arruda 803, Parque São Quirino,  
13088-300, Campinas, SP, Brasil  
manga@cnpm.embrapa.br

---

### **Abstract**

This paper uses Data Envelopment Analysis (DEA) to measure the efficiency of some agricultural producers from the Holambra district (São Paulo State, Brazil). The total area of each property is one of the variables of the model, with two possible values: the value declared by the producer and the value measured by using IKONOS II satellite images. This variable presents uncertainty in its measurement and can be represented as interval data. The efficiency frontier is constructed considering the limits of uncertainty (interval limits), that is, the smallest and greatest possible values to be assumed for the imprecise variable. In this way, a region is constructed in relation to which the DMUs have a certain membership degree. The theoretical model, called fuzzy DEA model, is presented, as well as its application to the case of the evaluation of the efficiency of the agricultural producers of Holambra.

### **Resumo**

Este artigo usa Análise de Envolvência de Dados (DEA) para medir a eficiência de alguns agricultores do município de Holambra (estado de São Paulo, Brasil).

A área total de cada propriedade é uma das variáveis do modelo. Entretanto, esta variável apresenta dois possíveis valores: o valor declarado pelo produtor e o valor medido com o uso de imagens do satélite IKONOS (de alta resolução espacial). Desta forma, como essa variável apresenta incerteza na medição, pode ser representada como um dado intervalar. Assim, constrói-se uma região em relação à qual as DMUs possuem um certo grau de pertença. O modelo teórico proposto é apresentado, chamado de modelo DEA de fronteira difusa, assim como os resultados de sua aplicação ao caso da avaliação eficiência dos agricultores de Holambra.

**Keywords:** Data envelopment analysis; Agriculture; Interval data; Fuzzy frontier; Satellite images.

**Title:** Fuzzy DEA Frontier in Agricultural Efficiency Evaluation

---

## 1 Introdução

Avaliar a eficiência de unidades produtivas tem importância tanto para fins estratégicos, quanto para o planejamento e para a tomada de decisão. A eficiência de uma unidade produtiva pode ser medida pela comparação entre os valores observados e os valores possíveis de seus produtos (saídas) e recursos (insumos). Esta comparação pode ser feita, em linhas gerais, pela razão entre a produção observada e a produção potencial máxima alcançável, dados os recursos disponíveis, ou pela razão entre a quantidade mínima necessária de recursos e a quantidade efectivamente empregada, dada a quantidade de produtos gerados. Combinações dessas razões podem igualmente prover informações importantes.

O uso da Análise de Envoltória de Dados (*Data Envelopment Analysis* – DEA) (Charnes et al., 1978; Cooper et al., 2000) para medir a eficiência relativa de unidades produtivas tem-se mostrado bastante atractivo em diversos sectores de aplicação. O emprego de modelos DEA em agricultura pode vir a apoiar as decisões dos agricultores (e dos órgãos de fomento ou de associações de agricultores), ao indicar as fontes de ineficiência e as unidades que podem servir de referência às práticas adoptadas (identificação de *benchmarks*). Revisões sobre o estado da arte do uso de fronteiras de produção em agricultura podem ser vistas em Battese (1992), que apresenta aplicações de modelos de fronteiras paramétricas; aplicações de métodos de fronteiras à agricultura de países em desenvolvimento são resumidas por Bravo-Ureta e Pinheiro (1993). Coelli (1995) completa o estudo desses autores e mostra que de 38 artigos encontrados sobre o tema “fronteiras de produção e avaliação de eficiência”, somente 3 utilizavam DEA até aquela data (Färe et al., 1985; Ray, 1985; Chavas e Aliber, 1993). Em Gomes et al. (2003) encontra-se uma breve revisão sobre o estado da arte do uso de modelos DEA em agricultura. Uma revisão completa do estado da arte no tema foge ao escopo deste artigo. Entretanto, uma busca bibliográfica não exaustiva revelou que existem mais de 80 artigos publicados sobre o tema, em periódicos científicos das áreas de Investigação Operacional, Economia e Ciências Agrárias, sem contabilizar dissertações de mestrado e teses de

doutoramento relativas ao assunto.

Este artigo faz uso de DEA para medir a eficiência de uma amostra de 71 agricultores do município de Holambra, estado de São Paulo, Brasil. Os dados utilizados são aqueles originalmente usados para tipificação dos produtores rurais (Mangabeira, 2002), referentes ao ano de 2002 e obtidos por levantamento de campo, via questionários socioeconômicos.

As variáveis usadas são área total, mão-de-obra, uso de máquinas, como *inputs*, e renda líquida, como *output*. Note-se que a área total apresenta dois possíveis valores: o valor declarado pelo produtor e aquele medido a partir de sensoriamento remoto, com o uso de imagens do satélite IKONOS II, de alta resolução espacial (1m) (Dial e Grodecki, 2003). Como essa variável apresenta incerteza na medição (variável incerta), pode ser representada como um dado intervalar. Dessa forma, a fronteira eficiente é construída ao considerarem-se os menores e maiores valores possíveis de serem assumidos pela variável incerta.

Constrói-se, assim, uma região em relação à qual as unidades tomadoras de decisão possuem um certo grau de pertença. O modelo teórico usado é o apresentado em Soares de Mello et al. (2005), chamado de modelo DEA de fronteira difusa. É feito um resumo desse modelo e são apresentados os resultados da sua aplicação à avaliação dos agricultores de Holambra.

Dada a importância econômica da actividade agrícola da floricultura no município de Holambra (actividade predominante), pretende-se com os resultados gerados pelo modelo proposto avaliar se há algum tipo de cultura predominante entre os agricultores mais eficientes. Em particular, deseja-se comparar a eficiência dos que produzem flores com a dos que não o fazem. Ainda, dentro dos que produzem flores, verificar-se-á se o uso de estufas pode aumentar a eficiência. Deseja-se também verificar se esses agricultores fazem uso adequado dos recursos de que dispõem, em especial em relação ao uso da terra.

## **2 Modelo DEA de fronteira difusa**

### **2.1 Revisão bibliográfica**

Um revisão do estado da arte em incorporação de incertezas em DEA pode ser vista em Zhu (2003) e em Soares de Mello et al. (2005), cujo modelo, com pequena adaptação, é usado neste estudo. Destacam-se, a seguir, alguns trabalhos correlatos.

Wang et al. (2005) estudaram a avaliação de eficiência com modelos DEA em situações nas quais as variáveis são intervalares ou difusas. Foram construídos novos modelos DEA, baseados em aritmética intervalar. Os modelos propostos medem os limites superior e inferior da eficiência de cada DMU com variáveis incertas.

Haghighat e Khorram (2005), com base no modelo de Cooper et al. (1999) que considera modelos DEA com dados intervalares, descreveram como deve ser

o conjunto de dados das DMUs para produzir, nessas situações, o número máximo e mínimo de unidades eficientes. Para encontrar o número máximo de DMUs eficientes, os autores definiram limites superiores e inferiores que os índices de eficiência deveriam assumir. Para obter o número mínimo de unidades eficientes, os autores propuseram uma procedimento *branch and bound*.

Modelos FDH (Deprins et al., 1984) com variáveis imprecisas foram propostos por Jahanshahloo et al. (2004). Os autores usaram transformações para converter o modelo não linear em um modelo linear. A medida de eficiência é um intervalo, com limites superior e inferior.

Lertworasirikul et al. (2003) trataram variáveis imprecisas como conjuntos difusos. Esses modelos foram formalizados através de programação linear *fuzzy*. Os autores associaram as variáveis difusas a distribuições de possibilidades (Zadeh, 1978). Nesse enfoque, os índices *fuzzy*-DEA são únicos, mas dependentes do nível de possibilidade utilizado, isto é, para vários níveis de possibilidade utilizados há diversos índices diferentes correspondentes.

Um modelo DEA para avaliar DMUs de forma otimista foi proposto por Entani et al. (2002). Com esses resultados, foi determinada a eficiência por intervalos, pela proposição de novos modelos DEA. Com base no modelo *Inverted DEA* (Yamada et al., 1994), os autores avaliaram cada DMU de forma pessimista e calcularam índices de ineficiência por intervalos. Consideraram ainda dados por intervalos e propuseram modelos para calcular a eficiência e a ineficiência por intervalos, tal como feito para os dados com valores únicos e exactos.

Despotis e Smirlis (2002) usaram o modelo IDEA (*Imprecise Data Envelopment Analysis*) para lidar com dados imprecisos e, como resultado, obtiveram um limite superior e inferior para a eficiência de uma determinada DMU, o que permite uma melhor discriminação entre as DMUs. O modelo IDEA, que é um problema de programação não linear, foi proposto por Cooper et al. (1999) para tratar de dados com limites superior e inferior.

Os departamentos académicos de uma Universidade foram avaliados por Lopes e Lanzer (2002), ao considerar como números difusos os resultados de DEA nas dimensões de ensino, pesquisa, extensão e qualidade.

Cooper et al. (2001) propuseram um modelo IDEA estendido, que permite não somente o uso de dados incertos, mas também o uso de restrições aos pesos do tipo regiões de segurança ou *cone-ratio*. O modelo foi aplicado à avaliação de eficiência de postos de uma companhia de telecomunicações coreana.

O modelo DEARA, proposto por Guo e Tanaka (2001), é uma extensão do modelo DEA CCR (Charnes et al., 1978), e usa conceitos da análise de regressão para gerar um modelo DEA difuso, que considera *inputs* e *outputs* difusos. Os índices de eficiência resultantes são índices de eficiência difusos ou intervalares.

Kao e Liu (2000) apresentam um modelo difuso que é transformado em uma família de modelos DEA convencionais, baseados em dados exactos. Os índices de eficiência difusos obtidos foram expressos por meio de funções intervalares, o que possibilitaria maior nível de informação para a gerência.



Para medir a eficiência técnica de DMUs, Triantis e Eeckaut (2000) usaram variáveis difusas (dados imprecisos) para realizar comparações pareadas e verificar a dominância ou não de uma DMU. Os resultados das comparações feitas par a par foram apresentados em uma matriz, que mostra dominância em dois sentidos. Este modelo não gera índices de eficiência, mas apenas indicações sobre quem domina quem.

Hougaard (1999) usou intervalos difusos para unir em um índice de eficiência a informação fornecida pelos índices de eficiência analíticos (DEA) e índices de eficiência subjectivos baseados em dados que reflectem aspectos qualitativos e organizacionais, expressos na forma de intervalos difusos. A forma de especificar a relação entre esses dois tipos de informação foi um índice corrigido, expresso na forma de um intervalo difuso.

Uma medida de eficiência técnica em ambiente difuso, em três estágios, foi proposta por Triantis e Girod (1998), com uso de modelos DEA clássicos que incorporaram conceitos desenvolvidos em programação paramétrica difusa (Carlsson e Korhonen, 1986).

Sengupta (1992) explorou a teoria dos conjuntos difusos no contexto de DEA, com três tipos de estatísticas difusas (programação matemática difusa, regressão difusa e entropia difusa) para ilustrar os tipos de decisão e de solução que podem ser alcançados quando os dados são vagos e a informação *a priori* é inexacta e imprecisa.

## 2.2 Abordagem proposta

A abordagem aqui usada para incorporar dados intervalares a modelos DEA é a proposta por Soares de Mello et al. (2005) e destaca-se das existentes na literatura por não fazer nenhuma suposição em relação à forma como cada *input* ou cada *output* varia. Ou seja, não é assumido que os valores obedecem a alguma distribuição de probabilidade. Nessa proposta é construído um índice único de eficiência para a situação em que as variáveis (*inputs* ou *outputs*) apresentam incerteza. Qualquer que seja a variação, a fronteira de eficiência é construída ao considerarem-se apenas os valores máximos e mínimos possíveis de serem assumidos, com o uso posterior de programação linear clássica e modelos DEA tradicionais para a determinação das fronteiras.

### 2.2.1 Criação da fronteira DEA difusa

Se em um modelo DEA não houver certeza sobre os valores assumidos por um *output* ou *input*, não haverá igualmente certeza sobre a exacta localização da fronteira DEA eficiente. Caso os valores do *output* (*input*) para algumas DMUs (unidades de tomada de decisão – *Decision Making Units*), sejam maiores que o suposto, a fronteira estará deslocada “mais acima” (“mais à direita”), isto é, em uma região de valores superiores para esse *output* (*input*). Caso os valores sejam inferiores ao suposto, a fronteira estará “mais abaixo” (“mais à esquerda”).

Portanto, neste caso, conforme destacado em Soares de Mello et al. (2005), a fronteira não é um conjunto no sentido clássico do termo, mas um conjunto nebuloso (Zadeh, 1965), para o qual não há sentido em dizer que um elemento pertence ou não ao conjunto; deve-se fazer referência ao grau de pertença desse elemento ao conjunto. Dessa forma, em vez de existirem DMUs na fronteira e outras fora da fronteira, haverá DMUs com diferentes graus de pertença à fronteira.

Pela abordagem aqui usada, diferentemente da lógica nebulosa clássica em que são postuladas funções de pertença, que determinam com que grau um certo elemento pertence a um determinado conjunto, o grau de pertença é calculado com base em propriedades geométricas das fronteiras geradas, em vez das funções clássicas.

No caso de um *output* com incerteza, a fronteira localizadas “mais acima” é aquela obtida por um modelo DEA clássico, CCR (Charnes et al., 1978) ou BCC (Banker et al., 1984) que considera o máximo valor do *output* incerto que cada DMU pode atingir, o que representa a melhor situação para todas as DMUs. Assim, essa fronteira é denominada de *Fronteira Optimista*. Analogamente, a fronteira obtida com o modelo DEA clássico que considere o menor valor de *output* para cada DMU é a *Fronteira Pessimista*, por considerar a situação menos favorável de produção.

No caso em que a variável incerta é um *input*, a *Fronteira Optimista* é aquela gerada pelos menores valores que os *inputs* podem assumir (localizada “mais à esquerda”). Já a fronteira composta pelos maiores valores assumidos pelos *inputs* (situação de uso de recursos não desejável) representa a *Fronteira Pessimista* (situada “mais à direita”).

A Figura 1 ilustra esses conceitos, para o modelo DEA BCC, em que a variável incerta é um *output*. A fronteira difusa é toda a região situada entre as fronteiras pessimista (fronteira inferior) e optimista (fronteira superior). Note-se ainda que uma DMU não é mais representada por um ponto; a incerteza na medição do *output* faz com que a representação da DMU seja um segmento de recta com extremidades determinadas pelos valores pessimista e optimista desse *output*. Em um caso mais geral, com várias variáveis incertas, a DMU ocupa um hiper volume, que pode ser representado pela sua diagonal (Soares de Mello et al., 2005).

Na Figura 1, conforme descrito em Soares de Mello et al. (2005),  $OP_{fo}$  e  $OP_{fp}$  referem-se ao *output* projectado na fronteira optimista e pessimista, respectivamente.  $c$  é o comprimento da DMU, ou seja, é a diferença entre os valores optimista e pessimista do *output*;  $l$  é a largura da faixa, isto é, representa para cada DMU a diferença entre o valor do *output* incerto para as fronteiras optimista e pessimista;  $p$  é parte que está na faixa, é a diferença entre o *output* optimista de cada DMU e a intersecção dessa DMU com a fronteira pessimista.

Uma vez definidos a fronteira difusa e os termos  $c$ ,  $l$  e  $p$ , deve-se definir o grau de pertença de cada DMU a essa fronteira. As DMUs que estão integralmente contidas na região que define a fronteira difusa devem ter grau de pertença 1 à fronteira. Já aquelas que apenas tocam a fronteira devem ter grau de pertença nulo. Entre esses dois casos extremos, as DMUs poderão ter graus de pertença intermédios.

Suponha-se a existência de uma DMU que contenha toda a largura da fronteira

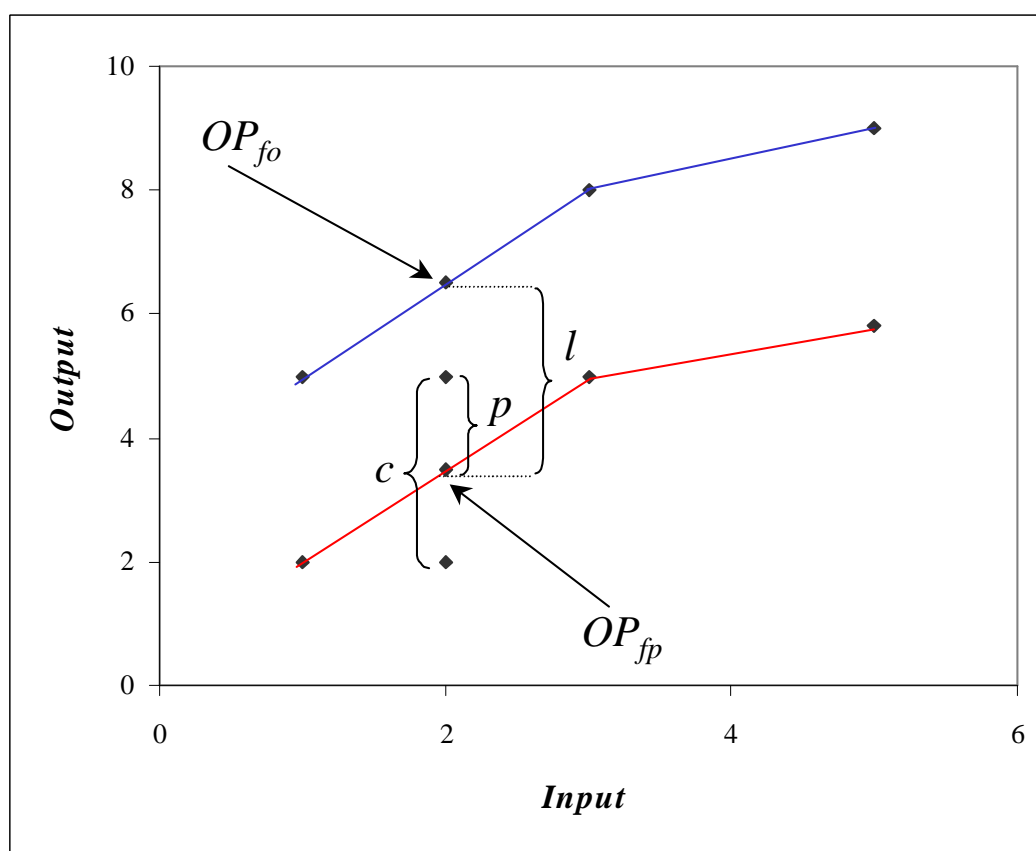


Figura 1: Fronteiras otimista e pessimista.

difusa, mas que tenha uma parte da largura da faixa externa à fronteira. Nesse caso, apesar de não estar totalmente excluída da fronteira, a sua pertença também não é total. Em situações semelhantes a essa, a pertença deveria obedecer à relação  $p/c$ , unitária quando  $p = c$ .

Por outro lado, suponha-se o caso em que uma DMU esteja totalmente contida na fronteira difusa, mas que exista uma região da fronteira que não contenha a DMU. Nessa situação, a pertença deveria ser  $p/l$ , unitária quando  $p = l$ .

Para garantir que uma DMU só tenha pertença 1 à fronteira difusa se ela for eficiente tanto na hipótese pessimista quanto na otimista, esses dois casos devem ser combinados. O produto das duas expressões anteriores satisfaz a essa propriedade e, dessa forma, a pertença à fronteira difusa,  $\wp$ , é definida por  $\wp = \frac{p^2}{lc}$ .

### 2.2.2 Cálculo algébrico da pertença

Para o caso em que apenas um *output* apresenta incerteza, para obter-se uma expressão que possa ser usada em modelos multidimensionais, faz-se necessário transformar as grandezas geométricas da relação  $\wp = p^2/lc$  em quantidades que possam ser extraídas dos modelos DEA clássicos: *outputs* otimistas e pessimistas, eficiências com *output* pessimista em relação à fronteira pessimista ( $Ef_{fp}$ ) e com *output* otimista em relação à fronteira otimista ( $Ef_{fo}$ ).

Os valores de  $Ef_{fp}$  e de  $Ef_{fo}$  são calculados, para o caso de um *output* com incerteza, das definições clássicas de DEA orientado a *outputs*, segundo as relações  $Ef_{fp} = \frac{OP_{fp}}{O_{fp}}$  e  $Ef_{fo} = \frac{OP_{fo}}{O_{fo}}$ , nas quais  $O_{fp}$  e  $O_{fo}$  são, respectivamente, os valores nas fronteiras pessimista e otimista deste *output*.

A largura da faixa  $l$  é a diferença entre o alvo da fronteira otimista e o alvo da fronteira pessimista, ou seja,  $l = OP_{fo} - OP_{fp} = O_{fo}Ef_{fo} - O_{fp}Ef_{fp}$ . O comprimento da DMU  $c$  é a diferença entre o *output* otimista e o pessimista, isto é,  $c = O_{fo} - O_{fp}$ . A parte da DMU que está na fronteira  $p$  é a diferença entre o *output* otimista e o alvo do *output* pessimista na fronteira pessimista, desde que a diferença seja positiva. Isto implica que o *output* otimista deve estar dentro da faixa da fronteira difusa; caso contrário,  $p$  deve ser igual a 0, ou seja,  $p = O_{fo} - O_{fp}Ef_{fp}$ , se  $O_{fo} - O_{fp}Ef_{fp} \geq 0$ ;  $p = 0$ , caso contrário.

Com a definição “não geométrica” de  $l$ ,  $c$  e  $p$ , é possível obter a expressão que representa algebricamente a pertença no caso de um *output* incerto,  $\wp_{output}$ , conforme apresentado em (I) (Soares de Mello et al., 2005).

$$\wp_{output} = \frac{(O_{fo} - O_{fp}Ef_{fp})^2}{(O_{fo}Ef_{fo} - O_{fp}Ef_{fp})(O_{fo} - O_{fp})}, \text{ se } O_{fo} - O_{fp}Ef_{fp} \geq 0 \quad (I)$$

$$\wp_{output} = 0, \text{ caso contrário}$$

Além do caso em que  $O_{fo} - O_{fp}Ef_{fp} < 0$ , onde o índice de pertença foi arbitrado como nulo, o cálculo algébrico dessa pertença também é zero caso  $O_{fo} - O_{fp}Ef_{fp} = 0$ . Por outro lado, como o numerador de (I) está elevado ao quadrado, nunca assume um valor negativo. Na orientação a *outputs*, o alvo na fronteira otimista é sempre

maior que o alvo na fronteira pessimista, e o *output* optimista é sempre maior que o *output* pessimista. Dessa forma, o denominador é, igualmente, sempre positivo. Portanto, o índice de pertença é sempre um número não negativo.

Caso as eficiências optimista e pessimista sejam unitárias, o índice de pertença é unitário. Entretanto, observa-se que  $O_p E f_{fp} \geq O_p$  (uma vez que na orientação a *outputs* a eficiência é maior ou igual a 1). Portanto, tem-se que  $O_{fo} - O_{fp} E f_{fp} \geq O_{fo} E f_{fo} - O_{fp} E f_{fp}$ . Verifica-se ainda que  $O_{fo} E f_{fo} - O_{fp} E f_{fo} \geq O_{fo} - O_{fp} E f_{fp}$  (haja visto que  $O_{fo} E f_{fo} \geq O_{fo}$ ). Logo, cada termo do denominador de (I) é maior ou igual à raiz quadrada do numerador. Em consequência, o numerador não é maior que o denominador e o índice de pertença não pode ser maior que a unidade.

Destaca-se ainda que a modelagem proposta só é válida nos casos em que os valores optimistas e pessimistas não sejam coincidentes, ou seja, nos casos em que haja incerteza na medição das variáveis.

Ainda deve-se observar que o índice de pertença (I) não é uma medida de eficiência, apesar de calculado a partir de eficiências e, dessa forma, não precisa respeitar as propriedades das medidas de eficiência. Em particular, o conjunto de DMUs com pertença unitária à fronteira difusa não é necessariamente um conjunto convexo.

De forma análoga ao que acontece com um *output* com incerteza, pode-se apresentar o caso em que haja um *input* com incerteza. Nesse caso, define-se como *input* optimista,  $I_{fo}$ , aquele com o menor valor que o *input* pode assumir, e *input* pessimista,  $I_{fp}$ , o de maior valor que o *input* pode assumir. Quando se consideram os *inputs* optimistas para todas as DMUs, tem-se a fronteira optimista; quando são considerados os *inputs* pessimistas para todas as DMUs, obtém-se a fronteira pessimista.

Deduções semelhantes às do caso de um *output* com incerteza permitem definir o índice de pertença para o caso de um *input* incerto,  $\varphi_{input}$ , conforme apresentado em (II) (Soares de Mello et al., 2005). Observe-se que em (II), as eficiências  $E f_{fp}$  (eficiência em relação à fronteira pessimista) e  $E f_{fo}$  (eficiência em relação à fronteira optimista) são orientadas a *inputs*.

$$\varphi_{input} = \frac{(I_{fp} E f_{fp} - I_{fo})^2}{(I_{fo} E f_{fo} - I_{fp} E f_{fp})(I_{fp} - I_{fo})}, \text{ se } I_{fp} E f_{fp} - I_{fo} \geq 0 \quad (II)$$

$$\varphi_{input} = 0, \text{ caso contrário}$$

Deve-se referir que as expressões (I) e (II) são exactas no caso em que o *input* (ou *output*) que apresenta incerteza é o único *input* (ou *output*). Caso contrário, esta é uma aproximação radial para uma situação não radial, conforme destacado em Soares de Mello et al. (2005). Redução radial refere-se à redução equiproporcional de todos os *inputs* ou aumento equiproporcional de todos os *outputs*. Entretanto, neste caso, a redução ou acréscimo é em apenas em um deles, naquele com incerteza na medição. As expressões derivadas usadas para a fronteira difusa são aquelas que seriam usadas caso houvesse redução ou acréscimo em todas as variáveis, sendo, portanto, uma aproximação radial de um caso não radial. Uma solução exacta pode ser obtida pela formulação multiobjectivo para DEA (Lins et al., 2004), como mostrado em Soares de Mello et al. (2005). No entanto, este modelo exige um grande esforço computacional, o que pode recomendar o uso da aproximação radial

em casos em que o intervalo de variação seja pequeno.

### **3 Estudo de caso: Avaliação de eficiência dos agricultores de Holambra com modelo DEA de fronteira difusa**

#### **3.1 Agricultura em Holambra**

O município de Holambra situa-se no Estado de São Paulo, Brasil, com uma dimensão territorial relativamente pequena em relação aos demais municípios brasileiros (65 km<sup>2</sup>). Holambra tem cerca de 8.300 habitantes, segundo estimativas do Instituto Brasileiro de Geografia e Estatística para o ano de 2005 (Instituto Brasileiro de Geografia e Estatística, 2005).

O município caracteriza-se pela recepção da imigração holandesa ao Brasil (final da década de 40) e tem economia fortemente baseada na actividade agro-pecuária, com predomínio das actividades hortícolas, citricolas, plantas decorativas, flores, suinocultura, avicultura e laticínios. Quase toda sua produção agrícola advém de um grupo aproximado de 287 produtores, com 20 ha em média de área (Holambra, 1998).

Mangabeira (2002) mapeou o uso das terras do município de Holambra com base em imagens do satélite IKONOS II (do ano de 2000) e levantamentos de campo (em 2001). Com o uso de questionários de campo, pesquisou informações de 266 variáveis socioeconómicas (que geraram 204 indicadores), para uma amostra de 74 produtores rurais.

De posse desses 204 indicadores, o autor tipificou essa amostra de produtores. Para tal foi usada a Análise de Correspondência Múltipla (Bishop et al., 1975), seguida da análise de *cluster* pelo método de Ward (Everitt, 1974), para classificar os tipos de produtores agrícolas. Mangabeira (2002) identificou 6 tipos de agricultores em Holambra, cujas características são brevemente descritas a seguir. Desses, três grupos são de produtores de flores (62% do total de agricultores analisado), a saber, agrofloricultores, florescitricultores e floricultores.

- Produtores de flores

- Agrofloricultores (7 agricultores): o sistema de produção é baseado na produção de flores em estufas. Caracteriza-se pelo baixo nível de capitalização, mas diferenciam-se dos demais grupos de produtores de flores pelo baixo uso de insumos agrícolas (fertilizantes, defensivos etc.) e pela não utilização de mecanização em estufas. Apresentam alta eficiência de emprego e renda por unidade de área.
- Florescitricultores (24): caracterizam-se por produzir citros e flores. São produtores com nível médio de capitalização, com nível baixo a médio de produtividade líquida do capital por hectare em relação à área total e alto consumo de insumos agrícolas.

- Floricultores (13): apresentam como sistema de produção flores em estufas. Caracterizam-se pelo alto nível de capitalização em relação aos outros grupos, alto uso de insumos agrícolas, geração de emprego e mecanização por hectare cultivado. Não há presença de vegetação natural em suas propriedades.
- Não produtores de flores
  - Agrocitricultores (14): têm o sistema de produção baseado na produção de citros e com metade deles produzindo também cultura anual. São produtores com nível médio de capitalização, com uso médio de insumos agrícolas e baixa geração de emprego.
  - Agropecuaristas (4): são criadores de bovinos e suínos e têm agricultura, principalmente anual, em suas propriedades. São produtores com baixo nível de capitalização, baixa geração de emprego e que não utilizam insumos agrícolas e mecanização. Possuem granjas em suas propriedades.
  - Produtores Agrianuais (9): são agricultores que, na sua maioria, têm a cultura anual como principal actividade nos seus sistemas de produção. Caracterizam-se pelo baixo nível de capitalização em relação aos outros grupos, baixo nível de emprego por hectare e renda por hectare trabalhado, pelo uso razoável de insumos agrícolas e pela mecanização em suas propriedades. Com relação ao meio ambiente, metade dos produtores deste grupo praticam rotação de cultura.

### 3.2 Estruturação do problema

Para estruturar um modelo DEA é necessário definir as unidades a avaliar, as variáveis a usar (inputs e outputs), o modelo DEA adoptado (CCR, BCC, entre outros; orientação a inputs radial ou não radial, a outputs radial ou não radial, não orientado etc.). Na construção do modelo DEA deve-se atentar quais variáveis serão consideradas *inputs* e quais serão consideradas *outputs*. Para a avaliação dos agricultores de Holambra foram seleccionados 4 dos 206 indicadores derivados por Mangabeira (2002). Essas variáveis indicam as relações clássicas entre produção, capital e trabalho. Como *inputs* foram seleccionados Área Total (representada pela área total da propriedade, em hectares), Emprego (expresso por homem-hora-ano) e Uso de Máquinas (calculado como número total de horas de uso de máquinas no ano); Renda Líquida anual da propriedade, em reais, é o *output*.

Em Gomes e Mangabeira (2004), para medir a eficiência dos agricultores de Holambra, foram usadas as mesmas variáveis de *output* e *input*, com excepção da variável Área Total que foi substituída pelo *input* Área Cultivada. Estes autores não consideraram incerteza nos dados. Enquanto no artigo citado foi considerado que o não aproveitamento de uma área não era uma ineficiência, no presente artigo considera-se que o agricultor deve procurar aproveitar toda a área à sua disposição. Ressalta-se que nas propriedades analisadas pode-se desprezar a área submetida a constrangimentos ambientais. O uso da variável “área total” ao invés de “área cultivada” pode causar alguns problemas, pois a variável “área total” penaliza aqueles agricultores quem têm áreas não incluídas como fonte de produção e renda.

Já o uso da “área cultivada” fornece resultados de eficiência técnica, mas não de uma eficiência global, pois desconsidera a razão entre terra disponível e em uso; o emprego da variável “área cultivada” aumenta a eficiência daqueles produtores que utilizam estufa e mascara o mau aproveitamento da terra, conforme mostrado em Gomes e Mangabeira (2004).

Os valores das variáveis Emprego, Uso de Máquinas e Renda Líquida são aqueles fornecidos pelos agricultores nos questionários de campo. Já a variável Área Total pôde ser obtida de duas fontes: dos questionários de campo (ou seja, da informação fornecida pelo agricultor) e das imagens de satélite do município. O uso da imagem do satélite IKONOS II bem como o conhecimento dos limites de cada propriedade, permitiu calcular, com apoio de um Sistema de Informações Geográficas, a área total de cada propriedade. Esse valor, no entanto, é diferente do valor fornecido pelo agricultor. Assim, a variável Área Total é uma variável incerta que pode ser representada como um dado intervalar, sendo um dos limites do intervalo a informação dada pelo agricultor e o outro, o dado calculado com base na imagem do satélite IKONOS II.

As unidades de avaliação (DMUs) são as propriedades agrícolas amostradas por Mangabeira (2002). Dos 74 produtores, 3 foram descartados desta análise por dois motivos: são produtores de leite (portanto, com tecnologia bem diferente das demais DMUs) e apresentaram valor negativo para a variável Renda Líquida (o que obrigaria ao desenvolvimento de modelos para lidarem com esta situação). Dentre estes 71 agricultores, 44 produzem flores, sendo 35 exclusivamente em estufas, 4 somente em campo aberto e 5 possuem produção mista. Os dados usados encontram-se no Apêndice 1.

Quanto à questão de selecção do modelo DEA no que se refere aos retornos de escala, Banker (1996) propôs um teste de hipóteses para confirmar ou não a existência de ganhos ou perdas de escala. No entanto, trabalhos recentes (Hollingsworth e Smith, 2003; Soares de Mello et al., 2004) levantaram a hipótese de que o modelo DEA BCC ou não tem relação com a variação dos retornos de escala ou apresentam sempre retornos decrescentes de escala. Assim, face às questões levantadas por estes autores, preferiu-se não optar pelo teste proposto por Banker (1996) para selecção do modelo DEA. Para a escolha do modelo, deve-se observar que o modelo DEA CCR exige proporcionalidade de causa e efeito. No presente estudo, a duplicação dos valores dos *inputs* uso de máquina e mão-de-obra não acarreta em duplicação da produção. Além disso, as propriedades têm dimensões bem diferentes, o que também recomenda o uso do modelo DEA BCC para que uma DMU seja comparada apenas com as suas semelhantes.

O modelo DEA BCC foi orientado a *input* por três motivos: pretende-se saber se há excesso no uso dos recursos; para comparar os resultados com aqueles obtidos por Gomes e Mangabeira (2004); e a estrutura matemática dos modelos DEA difusos quando apenas uma variável apresenta incerteza, obriga a que a orientação do modelo seja *input* ou *output* conforme essa variável o seja.

Todos os resultados do modelo DEA necessários ao cálculo do índice de pertença foram obtidos com o uso do *software* SIAD (Angulo Meza et al., 2005).



### 3.3 Resultados

#### 3.3.1 Cálculo das medidas de eficiência

Conforme apresentado em (II), para o cálculo da pertença à fronteira difusa, na presença de um *input* com incerteza, é necessário calcular as eficiência em relação às fronteiras otimista e pessimista. A fronteira pessimista é aquela gerada quando do uso dos maiores valores do *input* e a fronteira otimista é aquela que comporta os menores valores para o *input* incerto. Devido à pequena diferença entre os valores, foi usada a aproximação radial, apesar de o modelo ser não radial, como referido ao final do item 2.

Calculando-se os modelos DEA BCC, verifica-se que a eficiência média em relação às fronteiras otimista e pessimista foi de cerca de 34%. Tanto em relação à fronteira otimista quanto à pessimista, 10 DMUs foram DEA BCC eficientes, sendo as mesmas nos dois casos.

Dentre as 10 DMUs eficientes, 5 são produtores de flores, todos em estufas e sem produção de flores em campo aberto. A eficiência média dos agricultores que não produzem flores no modelo DEA BCC foi de 43% para as fronteiras otimista e pessimista. Para aqueles que produzem flores, a eficiência média foi menor, cerca de 29% em relação a ambas as fronteiras. Dentro desta categoria, os produtores de flores em campo aberto tiveram eficiência otimista e pessimista médias de 18% e os que produzem em estufas de 32%.

Esses resultados de eficiência estão de acordo com a hipótese levantada por pesquisadores e especialistas na agricultura da região, de que os produtores de flores são mais ineficientes que aqueles que se dedicam a outras actividades agrícolas. Confirmam igualmente que aqueles que cultivam flores em estufas possuem melhor desempenho em relação aos produtores de flores em campo aberto.

A menor eficiência daqueles que produzem flores deve-se ao fato de que estes usam excessivamente insumos agrícolas (e sem controle, seja por desconhecimento, falta de assistência técnica, de pesquisas, ou por outros factores), o que tem impacto na redução da renda líquida.

Para aqueles agricultores que produzem flores em estufas, a maior eficiência é explicada pelo tipo de cultivo, já que este requer maior cuidado operacional; o cultivo em estufas, por ser um ambiente fechado (com espaço reduzido e controlado), impõe maior controle climático, no uso de insumos, de máquinas e de mão-de-obra.

Os resultados de maior eficiência dos não produtores de flores são opostos àqueles obtidos em Gomes e Mangabeira (2004), onde 6 agricultores foram eficientes, sendo 4 produtores de flores. Cabe destacar que enquanto o resultado do modelo dos multiplicados apresentava para o modelo de Gomes e Mangabeira (2004) uma grande quantidade de pesos zero para a variável Área Cultivada (89% das DMUs), no modelo proposto neste artigo, o modelo dos multiplicadores indica que a maior quantidade de pesos zero foi dada para a variável Uso de Máquinas (68% das DMUs), enquanto que apenas 20% das DMUs desconsideram a Área Total no cálculo da medida de eficiência.

Adicionalmente, dentre os 6 agricultores vistos como eficientes por Gomes e Mangabeira (2004), 5 também o são pelo modelo aqui proposto. A exceção está em um agricultor, produtor de flor em estufa, que no modelo aqui proposto teve eficiência média de 28 e 27% em relação as fronteira otimista e pessimista, respectivamente.

Deve-se ser enfatizado que, como em qualquer modelo DEA, a escolha das variáveis é determinante na avaliação efectuada. No presente caso, além da escolha de variáveis, a forma como uma delas foi considerada mostrou-se fundamental. Considerar a variável “área” como sendo toda a área da propriedade ou apenas a área cultivada, traz importantes mudanças nos resultados. O uso da área total avalia melhor os agricultores que promovem uma total ocupação da terra com a actividade agrícola. Neste tipo de modelo, um agricultor pode ser eficiente, não por usar melhores técnicas mas apenas por empregar intensivamente toda a área à sua disposição. Já ao ser considerada como variável apenas a área cultivada, pode-se beneficiar indevidamente quem tenha uma grande quantidade de terra devoluta. Por outro lado, não se prejudica quem optou por manter reserva ambiental, construiu áreas de interesse social (lazer, educação, etc.) ou alojar outras actividades.

A escolha de uma das formas de considerar a área não é politicamente neutra. Assim, este artigo considera a área total, mas compara os resultados com os que foram obtidos com a área cultivada. Em tomada de decisão, as duas avaliações devem ser levadas em conta.

Ainda em relação à escolha das variáveis, o fato de o grupo dos agricultores que não produzem flores apresentar nível baixo a médio de capitalização e de emprego justifica também sua maior eficiência em relação ao grupo de produtores de flores, já que estas variáveis são *inputs* do modelo DEA.

### 3.3.2 Cálculo do índice de pertença

Ao calcular-se o índice de pertença  $\phi$  obtém-se como resultado que essas mesmas DMUs apresentam pertença unitária e 61 tiveram pertença zero à fronteira difusa. Não houve DMUs com pertença intermediária.

Devido ao grande número de empates, há necessidade de melhor discriminar as unidades em avaliação, em especial para aquelas com pertença unitária (melhor desempenho). Para tal, é proposto o uso da fronteira invertida. Conforme destacado em Leta et al. (2005), o cálculo de eficiência em relação a essa fronteira permite avaliar as unidades de modo pessimista. É possível ainda calcular um índice composto de avaliação, que agrega os resultados em relação à fronteira original e à invertida e, assim, realizar uma avaliação mais completa das unidades, já que esse índice considera não só os pontos fortes (fronteira original), como também os fracos (fronteira invertida).

### 3.3.3 Fronteira invertida e índice composto de avaliação

Como destacado no item 2.2.2., o grau de pertença à fronteira não é uma medida de eficiência; duas DMUs que tenham grau nulo de pertença à fronteira podem ter posições relativas bem diferentes não detectadas pelo índice aqui proposto. Ou seja, ao contrário dos modelos DEA clássicos que fornecem muitos empates nos índices 100% eficientes, o enfoque proposto por Soares de Mello et al. (2005) e usado neste artigo fornece empates para as DMUs totalmente não pertencentes à fronteira. Uma revisão de métodos para diminuir os empates em DEA nas DMUs eficientes, pode ser vista em Angulo Meza e Lins (2002).

Para distinguir entre essas DMUs, Soares de Mello et al. (2005) propuseram o uso da fronteira invertida, que consiste em estimar uma fronteira de eficiência DEA na qual os *outputs* são considerado como *inputs* e os *inputs* como *outputs*. Essa fronteira foi usada inicialmente por Yamada et al. (1994) e Entani et al. (2002) com o objectivo de ou avaliar as DMUs de maneira pessimista, já que essa fronteira consiste das DMUs com as piores práticas gerenciais (e poderia ser chamada de fronteira ineficiente), e por Lins et al. (2005) para avaliar essas unidades como sendo as de melhores práticas segundo um ponto de vista oposto. Destaca-se que ao fazer a avaliação pelo “inverso”, a orientação do modelo DEA seleccionado também deve ser invertida.

Leta et al. (2005) usaram o conceito de fronteira invertida para discriminar DMUs eficientes, com o cálculo de um índice que considera tanto a eficiência em relação à fronteira original quanto em relação à invertida. Soares de Mello et al. (2005) aproveitam essa abordagem para distinguir entre duas DMUs que apresentem grau nulo ou unitário de pertença à fronteira difusa original, mas que tenham posições relativas bem diferentes não detectadas pelo índice de pertença. Para esse caso, quanto maior o grau de pertença à fronteira invertida menor a eficiência da DMU.

O uso do conceito de fronteira invertida no contexto da fronteira difusa tenta resolver o problema do grande número de empates. No entanto, para valores de incerteza pequenos, pode provocar o surgimento de um grande número de empates para DMUs medianamente eficientes. Métodos para resolver esse problema necessitam de estudos e desenvolvimentos futuros como, por exemplo, o uso de camadas de iso-eficiência (Tavares, 1998; Gomes, 2003).

Para obter um índice único de eficiência, Soares de Mello et al. (2005) englobaram os dois graus de pertença e obrigaram a que a variação do índice fosse entre 0 e 1. Esse índice foi chamado de *eficiência difusa* (ou *eficiência fuzzy-DEA*) ( $Ef_{difusa}$ ) e é dado pela equação (III), na qual  $\varphi_{original}$  é o grau de pertença à fronteira original e  $\varphi_{invertida}$  é o grau de pertença à fronteira invertida.

$$Ef_{difusa} = \frac{(\varphi_{original} - \varphi_{invertida} + 1)}{2} \quad (III)$$

Destaca-se que, apesar de por simplicidade de linguagem a agregação dos dois índices de pertença tenha recebido a denominação de eficiência difusa, este não é um indicador de eficiência no sentido clássico do termo. É apenas uma ponderação normalizada entre um índice de pertença, e o complementar de outro índice de pertença. O uso do complemento justifica-se pelo fato de na fronteira invertida

avaliar-se ineficiência, e os índices usados avaliam eficiências.

### 3.3.4 Cálculo do índice fuzzy-DEA

Ao calcular-se o índice apresentado em (III), obteve-se como resultado que dentre as 10 DMUs que haviam apresentado  $\phi_{original}$  unitário, 8 apresentaram  $Ef_{difusa}$  igual a 1,0 (tiveram  $\phi_{invertida}$  igual a zero) e 2 (não produtores de flores) obtiveram  $Ef_{difusa}$  igual a 0,5. Dentre as demais 61 DMUs, 5 (4 produtores de flores; 2 em estufas, 2 em campo) apresentaram índice agregado zero e o restante (sendo 38 produtores de flores) 0,5. Os resultados encontram-se no Apêndice 2.

Esse resultado permite alocar os agricultores em três grupos: alto, médio e baixo desempenho, conforme o valor de  $Ef_{difusa}$  seja 1,0, 0,5 ou 0,0, respectivamente. Além disso, dentre as 10 unidades com  $\phi_{original}$  unitária, foi possível separar duas que poderiam ser consideradas “falsamente” eficientes, já que apresentaram  $\phi_{invertida}$  igual a 1. Ou seja, segundo esta análise, não basta ao produtor ter bom resultado nas características em que é bom; também não deve ter resultado ruim naquilo em que não é tão bom.

Esses resultados corroboram os anteriormente discutidos, do melhor desempenho dos agricultores que não produzem flores, já que este encontram-se majoritariamente no grupo de alto e médio desempenho. Igualmente, dentre os que produzem flores, aqueles que usam estufas em seu sistema de produção, têm desempenho superior (estão distribuídos nos três grupos, enquanto aqueles que produzem flor em campo, não aparecem no grupo de  $Ef_{difusa}$  unitária).

Para as unidades de desempenho intermediário, aqui denotado por  $Ef_{difusa}$  igual a 0,5, a discriminação não foi possível. Conforme anteriormente mencionado, o uso da fronteira invertida permite um melhor ajuste dos extremos, mas não separa as DMUs medianamente eficientes, o que pode ser conseguido, possivelmente, com métodos adicionais, como as já citadas camadas de iso-eficiência.

## 4 Conclusões

Ao não arbitrar, nem uma distribuição de probabilidade para as incertezas das variáveis, nem uma função nebulosa para as mesmas, a abordagem proposta neste artigo para incorporação de incertezas aos modelos DEA clássicos traz uma vantagem em relação aos modelos existentes na literatura. Além disso, é matematicamente simples, já que os resultados são obtidos por simples cálculos algébricos, sem necessidade de usar programação linear nebulosa. Ao determinar uma região onde se encontra a fronteira difusa e ao construir geometricamente uma função de pertença e, em consequência, a medida de eficiência difusa, os desenvolvimentos deste artigo situam-se próximos à origem dos conjuntos difusos sem, no entanto, usar suas funções características.

Para que uma DMU possua um valor elevado do índice fuzzy-DEA, deve ter um elevado grau de pertença em relação à fronteira otimista e baixo grau em relação

à fronteira pessimista. Dessa forma, todas as variáveis são levadas em conta no índice final. Assim, não basta a DMU ter bom desempenho naquilo em que ela é melhor; não deve ter também mau desempenho no critério em que for pior.

Essa característica permite eliminar outro dos inconvenientes dos modelos DEA BCC: o fato de a DMU de maior *output* ser eficiente independentemente dos valores dos *inputs* (Ali, 1993). No modelo aqui proposto, o valor da eficiência de tal DMU depende também de sua posição em relação à fronteira invertida.

Em relação ao estudo de caso, destacam-se dois aspectos interessantes. O primeiro refere-se à integração do sensoriamento remoto com técnicas de pesquisa operacional. O segundo diz respeito aos resultados obtidos, que confirmam a expectativa dos especialistas na agricultura praticada na região, de maior eficiência dos agricultores que não se dedicam à produção de flores. Ratificam, ainda, o sentimento de que os agricultores que usam essencialmente estufas têm melhor desempenho em relação ao que cultivam flores em campo aberto.

A comparação de resultados entre este artigo e o estudo que usou área cultivada permite a cada agricultor (ou ao seu órgão de assistência técnica) uma identificação mais precisa das suas deficiências: em particular se o seu problema é de uso de técnicas inadequadas ou de uso insuficiente da terra. Destaca-se, entretanto, que avaliações menos rigorosas, que levam em conta apenas a área cultivada, desconsiderando que tais propriedades rurais apresentam grandes extensões de terra sem uso.

Resultados obtidos com DEA são, via de regra, mais úteis para órgãos reguladores que para as unidades avaliadas. Ao comprovar hipóteses levantadas por pesquisadores, que não dispunham de ferramentas para testá-las, os resultados deste estudo (e de estudos similares) podem servir de parâmetro para pesquisadores e extensionistas, no apoio à busca de medidas que aumentem a racionalidade do uso dos insumos agrícolas, da mão-de-obra, de maquinarias e do uso eficiente da terra. Servem ainda para traçar políticas públicas mais eficientes nos planos, programas e projectos de desenvolvimento rural.

Destaca-se que apesar de considerar-se como variável incerta somente a área total de cada propriedade, os demais dados declarados pelo agricultor podem igualmente conter incertezas. Apesar de modelos que contemplem incertezas em várias variáveis já terem sido desenvolvidos (Soares de Mello et al., 2005), surge, neste caso, a questão de como quantificar essa incerteza e desenvolvimentos nesse sentido são vislumbrados pelos autores.

## 5 Agradecimentos

Ao CNPq, pelo apoio financeiro, por meio do Edital CNPq 19/2004 – Universal, processo n.º 472838/2004-0.

## 6 Referências

- Ali, A.I. Streamlined computation for data envelopment analysis. **European Journal of Operational Research**, v. 64, p. 61-67, 1993.
- Angulo Meza, L.; Biondi Neto, L.; Soares de Mello, J.C.C.B.; Gomes, E.G. ISYDS – Integrated System for Decision Support (SIAD – Sistema Integrado de Apoio a Decisão): A Software Package for Data Envelopment Analysis Model. **Pesquisa Operacional**, v. 25, n. 3, 2005.
- Angulo Meza, L.; Lins, M.P.E. Review of methods for increasing discrimination in data envelopment analysis. **Annals of Operations Research**, v. 116, p. 225-242, 2002.
- Banker, R.D. Hypothesis tests using DEA. **Journal of Productivity Analysis**, v. 7, n.2-3, p. 139-159, 1996.
- Banker, R.D.; Charnes, A.; Cooper, W.W. Some models for estimating technical scale inefficiencies in data envelopment analysis. **Management Science**, v. 30, n. 9, p. 1078-1092, 1984.
- Battese, G.E. Frontier production functions and technical efficiency: a survey of empirical applications in agricultural economics. **Agricultural Economics**, v. 7, p. 185-208, 1992.
- Bishop, Y.; Fienberg, S.; Holland, P. **Discrete multivariate analysis: theory and practice**. Cambridge: MIT Press, 1975.
- Bravo-Ureta, B.E.; Pinheiro, A.E. Efficiency analysis of developing country agriculture: a review of the frontier function literature. **Agricultural and Resource Economics Review**, v. 22, p. 88-101, 1993.
- Carlsson, C.; Korhonen, P. A parametric approach to fuzzy linear programming. **Fuzzy Sets and Systems**, v. 20, p. 17-33, 1986.
- Charnes, A.; Cooper, W.W.; Rhodes, E. Measuring the efficiency of decision-making units. **European Journal of Operational Research**, v. 2, p. 429-444, 1978.
- Chavas, J.; Aliber, M. An analysis of economic efficiency in agriculture: a nonparametric approach. **Journal of Agriculture and Resource Economics**, v. 18, p. 1-16, 1993.
- Coelli, T.J. Recent developments in frontier modelling and efficiency measurement. **Australian Journal of Agricultural Economics**, v. 39, n. 3, p. 219-245, 1995.
- Cooper, W.W.; Park, K.S.; Yu, G. IDEA and AR-IDEA: models for dealing with imprecise data in DEA. **Management Science**, v. 45, p. 597-607, 1999.
- Cooper, W.W.; Park, K.S.; Yu, G. An illustrative application of IDEA (Imprecise Data Envelopment Analysis) to a Korean mobile telecommunication company. **Operations Research**, v. 49, n. 6, p. 807-820, 2001.
- Cooper, W.W.; Seiford, L.M.; Tone, K. **Data Envelopment Analysis: a comprehensive text with models, applications, references and DEA-Solver software**. Boston: Kluwer Academic Publishers, 2000.
- Deprins, D.; Simar, L.; Tulkens, H. Measuring labor inefficiency in post offices. In: Marchand, M.; Pestieau, P.; Tulkens, H. (eds.) **The Performance of Public Enterprises: Concepts and Measurements**. Amsterdam: North-Holland, p. 243-267, 1984.
- Despotis, D.K.; Smirlis, Y.G. Data envelopment analysis with imprecise data. **European Journal of Operational Research**, v. 140, p. 24-36, 2002.

Dial, G.; Grodecki, J. Applications of IKONOS imagery. In: American Society for Photogrammetry and Remote Sensing 2003 Annual Conference, Anchorage, Alaska. **Proceedings...**, 2003.

Entani, T.; Maeda, Y.; Tanaka, H. Dual models of interval DEA and its extensions to interval data. **European Journal of Operational Research**, v. 136, p. 32-45, 2002.

Everitt, B. **Cluster analysis**. London: Heinemann Educational, 1974.

Färe, R.; Grabowski, R.; Grosskopf, S. Technical efficiency in Philippine agriculture. **Applied Economics**, v. 17, p. 205-214, 1985.

Gomes, E.G. **Modelos de análise de envoltória de dados com ganhos de soma zero**. Rio de Janeiro, 2003. Tese (Doutorado em Engenharia de Produção), Universidade Federal do Rio de Janeiro.

Gomes, E.G.; Mangabeira, J.A.C. Uso de análise de envoltória de dados em agricultura: o caso de Holambra. **Engevista**, v. 6, n. 1, p. 19-27, 2004.

Gomes, E.G.; Soares de Mello, J.C.C.B.; Biondi Neto, L. Avaliação de eficiência por análise de envoltória de dados: conceitos, aplicações à agricultura e integração com sistemas de informação geográfica. **Documentos**, n. 28. Campinas: Embrapa Monitoramento por Satélite, 2003. Disponível em: <[http://www.cnpm.embrapa.br/publica/download/d28\\_envoltdados.vf.pdf](http://www.cnpm.embrapa.br/publica/download/d28_envoltdados.vf.pdf)>. Acesso em: 14 dez. 2005.

Guo, P.; Tanaka, H. Fuzzy DEA: a perceptual evaluation method. **Fuzzy Sets and Systems**, v. 119, p. 149-160, 2001.

Haghighat, M.S.; Khorram, E. The maximum and minimum number of efficient units in DEA with interval data. **Applied Mathematics and Computation**, v. 163, p. 919-930, 2005.

Holambra em números. **Holambra Hoje**, agosto, 1998. (3<sup>a</sup> capa).

Hollingsworth, G.; Smith, P. Use of ratios in data envelopment analysis. **Applied Economics Letters**, v. 10, p. 733-735, 2003.

Hougaard, J.L. Fuzzy scores of technical efficiency. **European Journal of Operational Research**, v. 115, p. 529-541, 1999.

Instituto Brasileiro de Geografia e Estatística. **Cidades@**. Disponível em: <<http://www.ibge.gov.br/cidadesat/default.php>>. Acesso em: 13 dez. 2005.

Jahanshahloo, G.R.; Matin, R.K.; Vencheh, A.H. On FDH efficiency analysis with interval data. **Applied Mathematics and Computation**, v. 159, p. 47-55, 2004.

Kao, C.; Liu, S.T. Fuzzy efficiency measures in data envelopment analysis. **Fuzzy Sets and Systems**, v. 113, p. 427-437, 2000.

Lertworasirikul, S.; Fang, S.C.; Joines, J.A.; Nuttle, H.L.W. Fuzzy data envelopment analysis (DEA): a possibility approach. **Fuzzy Sets and Systems**, v. 139, n. 2, p. 379-394, 2003.

Leta, F.R.; Soares de Mello, J.C.C.B.; Gomes, E.G.; Angulo Meza, L. Métodos de melhora de ordenação em DEA aplicados à avaliação estática de tornos mecânicos. **Investigação Operacional**, v. 25, n. 2, 2005.

Lins, M.P.E.; Angulo Meza, L.; Silva, A.C.M. A multi-objective approach to determine alternative targets in data envelopment analysis. **Journal of the Operational Research Society**, v. 55, p. 1090-1101, 2004.

Lins, M.P.E.; Novaes, L.F.L.; Legey, L.F.L. Real estate appraisal: a double perspective data envelopment analysis approach. **Annals of Operations Research**, v. 138, p. 79-96, 2005.

Lopes, A.L.M.; Lanzer, E.A. Data envelopment analysis – DEA and fuzzy sets to assess the performance of academic departments: a case study at Federal University of Santa Catarina – UFSC. **Pesquisa Operacional**, v. 22, n. 2, p. 217-230, 2002.

Mangabeira, J.A. de C. **Tipificação de produtores rurais apoiada em imagens de alta resolução espacial, geoprocessamento e estatística multivariada: uma proposta metodológica**. Campinas, 2002. Dissertação (Mestrado em Planejamento e Desenvolvimento Rural Sustentável), Universidade Estadual de Campinas.

Ray, S. Measurement and test of efficiency of farmers in linear programming models: a study of West Bengal farms. **Oxford Bulletin of Economics and Statistics**, n. 47, p. 371-386, 1985.

Sengupta, J.K. A fuzzy systems approach in data envelopment analysis. **Computers & Mathematics with Applications**, v. 24, n. 8-9, p. 259-266, 1992.

Soares de Mello, J.C.C.B.; Angulo Meza, L.; Gomes, E.G.; Biondi Neto, L. Paradoxos em modelos DEA BCC. In: 11º Congresso da Associação Portuguesa de Investigação Operacional, 2004, Porto. **Actas...**, p. 62-62, 2004.

Soares de Mello, J.C.C.B.; Gomes, E.G.; Angulo Meza, L.; Biondi Neto, L.; Sant'anna, A.P. Fronteiras DEA difusas. **Investigação Operacional**, v. 25, n. 1, p. 85-103, 2005.

Tavares, G. **Data envelopment analysis: basic models and their main extensions. A model to analyse the modernisation of telecommunication services in OECD countries**. Coimbra, 1998. Tese (Mestrado em Ciências e Tecnologia), Universidade de Coimbra.

Triantis, K.; Eeckaut, P.V. Fuzzy pair-wise dominance and implications for technical efficiency performance assessment. **Journal of Productivity Analysis**, v. 13, p. 207-230, 2000.

Triantis, K.; Girod, O. A mathematical programming approach for measuring technical efficiency in a fuzzy environment. **Journal of Productivity Analysis**, v. 10, n. 1, p. 85-102, 1998.

Wang, Y.M.; Greatbanks, R.; Yang, J.B. Interval efficiency assessment using data envelopment analysis. **Fuzzy Sets and Systems**, v. 153, p. 347-370, 2005.

Yamada, Y.; Matui, T.; Sugiyama, M. New analysis of efficiency based on DEA. **Journal of the Operations Research Society of Japan**, v. 37, n. 2, p. 158-167, 1994.

Zadeh, L. Fuzzy sets as a basis for a theory of possibility. **Fuzzy Sets and Systems**, v. 1, p. 3-28, 1978.

Zadeh, L. Fuzzy Sets. **Information and Control**, v. 8, n. 3, p. 338-353, 1965.

Zhu, J. Imprecise data envelopment analysis (IDEA): a review and improvement with an application. **European Journal of Operational Research**, v. 144, p. 513-529, 2003.



**A Dados do estudo de caso.**

DMU	Área total (ha) limite inferior	Área total (ha) limite superior	Emprego (homem-hora-ano)	Uso de máquinas (hora-ano)	Renda líquida (R\$-ano)
DMU1	32,41	34,30	1.005	151,4	39.955
DMU2	34,00	34,90	41.571	1.645	481.536
DMU3	39,60	41,10	23.130	1.775	192.920
DMU4	6,09	6,10	16.380	720	125.000
DMU5	7,28	7,30	219	36,5	21.028
DMU6	11,80	11,90	66	55	1.700
DMU7	7,30	7,36	219	36,5	13.728
DMU8	6,00	6,40	12	20	1.340
DMU9	41,00	42,50	30.000	3.236	15.000
DMU10	5,30	5,70	312	12	3.370
DMU11	8,00	9,00	750	56	15.120
DMU12	16,70	17,00	2.772	72	51.360
DMU13	5,40	5,80	300	0	9.600
DMU14	14,50	14,60	42	84	2.920
DMU15	31,40	33,70	1.800	150	63.200
DMU16	19,50	19,80	2.040	54	319.050
DMU17	15,90	16,70	15.600	752	112.000
DMU18	19,40	19,80	11.400	152	9.000
DMU19	72,00	74,70	23.640	5.316	892.000
DMU20	14,60	15,00	9.000	120	7.500
DMU21	3,80	4,20	9.000	225	224.000
DMU22	26,30	27,20	10.458	1.788	805.070
DMU23	24,50	25,60	5.700	1.230	94.000
DMU24	18,30	18,50	4.350	276	142.621
DMU25	27,60	28,20	4.470	330	179.500
DMU26	15,30	15,70	4.692	1.100	139.700
DMU27	29,80	30,00	840	280	112.000
DMU28	12,20	13,00	25.500	1.820	390.000
DMU29	10,00	10,70	1.200	800	76.000
DMU30	9,20	10,00	5.526	170	34.696
DMU31	20,30	20,36	2.736	190	84.923
DMU32	20,70	21,50	15.612	744	37.217
DMU33	16,00	16,30	8.400	360	16.000
DMU34	25,60	28,00	16.830	900	3.000
DMU35	8,20	8,40	4.500	1.098	105.000
DMU36	32,80	38,70	76.55	553,5	21.097
DMU37	17,10	18,00	9.030	1.100	19.500
DMU38	28,70	29,00	23.400	1.122	32.512
DMU39	3,80	3,81	3.600	900	1.200.000
DMU40	8,30	8,40	16.200	2.025	250.000
DMU41	30,10	30,20	4.872	645	88.000

DMU	Área total (ha) limite inferior	Área total (ha) limite superior	Emprego (homem-hora-ano)	Uso de máquinas (hora-ano)	Renda líquida (R\$-ano)
DMU42	35,00	36,80	1.401	205	72.250
DMU43	30,00	30,20	14.400	72	40.080
DMU44	9,80	10,00	1.848	186	13.200
DMU45	34,00	37,00	12.960	900	465.000
DMU46	25,70	26,00	19.320	725	303.700
DMU47	24,00	24,50	402	80	22.000
DMU48	25,00	25,20	480	54	1.825
DMU49	15,00	15,30	1.263	113,6	14.588
DMU50	19,80	20,00	693	82,5	35.360
DMU51	5,80	6,50	918	9	63.875
DMU52	2,47	2,50	300	0	66.000
DMU53	16,90	17,00	1.500	90	4.016
DMU54	10,20	10,50	600	0	576
DMU55	30,00	31,60	300	0	2.880
DMU56	33,00	36,80	5.886	832	59.000
DMU57	21,20	22,50	1.335	99,5	75.850
DMU58	24,00	25,20	12.780	1.148	129.400
DMU59	1,46	1,50	2.400	0	52.000
DMU60	1,98	2,00	1.500	0	27.000
DMU61	22,50	23,00	6.540	204	65.998
DMU62	27,00	28,20	4.830	546	32.633
DMU63	9,50	10,30	13.200	200	108.000
DMU64	2,30	2,40	9.000	0	120.000
DMU65	32,70	32,90	6.540	292	40.700
DMU66	4,30	4,40	3.000	100	18.000
DMU67	16,80	17,00	11.475	435	31.000
DMU68	7,30	7,39	1.350	50	16.000
DMU69	13,00	13,10	1.710	42	22.000
DMU70	10,30	12,00	3.240	0	15.000
DMU71	3,60	4,20	3.600	150	6.000

## B Resultados do modelo DEA de fronteira difusa.

	Fronteira original			Fronteira invertida			$Ef_{difusa}$
	$Eff_o$	$Eff_p$	$\varphi_{original}$	$Eff_o$	$Eff_p$	$\varphi_{invertida}$	
DMU1	0,1605	0,1605	0	0,7900	0,7738	0	0,5
DMU2	0,1710	0,1709	0	1,0000	1,0000	1	0,0
DMU3	0,0585	0,0581	0	0,8382	0,8374	0	0,5
DMU4	0,2642	0,2700	0	0,5005	0,5005	0	0,5
DMU5	0,6169	0,6393	0	0,1709	0,1766	0	0,5
DMU6	0,4875	0,5129	0	0,5189	0,5228	0	0,5
DMU7	0,6157	0,6356	0	0,1780	0,1832	0	0,5

	Fronteira original			Fronteira invertida			<b>Ef</b> <sub>difusa</sub>
	<i>Ef<sub>fo</sub></i>	<i>Ef<sub>fp</sub></i>	<i>φ<sub>original</sub></i>	<i>Ef<sub>fo</sub></i>	<i>Ef<sub>fp</sub></i>	<i>φ<sub>invertida</sub></i>	
DMU8	1,0000	1,0000	1	0,3347	0,3199	0	1,0
DMU9	0,0472	0,0465	0	1,0000	1,0000	1	0,0
DMU10	0,6737	0,6612	0	0,1782	0,1745	0	0,5
DMU11	0,3575	0,3426	0	0,2117	0,1951	0	0,5
DMU12	0,1450	0,1443	0	0,3878	0,3949	0	0,5
DMU13	1,0000	1,0000	1	0,1541	0,1496	0	1,0
DMU14	0,4121	0,4340	0	0,4722	0,4964	0	0,5
DMU15	0,1437	0,1437	0	0,7612	0,7353	0	0,5
DMU16	1,0000	1,0000	1	0,3690	0,3768	0	1,0
DMU17	0,1178	0,1153	0	0,4814	0,4814	0	0,5
DMU18	0,1051	0,1048	0	0,5518	0,5669	0	0,5
DMU19	0,1198	0,1198	0	1,0000	1,0000	1	0,0
DMU20	0,1381	0,1370	0	0,4440	0,4570	0	0,5
DMU21	0,5303	0,5161	0	0,2558	0,2558	0	0,5
DMU22	0,3039	0,3032	0	0,3985	0,3991	0	0,5
DMU23	0,0987	0,0960	0	0,5639	0,5595	0	0,5
DMU24	0,1878	0,1880	0	0,3921	0,4021	0	0,5
DMU25	0,1966	0,1960	0	0,5810	0,5896	0	0,5
DMU26	0,1599	0,1580	0	0,3335	0,3369	0	0,5
DMU27	0,4086	0,4086	0	0,6513	0,6707	0	0,5
DMU28	0,1761	0,1677	0	0,7097	0,7097	0	0,5
DMU29	0,2694	0,2670	0	0,2392	0,2366	0	0,5
DMU30	0,2205	0,2092	0	0,2314	0,2206	0	0,5
DMU31	0,1286	0,1288	0	0,4518	0,4670	0	0,5
DMU32	0,0927	0,0913	0	0,5110	0,5110	0	0,5
DMU33	0,1305	0,1302	0	0,3832	0,3899	0	0,5
DMU34	0,0776	0,0734	0	1,0000	1,0000	1	0,0
DMU35	0,2632	0,2613	0	0,3183	0,3183	0	0,5
DMU36	0,0717	0,0624	0	0,9058	0,7958	0	0,5
DMU37	0,1219	0,1185	0	0,4219	0,4155	0	0,5
DMU38	0,0654	0,0658	0	0,7689	0,7689	0	0,5
DMU39	1,0000	1,0000	1	0,1693	0,1693	0	1,0
DMU40	0,2245	0,2260	0	0,5560	0,5560	0	0,5
DMU41	0,0823	0,0830	0	0,6684	0,6907	0	0,5
DMU42	0,1926	0,1926	0	0,8251	0,8135	0	0,5
DMU43	0,0708	0,0713	0	0,6955	0,7162	0	0,5
DMU44	0,2446	0,2429	0	0,2446	0,2489	0	0,5
DMU45	0,2781	0,2736	0	0,6269	0,5975	0	0,5
DMU46	0,2080	0,2082	0	0,5839	0,5897	0	0,5
DMU47	0,2394	0,2394	0	0,5730	0,5819	0	0,5
DMU48	0,2386	0,2386	0	1,0000	1,0000	1	0,0
DMU49	0,2017	0,2025	0	0,3632	0,3692	0	0,5
DMU50	0,2242	0,2282	0	0,4625	0,4746	0	0,5
DMU51	0,4189	0,3810	0	0,1467	0,1357	0	0,5
DMU52	1,0000	1,0000	1	0,0563	0,0577	0	1,0
DMU53	0,1742	0,1759	0	0,5211	0,5446	0	0,5
DMU54	1,0000	1,0000	1	1,0000	1,0000	1	0,5

	Fronteira original			Fronteira invertida			<b>Ef</b> <sub>difusa</sub>
	$Ef_{fo}$	$Ef_{fp}$	$\varphi_{original}$	$Ef_{fo}$	$Ef_{fp}$	$\varphi_{invertida}$	
DMU55	1,0000	1,0000	1	1,0000	1,0000	1	0,5
DMU56	0,0730	0,0667	0	0,8342	0,7755	0	0,5
DMU57	0,2304	0,2304	0	0,5030	0,4913	0	0,5
DMU58	0,0925	0,0900	0	0,5396	0,5328	0	0,5
DMU59	1,0000	1,0000	1	0,0776	0,0776	0	1,0
DMU60	0,9677	0,9737	0	0,0495	0,0495	0	0,5
DMU61	0,1022	0,1014	0	0,5183	0,5257	0	0,5
DMU62	0,0892	0,0867	0	0,6536	0,6487	0	0,5
DMU63	0,1905	0,1854	0	0,4086	0,4086	0	0,5
DMU64	1,0000	1,0000	1	0,2760	0,2760	0	1,0
DMU65	0,0729	0,0734	0	0,7573	0,7803	0	0,5
DMU66	0,4552	0,4534	0	0,1033	0,1046	0	0,5
DMU67	0,1171	0,1176	0	0,3945	0,4042	0	0,5
DMU68	0,3289	0,3290	0	0,1737	0,1779	0	0,5
DMU69	0,1891	0,1899	0	0,3064	0,3152	0	0,5
DMU70	1,0000	1,0000	1	0,2824	0,2512	0	1,0
DMU71	0,4904	0,4469	0	0,1789	0,1789	0	0,5

# Inferência Estatística dos Estimadores de Eficiência obtidos com a Técnica Fronteira Não Paramétrica de DEA. Uma Metodologia de *Bootstrap*

Rui Cunha Marques <sup>†</sup>  
Duarte Silva <sup>‡</sup>

<sup>†</sup> Centro de Sistemas Urbanos e Regionais  
Instituto Superior Técnico  
Universidade Técnica de Lisboa  
rcmar@civil.ist.utl.pt

<sup>‡</sup> Rave  
dnsilva@rave.pt

---

## Abstract

The efficiency measurement of decision making units is a key-task in contemporary societies. Despite the existence of several methodologies with this aim none of them is clearly superior. One of the most used is the non-parametric technique of data envelopment analysis (DEA). The DEA method has many advantages, therefore, being very popular and widely adopted. However, it doesn't allow for the results statistical inference, which can constrain its use. This document provides some insights for that debate, discussing and applying the DEA technique to 70 water services in Portugal. Later, a re-sampling (bootstrap) methodology is applied to the results attained by DEA, enabling its statistical inference. The quality of the values obtained by the bootstrap depends much on the average efficiency level and on its dispersion. That is, if the average inefficiency gap and its dispersion are considerable, the results will be far from the ones wished. Nevertheless, under the opposite circumstances, the results are interesting and satisfactory, surpassing the traditional drawbacks pointed at DEA.

## Resumo

A medição de eficiência de unidades organizatórias ou unidades de decisão constitui uma tarefa-chave nas sociedades contemporâneas. Para a sua prossecução existem várias metodologias disponíveis embora nenhuma seja claramente superior. Um dos métodos mais conhecidos, e aqui analisado, consiste na técnica não paramétrica *data envelopment analysis* (DEA). Esta técnica possui diversas vantagens, sendo, portanto muito popular e largamente adoptada. Todavia, a DEA não permite a inferência estatística dos resultados obtidos, o que condiciona o seu uso empírico. Este documento fornece alguns contributos para esse debate, discutindo e aplicando a técnica DEA a 70 serviços de água em Portugal. Posteriormente, é empregue uma metodologia de *bootstrap* (reamostragem) aos estimadores de eficiência obtidos com a DEA, permitindo a sua

inferência estatística. A qualidade dos resultados alcançados com o *bootstrap* depende muito do nível de eficiência médio e da sua dispersão. Isto é, quando a ineficiência média é considerável, tal como a sua dispersão, os resultados estão longe do desejável. Porém, quando o valor da ineficiência é reduzido e a sua amplitude é baixa os resultados são satisfatórios e bem interessantes, ultrapassando alguns dos problemas tradicionais apontados à DEA.

**Keywords:** Efficiency, DEA, Statistical Inference, Bootstrap, Water Services

**Title:** Statistical Inference of Efficiency Estimators obtained with the DEA Nonparametric Frontier Technique. A Bootstrap Methodology

---

## 1 Introdução

A noção de eficiência corresponde à comparação entre os valores dos *inputs* e dos *outputs* observados associados à realização de uma certa actividade, e os seus valores óptimos. A sua obtenção pode ser efectuada numa perspectiva de minimização de *inputs*, isto é, do rácio entre a quantidade observada de *inputs* consumidos e a quantidade mínima de *inputs* para uma quantidade idêntica de *outputs* produzida ou, numa lógica complementar, de maximização de *outputs*, do rácio entre os *outputs* observados e os *outputs* máximos possíveis para uma quantidade fixa de *inputs* utilizados. Em ambas as situações é calculada a eficiência técnica (ET) que mede a intensidade dos desperdícios verificados no processo produtivo. Para uma unidade organizatória ou unidade de decisão (DMU) pode ainda determinar-se a eficiência alocativa de preços (EA) que reflecte, no modelo com orientação de *inputs*, qual a percentagem de custo que pode ser reduzida passando de uma posição da DMU tecnicamente eficiente para uma posição também sobre a fronteira mas que corresponda ao custo de produção mínimo, tendo em conta os preços actuais dos *inputs*. O produto da ET pela EA corresponde à eficiência económica ou global (Farrell, 1957).

Os métodos de cálculo da eficiência têm sido classificados em métodos paramétricos e em não paramétricos, consoante admitam, ou não, uma forma funcional, definida *a priori*, para a tecnologia de produção. Os métodos paramétricos possibilitam a medição do erro, mas introduzem a dificuldade adicional de conhecer a especificação associada ao comportamento admitido (Coelli *et al.*, 1998). Entre os métodos mais aplicados, a *data envelopment analysis* (DEA) e os números índices são considerados métodos não paramétricos, ao passo que as fronteiras estocásticas (SFA) e os modelos de regressão (e. g. OLS e COLS) são paramétricos. Na figura 1, onde o processo produtivo é caracterizado por um único *input* ( $x$ ) e um único *output* ( $y$ ), observa-se a diferença entre estas duas classes de métodos. A DEA é um método não paramétrico, ou seja, é empiricamente baseado, enquanto os restantes (OLS, COLS e SFA) requerem a especificação de uma função para a tecnologia de produção. Os métodos paramétricos ou não paramétricos podem ainda ser classificados em não fronteira ou em fronteira, em função de pressuporem, ou não, que as DMU sejam tecnicamente eficientes ou, de outra forma, dependendo se os *benchmarks* assentam nas aproximações médias ou nas *best practices* (Coelli *et al.*, 1998). A DEA e as SFA constituem métodos fronteira, ao passo que os números índices e os modelos de regressão (e. g. OLS), desde que não corrigidos (COLS), são não fronteira. Uma vez mais, a figura 1 permite clarificar esta classificação. A DEA, a SFA e a COLS são métodos fronteira, dado que se baseiam nas DMU com valores limite da amostra (*best practices*), enquanto a OLS é não fronteira, apoiando-se no seu ajustamento médio.

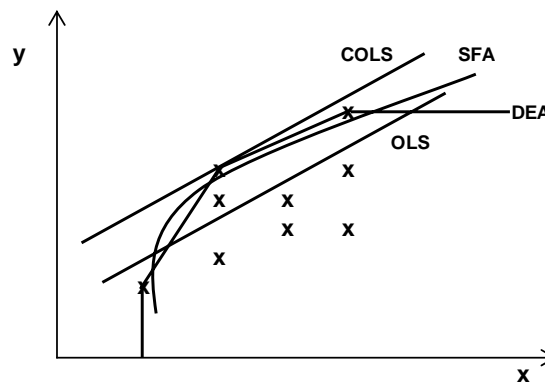


Figura 1: Tecnologias de produção obtidas por diferentes métodos

A superioridade dos métodos fronteira em relação aos não fronteira parece evidente e relativamente consensual. O mesmo não ocorre, porém, dentro dos métodos fronteira, em relação aos não paramétricos e aos paramétricos ou vice-versa, nomeadamente entre a DEA e SFA. Não obstante o argumento da maior aplicabilidade da técnica DEA possa sustentar a sua supremacia (Tavares, 2002), o facto de esta não permitir a inferência estatística minimiza o seu potencial de aplicação, sobretudo no domínio empírico. Por outro lado, a SFA é de cálculo complexo e requer fortes assunções sobre a distribuição estatística do erro sem justificação teórica. Os reguladores batem-se, comumente, com essa dificuldade, quando pretendem utilizar a técnica DEA de uma forma explícita, mas são confrontados com a oposição das entidades gestoras, em virtude de os resultados obtidos e dos modelos usados não estarem sujeitos a inferência estatística (*vide* Bosworth *et al.*, 1996; Burns *et al.*, 2000 e Sarafadis, 2002).

Este documento, assente num trabalho de investigação recente (Marques, 2005a), pretende providenciar contributos nesse sentido, analisando e aplicando a técnica de *bootstrap* aos estimadores da eficiência obtidos com a DEA, usando como caso de estudo os serviços de água portugueses. Os serviços de água, para além de prestarem um serviço de interesse económico geral, funcionam em regime de monopólio natural, exigem investimentos irrecuperáveis e apresentam economias de escala, de gama e de densidade relevantes. Em Portugal, a indústria da água envolve um valor de negócio anual superior a 1500 milhões de euros, acreditando-se que, pelas suas características e pelo seu histórico (mercado muito desregulado), apresenta níveis de ineficiência muito significativos (Marques, 2005b). O emprego de DEA neste sector é, por isso, da maior pertinência, embora o seu uso empírico (e. g. na regulação económica deste sector) requeira uma elevada robustez da análise. Para esse fim, existe a ferramenta de *bootstrap* que, sendo um instrumento de simulação estatística, possibilita atestar a qualidade dos modelos concebidos e dos resultados alcançados. O artigo, após esta introdução, expõe, numa primeira parte, sumariamente, a técnica de DEA. Seguidamente, esta ferramenta é aplicada a 70 serviços de água. Numa terceira parte, a metodologia de *bootstrap* é discutida e testada, usando os resultados alcançados anteriormente com a DEA.

## 2 Data Envelopment Analysis

### 2.1 Introdução

A DEA é um método assente em programação linear para avaliação da eficiência relativa de unidades mais ou menos homogêneas (DMU), tal como, *inter alia*, escolas, hospitais e

serviços de infra-estruturas, que transformam um conjunto de recursos ou factores de produção (*inputs*) num conjunto de produtos ou resultados desejados (*outputs*). A DEA não estabelece uma forma funcional para a fronteira eficiente e não emprega valores específicos para os pesos. A técnica de DEA é empiricamente baseada (a construção da fronteira), ao contrário das abordagens paramétricas concorrentes.

A aplicação da técnica de DEA possui várias vantagens, das quais se destacam: a) identificação para cada DMU ineficiente de um conjunto de DMU (*peers*) eficientes com combinação de *inputs* e *outputs* semelhante; b) facilidade em lidar com múltiplos *inputs* e *outputs*; c) adopção dos melhores resultados como elementos de comparação; d) não admissão de uma forma paramétrica para a fronteira ou para a ineficiência quando associada ao erro; e) natureza conservativa das avaliações; e f) decomposição da natureza da eficiência em várias componentes. Todavia, a técnica de DEA revela alguns inconvenientes, a saber, a sensibilidade elevada aos *outliers*, a exigência em termos de informação requerida e a dificuldade de medir o erro associado ou de testar estatisticamente os resultados e os modelos adoptados. Desta forma, também, na DEA, a análise dos factores explanatórios é complexa, dependendo da correlação existente.

## 2.2 Formulação Base da DEA

Tendo como base o trabalho de Farrell (Farrell, 1957), Charnes, Cooper e Rhodes desenvolveram o modelo CCR, caracterizado pela assunção de rendimentos constantes à escala (RCE) e da sobreutilização (congestão) das variáveis (Charnes *et al.*, 1978). Considere que uma DMU produz os *outputs*  $y_i$ ,  $i = 1, \dots, I$  a partir dos *inputs*  $x_j$ ,  $j = 1, \dots, J$ , com os respectivos pesos  $a_i$  e  $b_j$  e que existem  $M$  DMU. O problema primal do modelo de DEA, também designado por modelo dos multiplicadores, corresponde à seguinte formulação para a orientação de minimização de *inputs*, no qual  $W_k$  diz respeito à eficiência da DMU  $k$  e  $\varepsilon$  é uma quantidade não Arquimediana (na ordem de  $10^{-5}$  ou  $10^{-6}$ ):

$$\begin{aligned} \text{Max: } W_k &= \sum_{i=1}^I a_i y_{im} \\ \text{s.a.} \\ \sum_{j=1}^J b_j x_{jm} &\geq \sum_{i=1}^I a_i y_{im} & m = 1, \dots, k, \dots, M \\ \sum_{j=1}^J b_j x_{jk} &= 1 \\ a_i, b_j &> \varepsilon & i = 1, \dots, I; j = 1, \dots, J \end{aligned} \quad (1)$$

Na programação linear existe a formulação dual, denominada de problema da envolvente. O dual constrói uma aproximação à fronteira verdadeira, formada pela união das DMU eficientes que minimizam a quantidade *inputs* para um nível desejado de produção de *outputs*. No primal a escolha dos pesos reside sobre os *inputs* e os *outputs*, enquanto no dual a selecção dos “pesos” recai sobre as DMU. O problema dual consiste em:



$$\begin{aligned}
& \text{Min: } h_k - \varepsilon \left( \sum_{i=1}^I s_i + \sum_{j=1}^J s_j \right) \\
& \text{s.a.} \\
& \sum_{m=1}^M \lambda_m x_{jm} = x_{jk} h_k - s_j \quad j = 1, \dots, J \\
& \sum_{m=1}^M \lambda_m y_{im} = y_{ik} + s_i \quad i = 1, \dots, I \\
& \lambda_m \geq 0 \quad m = 1, \dots, k, \dots, M \\
& s_i, s_j \geq 0 \quad i = 1, \dots, I; j = 1, \dots, J
\end{aligned} \tag{2}$$

onde  $\lambda_m$  são as ponderações das DMU de referência,  $S_i$  são as folgas dos *outputs* e  $S_j$  são as folgas dos *inputs*. Os  $\lambda_m$  representam a proporção de *inputs* e de *outputs* da DMU  $m$  utilizados para construir os *targets* da DMU  $k$  em avaliação. As folgas, introduzidas como variáveis do modelo, representam fontes adicionais de ineficiência, isto é, traduzem em que medida os *inputs* podem ser reduzidos após as contracções radiais terem sido efectuadas. A DMU  $k$  é eficiente se e só se  $h_k$  for igual à unidade e as folgas (*slacks*) forem iguais a zero. Neste caso, as DMU designam-se por eficientes e são exemplos das melhores práticas (*best practices*).

O problema dual é computacionalmente menos exigente do que o primal, visto que é munido de um número inferior de restrições ( $I+J < M+1$ ). Todavia, um algoritmo de programação linear é resolvido para cada DMU da amostra, obtendo-se o valor de  $h_k$  respeitante à sua eficiência. Este valor é alcançado através da contracção radial do vector de *inputs*, tanto quanto possível, para uma dada quantidade de *outputs* que ainda pertence à tecnologia de produção, o que resulta na projecção da DMU na fronteira eficiente. A fronteira de produção é determinada pela isoquântica formada pelas DMU da amostra que conduzem ao menor consumo de *inputs* para a produção de *outputs* desejados. O *target* ou o ponto projectado corresponde a uma combinação linear das DMU que pertencem à fronteira e que constituem os seus pares.

### 2.3 Extensões ao Modelo Base de DEA

O modelo CCR foi estendido em 1984, procurando contabilizar a hipótese de RVE (Banker *et al.*, 1984). Nesta situação, a natureza da ET pode ser separada em eficiência técnica pura (ETP) e em eficiência de escala (EE), se se relacionarem os seus resultados com os dos modelos CCR. Este modelo ficou conhecido como modelo BCC. No modelo CCR a ET obtida poderia dever-se à ineficiência técnica propriamente dita ou à operação a uma escala não óptima, não se distinguindo esse efeito. O modelo BCC pode ser resolvido modificando ligeiramente as formulações 1 e 2, adicionando-lhes uma restrição que garanta que a DMU em análise seja comparada com uma combinação convexa das DMU da amostra. Esta abordagem, mais restritiva, forma um invólucro convexo em vez de cónico sobre as DMU, assegurando, assim, que as DMU só sejam comparadas com DMU de tamanho equivalente. A medida de eficiência, obtida por esta via (ETP), será sempre superior ou igual à alcançada com o modelo CCR. A resolução da formulação BCC

determina a ETP, a qual, conjugada com a ET auferida por 1 (ou 2), permite obter a EE. A EE mede a economia de recursos que se verificaria se a escala de operação fosse ótima.

Em meados dos anos 80, outros modelos foram introduzidos como os denominados modelos multiplicativos, os quais admitem uma forma particular para a envolvente, como a função log-linear ou a Cobb-Douglas, o modelo aditivo, baseado no conceito de ótimo de Pareto e o modelo BCC, abandonando o pressuposto da sobreutilização das variáveis. Os modelos CCR e BCC foram também adaptados para integrar variáveis que são não discriminárias, isto é, que não estão dependentes da acção dos decisores e permitem a presença de variáveis ordinais. Estes e outros modelos estão explanados em (*vide* Charnes *et al.*, 1994 e Cooper *et al.*, 2000), não se apresentando aqui os artigos originais por uma questão de espaço.

## 2.4 Especificação dos Modelos de DEA

O cálculo da eficiência através da DEA é efectuado sobre os dados de uma determinada amostra não deixando, por isso, de constituir uma estimação. Logo, não obstante a técnica de DEA ser considerada, vulgarmente, como determinística, desde muito cedo existiu a vontade de lhe instituir atributos estatísticos. Esta preocupação tornou-se mais visível com o número crescente de aplicações empíricas, que demonstraram que os resultados da DEA eram falíveis e que estavam muito dependentes da qualidade da amostra e da especificação dos modelos. Contudo, a DEA “estatística” está ainda na sua infância (Grosskopf, 1996; Simar e Wilson, 2000a e Marques, 2005a). Com efeito, nas situações de dimensionalidade elevada (Kittelsen, 1999), de correlação significativa entre variáveis (Pedraja-Chaparro *et al.*, 1999) e para as DMU com valores extremos (*input* mínimo ou *output* máximo no caso de RVE), e quando a amostra é reduzida, os resultados obtidos são enviesados (Banker, 1996 e Pedraja-Chaparro *et al.*, 1999). Ademais, a especificação do modelo pode não ser robusta, isto é, omitir variáveis importantes, incluir variáveis irrelevantes ou, ainda, escolher uma forma funcional inadequada. Por exemplo, a adição de uma nova variável aumenta o enviesamento médio nas estimativas da eficiência (Kittelsen, 1999).

Em 1987, Färe e Primont demonstraram que a desagregação de um *input* do modelo em dois *inputs* só pode piorar o resultado da função objectivo, isto é, melhorar a eficiência das DMU (Färe e Primont, 1987). Um resultado mais generalizado consiste no facto de a adição de uma variável ao modelo nunca poder piorar a eficiência das DMU. Este resultado tem consequências importantes, permitindo a realização de testes de hipóteses entre dois modelos, sobretudo quando existe desagregação ou quando um deles é alojado (*nested*) no outro, caso típico quando se acrescenta uma variável ao modelo e se quer estudar o seu impacto. Assim, testando se a mudança da especificação do modelo de A para B, onde B tem mais uma variável, altera os resultados significativamente, isto é, se se aceita a hipótese  $H_0$  na qual a estimativa da eficiência de A ( $\hat{E}_A$ ) é equivalente à estimativa da eficiência de B ( $\hat{E}_B$ ), vem:

$$H_0: \hat{E}_{A,I} = \hat{E}_{B,i}; H_1: \hat{E}_{A,I} < \hat{E}_{B,i} \quad (3)$$

O mesmo teste pode ser efectuado, admitindo a mesma especificação das variáveis do modelo, mas a existência de tecnologias distintas, porque se sabe que (Grosskopf, 1986):

$$\hat{E}_{A,I} (RCE) \leq \hat{E}_{A,I} (RVE) \quad (4)$$

As questões que se colocam são se existem e quais os testes estatísticos a adoptar. Banker sugeriu dois testes assintóticos, assumindo duas especificações da ineficiência diferentes, designadamente, a meia normal ( $F^m$ ) e a exponencial ( $F^e$ ) (Banker, 1993 e 1996). Se a assunção da distribuição da ineficiência não for possível, Banker sugere o uso do teste não paramétrico de Kolmogorov-Smirnov (K-S) para verificar a igualdade das duas distribuições (Banker, 1993). Um outro teste sugerido, que evita o problema da especificação da distribuição das eficiências, consiste no teste T para comparação de médias (Kittelsen, 1998).

A metodologia baseada nos testes estatísticos dá origem à conhecida abordagem *stepwise* (ST) de Kittelsen da especificação do modelo. Esta técnica, cuja aplicação será efectuada de seguida, é sintetizada pelos seguintes passos (Kittelsen, 1998): a) definir um modelo base que contenha todas as variáveis consideradas essenciais por razões empíricas ou teóricas; b) obter a estimativa da eficiência para o modelo; c) calcular a estimativa da eficiência com uma variável candidata ou com as variáveis desagregadas; e d) determinar os 4 testes atrás referidos ( $F^m$ ,  $F^e$ , K-S e T) e aceitar, ou não, o novo modelo. Caso existam mais variáveis a desagregar ou novas variáveis candidatas, voltar a b). Embora o modelo ST exija, do ponto de vista teórico, requisitos que, em geral, não se verificam, como a dimensão da amostra ou a independência (Simar e Wilson, 2002), Kittelsen demonstrou, através de simulação de Monte Carlo, que, no mundo empírico, as distorções ocorridas pelo seu incumprimento não têm os efeitos que, à partida, se poderiam esperar (Kittelsen, 1998).

### 3 Caso de Estudo

#### 3.1 Dados

A aplicação envolveu 70 serviços de água, abrangendo cerca de 63 % da população portuguesa. As DMU incluídas compreendiam as várias formas de organização presentes em Portugal e uma gama muito diversa de população. A tabela 1 sistematiza as características da amostra. A informação usada é respeitante ao ano de 2001.

Tabela 1: Número de DMU por forma de organização e intervalo populacional

Forma organizatória	Número de DMU	Número de pessoas	Número de DMU
Serviço municipal	11	< 25 000	6
Serviço municipalizado	39	25 000 < P < 50 000	8
Empresa municipal	4	50 000 < P < 100 000	21
Empresa pública	1	100 000 < P < 150 000	21
Empresa privada	15	150 000 < P < 250 000	8
		P > 250 000	6

#### 3.2 Modelos de Cálculo

A concepção dos modelos de DEA deve ter em conta as características das DMU e as suas especificidades, a recolha bibliográfica, a opinião dos especialistas, os dados disponíveis e os objectivos do estudo. Na DEA é desejável, que tanto quanto possível, contemplem os aspectos mais abrangentes que caracterizam a actividade produtiva, isto é, os recursos consumidos e os produtos obtidos. As variáveis devem ser quantificáveis e consistentes entre DMU. Pode também existir uma segunda fase de análise, para integrar o ambiente operacional, onde o processo produtivo se desenvolve. No presente caso a orientação dos

modelos deve ser a de minimização de *inputs*, dado que as DMU estão, não só obrigadas a satisfazer todas as necessidades dos seus clientes, como também estão sujeitas a políticas de gestão de procura (*demand side management*) que condicionam a sua expansão. A especificação das variáveis em quantidades ou unidades monetárias deve também ser atentamente ponderada, tendo em conta os prós e os contras de cada opção (Marques, 2005a). Na tabela 2 evidenciam-se as variáveis mais adoptadas nos estudos de DEA aplicados ao sector da água (Marques e Monteiro, 2004).

Tabela 2: Variáveis mais usadas nos estudos de DEA aplicados aos serviços de água

<b>Inputs</b>	<b>Outputs</b>	<b>Factores explanatórios</b>
Pessoal (n.º, €)	Volume facturado (m <sup>3</sup> )	Origem de água / tratamento / custo de produção (% , €)
OPEX (€)	Número de clientes (n.º)	Volume facturado por natureza (% , m <sup>3</sup> )
Energia (€, kwh)	Extensão de redes (km)	Densidade populacional / clientes (n.º km <sup>-2</sup> , n.º km <sup>-1</sup> )
Extensão de redes (km)	Receitas (€)	Perdas (% , m <sup>3</sup> )
Custo total (€)		Factor de ponta / capitação (-, 1 hab <sup>-1</sup> dia <sup>-1</sup> )
CAPEX (€)		
Outros OPEX (€)		
Custos miscelânea (€)		

A discricionariedade da escolha de um modelo ou a adopção de vários modelos pode não ser prática, podendo mesmo fornecer resultados inconsistentes, não sendo, como tal, viável no mundo empírico. Assim, e tendo em atenção as variáveis presentes na tabela 2, o modelo é obtido estatisticamente pelo modelo ST com vista à selecção da tecnologia mais adequada, ou seja, mais representativa da realidade. Estipulou-se, como critério de paragem, o não cumprimento da hipótese  $h_0$  em mais de 50 % dos testes ( $F^m$ ,  $F^e$ , K-S e T) com um nível de significância superior a 1 %. O processo foi iniciado adoptando um modelo com um único *input* e um único *output*, respectivamente, o custo total e o volume de água facturado e a presença de RVE. Os valores limite para os níveis de significância de 1 %, 5 % e 10 % para os quatro testes considerados são evidenciados na tabela 3. A tabela 4 apresenta os resultados da primeira iteração. Nesta tabela verifica-se que a variável que conduz aos resultados dos testes estatísticos mais convincentes é a variável perdas de água, a qual permite rejeitar a hipótese  $h_0$  de que as estimativas das duas eficiências são equivalentes a um nível de significância de 1 % para os quatro testes. Esta variável tem justificação plena em Portugal, visto que as perdas de água atingem um valor muito elevado, o que, conjugado com o custo muito significativo de produção de algumas DMU, acentua este fenómeno, constituindo a variável mais condicionante na estimativa da eficiência das DMU. Na segunda iteração, com a contemplação do *input* perdas de água no modelo base, outras variáveis foram candidatas potenciais à inclusão. Verificou-se que a inserção de qualquer variável, ou a sua desagregação, cumpria todos os testes com um grau de significância de 1 %. A inclusão da variável número de clientes como *output* e a desagregação do *input* custo total em OPEX e em extensão de redes permitem rejeitar três dos quatro testes estatísticos a um nível de significância de 5 %, sendo o quarto não cumprido com um grau de 10 %. Ambas as hipóteses  $h_0$  são aceites à luz dos resultados obtidos e de acordo com o critério de paragem adoptado. Assim, em relação à adição ou à desagregação de variáveis, o processo iterativo termina aqui. Porém, falta atestar a validade acerca do tipo de rendimentos à escala. A tabela 5 testa a especificação do modelo admitindo RCE. Os resultados permitem rejeitar a hipótese  $h_0$ , na qual não existe diferença entre as distribuições das eficiências de ambas as tecnologias. Logo, segundo a metodologia ST, a tecnologia adoptada para a amostra dos serviços de água portugueses é caracterizada por RVE e por dois *inputs*, o custo total e as perdas de água, e um *output*, o volume de água facturado.

Tabela 3: Valores críticos de significância para os testes estatísticos considerados

Nível de significância	F <sub>m</sub>	F <sub>e</sub>	D (K-S)	T
1 %	1.755	1.485	0.257	2.356
5 %	1.486	1.321	0.207	1.657
10 %	1.361	1.243	0.182	1.288
<b>Graus de liberdade:</b>	70; 70	140; 140	70	138

Tabela 4: Escolha da especificação do modelo (1ª iteração)

	ETP	DMU eficiente	F <sub>m</sub>	F <sub>e</sub>	(K-S)	T
Modelo base:						
Custo total (CT)	0.626	5				
Volume facturado						
Novas variáveis:	Variação					
Clientes	0.054	2	1.592**	1.264*	0.171	1.719**
Extensão da rede	0.038	2	1.690**	1.223	0.143	1.235
Densidade de clientes	0.061	5	1.203	1.217	0.186*	1.845**
Perdas de água	0.106	8	2.593***	1.636***	0.271***	3.388***
Capitação	0.040	5	1.186	1.152	0.114	1.248
Desagregação:	Variação					
CT sem CP/ CP	0.069	5	1.335	1.283*	0.243**	2.149**
OPEX/CAPEX	0.050	2	1.551**	1.259*	0.200*	1.617*
OPEX/Extensão	0.093	6	1.515**	1.436**	0.271***	2.916***
VD / VND	0.048	3	1.224	1.186	0.171	1.504*

\* sig. a 10 %; \*\* sig. a 5 %; \*\*\* sig. a 1 %; CP – custo de produção; VD – volume doméstico; VND – volume não doméstico

Tabela 5: Escolha da especificação da tecnologia do modelo (tipo de rendimentos à escala)

	ETP	DMU eficiente	F <sub>m</sub>	F <sub>e</sub>	(K-S)	T
<b>Modelo base:</b>						
Custo total	0.731	13				
Perdas						
Volume						
<b>RCE</b>	<b>0.645</b>	<b>3</b>	<b>1.691**</b>	<b>1.412**</b>	<b>0.257***</b>	<b>2.924***</b>

\* significativa a 10 %; \*\* significativa a 5 %; \*\*\* significativa a 1 %

### 3.3 Resultados

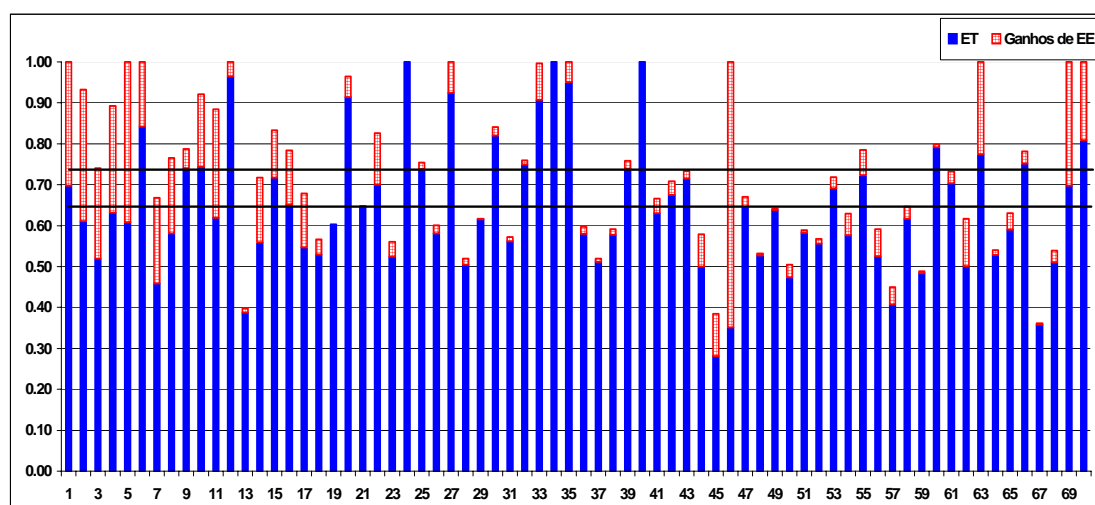
A figura 2 evidencia o valor da ET e a sua decomposição para o modelo ST por DMU. O valor de ET mais os ganhos de EE corresponde ao valor da ETP. A relação entre a ET e a ETP permite obter o valor da EE. Na figura estão ainda ilustrados os seus valores médios (ET e ETP). As DMU estão organizadas por ordem decrescente de população residente e por forma organizatória.

Na tabela 6 representam-se os valores médios e ponderados, quer pelo volume facturado, quer pela população servida, da ET e da sua decomposição, o número de DMU eficientes e os valores mínimos de eficiência. O cálculo foi efectuado por intermédio do *software* in2DEA desenvolvido pelos autores.

Tabela 6: Resumo estatístico dos resultados do modelo preferido, segundo a abordagem *stepwise*

	Média	Desvio padrão	Mínimo	Máximo	Média ponderada*	Média ponderada**	DMU eficiente
ET	0.645	0.161	0.282	1.000	0.623	0.654	3
ETP	0.731	0.184	0.362	1.000	0.804	0.769	13
EE	0.894	0.123	0.351	1.000	0.798	0.863	4
Slack do input custo ( $10^4$ €)	7.569	40.938	0.000	246.726	13.945	10.298	66
Slack do input perdas ( $10^4$ m <sup>3</sup> )	15.876	41.169	0.000	255.271	9.123	8.044	46
Slack do output volume (m <sup>3</sup> )	0	0	0	0	0	0	70

\* ponderada pelo volume facturado; \*\* ponderada pela população servida

Figura 2: Resultados por DMU do modelo preferido, segundo a abordagem *stepwise*

### 3.4 Análise de Resultados

Os níveis de ineficiência das DMU portuguesas para o ano de 2001 foram significativos, apresentando uma ET média de 0,645. No modelo, o factor determinante na origem da ET foi o valor da ETP, ou seja, precisamente, aquela parcela que é mais controlável pelos gestores das DMU. A ineficiência de escala apresenta também valores não desprezáveis. Esta origem de ineficiência não é, contudo, controlável a curto prazo. Os serviços de água de maior dimensão são, em geral, penalizados pela EE, apresentando rendimentos decrescentes à escala. Em Portugal, os serviços de água têm um potencial de melhoria da ET média de 35,5 %, da qual 10,6 % correspondem a ganhos de escala (EE) e 26,9 % à melhoria da ETP. Isto significa que, em média, cada DMU pode reduzir em 35,5 % os *inputs* custo total e perdas de água, apresentando ainda o mesmo volume de água facturado. Das 70 DMU, e para RVE, 27 têm *slacks* em relação a um dos *inputs*, isto é, podem produzir a mesma quantidade de *output* consumindo menos recursos. No modelo ST, em média, os *inputs* custo total e perdas de água têm um *slack*, respectivamente, de 75 690 euros e 158 760 m<sup>3</sup>. Conforme referido, se a escala fosse óptima, ou seja, se existissem RCE, os serviços de água consumiam menos 10,6 % dos *inputs* para a mesma quantidade de *outputs* produzidos. Quando determinado o modelo DEA para RVE resulta que das 70 DMU, 27 DMU apresentam RVED, 39 DMU RVEC e 4 DMU RCE. Se se considerar a população servida, o valor da ET é ainda mais reduzido, em virtude da penalização da EE das DMU mais populosas. Para o volume facturado ocorre a situação contrária. As figuras 3 e 4 evidenciam a ET e a ETP hierarquizada por classes de pontuação para o modelo ST.

Considerando como exemplo as 5 DMU de maior dimensão, a tabela 7 evidencia os valores dos *inputs* actuais e dos seus *targets*, sem e com *slacks*, bem como os seus *peers*. Esta tabela atesta, de forma peremptória, as potencialidades da técnica de DEA como instrumento de *benchmarking*, identificando não só as poupanças potenciais destas DMU (radiais e não radiais), mas também os pares de referência que podem constituir seus *benchmarks*.

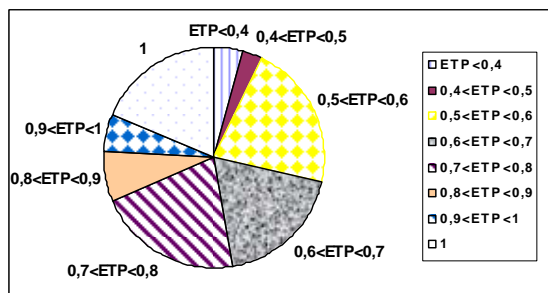
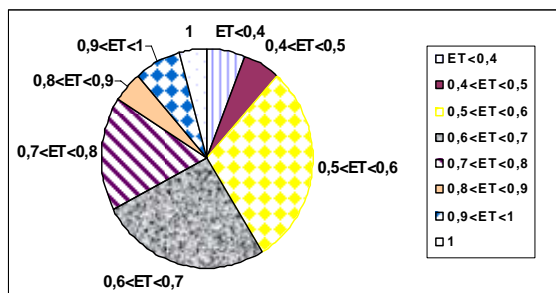


Figura 3: Número de DMU por classes da ET

Figura 4: Número de DMU por classes da ETP

Tabela 7: *Targets* e *peers* para as 5 DMU de maior dimensão sem e com *slacks*

DMU	Tipo	Variável	DMU 1	DMU 2	DMU 3	DMU 4*	DMU 5
Com <i>slacks</i>	Input	Custo total (10 <sup>4</sup> €)	2187 / 2187	2092 / 1950	2433 / 1800	1770 / 1342	2583 / 2583
		OPEX (10 <sup>4</sup> €)	8830 / 8830	9930 / 9260	10030 / 7440	5280 / 4710	14840 / 14840
	Output	Volume (10 <sup>4</sup> m <sup>3</sup> )	2404 / 2404	2233 / 2233	2020 / 2020	1429 / 1429	2909 / 2909
Peers	-	-	1	12, 1, 5	1, 6, 35	24, 1	5

\* Sem *slacks* apenas para a DMU 4 existirão diferenças. Para esta DMU o *input* custo total terá um *target* de 1580.

## 4 Análise de Sensibilidade aos Estimadores da DEA

### 4.1 Metodologia de *Bootstrap*

A DEA é encarada, habitualmente, como uma técnica não paramétrica, na qual a inferência estatística tradicional é difícil de prosseguir. Esta visão da DEA, muitas vezes considerada como uma das suas vantagens, não deixa, simultaneamente, de constituir uma das suas maiores fraquezas. Uma abordagem recente, que permite o desenvolvimento da desejada inferência estatística, consiste na aplicação da metodologia de *bootstrap* (reamostragem) aos estimadores da DEA. O *bootstrap* permite estimar o enviesamento e conduzir à inferência sobre os resultados da DEA. Esta metodologia foi introduzida por Bradley Efron em 1979 (Efron, 1979), mas só cerca de 10 anos mais tarde, em 1992, foi empregue, pela primeira vez, aos estimadores de DEA (Simar, 1992; Hall *et al.*, 1995; Ferrier e Hirschberg, 1997 e Simar e Wilson, 1998). A exposição e discussão aqui efectuada segue a abordagem desenvolvida por Simar e Wilson, que aplicam um algoritmo de *bootstrap* suave (*smooth*), adoptado com base no argumento da consistência (em alternativa poderia ser usado o *bootstrap naive*, mas que é inconsistente), baseado num processo de geração de dados (PGD) onde os *inputs* empregues são obtidos pelos desvios aleatórios da fronteira eficiente de *inputs* formada pelos *targets* (Simar e Wilson, 1998). Com efeito, de um ponto de vista formal, cada *input* no conjunto de observações de *input* / *output*  $x_w = \{(x_k, y_k); k=1, \dots, w\}$  pode ser representado por:

$$(\mathbf{x}_k, y_k) = (\mathbf{x}_k^F / \theta_k, y_k) \quad (5)$$

onde  $\mathbf{x}_k^F$  constitui a fronteira de *inputs* não observável da DMU  $k$ . As medidas de eficiência “absolutas” são estimadas a partir da mesma distribuição, ou seja,  $\theta_k \sim F_0$ ,  $k = 1, \dots, w$ . O PGD segue o princípio de que, restringido às relações entre *inputs* e *outputs*, os elementos estocásticos no processo produtivo são englobados totalmente pelas medidas aleatórias de eficiência dos *inputs*. A ideia principal da simulação de *bootstrap* consiste na imitação do PGD. O procedimento para o algoritmo em cada reamostra, limitado pelas proporções entre *inputs* e *outputs* observadas, é construído segundo duas fases. Na primeira fase a fronteira de *inputs* é estimada e os *pseudo-inputs* do *bootstrap* são criados através da aplicação do PGD na fronteira estimada de *inputs* e das pseudo-eficiências, desenvolvido a partir de alguma estimativa da distribuição  $F_0$ . O algoritmo emprega um procedimento de reamostragem alisado (*smooth*), baseado no argumento da consistência (Silverman e Young, 1987). Na segunda fase a estimativa da eficiência de *bootstrap* é obtida por intermédio da contabilização da distância do *input* original em relação à estimativa do *bootstrap* da fronteira.

O algoritmo do *bootstrap* sistematiza-se em cinco passos, que são de seguida apresentados (Simar e Wilson, 1998 e Löthgren, 1998).

Em primeiro lugar, constroem-se os vectores de *inputs* e de *outputs* da fronteira usando as estimativas iniciais da eficiência  $[(\mathbf{x}_k^F, y_k) = (\mathbf{x}_k \times \hat{\theta}_{kw}^*, y_k)]$ .

Em segundo lugar, utiliza-se um alisamento de Kernel da distribuição empírica das estimativas originais da eficiência para gerar uma reamostra com substituição de pseudo-eficiências  $\gamma_k^*$ . O procedimento de alisamento baseia-se no método da reflexão que consiste em duas etapas, respectivamente, na adição de uma pequena perturbação a  $\delta_k^*$  e na correcção sequente da reamostragem, onde  $\delta_k^*$  representa a reamostra não alisada obtida independentemente com substituição a partir da distribuição empírica das estimativas originais da eficiência (Silverman, 1986). Na primeira etapa acrescenta-se a  $\delta_k^*$  a perturbação  $h\varepsilon_k^*$ , na qual  $h$  representa a largura da banda ou da janela (*bandwidth* ou *window*) e o  $\varepsilon_k^*$  é obtido i.i.d. a partir de uma distribuição normal para gerar a pseudo-eficiência  $\tilde{\delta}_k^*$ . Como os valores das medidas de eficiência são limitados ao intervalo unitário, usa-se o procedimento de reflexão para gerar  $\tilde{\delta}_k^*$  a partir de:

$$\tilde{\delta}_k^* = \begin{cases} \delta_k^* + h\varepsilon_k^* & \delta_k^* + h\varepsilon_k^* \leq 1 \\ 2 - (\delta_k^* + h\varepsilon_k^*) & \delta_k^* + h\varepsilon_k^* > 1 \end{cases} \quad (6)$$

Nesta acção torna-se fulcral a escolha da largura da banda, a qual se pode obter, de modo expedito, pela fórmula  $h = 1.06 \times \sigma(\hat{\theta}_{kw}^*) \times w^{-0.2}$  (Silverman, 1986). As eficiências finais das reamostras alisadas representadas por  $\gamma_k^*$  são obtidas por correcção da  $\tilde{\delta}_k^*$  através de:



$$\gamma_k^* = \tilde{\delta}_k^* + \frac{(\tilde{\delta}_k^* - \bar{\delta}_k^*)}{\sqrt{1 + \frac{h^2}{\hat{\sigma}_{\hat{\theta}}^2}}} \quad (7)$$

onde  $\bar{\delta}_k^*$  é a média da reamostra das eficiências originais. Esta correcção garante que a eficiência assintótica da reamostra tenha os mesmos dois primeiros momentos em relação às estimativas de eficiência originais.

Em terceiro lugar, as pseudo-eficiências alisadas são usadas para gerar os pseudo-dados do *bootstrap*.

Em quarto lugar, calcula-se novamente as eficiências utilizando o modelo de DEA para estimar as eficiências das amostras obtidas pelo *bootstrap*, usando o conjunto de pseudo-dados obtido.

Por fim, repetem-se B vezes os passos 2 a 4 para estimar um conjunto de B valores de estimativas de eficiência para DMU. Segundo alguns autores, o valor mínimo de B deve ser igual a 1000 (Hall, 1986).

A estimativa de *bootstrap* do enviesamento do estimador DEA é obtida pela fórmula:

$$\hat{\text{Env}}_k = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_{kb}^* - \hat{\theta}_k \quad (8)$$

na qual o primeiro termo (do lado direito) corresponde à média do resultado da eficiência do *bootstrap* ( $\bar{\theta}_k^*$ ) e o segundo termo ao resultado inicial da estimativa de DEA. O estimador de  $\theta_k$  corrigido do enviesamento é calculado por:

$$\tilde{\theta}_k = \hat{\theta}_k - \hat{\text{Env}}_k \quad (9)$$

O erro padrão de  $\hat{\theta}_{kw}^{*b}$  é determinado pela expressão:

$$\hat{\text{EP}} = \left\{ \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_{kb}^* - \bar{\theta}_k^*)^2 \right\}^{1/2} \quad (10)$$

Conhecendo a função de distribuição empírica das eficiências de *bootstrap*  $\hat{\theta}_{kb}^*$  com  $b = 1, \dots, B$  e depois de uma nova correcção do enviesamento, que resulta na expressão 11 (Simar e Wilson, 1998), os intervalos de confiança podem ser obtidos, por exemplo, por intermédio do método percentil (Hall, 1998).

$$\tilde{\theta}_{kb}^* = \hat{\theta}_{kb}^* - 2 \times \hat{\text{Env}}_k \quad (11)$$

O método percentil é baseado na função de distribuição empírica das eficiências de *bootstrap*  $\hat{\theta}_{kw}^{*b}$ ,  $b = 1, \dots, B$ , definido como  $\hat{F}_k(s) = \frac{1}{B} \sum_{b=1}^B I(\hat{\theta}_{kw}^{*b} \leq s)$  para algum valor real  $s$ , onde  $I(\cdot)$  representa uma função indicatriz normal (Efron e Tibshirani, 1993). Um intervalo de confiança  $(1-2\alpha)$  para o valor real da eficiência  $\theta$  para a DMU  $k$  é:

$$(\hat{\theta}_{kb}^{*(\alpha)}, \hat{\theta}_{kb}^{*(1-\alpha)}) \quad (12)$$

onde  $\hat{\theta}_{kb}^{*(\alpha)}$  é o  $\alpha$  quantil de  $\hat{F}_k$ , isto é,  $\hat{\theta}_{kw}^{*(\alpha)} = \hat{F}_k^{-1}(\alpha)$ . Se a função de distribuição empírica de  $\hat{\theta}_{kw}^*$  for enviesada, é preferível centrar a mediana da distribuição sobre  $\tilde{\theta}_k$ , levando à obtenção de intervalos de confiança distintos, baseados na correcção do enviesamento das medianas (Simar e Wilson, 1998).

Simar e Wilson, em publicações posteriores, sugerem a realização de um artifício, baseado na técnica de “*trimming*”, para melhorarem os intervalos de confiança obtidos com o *bootstrap* (Simar e Wilson, 1998 e 2000b). Estes autores avançam uma outra justificação para alguma distorção dos resultados, assente na ideia de que o uso da estimativa do enviesamento acrescenta, necessariamente, erro aos intervalos de confiança estimados (*i. e.* o erro médio quadrado é maior no *bootstrap* corrigido do que na DEA). Argumentam também que só na circunstância onde a variância das eficiências de *bootstrap* seja menor que 1/3 do enviesamento ao quadrado a correcção do enviesamento deverá ser implementada ( $R_k = \frac{1}{3\sigma^2} \times (\hat{\text{Env}}_B(\hat{\theta}(x_k, y_k)))^2$ ). Mais ainda, os mesmos autores

alegam que, por essa relação se basear apenas nos valores estimados do enviesamento e das variâncias, e não nos valores reais, a opção pela correcção do enviesamento só deverá ser efectuada quando o  $R_k$  obtido estiver bem acima da unidade (Simar e Wilson, 1998 e 2000b).

## 4.2 Estimação dos Resultados do Caso de Estudo através de *Bootstrap*

Na análise da robustez da especificação do modelo ST aplicou-se o método de *bootstrap* atrás descrito. A diferença entre as medidas de eficiência estimadas pela DEA e as obtidas pelo *bootstrap*, ou melhor, o enviesamento, tal como o erro padrão, constituem medidas relevantes de análise de fiabilidade dos modelos. Efron sugere que uma relação entre a estimativa do enviesamento e do erro padrão superior a 0.25 é significativa, pelo que não devem ser rejeitados os resultados da eficiência obtidos por *bootstrap* (Efron, 1982). Simar e Wilson defendem que os resultados do *bootstrap* para cada DMU só devem ser aceites e preferíveis em relação às estimativas da DEA quando a variância dos valores obtidos pelo *bootstrap* for inferior a um terço do enviesamento ao quadrado (Simar e Wilson, 1998 e 2000b).

Nesta secção procurou-se não só examinar a sensibilidade dos resultados como testar a própria metodologia *bootstrap*, cuja aplicação empírica na DEA é recente e, embora defendida, por alguns autores, pode não ser muito benéfica, pelo menos em certas situações. Ademais, os casos práticos, referidos em (Simar e Wilson, 1998, 2000a e

2000b), ilustram distribuições de eficiência bem “comportadas” para este fim, isto é, com uma amplitude diminuta entre a eficiência máxima e a mínima, razão pela qual os intervalos de confiança, assim como os enviesamentos, são reduzidos. Como tal, podem constituir, ficticiamente, bons resultados. Todavia, se a relação entre o enviesamento e a amplitude referida se utilizar como indicador, verifica-se que a aplicação do *bootstrap*, de acordo com as metodologias seguidas, conduz a resultados pouco satisfatórios em relação a algumas DMU, em particular para as consideradas eficientes pela DEA. Pensa-se que esse fenómeno poderá ter origem na grande amplitude das eficiências determinadas pela DEA, as quais, recorde-se, definem a distribuição para a reamostragem no *bootstrap*, nos dados atípicos de algumas DMU, posicionando-as em regiões mais dispersas em termos de pontos que possam constituir referência para a determinação das eficiências em cada iteração do método, ou na conjugação desses dois efeitos. Na situação actual dos serviços de água portugueses, onde ocorrem DMU com eficiências inferiores a 0.4, muito embora os resultados do *bootstrap* possam, e devam, ser encarados como indicações de uma menor robustez das eficiências obtidas pela DEA, não é expectável que os valores alcançados pelo *bootstrap* sejam aceitáveis em aplicações práticas.

Para a aplicação do *bootstrap* empregou-se o *software* in2DEA, seguindo o procedimento exposto no trabalho de referência de Simar e Wilson. Numa primeira fase adoptou-se a linha sugerida por estes autores em 1998 (Simar e Wilson, 1998) e, posteriormente, numa segunda fase, introduziram-se as alterações subsequentes que os mesmos propuseram em 2000 referentes à obtenção dos intervalos de confiança (Simar e Wilson, 2000a e b).

A tabela 8 ilustra os resultados globais alcançados com a aplicação do *bootstrap* para o modelo ST (RVE). Como seria de esperar, o valor obtido pelo *bootstrap* é inferior ao da DEA, porque a DEA é um estimador da eficiência por excesso (conservativo). Também, por isso, quando se determinam os intervalos de confiança pelo *bootstrap* pode aparecer valores superiores a 1.

Tabela 8: Resultados globais do *bootstrap* do modelo ST

Modelo / estimativa	Média	Desvio padrão	Mínimo	Máximo
ST / DEA original	0.731	0.184	0.362	1.000
ST / DEA <i>bootstrap</i>	0.626	0.136	0.295	0.890

Na tabela 9 apresentam-se os resultados por DMU para o modelo ST (RVE). Na tabela são também estimados os intervalos de confiança com um nível de significância de 5 %, tanto para a média como centrados na mediana, para ter em conta o enviesamento. Conforme já tinha adiantado, os resultados alcançados não são muito úteis em relação às DMU eficientes. Parece que estas entidades, na maioria das situações, saem penalizadas com a aplicação do *bootstrap*, sendo mesmo alterada a sua ordem de eficiência. Mesmo usando o artifício sugerido por Simar e Wilson atrás referido os resultados obtidos continuam a ser algo limitativos do ponto de vista empírico.

Efectuaram-se diversos procedimentos de análise da sensibilidade da metodologia adoptada e apuraram-se as seguintes conclusões:

- O modelo é pouco sensível à fórmula de cálculo da largura da banda (h). Testaram-se vários métodos de cálculo deste parâmetro e variou-se, arbitrariamente, o seu valor, constatando-se que os resultados oscilaram pouco;

- Embora se tenha adoptado um  $B=1000$ , observou-se que um  $B$  inferior ( $B=500$  ou  $B=800$ ) ou superior ( $B=2000$  ou  $B=5000$ ) não conduzia a resultados muito díspares;
- Testou-se o modelo considerando os *pseudo-inputs* obtidos pelo *bootstrap* nas novas DEA em vez dos *inputs* originais (Löthgren, 1998), no qual se registaram diferenças significativas, mas não relevantes;
- Verificou-se que os valores das eficiências alcançados pelo *bootstrap* dependiam muito do passo inicial de geração de valores aleatórios com substituição das eficiências originais, obtidas pela DEA, com base na sua distribuição empírica. Tal pode ser explicado pela amplitude elevada existente entre o *input* menos eficiente e o mais eficiente. Este efeito é menos significativo quando esta diferença é reduzida;
- Por fim, analisou-se a influência da adopção de um *bootstrap* suave (através da estimativa da função de densidade de probabilidades de Kernel Gaussiana) em detrimento do *bootstrap naïve* inconsistente, cujo objectivo era o de melhorar a estimação, tendo em conta a existência de uma mancha e de vários valores no limite da amostra. Constatou-se que os resultados para o *bootstrap naïve*, embora distintos, eram inconsistentes e sem significado para a amostra estudada.

Recalculou-se o modelo de Simar e Wilson atrás aplicado com as alterações posteriores, sugeridas pelos mesmos autores, relativas à correcção do enviesamento do *bootstrap* e à melhoria dos intervalos de confiança (Simar e Wilson, 2000a e b). Adoptou-se ainda um nível de significância de 90 %, uma correcção por *trimming* de 10 % e um  $B=2000$ . Os resultados obtidos, que revelam algumas melhorias, mas ainda aquém do desejável, para efeitos da aplicação prática apresentam-se na tabela 10.

Por último, experimentou-se efectuar novamente todos os cálculos (representados nas tabelas 9 e 10), mas considerando apenas as 20 DMU mais eficientes para o modelo ST (RVE). Esta iniciativa dos autores, original na literatura, permitiu atestar que a eficiência oscilava entre 0.826 para a DMU 22 e 1.0 para as DMU tecnicamente eficientes da amostra. Como a amplitude entre a eficiência mínima e máxima era reduzida, seria esperado que os resultados alcançados pelo *bootstrap* fossem mais aceitáveis, o que se veio a confirmar. As tabelas 11 e 12 apresentam os desfechos desses exercícios.

Em suma, a técnica de *bootstrap*, que mantém a mesma linha não paramétrica da DEA e que será testada no caso de estudo, permite avaliar a robustez das medidas de eficiência obtidas pela DEA às variações da amostra e das variáveis não observáveis. Quando se repete o processo de reamostragem  $B$  vezes, estar-se-á em cada uma a determinar uma fronteira fictícia particular correspondente a um conjunto específico de *peers* (em relação ao qual cada DMU efectua o seu *benchmarking*), representando um nível particular de variáveis não observáveis ou não consideradas. A eficácia desta metodologia está muito dependente das características da amostra em análise, designadamente, do *gap* de eficiência existente. Quando a diferença entre as DMU mais eficientes e as menos eficientes não é substancial, a implementação da técnica *bootstrap* justifica-se possibilitando a inferência estatística, sendo compensador o esforço do seu cálculo. Caso contrário, quando o nível de ineficiência é elevado, com uma amplitude significativa (diferença entre a eficiência mínima e a eficiência máxima), o emprego da técnica *bootstrap*, ainda assim pode ser interessante para a aferição da robustez das medidas de eficiência da DEA, mas não os resultados que terão uma utilização condicionada num contexto prático.

Tabela 9: Resultados do *bootstrap* do modelo ST

DMU	$\hat{\theta}$ (DEA)	$\tilde{\theta}$ (Boot corr.)	$\hat{Env}_k$	Médiana <sub>i</sub>	$\hat{EP}$ (Erro pad.)	2.5 % (Média)	97.5 % (Média)	2.5 % (Mediana)	97.5 % (Mediana)
1	1.000	0.677	0.323	0.674	0.199	0.387	1.101	0.292	1.062
2	0.932	0.755	0.177	0.722	0.130	0.611	1.089	0.510	1.000
3	0.74	0.592	0.148	0.567	0.111	0.467	0.843	0.385	0.799
4	0.893	0.74	0.153	0.710	0.107	0.617	0.983	0.543	0.936
5	1.000	0.634	0.366	0.618	0.253	0.301	1.205	0.144	1.123
6	1.000	0.877	0.123	0.873	0.060	0.791	0.984	0.782	0.972
7	0.668	0.583	0.085	0.572	0.057	0.522	0.715	0.492	0.674
8	0.765	0.676	0.089	0.670	0.053	0.615	0.774	0.596	0.756
9	0.787	0.628	0.16	0.619	0.103	0.494	0.817	0.438	0.817
10	0.921	0.773	0.148	0.760	0.087	0.655	0.957	0.616	0.929
11	0.884	0.786	0.098	0.774	0.062	0.716	0.929	0.683	0.890
12	1.000	0.69	0.31	0.718	0.169	0.414	0.958	0.368	1.012
13	0.398	0.364	0.034	0.361	0.039	0.342	0.401	0.335	0.393
14	0.717	0.648	0.069	0.641	0.049	0.603	0.735	0.582	0.714
15	0.833	0.754	0.079	0.747	0.052	0.702	0.844	0.681	0.826
16	0.784	0.699	0.085	0.696	0.050	0.640	0.775	0.632	0.766
17	0.678	0.608	0.07	0.602	0.050	0.561	0.697	0.540	0.676
18	0.566	0.508	0.058	0.502	0.045	0.469	0.576	0.452	0.563
19	0.603	0.54	0.064	0.538	0.046	0.494	0.604	0.486	0.593
20	0.964	0.855	0.109	0.850	0.060	0.781	0.969	0.761	0.949
21	0.648	0.573	0.075	0.567	0.052	0.519	0.655	0.503	0.642
22	0.826	0.726	0.1	0.707	0.074	0.651	0.890	0.598	0.853
23	0.561	0.512	0.049	0.507	0.043	0.482	0.569	0.467	0.558
24	1.000	0.674	0.326	0.659	0.196	0.381	1.012	0.296	1.051
25	0.754	0.644	0.11	0.627	0.075	0.560	0.795	0.514	0.775
26	0.601	0.542	0.058	0.536	0.047	0.503	0.617	0.481	0.603
27	1.000	0.89	0.11	0.884	0.060	0.814	1.006	0.797	0.983
28	0.52	0.47	0.05	0.464	0.044	0.436	0.532	0.420	0.520
29	0.617	0.566	0.051	0.562	0.043	0.532	0.624	0.521	0.611
30	0.841	0.796	0.045	0.792	0.040	0.774	0.843	0.761	0.832
31	0.572	0.523	0.049	0.519	0.042	0.491	0.576	0.480	0.565
32	0.76	0.693	0.067	0.690	0.046	0.651	0.759	0.639	0.747
33	0.997	0.853	0.144	0.840	0.086	0.745	1.042	0.701	1.005
34	1.000	0.873	0.127	0.868	0.068	0.780	1.004	0.758	0.988
35	1.000	0.805	0.195	0.791	0.116	0.644	1.028	0.592	1.019
36	0.597	0.564	0.033	0.561	0.038	0.547	0.596	0.538	0.589
37	0.52	0.476	0.044	0.472	0.041	0.448	0.523	0.438	0.513
38	0.592	0.559	0.033	0.556	0.038	0.542	0.593	0.533	0.585
39	0.759	0.718	0.041	0.714	0.039	0.698	0.761	0.686	0.750
40	1.000	0.89	0.11	0.885	0.059	0.813	0.995	0.799	0.980
41	0.666	0.571	0.095	0.566	0.059	0.500	0.679	0.480	0.663
42	0.708	0.663	0.045	0.660	0.040	0.639	0.706	0.628	0.698
43	0.733	0.666	0.066	0.662	0.047	0.624	0.737	0.610	0.723
44	0.579	0.519	0.06	0.511	0.047	0.480	0.593	0.459	0.579
45	0.385	0.295	0.089	0.274	0.080	0.218	0.462	0.158	0.433
46	1.000	0.511	0.489	0.446	0.393	0.058	1.422	0.258	1.280
47	0.671	0.556	0.115	0.540	0.079	0.463	0.704	0.420	0.692
48	0.532	0.47	0.062	0.465	0.045	0.427	0.535	0.417	0.523
49	0.642	0.532	0.109	0.519	0.076	0.444	0.674	0.403	0.662
50	0.505	0.469	0.037	0.466	0.039	0.448	0.502	0.440	0.497
51	0.589	0.537	0.052	0.532	0.042	0.506	0.593	0.493	0.581
52	0.568	0.521	0.047	0.517	0.041	0.493	0.574	0.481	0.562
53	0.719	0.675	0.044	0.671	0.040	0.652	0.717	0.641	0.709
54	0.629	0.586	0.043	0.581	0.042	0.560	0.640	0.545	0.626
55	0.785	0.728	0.057	0.722	0.044	0.695	0.792	0.678	0.777
56	0.592	0.522	0.07	0.515	0.052	0.472	0.618	0.449	0.596
57	0.45	0.405	0.045	0.401	0.042	0.374	0.454	0.364	0.445
58	0.646	0.612	0.035	0.608	0.039	0.594	0.647	0.585	0.639
59	0.489	0.454	0.035	0.452	0.039	0.434	0.486	0.427	0.481
60	0.799	0.712	0.087	0.704	0.057	0.650	0.817	0.625	0.799
61	0.732	0.695	0.037	0.692	0.039	0.677	0.733	0.667	0.724
62	0.617	0.529	0.088	0.511	0.069	0.462	0.670	0.415	0.643
63	1.000	0.784	0.216	0.752	0.163	0.599	1.154	0.476	1.092
64	0.54	0.509	0.031	0.506	0.038	0.493	0.541	0.485	0.533
65	0.631	0.595	0.037	0.592	0.040	0.576	0.637	0.564	0.626
66	0.782	0.737	0.045	0.734	0.040	0.714	0.782	0.703	0.772
67	0.362	0.319	0.043	0.317	0.041	0.288	0.365	0.278	0.360
68	0.539	0.505	0.034	0.503	0.038	0.486	0.537	0.478	0.532
69	1.000	0.645	0.355	0.586	0.249	0.328	1.213	0.164	1.127
70	1.000	0.766	0.234	0.735	0.164	0.564	1.124	0.455	1.076

Tabela 10: Resultados das alterações efectuadas ao *bootstrap* do modelo ST

DMU	$\hat{\theta}$ (DEA)	$\tilde{\theta}$ (Boot. corr.)	$\hat{\text{Env}}_k$	$\sigma(\hat{\theta})$	$R_k^{(*)}$	Intervalo de confiança 90 %	
1	1.000	0.677	0.323	0.196	0.903	0.416	0.935
2	0.932	0.755	0.177	0.126	0.661	0.577	0.880
3	0.74	0.592	0.148	0.107	0.622	0.439	0.707
4	0.893	0.74	0.153	0.099	0.785	0.593	0.846
5	1.000	0.634	0.366	0.244	0.743	0.301	0.938
6	1.000	0.877	0.123	0.049	2.089	0.812	0.939
7	0.668	0.583	0.085	0.047	1.049	0.518	0.633
8	0.765	0.676	0.089	0.041	1.521	0.621	0.724
9	0.787	0.628	0.16	0.098	0.883	0.491	0.747
10	0.921	0.773	0.148	0.078	1.144	0.664	0.872
11	0.884	0.786	0.098	0.053	1.121	0.712	0.842
12	1.000	0.69	0.31	0.165	1.216	0.485	0.935
13	0.398	0.364	0.034	0.015	1.755	0.343	0.380
14	0.717	0.648	0.069	0.033	1.413	0.606	0.684
15	0.833	0.754	0.079	0.036	1.575	0.705	0.795
16	0.784	0.699	0.085	0.034	1.968	0.654	0.746
17	0.678	0.608	0.07	0.034	1.357	0.565	0.647
18	0.566	0.508	0.058	0.028	1.461	0.467	0.538
19	0.603	0.54	0.064	0.027	1.862	0.503	0.574
20	0.964	0.855	0.109	0.048	1.661	0.791	0.915
21	0.648	0.573	0.075	0.035	1.540	0.522	0.615
22	0.826	0.726	0.1	0.067	0.758	0.622	0.790
23	0.561	0.512	0.049	0.023	1.462	0.479	0.538
24	1.000	0.674	0.326	0.187	1.016	0.424	0.930
25	0.754	0.644	0.11	0.066	0.913	0.547	0.717
26	0.601	0.542	0.058	0.030	1.180	0.498	0.577
27	1.000	0.89	0.11	0.046	1.809	0.830	0.947
28	0.52	0.47	0.05	0.025	1.275	0.433	0.498
29	0.617	0.566	0.051	0.022	1.728	0.535	0.592
30	0.841	0.796	0.045	0.017	2.172	0.773	0.815
31	0.572	0.523	0.049	0.022	1.632	0.494	0.548
32	0.76	0.693	0.067	0.028	1.940	0.656	0.727
33	0.997	0.853	0.144	0.077	1.182	0.744	0.942
34	1.000	0.873	0.127	0.058	1.572	0.794	0.946
35	1.000	0.805	0.195	0.105	1.123	0.662	0.937
36	0.597	0.564	0.033	0.013	2.230	0.547	0.578
37	0.52	0.476	0.044	0.019	1.744	0.449	0.498
38	0.592	0.559	0.033	0.013	2.223	0.541	0.573
39	0.759	0.718	0.041	0.016	2.183	0.696	0.736
40	1.000	0.89	0.11	0.046	1.956	0.827	0.945
41	0.666	0.571	0.095	0.046	1.436	0.507	0.627
42	0.708	0.663	0.045	0.018	2.133	0.639	0.684
43	0.733	0.666	0.066	0.029	1.715	0.627	0.701
44	0.579	0.519	0.06	0.031	1.258	0.474	0.551
45	0.385	0.295	0.089	0.071	0.550	0.189	0.367
46	1.000	0.511	0.489	0.388	0.518	-	0.936
47	0.671	0.556	0.115	0.068	0.938	0.455	0.636
48	0.532	0.47	0.062	0.027	1.661	0.433	0.503
49	0.642	0.532	0.109	0.065	0.933	0.437	0.610
50	0.505	0.469	0.037	0.015	2.149	0.448	0.485
51	0.589	0.537	0.052	0.023	1.734	0.507	0.561
52	0.568	0.521	0.047	0.021	1.652	0.492	0.545
53	0.719	0.675	0.044	0.017	2.252	0.651	0.694
54	0.629	0.586	0.043	0.020	1.465	0.558	0.608
55	0.785	0.728	0.057	0.025	1.668	0.693	0.756
56	0.592	0.522	0.07	0.038	1.191	0.468	0.563
57	0.45	0.405	0.045	0.021	1.526	0.373	0.428
58	0.646	0.612	0.035	0.013	2.247	0.592	0.625
59	0.489	0.454	0.035	0.013	2.282	0.436	0.470
60	0.799	0.712	0.087	0.043	1.319	0.648	0.763
61	0.732	0.695	0.037	0.014	2.281	0.676	0.710
62	0.617	0.529	0.088	0.060	0.753	0.435	0.587
63	1.000	0.784	0.216	0.158	0.636	0.554	0.957
64	0.54	0.509	0.031	0.012	2.351	0.492	0.522
65	0.631	0.595	0.037	0.015	1.839	0.574	0.611
66	0.782	0.737	0.045	0.017	2.268	0.713	0.756
67	0.362	0.319	0.043	0.021	1.395	0.291	0.345
68	0.539	0.505	0.034	0.014	2.059	0.487	0.521
69	1.000	0.645	0.355	0.249	0.697	0.273	0.932
70	0.668	0.766	0.234	0.159	0.738	0.537	0.939

$$(*) R_k = \frac{1}{3\sigma^2} \times (\hat{\text{Env}}_B(\hat{\theta}(x_k, y_k)))^2$$

Tabela 11: Resultados do *bootstrap* do modelo ST para as 20 DMU mais eficientes

DMU	$\hat{\theta}$ (DEA)	$\tilde{\theta}$ (Boot corr.)	$\hat{Env}_k$	Médiana <sub>i</sub>	2.5 % (Média)	97.5 % (Média)	2.5 % (Mediana)	97.5 % (Mediana)
1	1.000	0.956	0.044	0.941	0.917	1.086	0.873	1.039
2	0.932	0.91	0.023	0.905	0.891	0.959	0.877	0.941
4	0.893	0.871	0.022	0.866	0.853	0.915	0.838	0.903
5	1.000	0.957	0.043	0.940	0.918	1.086	0.869	1.045
6	1.000	0.964	0.036	0.960	0.933	1.012	0.919	1.009
10	0.921	0.898	0.023	0.895	0.879	0.936	0.869	0.928
11	0.884	0.865	0.02	0.863	0.848	0.890	0.841	0.887
12	1.000	0.954	0.046	0.937	0.913	1.093	0.860	1.047
20	0.964	0.939	0.025	0.935	0.918	0.979	0.907	0.971
22	0.826	0.805	0.021	0.801	0.788	0.851	0.775	0.835
24	1.000	0.952	0.048	0.934	0.909	1.096	0.855	1.050
27	1.000	0.975	0.025	0.973	0.954	1.010	0.947	1.004
33	0.997	0.973	0.024	0.968	0.953	1.032	0.935	1.011
34	1.000	0.958	0.042	0.945	0.921	1.044	0.887	1.029
35	1.000	0.966	0.034	0.963	0.937	1.025	0.920	1.012
40	1.000	0.963	0.037	0.955	0.930	1.034	0.907	1.018
46	1.000	0.953	0.047	0.935	0.911	1.092	0.861	1.045
63	1.000	0.978	0.022	0.973	0.959	1.035	0.940	1.016
69	1.000	0.952	0.048	0.934	0.909	1.084	0.857	1.047
70	1.000	0.966	0.034	0.961	0.936	1.029	0.917	1.015

Tabela 12: Resultados das alterações efectuadas ao modelo ST para as 20 DMU mais eficientes

DMU	$\hat{\theta}$ (DEA)	$\tilde{\theta}$ (Boot. Corr.)	$\hat{Env}_k$	$\sigma(\hat{\theta})$	$R_k^{(*)}$	Intervalo de confiança 90 %	
1	1.000	0.954	0.046	0.044	0.355	0.885	0.992
2	0.932	0.909	0.023	0.017	0.620	0.888	0.925
4	0.893	0.871	0.022	0.017	0.548	0.850	0.887
5	1.000	0.955	0.045	0.045	0.327	0.885	0.992
6	1.000	0.964	0.036	0.024	0.764	0.930	0.992
10	0.921	0.898	0.023	0.015	0.770	0.875	0.914
11	0.884	0.865	0.02	0.012	0.956	0.848	0.877
12	1.000	0.954	0.046	0.048	0.314	0.880	0.992
20	0.964	0.939	0.025	0.017	0.728	0.916	0.957
22	0.826	0.804	0.021	0.016	0.622	0.784	0.820
24	1.000	0.952	0.048	0.048	0.330	0.874	0.992
27	1.000	0.975	0.025	0.014	1.044	0.955	0.992
33	0.997	0.972	0.024	0.020	0.511	0.950	0.990
34	1.000	0.957	0.043	0.036	0.457	0.898	0.992
35	1.000	0.966	0.034	0.023	0.733	0.937	0.992
40	1.000	0.964	0.036	0.027	0.561	0.927	0.992
46	1.000	0.953	0.047	0.047	0.326	0.886	0.992
63	1.000	0.977	0.023	0.020	0.424	0.953	0.994
69	1.000	0.952	0.048	0.049	0.322	0.875	0.992
70	1.000	0.965	0.035	0.026	0.611	0.933	0.992

$$(*) R_k = \frac{1}{3\sigma^2} \times (\hat{Env}_B(\hat{\theta}(x_k, y_k)))^2$$

## 4 Conclusões

Os objectivos do presente artigo consistiram na análise e discussão da técnica fronteira não paramétrica de DEA para cálculo da eficiência de DMU e no emprego do método de *bootstrap* sobre os resultados obtidos no sentido do fornecimento de inferência estatística. Após a introdução, o artigo iniciou-se com uma explanação sumária sobre a técnica DEA, a qual foi aplicada a um conjunto de 70 serviços de água portugueses empregando o método de ST na especificação do modelo adoptado. Ulteriormente, a metodologia de *bootstrap* foi introduzida e discutida tendo em conta o *state-of-the-art*. Esta ferramenta foi testada usando como caso de estudo o sector da água em Portugal. Como conclusões mais relevantes deste documento refiram-se as seguintes:

- O uso da técnica de DEA apresenta várias vantagens, das quais se destacam a identificação de um conjunto de DMU (*peers*) eficientes com combinação de *inputs* e *outputs* semelhante para cada DMU ineficiente, a facilidade em lidar com múltiplos *inputs* e *outputs*, a adopção dos melhores resultados como elementos de comparação, a não assunção de uma forma funcional para a fronteira ou para a ineficiência quando associada ao erro, a natureza conservativa das avaliações e a decomposição da natureza da eficiência em várias componentes. Porém, a técnica de DEA não permite com facilidade a inferência estatística dos resultados obtidos;
- A primeira dificuldade do emprego da técnica de DEA diz respeito à selecção do tipo de modelo a considerar e do nível de detalhe adequado à respectiva aplicação. A abordagem *stepwise*, que permite determinar, passo a passo, a especificação dos modelos, constitui uma metodologia muito útil na selecção, não só das variáveis a adoptar como *inputs* e *outputs*, mas também da tecnologia de produção adequada, embora requeira alguns pressupostos que, por vezes, não se verificam (e. g. dimensão da amostra e independência);
- As DMU portuguesas apresentam, para o modelo ST adoptado, níveis de ineficiência significativos, sobretudo relacionados com a ETP, precisamente aquela parcela mais controlável pelos gestores das DMU. Os níveis de ineficiência de escala são também relevantes;
- A técnica de simulação de *bootstrap* faculta a análise de sensibilidade e da robustez dos resultados alcançados com a DEA. Esta técnica, de índole estatística, possibilita a avaliação da robustez das medidas de eficiência alcançadas com a DEA às variações da amostra e das variáveis não observáveis. De facto, a técnica de *bootstrap*, apesar de algo complexa e de produzir resultados ainda pouco apurados, constitui uma orientação claramente não determinística da DEA, no sentido da desejada inferência estatística;
- A aplicação da metodologia de *bootstrap* ao modelo ST demonstrou que, independentemente dos artifícios usados e dos avanços e desenvolvimentos recentes, os valores obtidos para algumas DMU (tendencialmente as que apresentam uma eficiência unitária e/ou a informação que as caracteriza é atípica) são ainda pouco satisfatórios, ou melhor, pouco úteis do ponto de vista empírico, pelo menos para a realidade dos serviços de água em Portugal. Esta realidade evidencia, por um lado, a falta de robustez e significância estatística das estimativas de eficiência obtidas pelo modelo DEA e, por outro lado, que a técnica de *bootstrap* tem um valor algo limitado quando a dispersão dos valores da eficiência do modelo de DEA é muito grande. Alguns autores têm proposto artifícios para testar ou melhorar os valores obtidos pelo *bootstrap*, procurando melhorar a viabilidade da implementação desta metodologia. No presente trabalho, procurou-se mitigar os valores das ineficiências que definem o processo de reamostragem, ensaiando-se a metodologia de *bootstrap* para os 20 serviços de água portugueses mais eficientes, tendo-se alcançado resultados para os intervalos de confiança mais interessantes e com um potencial de uso empírico relevante, havendo, contudo, ainda um caminho longo a percorrer neste domínio.

## 5 Referências

- Banker, R. (1993) Maximum Likelihood, Consistency and Data Envelopment Analysis: a Statistical Foundation, *Management Science*, Vol 39, No 10, pp. 1265-1273.
- Banker, R. (1996) Hypothesis Tests using Data Envelopment Analysis, *Journal of Productivity Analysis*, Vol 7, No 2/3, pp. 139-159.
- Banker, R., Charnes, A. e Cooper, W. (1984) Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science*, Vol 30, No 9, pp. 1078-1092.



- Bosworth, D., Stoneman, P. e Thanassoulis, E. (1996) The Measurement of Comparative Total Efficiency in the Sewerage and Water industry: an Exploratory Study, Office of Water Services, UK.
- Burns, P., Huggins, M., Riechmann, C. e Weyman-Jones, T. (2000) Choice of Model and Availability of Data for the Efficiency Analysis of Dutch Network and Supply Businesses in the Electricity Sector, *Frontier Economics*, Dienst uitvoering en toezicht Elektriciteitswet, Holanda.
- Charnes, A., Cooper, W. e Rhodes, E. (1978) Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, Vol 2, No 6, pp. 429-444.
- Charnes, A., Cooper, W., Lewin, A. e Seiford, L. (1994) *Data Envelopment Analysis. Theory, methodology and applications*, Kluwer Academic Publishers, USA.
- Coelli, T., Prasada, R. e Battese, G. (1998) *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishers, USA.
- Cooper, W., Seiford, L. e Tone, K. (2000) *Data Envelopment Analysis. A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, Kluwer Academic Publishers, USA.
- Efron, B. (1979) Bootstrap Methods: another Look at the Jackknife, *Annals of Statistics*, Vol 7, No 1, pp. 1-26.
- Efron, B. (1982) *The Jackknife, the Bootstrap and other Resampling Plans*, Monograph No. 38, Society for Industrial and Applied Mathematics, USA.
- Efron, B. e Tibshirani, J. (1993) *An Introduction to the Bootstrap*, Chapman and Hall, UK.
- Färe, R. e Primont, D. (1987) Efficiency Measures for Multiplant Firms with Limited Data, In *Measurement in Economics: Theory and Applications in Economic Indices*, Ed. por W. Eichhorn e E. Diewert, Physica-Verlag, Alemanha, pp. 177-186.
- Farrell, M. (1957) The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society*, Vol 120A, No 3, pp. 353-290.
- Ferrier, G. e Hirschberg, J. (1997) Bootstrap Confidence Intervals for Linear Programming Efficiency Scores: with an Illustration using Italian Banking Data, *Journal of Productivity Analysis*, Vol 8, No 1, pp. 19-33.
- Grosskopf, S. (1986) The Role of the Reference Technology in Measuring Productive Efficiency, *Economic Journal*, Vol 96, No 382, pp. 499-513.
- Grosskopf, S. (1996) Statistical Inference and Nonparametric Efficiency: a Selective Survey, *Journal of Productivity Analysis*, Vol 7, No 2/3, pp. 161-176.
- Hall, P. (1986) On the Number of Bootstrap Simulations Required to Construct a Confidence Interval, *Annals of Statistics*, Vol 14, No 4, pp. 1453-1462.
- Hall, P. (1988) Theoretical Comparison of Bootstrap Confidence Intervals, *Annals of Statistics*, Vol 16, No 3, pp. 927-953.
- Hall, P., Hardle, W. e Simar, L. (1995) Iterated Bootstrap with Applications to Frontier Models, *Journal of Productivity Analysis*, Vol 6, No 1, pp. 63-76.
- Kittelsen, S. (1998) Stepwise DEA: Choosing Variables for Measuring Technical Efficiency in Norwegian Electricity Distribution, Memorandum, No. 6, Oslo University.
- Kittelsen, S. (1999) Monte Carlo Simulations of DEA Efficiency Scores and Hypothesis Tests, Memorandum, No. 9, Oslo University.
- Löthgren, M. (1998) How to Bootstrap DEA Estimators: a Monte Carlo Comparison, Working Paper, No. 223, Stockholm University.
- Marques, R. (2005a) *A Regulação dos Serviços de Água e de Águas Residuais. Aplicação de Técnicas Não Paramétricas de Benchmarking*, Dissertação de Doutoramento, Instituto Superior Técnico, Universidade Técnica de Lisboa.
- Marques, R. (2005b) *A Regulação dos Serviços Públicos*, Edições Silabo, Lisboa.

Marques, R. e Monteiro, A. (2004) Benchmarking the Economic Performance of Portuguese Water and Sewerage Services, In *Data Envelopment Analysis and Performance Measurement*, Ed. por Ali Emrouznejad e Victor Podinovski, Warwick Print, UK, pp. 65-72.

Pedraja-Chaparro, F., Salinas-Jiménez, J. e Smith, P. (1999) On the Quality of the Data Envelopment Analysis Model, *Journal of the Operations Research Society*, Vol 50, No 6, pp. 636-644.

Sarafidis, V. (2002) *An Assessment of Comparative Efficiency Measurement Techniques*, Europe Economics, Office of Water Services, UK.

Silverman, B. (1986) *Density Estimation of Statistics and Data Analysis*, Chapman and Hall, UK.

Silverman, B. e Young, G. (1987) The Bootstrap: to Smooth or not to Smooth?, *Biometrika*, Vol 74, No 3, pp. 469-479.

Simar, L. (1992) Estimating Efficiencies from Frontier Models with Panel Data: a Comparison of Parametric, Non-parametric and Semi-parametric Methods with Bootstrapping, *Journal of Productivity Analysis*, Vol 3, No 1/2, pp. 171-191.

Simar, L. e Wilson, P. (1998) Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Models, *Management Science*, Vol 44, No 1, pp. 46-61.

Simar, L. e Wilson, P. (2000a) Statistical Inference in Non-parametric Frontier Models: the State of the Art, *Journal of Productivity Analysis*, Vol 13, No 1, pp. 49-78.

Simar, L. e Wilson, P. (2000b) A General Methodology for Bootstrapping in Nonparametric Frontier Models, *Journal of Applied Statistics*, Vol 27, No 6, pp. 779-802.

Simar, L. e Wilson, P. (2002) Non-parametric Tests of Returns to Scale, *European Journal of Operational Research*, Vol 139, No 1, pp. 115-132.

Tavares, G. (2002) *A bibliography of data envelopment analysis (1978-2001)*, RUTCOR Research Report, Rutgers University.

## An integer programming model for truss topology optimization

Ana M. Faustino \*

Joaquim J. Júdice †

Isabel M. Ribeiro \*

A. Serra Neves ‡

\* This paper reports research developed under financial support provided by *FCT - Fundação para a Ciência e Tecnologia*, Portugal

Secção de Matemática do Departamento de Engenharia Civil, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal

afausti@fe.up.pt

iribeiro@fe.up.pt

† Support for this author was provided by *Instituto de Telecomunicações* and by *FCT* under grant POCTI/35059/MAT/2000.

Departamento de Matemática da Universidade de Coimbra and Instituto de Telecomunicações, Coimbra, Portugal

joaquim.judice@co.it.pt

‡ Secção de Materiais de Construção do Departamento de Engenharia Civil, Faculdade de Engenharia da Universidade do Porto, Porto

asneves@fe.up.pt

---

### Abstract

In this paper a truss-structure model is described for finding a kinematically stable structure with optimal topology and cross-sectional size and minimum volume. The underlying model finds applications in some civil engineering structural design problems and takes into consideration all the conditions associated with the limit states usually presented in structural safety codes. Ultimate limit states are treated applying plasticity theory, while serviceability limit states are dealt with via elasticity theory. The admissible solution space is discretised using bar elements. A  $0 - 1$  variable is assigned to each one of these elements, in order to indicate if it is or not included in the solution. The mathematical formulation of the model leads to a mixed  $0 - 1$  integer nonlinear program with a nonlinear objective function and linear and bilinear constraints. It is shown that this problem can be reduced into a mixed  $0 - 1$  integer linear program by exploiting the so-called reformulation-linearization technique. Some computational experience is included to highlight the importance of these formulations in practice.

**Keywords:** Truss topology optimization, integer programming, reformulation-linearization technique.

## 1 Introduction

A decision support system to get the optimal skeleton form of the structure that supports the applied loads can become a quite useful tool for an engineer in the design of a structure. This has motivated an area of intense research [5, 6, 15, 16, 17, 18, 20, 25, 26, 28]. Most of these papers describe trusses that, due to its simplicity, are ideal for the study of properties and characteristics associated to optimal structures. The results of this research have found interest in coverings, bridges and towers high-voltage.

Some optimization approaches have been developed for finding optimal trusses. The most challenged one is Topological Optimization, where some elements of an initial structure may be removed in order to get an optimal subset of elements. The structural model changes during the process and this turns out to be one of the most interesting problems in structural optimization. Another important characteristic of these models for civil engineering structures is the fact that the matrix associated with the connection of the bars in the resultant optimal substructure must have full row rank. This condition can not be analytically represented as a constraint, turning the optimization problem even more difficult to process.

Some research has been developed in the last four decades in this area, the majority related with sizing and geometrical optimization of optimal structural design. Due to its complexity, few contributions have been done in optimal topology. However, its importance in structural optimization is recognized since it allows a substantial gain of material and a significant improvement in the design of the structure.

The topological optimization of discrete structures was introduced by Dorn et al. (1964) [9], who applied a linear programming method to optimize truss topology. Since then much research has been developed in this area [19, 28]. In most of these studies, the cross-sectional area of each bar may take a zero value, in which case the bar is eliminated from the structure. Moreover, these variables assume real values in order to make easier the solution process. As is shown in [8, 24] some of these bars cannot be eliminated from the structure, whence the structural problem should have a combinatorial nature.

Another drawback of these studies is the possibility of the optimal topology to correspond to a singular point of the design admissible domain. This fact is a consequence of the discontinuity of some constraints when the cross-sectional area is zero. The singularity of the optimal topology in trusses was firstly shown in [24] and since then it has been a subject of intense research [14, 15, 16, 17, 18]. The substitution of a discrete variable  $x_i \in \{0, 1\}$  by a continuous ones  $0 \leq x_i \leq 1$  is much too strong. Thus for these models to be useful in practice, the variables associated to the cross-sectional area of each bar must be considered as discrete. In this case the difficulty associated with the discontinuity of some constraints completely disappear. Some algorithms have been developed for processing this type of problems [3, 4, 25, 26]. More recently, Bollapragada et al. [6] presented a topological and sizing optimization model that has been formulated as a mixed integer linear program [11].

Another aspect that distinguishes the different contributions in this area is the manner how the limit states for the structural safety codes are defined. In civil engineering structural problems, both ultimate and serviceability limit states have to be considered. The former corresponds to collapse or other forms of structural ruin, while the latter are related to scenarios that should be only reached in extreme circumstances. Such scenarios involve excessive displacements and deformations, vibrations that may cause discomfort or alarm, and damage affecting the form, durability or the use of the structure.

In the ultimate limit states the effect of the design loads, forces or stresses should not

exceed the design values of the structural resistance, admissible forces or stresses. On the other hand in the serviceability limit states the effects should comply with criteria of good performance, that give limits to the displacements to their maximum admissible values.

Structural optimization usually reports to plastic models or to elastic models. In the former case, the ultimate limit states are treated using the Plasticity Theory and serviceability limit states are not considered [22, 27]. The elastic models deal with both ultimate and serviceability limit states by means of the Elasticity Theory [29, 6].

In this paper a mixed model has been adopted using Elasticity Theory for serviceability limit states and Plasticity Theory for ultimate limit states. This model, here designed by elastoplastic, is in our opinion the best suited to the current trend of the safety codes [1]. As in [14, 15, 16, 17, 19, 24], the analysis presented in this paper is based on the trusses. Since the application of kinematically unstable trusses is confined to particular structures and special forces, one of our most important objectives is to eliminate from the feasible set the solutions associated with kinematically unstable trusses. So, each admissible solution is characterized by a vector  $x$ , whose components are assigned to the value 1 or 0, depending on the corresponding bar to be or not to be included in the feasible solution under consideration. The mathematical formulation leads to a mixed integer 0 – 1 nonlinear program. By exploiting the so-called reformulation–linearization technique [21] it is possible to reduce this program into a mixed integer 0 – 1 linear programming problem. Computational experience with the solution of some instances of this integer program shows that the formulation is a quite interesting tool for the design of a truss structure.

This paper is organized as follows. In section 2 the topological optimization model is introduced. Conditions for a truss to be kinematically stable are discussed in Section 3. A mixed-integer linear programming formulation of the model is fully presented in Section 4. The solution of some illustrative examples and some conclusions concerning the validity of the formulation are reported in the last section of the paper.

## 2 A topological optimization model

The admissible structural domain is referenced by a bidimensional cartesian system  $Oxy$ , in which the various alternative solutions for the problem under consideration can be developed. A discretisation [30] of this domain is then considered in which the mesh is composed by bar elements joined at the nodal points.

The structural domain is submitted to the various actions defined in the safety code [1] such as the structural self-weight, wind, earthquake and so on. These actions lead to different  $l$  loading conditions, each of them is represented by nodal point loads

$$f^l = \begin{bmatrix} f_x^l \\ f_y^l \end{bmatrix}.$$

Some of these loads are reactions  $r^l$ , when the associated nodes are connected to the exterior. The nodal displacements

$$u^l = \begin{bmatrix} u_x^l \\ u_y^l \end{bmatrix}$$

are associated to these nodal forces. The stress field within each bar element  $i$  for loading condition  $l$  can be determined from its axial load  $e_i^l$ , while the strain field is given by the axial deformation  $d_i^l$ .

The fundamental conditions to be satisfied in the serviceability limit states are equilibrium, compatibility, boundary conditions and elastic constitutive relations of the structural

material. As in structural civil engineering problems the displacements are generally accepted to be small, the fundamental conditions can be performed on the initial structures.

Equilibrium has to be verified at a nodal level and relates the elastic axial bar forces  $e_e^l$  with support reactions  $r_e^l$  and applied nodal loads  $f^l$  by

$$C^T e_e^l - B r_e^l - f^l = 0, \quad (1)$$

where  $C$  and  $B$  are matrices depending on the structural topology.

The compatibility conditions imply equal displacement for all the bar ends joining at the same node and can be expressed as

$$d_e^l = C u^l, \quad (2)$$

where  $d_e^l$  is the bar deformation vector,  $u^l$  is the nodal displacement vector and  $C$  is the connectivity matrix already used in (1).

The forces  $e_e^l$  in the structural bars are related to the bar deformations  $d_e^l$  by linear elastic constitutive relations given by the so-called Hooke's Law

$$e_e^l = K D_A d_e^l, \quad (3)$$

where  $D_A = \text{diag}\{A_i\}$ , with  $A_i$  a discrete variable associated to the cross-sectional area of bar  $i$  and  $K = \text{diag}\{E_i h_i^{-1}\}$ , with  $E_i > 0$  the Young's modulus of bar  $i$  and  $h_i$  its length. It follows from (1), (2) and (3) that

$$C^T K D_A C u^l - B r_e^l - f^l = 0. \quad (4)$$

The structural boundary conditions are given by

$$u_m^l = 0 \quad (5)$$

for the nodes  $m$  connected to supports with zero displacement.

The nodal displacements should comply with the upper and lower bounds defined in the safety codes

$$u_{min} \leq u^l \leq u_{max}. \quad (6)$$

In structural civil engineering problems, ultimate limit states can be considered on the basis of the Plasticity Analysis. According to the Static Theorem of the Plasticity Theory, the fundamental conditions to be fulfilled by the structure are equilibrium, plasticity conditions and boundary conditions.

The equilibrium conditions are given in a similar form to (1) by

$$C^T e_p^l - B r_p^l - \lambda f^l = 0, \quad (7)$$

where  $e_p^l$  is the plastic force vector,  $r_p^l$  the plastic reaction vector and  $\lambda$  is a partial safety majoration factor for the nodal forces corresponding to the applied actions, prescribed in structural safety codes [1, 2].

The plasticity conditions can be expressed as

$$e_{min} \leq e_p^l \leq e_{max}, \quad (8)$$

where  $e_{min}$  and  $e_{max}$  are the minimum and maximum admissible values for the element forces defined in the code [2].

The conditions (4), (5), (6), (7) and (8) considered so far are satisfied by many solutions in which some bars have zero force. A vector  $x$  is further introduced in the model such that each variable  $x_i$  is associated with bar  $i$  and takes value 1 or 0, depending on the bar  $i$  to be or not to be included in the solution.

The force in a generic bar  $i$  can then be replaced by the product  $x_i e_{p_i}^l$  yielding a null force in non-existing bars. So the axial bar force must verify the following conditions

$$D_x e_{min} \leq e_p^l \leq D_x e_{max}, \quad (9)$$

where

$$D_x = \text{diag}(x_i). \quad (10)$$

Furthermore the diagonal matrix  $D_A$  takes the form  $D_A D_x$ . The model seeks an optimal solution corresponding to the minimum use of structural material  $V$ . If  $A_i$  is the cross-sectional area of bar  $i$  and  $h_i$  is its length, then the objective function takes the form

$$V = \sum_i x_i A_i h_i. \quad (11)$$

### 3 Kinematic Stability

It follows from the description of the constraints and objective function presented in the previous section that the model can generate trusses which are not kinematically stable. In structural civil engineering problems the trusses must be kinematically stable that is, a mechanism can not be generate independently of the loads applied set.

A simple criterion to check a mechanism is the use of the Grubler's Criterion [12]. Let

$$DOF = 2nn^* - nb^* - m^* \quad (12)$$

be the degree of freedom of the truss with  $nn^*$  nodes,  $nb^*$  bars and  $m^*$  simple supports. If the corresponding truss is not a mechanism then  $DOF \leq 0$ .

In order to incorporate this criterion in the model, another vector  $z \in \mathbb{R}^{nn}$  has to be introduced such that each variable  $z_n$  is associated with node  $n \in \{1, \dots, nn\}$  belonging to the initial mesh and takes a value 1 or 0 if it is or not included in the solution, that is, if there are or not bars connected with this node.

So the existence of such a node  $n \in \{1, \dots, nn\}$  can be stated as

$$z_n \leq \sum_{i \in I(n)} x_i \leq |I(n)| z_n, \quad (13)$$

where  $I(n) \subset \{1, \dots, nb\}$  is the set of bar indices  $i$  occurring at node  $n$  and its cardinal  $|I(n)|$  is always different from zero. Note that  $nn$  and  $nb$  are the number of nodes and the number of bars of the initial mesh. So, if in the solution no bars occur in node  $n$ ,  $\sum_{i \in I(n)} x_i = 0$  and by the first inequality of (13)  $z_n = 0$ . Otherwise, if  $\sum_{i \in I(n)} x_i \geq 1$  then by the second inequality of (13)  $z_n = 1$ .

Therefore, the Grubler's Criterion can be stated as

$$2 * \sum_{n=1}^{nn} z_n - \sum_{i=1}^{nb} x_i - \sum_{n=1}^{nn} s_n z_n \leq 0, \quad (14)$$

where  $s_n$  is the number of simple supports associated with node  $n$ .

This criterion eliminates many solutions associated to mechanisms. However, it does not give a sufficient condition for a structure to be kinematically stable. A further condition is required, that is associated with the rank of the matrix  $C^*$  of order  $2nn^* \times (nb^* + na^*)$  matrix (with  $nb^* = (\sum_{i=1}^{nb} x_i)$ ,  $nn^* = \sum_{n=1}^{nn} z_n$  and  $na^* = \sum_{n=1}^{nn} s_n z_n$ ), obtained by elimination of the rows and columns associated to  $z_n = 0$  and the zero columns from the matrix  $[-C^T D_x B]$ . Then [10] the structure is kinematically stable if  $C^*$  has full row rank. In order this condition to be satisfied by the optimal structure, the two following constraints similar to (7) and (8) are incorporated in the model

$$-C^T(D_x e_a) + Br_a + Zf_a = 0 \quad (15)$$

$$D_x e_{min} \leq D_x e_a \leq D_x e_{max}, \quad (16)$$

where  $Z$  is a  $2nn \times 2nn$  diagonal matrix, with diagonal elements  $z_{jj}$  equal to  $z_n$  of the node  $n$  associated to the direction  $j$  and  $f_a$  is a vector of the perturbed nodal load applied in all directions.

Let  $f_a^*$  be the vector obtained by elimination of the rows associated to  $z_n = 0$  from vector  $Zf_a$ . If the rank of matrix  $C^*$  is smaller than  $2nn^*$ , the rank of the augmented matrix  $[C^* f_a^*]$  is  $rank(C^*) + 1$ , since  $f_a^*$  is randomly generated and the constraints (15) are unfeasible. Thus the probability of the matrix  $C^*$  to have linearly dependent rows is null and matrix  $C^*$  should have full row rank in the set of feasible solutions.

To show the need for incorporating constraints (15) and (16) in the model, let us consider the following truss with only one nodal load applied on node II in direction Ox.

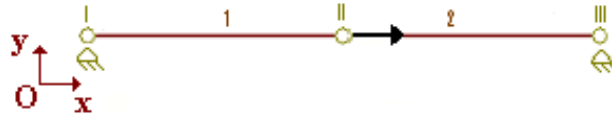


Figure 1: Truss kinematically unstable

The truss, shown in Figure 1, is not a mechanism, since  $DOF = 0$ . If the constraints (15) and (16) are not considered in the formulation of the model, this truss can be an optimal solution for the problem. However, the matrix  $C^*$  associated with this truss

$$C^* = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

has not full row rank and this truss is kinematically unstable.

Since, in constraint (15),  $f_a \neq 0$  is a vector of perturbed nodal load applied in all direction and  $C^*$  has not full row rank, the constraints (15) are unfeasible. Thus kinematically unstable trusses are not feasible solutions for our model.

The model underlying to this problem can be formulated as the following mixed-integer



0 – 1 nonlinear program:

$$\begin{aligned}
 \text{(P1)} \quad & \text{Minimize} \quad V = \sum_{i=1}^{nb} x_i A_i h_i \\
 & \text{subject to} \\
 & M(D_A D_x) d^l - B r_e^l - f^l = 0 \quad (17) \\
 & d^l = C u^l \quad (18) \\
 & u_{\min_j} \leq u_j^l \leq u_{\max_j} \quad (19) \\
 & u_{j_m}^l = 0 \quad (20) \\
 & -C^T e_p^l + B r_p^l + \lambda f^l = 0 \quad (21) \\
 & D_x D_A t_{\min} \leq e_p^l \leq D_x D_A t_{\max} \quad (22) \\
 & z_n \leq \sum_{i \in I(n)} x_i \leq |I(n)| z_n \quad (23) \\
 & 2 * \sum_{n=1}^{nn} z_n - \sum_{i=1}^{nb} x_i - \sum_{n=1}^{nn} s_n z_n \leq 0 \quad (24) \\
 & -C^T e_a + B r_a + Z f_a = 0 \quad (25) \\
 & D_x D_A t_{\min} \leq e_a \leq D_x D_A t_{\max} \quad (26) \\
 & A_i \in S_i = \{A_{i1}, \dots, A_{iN_i}\}, \quad (27) \\
 & x_i \in \{0, 1\}, \quad i = 1, \dots, nb, \quad (28)
 \end{aligned}$$

with  $l = \{1, \dots, nc\}$ ,  $j = \{1, \dots, 2nn\}$ ,  $j_m = \{1, \dots, na\}$ ,  $n = \{1, \dots, nn\}$  and  $i = \{1, \dots, nb\}$ .

The meanings of the parameters in this program are presented below:

$nb$	number of bars;
$na$	number of simple supports;
$nn$	number of nodes;
$nc$	number of loading conditions;
$N_i$	number of discrete sizes available for cross-sectional area of bar $i$ ;
$A_{ik}$	$k$ -th discrete size for bar $i$ ;
$S_i$	set of possible discrete cross-sectional area available for bar $i$ ;
$C$	$nb \times 2nn$ matrix of direction cosines relating bar forces with nodal directions;
$B$	$2nn \times na$ matrix of direction cosines relating nodal directions with nodal supports directions;
$D_x$	diagonal matrix, $[diag(x_i)]$ ;
$D_A$	diagonal matrix, $[diag(A_i)]$ ;
$M$	matrix $\left[ C^T diag \left( \frac{E_i}{h_i} \right) \right]$ ;
$E_i$	Young's modulus of bar $i$ ;
$h_i$	length of bar $i$ ;
$f_j^l$	applied nodal loads in direction $j$ for loading condition $l$ ;
$I(n)$	set of bars indices which occur in node $n$ ;

$\lambda$	safety factor;
$ I(n) $	cardinal of set $I(n)$ ;
$s_n$	number of simple supports associated with node $n$ ;
$Z$	$2nn \times 2nn$ diagonal matrix, with $z_{jj}$ equal to $z_n$ of the node $n$ associated to the direction $j$ ;
$f_a$	perturbed nodal load applied in all directions;
$t_{min_i}$	minimum stress in compression of bar $i$ , where $t_{min_i} = \frac{e_{min_i}}{A_i}$ ;
$t_{max_i}$	maximum stress in tension of bar $i$ , where $t_{max_i} = \frac{e_{max_i}}{A_i}$ ;
$u_{min_j}, u_{max_j}$	minimum and maximum nodal displacement in direction $j$ .

The variables have the following meanings:

$A_i$	cross-sectional area of bar $i$ ;
$x_i$	0 – 1 variable stating whether the bar $i$ exists or not;
$e_{p_i}^l$	bar force of bar $i$ for loading condition $l$ ;
$r_{p_m}^l$	plastic reaction in supports $m$ for loading condition $l$ ;
$r_{e_m}^l$	elastic reaction in supports $m$ for loading condition $l$ ;
$d_i^l$	deformation of bar $i$ for loading condition $l$ ;
$u_j^l$	nodal displacement in the direction $j$ for loading condition $l$ ;
$z_n$	0 – 1 variable stating whether the node $n$ exists or not;
$e_{a_i}$	bar force of bar $i$ for the perturbed nodal load;
$r_{a_m}$	plastic reaction in supports $m$ for the perturbed nodal load.

This mixed-integer 0 – 1 nonlinear problem has  $nc \times (4nn + 3nb) + 4nn + 2nb + 1$  constraints and  $nc \times (2nb + 2nn + 2na) + 3nb + nn + na$  variables. A similar formulation has been used in [6, 13]. However, as discussed in [23], it represents a sizing problem rather than a topological optimization problem.

## 4 A mixed-integer linear programming formulation

The formulation (P1) described in Section 3 contains a bilinear objective function on the variables  $x_i$  and  $A_i$  and some linear and bilinear constraints. The existence of these bilinear functions imposes some limitations to the use of commercial software for processing mixed-integer programs. In this section formulation (P1) is reduced to a mixed-integer 0 – 1 linear program by using the so-called reformulation-linearization technique (RLT) [21].

The RLT consists of two steps, namely the reformulation and the linearization. The reformulation phase defines a set of nonnegative variable factors, based on the various bound restrictions, and then forms products of these factors with the original constraints to generate another implied nonlinear constraints. In the linearization phase, an appropriate technique of substitution of variables is used to linearize these nonlinear constraints. In general, the original nonlinear program and the resulting linear program do not have the same optimal solutions. However, there are certain special cases in which those problems possess exactly the same optimal solution. In this section the mixed nonlinear 0 – 1 integer program (P1) is shown to be equivalent to a mixed-integer 0 – 1 linear program. To do this, the vectors  $d^l$  are assumed to be bounded, that is, there exist fixed constant vectors  $d_{min}$  and  $d_{max}$  such that

$$d_{min} \leq d^l \leq d_{max} \quad l = 1, \dots, nc. \quad (29)$$

The existence of these bounds is a consequence of the conditions (18), (19) and (20). In the Reformulation Phase the constraints (29) are multiplied by  $x_i$  and  $1 - x_i$ ,  $i = 1, \dots, nb$  and the further constraints

$$\begin{aligned} x_i d_{\min_i} &\leq x_i d_i^l \leq x_i d_{\max_i} \\ d_{\min_i}(1 - x_i) &\leq d_i^l - x_i d_i^l \leq d_{\max_i}(1 - x_i) \end{aligned}$$

are incorporated in (P1). In the Linearization Phase new variables  $v_i^l$  defined by

$$v_i^l = x_i d_i^l \quad (30)$$

are introduced. Using these two phases of *RLT* technique, the problem (P1) can be stated as follows:

$$\begin{aligned} \text{(P2)} \quad & \text{Minimize} \quad V = \sum_{i=1}^{nb} x_i A_i h_i \\ & \text{subject to} \\ & M(D_A)v^l - Br_e^l - f^l = 0 \quad (31) \\ & d^l = Cu^l \quad (32) \\ & u_{j_m}^l = 0 \quad (33) \\ & u_{\min_j} \leq u_j^l \leq u_{\max_j} \quad (34) \\ & d_{\min_i} x_i \leq v_i^l \leq d_{\max_i} x_i \quad (35) \\ & d_{\min_i}(1 - x_i) \leq d_i^l - v_i^l \leq d_{\max_i}(1 - x_i) \quad (36) \\ & -C^T e_p^l + Br_p^l + \lambda f^l = 0 \quad (37) \\ & D_x D_A t_{\min} \leq e_p^l \leq D_x D_A t_{\max} \quad (38) \\ & z_n \leq \sum_{i \in I(n)} x_i \leq |I(n)| z_n \quad (39) \\ & 2 * \sum_{n=1}^{nn} z_n - \sum_{i=1}^{nb} x_i - \sum_{n=1}^{nn} s_n z_n \leq 0 \quad (40) \\ & -C^T e_a + Br_a + Z f_a = 0 \quad (41) \\ & D_x D_A t_{\min} \leq e_a \leq D_x D_A t_{\max} \quad (42) \\ & A_i \in S_i = \{A_{i1}, \dots, A_{iN_i}\} \quad (43) \\ & x_i \in \{0, 1\}, \quad (44) \end{aligned}$$

where  $l = 1, \dots, nc$ ,  $j = 1, \dots, 2nn$ ,  $j_m = 1, \dots, na$ ,  $n = 1, \dots, nn$  and  $i = 1, \dots, nb$ .

Next, the equivalence between the programs (P1) and (P2) is established. Since the objective functions of the two problems are the same, then it is enough to prove that there is a bijective correspondence between the feasible solutions of the two problems. If the vector  $\bar{v}^l$ , of components  $\bar{v}_i^l = \bar{x}_i \bar{d}_i^l$ , is introduced then by definition of the variables  $x_i$  it is easy to conclude that  $(\bar{x}, \bar{a}, \bar{e}_p^l, \bar{r}_p^l, \bar{d}^l, \bar{u}^l, \bar{r}_e^l)$  is a feasible solution of (P1) if and only if  $(\bar{x}, \bar{a}, \bar{e}_p^l, \bar{r}_p^l, \bar{v}^l, \bar{d}^l, \bar{u}^l, \bar{r}_e^l)$  is feasible for (P2). So the two formulations are equivalent.

In problem (P2), for each bar  $i$ , the cross-sectional area  $A_i$  takes a discrete value in the set  $S_i$ . According to the cardinal of these sets, two cases may occur and are discussed below.

- If  $\#S_i = 1$  for all  $i$ , then (P2) is a mixed-integer linear program (MILP).
- If there exists an  $i$  such that  $\#S_i > 1$ , then (P2) is a mixed-integer nonlinear program (MINLP).

In the latter case, problem (P2) can be also reformulated as a MILP. For that purpose, let

$$A_i x_i = \sum_{k=1}^{N_i} A_{ik} y_{ik}, \quad (45)$$

where  $A_i \in \{A_{i1}, \dots, A_{iN_i}\}$  and  $y_{ik}$  are binary variables such that

$$\sum_{k=1}^{N_i} y_{ik} = x_i. \quad (46)$$

Since  $x_i$  is a binary variable for each  $i$ , then the last expression is equivalent to

$$\sum_{k=1}^{N_i} y_{ik} \leq 1. \quad (47)$$

So, for each bar  $i$ ,  $A_i x_i$  is zero or assumes the value of only one discrete size in the set  $S_i$ . Moreover,

$$A_i v_i^l = \sum_{k=1}^{N_i} A_{ik} q_{ik}^l, \quad (48)$$

where  $q_{ik}^l$  is the deformation corresponding to the  $k$ -th possible discrete size for the cross-sectional area of the  $i$ -th bar under  $l$ -th loading condition. The variables  $q_{ik}^l$  satisfy the following expressions

$$\sum_{k=1}^{N_i} q_{ik}^l = v_i^l \quad (49)$$

$$d_{\min_i} y_{ik} \leq q_{ik}^l \leq d_{\max_i} y_{ik}. \quad (50)$$

By using the expressions (45), (46), (48) and (49) and adding the constraints (47) and (50)

the following mixed-integer 0 – 1 linear program is obtained:

$$\begin{aligned}
 \text{(P3)} \quad & \text{Minimize} \quad V = \sum_{i=1}^{nb} \left( \sum_{k=1}^{N_i} A_{ik} y_{ik} \right) h_i \\
 & \text{subject to} \\
 & \sum_{i=1}^{nb} M_{ji} \left( \sum_{k=1}^{N_i} A_{ik} q_{ik}^l \right) - \sum_{m=1}^{na} B_{jm} r_{e_m}^l - f_j^l = 0 \quad (51) \\
 & d^l = C u^l \quad (52) \\
 & u_{min} \leq u^l \leq u_{max} \quad (53) \\
 & d_{min_i} y_{ik} \leq q_{ik}^l \leq d_{max_i} y_{ik} \quad (54) \\
 & d_{min_i} \left( 1 - \sum_{k=1}^{N_i} y_{ik} \right) \leq d_i^l - \sum_{k=1}^{N_i} q_{ik}^l \leq d_{max_i} \left( 1 - \sum_{k=1}^{N_i} y_{ik} \right) \quad (55) \\
 & u_{j_m}^l = 0 \quad (56) \\
 & -C^T e_p^l + B r_p^l + \lambda f^l = 0 \quad (57) \\
 & t_{min_i} \sum_{k=1}^{N_i} A_{ik} y_{ik} \leq e_{p_i}^l \leq t_{max_i} \sum_{k=1}^{N_i} A_{ik} y_{ik} \quad (58) \\
 & z_n \leq \sum_{i \in I(n)} \sum_{k=1}^{N_i} y_{ik} \leq |I(n)| z_n \quad (59) \\
 & 2 * \sum_{n=1}^{nn} z_n - \sum_{i=1}^{nb} \sum_{k=1}^{N_i} y_{ik} - \sum_{n=1}^{nn} s_n z_n \leq 0 \quad (60) \\
 & -C^T e_a + B r_a + Z f_a = 0 \quad (61) \\
 & t_{min_i} \sum_{k=1}^{N_i} A_{ik} y_{ik} \leq e_{a_i} \leq t_{max_i} \sum_{k=1}^{N_i} A_{ik} y_{ik} \quad (62) \\
 & y_{ik} \in \{0, 1\} \quad (63) \\
 & \sum_{k=1}^{N_i} y_{ik} \leq 1, \quad (64)
 \end{aligned}$$

where  $l = 1, \dots, nc$ ,  $j = 1, \dots, 2nn$ ,  $j_m = 1, \dots, na$ ,  $k = 1, \dots, N_i$ ,  $n = 1, \dots, nn$  and  $i = 1, \dots, nb$ .

This problem is a mixed-integer linear program with some 0 – 1 variables with

$$nc \times \left( 4nn + 5nb + 2 \sum_{i=1}^{nb} N_i \right) + 3nb + 4nn + 1$$

constraints and

$$nc \times \left( 2nb + 2nn + 2na + \sum_{i=1}^{nb} N_i \right) + \sum_{i=1}^{nb} N_i + nn + nb + na$$

variables. This dimension is considerably larger than the dimension of the problem (P1). However, all the functions involved in this formulation are linear and this enables the use of an integer linear programming code for its solution.

## 5 Computational Experience and Conclusions

In this section computational experience is reported on the solution of some structural models by using the mixed-integer formulation (P3). These experiences have been performed on a Pentium IV 2.4GHz with 256 MB of RAM. Moreover, the commercial program OSL of the GAMS collection [7] has been used to process the mixed-integer linear programs (P3).

### (I) Test Problems

In each test problem the corresponding initial structure consists of nodal points and bars and takes a similar form to the type mesh displayed in Figure 2.

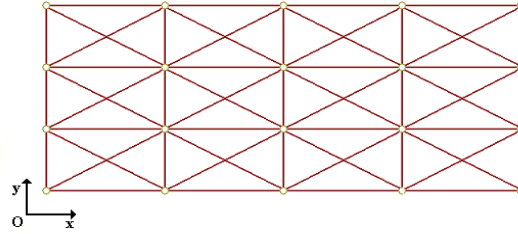


Figure 2: Initial mesh

The main goal of this model is to find the set of included bars in the so-called optimal shape of the structure, which is given by the values of the 0 – 1 variables  $x_i$  in the optimal solution of the problem.

Different types of sizes of initial meshes, and different applied nodal forces have been taken in consideration in the construction of the test problems. Four sizes of initial meshes,  $M_i$ ,  $i = 0, \dots, 3$ , have been considered, whose topologies are presented in Table 1 leading to five test problems Pt0 to Pt4, according to the following definitions:

- Pt0 - mesh  $M0$  and only one nodal load is applied ( $f_{x_4}^1 = 65$ ,  $f_{y_4}^1 = 0$ ).
- Pt1, Pt2 - mesh  $M1$  and two types of applied nodal loads are applied. In Pt1 only one nodal load ( $f_{x_8}^1 = 0$ ,  $f_{y_8}^1 = -65$ ) is applied, while two nodal loads ( $f_{x_8}^1 = 0$ ,  $f_{y_8}^1 = -65$ ,  $f_{x_9}^2 = -40$ ,  $f_{y_9}^2 = -40$ ) are applied in Pt2.
- Pt3 - mesh  $M2$  and two nodal loads are simultaneously applied ( $f_{x_3}^1 = -45.9619$ ,  $f_{y_3}^1 = 45.9619$ ,  $f_{x_{12}}^1 = -45.9619$ ,  $f_{y_{12}}^1 = -45.9619$ ).
- Pt4 - mesh  $M3$  and only one nodal load is applied ( $f_{x_{23}}^1 = 0$ ,  $f_{y_{23}}^1 = -65$ ).

In these definitions the following parameters are used:

- $f_{x_n}^l$  nodal load in (kN) applied in node  $n$  in direction  $Ox$  for loads combination  $l$ ;  
 $f_{y_n}^l$  nodal load in (kN) applied in node  $n$  in direction  $Oy$  for loads combination  $l$ .

Table 1: Test Problems Meshes.

	MESH	$h_x$	$h_y$	$nal$	$nb$	$nn$	$na$	$N_i$	$S_i$
GROUP I	M0	4	3	$2 \times 2$	6	4	3	1	3
	M1	8	6	$3 \times 3$	20	9	3	1	3
	M2	6	9	$3 \times 4$	29	12	8	1	3
	M3	16	12	$5 \times 5$	72	25	3	1	3
GROUP II	SM1	8	6	$3 \times 3$	20	9	3	2	0.5;3
	SM2	8	6	$3 \times 3$	20	9	3	3	0.5;1;2
	SM3	6	9	$3 \times 4$	29	12	8	2	0.5;3
	SM4	6	9	$3 \times 4$	29	12	8	3	0.5;2;3

In Table 1 the following notations are included.

- $nal$  dimension of the mesh in terms of number of nodal in  $Ox$  and  $Oy$  axes, respectively (in Figure 2,  $nal = 5 \times 4$ )
- $h_x$  total length (in  $m$ ) to the  $Ox$  axis
- $h_y$  total length (in  $m$ ) to the  $Oy$  axis
- $nb$  number of bars
- $nn$  number of nodes
- $na$  number of simple supports
- $S_i$  set of discrete sizes available for cross-sectional area of bar  $i$  (in  $cm^2$ )
- $N_i$  number of discrete sizes available for cross-sectional area of bar  $i$

In the first group of test problems, structures have been considered for which an unique discrete value is given in each cross-sectional area of each bar. In the second set of problems it is allowed that each bar of the structure assumes one of the values in a finite set of discrete sizes available for its cross-sectional area. This last group leads to four additional test problems, denoted by St1, St2, St3 and St4, and whose associated initial meshes are SM1, SM2, SM3 and SM4, respectively. The meshes SM1 and SM2 have the same dimensions of the M1 mesh, while SM3 and SM4 have the same dimensions of the ones in M2. The nodal loads applied in St1 and St2 are the same as in Pt1, while in St3 and St4 are the same as in Pt3. The number of constraints ( $nr$ ) and the number of variables ( $nv$ ) of formulations P3 associated to these test problems are presented in the Table 2.

Table 2: Dimensions of Test Problems.

		P3	
	PROB	$nr$	$nv$
GROUP I	Pt0	93	51
	Pt1	273	136
	Pt2	449	220
	Pt3	387	205
	Pt4	921	444
GROUP II	St1	313	176
	St2	353	216
	St3	445	263
	St4	503	321

In all test problems the displacements and bars stress limits considered are  $u_{max} = -u_{min} = 50cm$ ,  $t_{max} = -t_{min} = 355MPa$ , respectively and the partial safety factor

$\lambda$  is equal to 1.5.

## (II) Numerical Results

Tables 3 and 4 display the numerical results corresponding to the performance of the OSL code for finding an optimal solution of each one of the test problems associated to formulation (P3). This performance is evaluated in terms of number of pivot step iterations (NI), nodes (ND) and CPU time in seconds (T). In this table, the notation "> 25000000" is used whenever the code has been unable to find an optimal solution for a particular test problem after 25000000 iterations (pivot step iterations). The optimal objective function value (OBJ), in  $dm^3$ , found by the solver is also included. In the failure case (NI > 25000000), (OBJ) corresponds to the best upper bound computed by the algorithm. Note that best solution found may be not optimal for the problem (P3) in this last case.

Table 3: Numerical results with only one available value for the cross-sectional area.

PROB	OSL			OBJ ( $dm^3$ )
	NI	T	ND	
PT0	53	0.04	7	3.60
PT1	3033	0.69	311	10.80
PT2	5579	1.77	497	12.90
PT3	891143	325.64	82075	11.92
PT4	>25000000	15018.78	347541	27.30

Table 4: Numerical results with more than one available value for the cross-sectional area.

PROB	OSL			OBJ ( $dm^3$ )
	NI	T	ND	
ST1	64943	22.80	8132	7.05
ST2	57473	30.01	10052	4.90
ST3	4788682	3996.54	411084	6.29
ST4	20606789	61486.08	1496081	5.46

The results presented in Tables 3 and 4 show that for meshes of small and average dimension the integer programming code OSL is able to find an optimal solution that leads to structural shapes containing a smaller number of nodes and bars. This is in accordance to the objective of finding an optimal structure with the smallest possible volume. However, for meshes of larger dimensions we may not guarantee that the best solution found is optimal. But, even in this case the corresponding solution leads into a structure of small volume. So the results clearly indicate the validity of the new formulation, as in general the solution of the integer program corresponds to optimal structural shapes that are kinematically stable and involve a small amount of material. This conclusion is well illustrated in Figure 3, which includes the initial mesh M2 and the optimal and deformed forms obtained for the problems PT3, ST3 and ST4 associated to this mesh. The bar forces associated to the bars as well as the elastic and the plastic reactions are indicated under parenthesis. It is not difficult to see that the optimal structure requires a small amount of material and is kinematically stable.



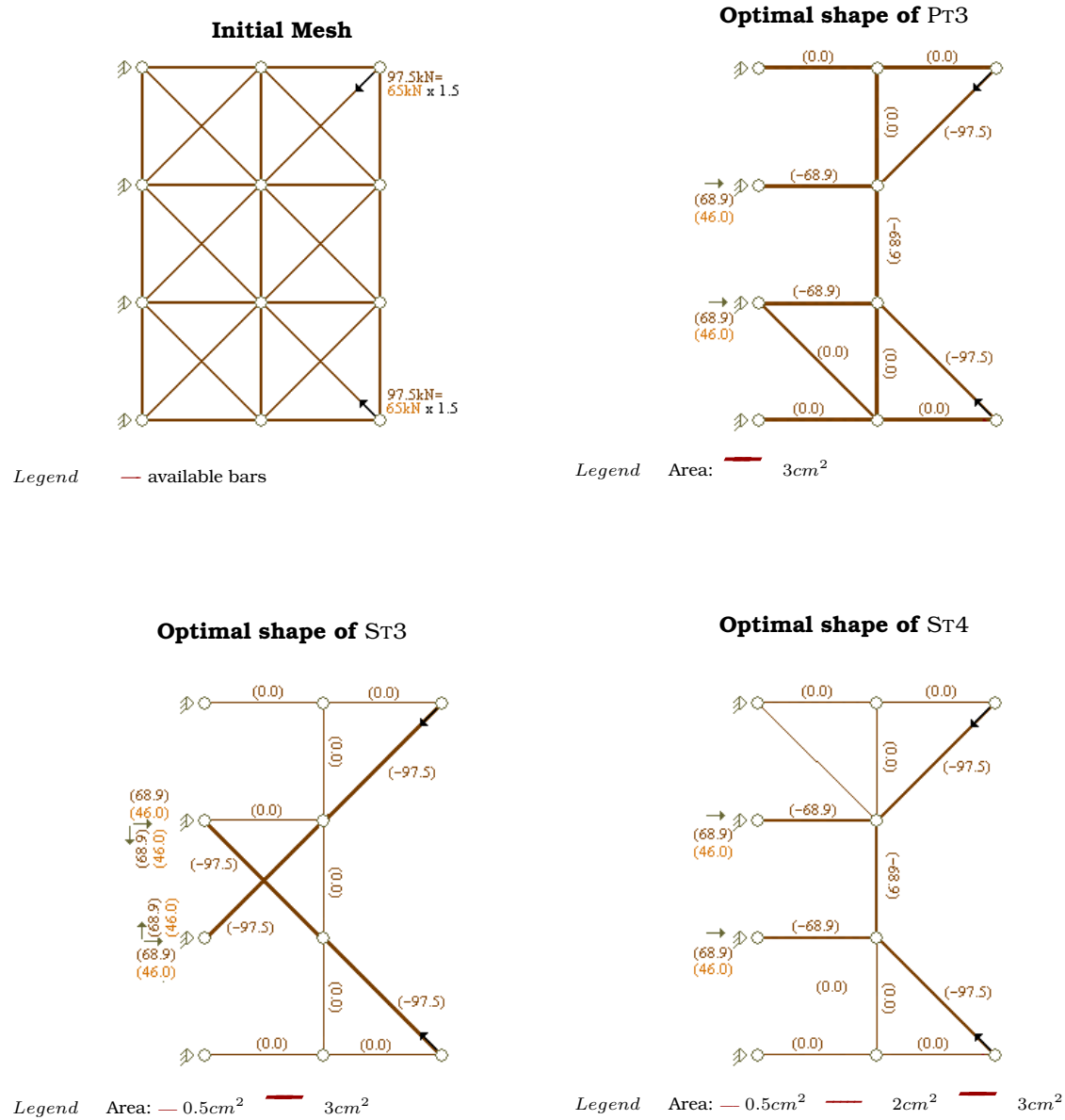


Figure 3: Initial and optimal shape structures of PT3, ST3 and ST4

The numerical results also show that a branch-and-bound algorithm, such as OSL, may face difficulties to find the optimal solution for the integer program corresponding to structures with large number of nodes and bars. An alternative technique for processing the linear integer program in this instance is then required. This algorithm should be designed to find the global minimum for the associated integer program or at least a good feasible solution requiring a small amount of material. The design of such algorithm will certainly be a topic for future research.

## References

- [1] Eurocode 1 EN 1991. *Basis of Design and Actions on Structures*. CEN, Brussels, 1998.
- [2] Eurocode 2 EN 1992. *Design of Concrete Structures - Part 1: General Rules and Rules for Buildings*. CEN, Brussels, 1999.
- [3] I. Arora and M. Haung. Methods for optimization of nonlinear problems with discrete variables: a review. *Structural Optimization*, 8:69–85, 1994.
- [4] J. Bauer. A survey of methods for discrete optimum structural design. *Computer Assisted Mechanics and Engineering Sciences*, 1:27–38, 1994.
- [5] M. P. Bendsoe and O. Sigmund. *Topology Optimization - Theory, Methods and Applications*. Springer, 2003.
- [6] S. Bollapragada, O. Ghattas, and J. Hoocher. Optimal design of truss structures by logic-based branch and cut. *Operations Research*, 49:42–51, 2001.
- [7] A. Brooke, D. Kendrick, A. Meeraus, and R. Raman. *GAMS a User's Guide*. GAMS Development Corporation, New York, 1998.
- [8] W. Dobbs and L. Felton. Optimization of truss geometry. *ASCE Journal of Structural Division*, 95:2105–2118, 1969.
- [9] W. Dorn, R. Gomory, and H. Greenberg. Automatic design of optimal structures. *Journal de Mécanique*, 3:25–52, 1964.
- [10] A. Ghali and A. M. Neville and T. G. Brown. *Structural analysis: A unified classical and matrix approach*. Spon Press, London, 2003.
- [11] O. Ghattas and I. Grossmann. MINLP and milp strategies for discrete sizing structural optimization problems. In *Proceedings of ASCE 10th Conference on Electronic Computation*. Indianapolis, 1991.
- [12] A. Ghosh and A. Mllik. *Theory of Mechanisms and Machines*. Affiliated East-West Press, New Delhi, 1988.
- [13] I. Grossmann, V. T. Voudouris, and O. Ghattas. Mixed-integer linear programming formulations of some nonlinear discrete design optimization problems. In C. A. Floudas and P. M. Pardalos, editors, *Recent Advances in Global Optimization*. Princeton University Press, 1992.
- [14] X. Guo and G. Cheng. An extrapolation approach for the solution of singular optima. *Structural and Multidisciplinary Optimization*, 19:255–262, 2000.
- [15] X. Guo, G. Cheng, and K. Yamazaki. A new approach for the solution of singular optima in truss topology optimization with tress and local buckling constraints. *Structural and Multidisciplinary Optimization*, 22:364–372, 2001.

- [16] A. Hoback. Optimization of singular problems. *Structural Optimization*, 12:93–97, 1996.
- [17] U. Kirsch. On singular topologies in optimum structural design. *Structural Optimization*, 2:133–142, 1990.
- [18] G. Rozvany. Difficulties in truss topology optimization with stress, local bucking and system stability constraints. *Structural Optimization*, 11:213–217, 1996.
- [19] G. Rozvany, M. Bendsoe, and Kirsch. Layout optimization of structures. *Applied Mechanics Reviews*, 48:41–118, 1995.
- [20] E. Salajegheh and G. Vanderplaats. Optimum design of trusses with discrete sizing and shape variables. *Structural Optimization*, 6:79–85, 1993.
- [21] H. Sherali and W. Adams. *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*. Kluwer Academic Publishers, Boston, 1999.
- [22] H. Simon. *The Sciences of the Artificial*. MIT Press, Massachusetts, 1969.
- [23] M. Stolpe and k. Svanberg. Modeling topology optimization problems as mixed linear programs. In *Optimization and Systems Theory*. Department of Mathematics, Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden, 2001.
- [24] G. Sved and Z. Ginos. Structural optimization under multiple loading. *International Journal of Mechanical Sciences*, 10:803–805, 1968.
- [25] A. Templeman. Heuristic methods in discrete structural optimization. In W. Gutkowski, editor, *Discrete Structural Optimization*, pages 135–165. CISM Courses and Lectures No. 373, Udine, Italy, 1997.
- [26] P. Thander and G. Vanderplaats. Survey of discrete variable optimization for structural design. *Journal of the Structural Engineering*, 121:301–306, 1995.
- [27] A. Toakley. Optimum design using available sections. *Journal of the Structural Division, American Society of Civil Engineers*, 34:1219–1241, 1968.
- [28] B. Topping. Topology design of discrete structures. In M. Bendsoe and C. Soares, editors, *Topology design of structures*, pages 517–534. Kluwer Academic Publishers, Amsterdam, 1993.
- [29] G. Vanderplaats and P. Thanedar. A survey of discrete variable optimization for structural design. In *Proceedings of the ASCE Tenth Conference on Electronic Computations*, pages 173–180. Dayton, Ohio, 1991.
- [30] O. Zienkiewicz. *The Finite Element Method*. MacGraw-Hill, Berkshire, 1997.



## **REVISTA INVESTIGAÇÃO OPERACIONAL**

### **Política Editorial**

*Investigação Operacional (IO)* é a revista científica da APDIO - Associação Portuguesa de Investigação Operacional. A política editorial da IO é publicar artigos originais e de elevada qualidade que contribuam para a teoria, metodologia, técnicas e software de Investigação Operacional e a sua aplicação a diferentes campos. A Revista também publica artigos com revisões relevantes de temas de IO. Casos de sucesso na aplicação a problemas práticos são especialmente bem vindos.

### **Processo de Aceitação**

Todos os manuscritos submetidos para publicação são revistos e aceites apenas com base na avaliação da sua qualidade, importância e adequação à política editorial. Será responsabilidade do Editor interpretar a avaliação dos revisores. A contribuição de cada artigo deve estar claramente evidenciada na Introdução. Critérios como a relação com literatura existente, comprimento e estilo do artigo são tidos em consideração. Uma indicação clara da viabilidade de aceitação do artigo é habitualmente dada na primeira fase de revisão do artigo.

Será requerido aos autores de um artigo aceite que transfiram os direitos de autoria para a APDIO, que assegurará a mais ampla disseminação possível de informação. Os volumes da Revista são publicados em papel, e distribuídos a todos os associados da APDIO, e em formato electrónico na rede SciELO - Scientific Electronic Library Online.

### **Resumos dos Artigos indexados em**

IAOR - International Abstracts in Operations Research

### **Instruções aos Autores**

1. *Submeter artigos* para publicação ao editor principal, de preferência por e-mail em Microsoft Word ou "Portable Document Format" (PDF) para [jfo@fe.up.pt](mailto:jfo@fe.up.pt), ou por correio normal (quatro cópias) para o seguinte endereço: Prof. José Fernando Oliveira, Departamento de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 PORTO, Portugal.
2. *Língua.* Os artigos devem ser escritos em Português, Inglês ou Espanhol.
3. *Os Manuscritos* devem ser impressos. Numerar as páginas consecutivamente.
4. *A primeira página* do manuscrito escrito em português ou em espanhol deve ter a seguinte informação: (a) Título; (b) nome, e-mail e afiliação institucional dos autores; (c) um resumo; (d) palavras-chave; (e) título em inglês (f) um resumo em inglês; (g) palavras-chave em inglês; (h) identificação do autor correspondente. Se o manuscrito for escrito em inglês, a primeira página deve ter a seguinte informação: (a) Título em inglês; (b) nome, e-mail e afiliação institucional dos autores; (c) um resumo em inglês; (d) palavras-chave em inglês; (e) identificação do autor correspondente.

5. *Agradecimentos*, incluindo informação sobre apoios, dever ser colocados imediatamente antes da secção de referências.
6. *Notas de rodapé* devem ser evitadas.
7. *Formulas* que são referenciadas devem ser numeradas consecutivamente ao longo do manuscrito como (1), (2), etc. do lado direito.
8. *Figuras*, incluindo grafos e diagramas, devem ser numerados consecutivamente em numeração árabe.
9. *Tabelas* devem ser numeradas consecutivamente em numeração árabe.
10. *Referências*. Citar apenas as mais relevantes e listar só as que são citadas no texto. Indicar as citações no texto através de parênteses rectos, e.g., [4]. No final do artigo listar as referências alfabeticamente por apelido do primeiro autor e numerá-las consecutivamente, de acordo com o seguinte formato: Artigos: autore(s), título, nome e volume da revista (ou livro, mas neste caso incluir o nome dos editores), ano e páginas. Livros: Autor(es), título, editor, ano.
11. *Artigos aceites* devem ser enviados pelo autor ao editor, de preferência na forma de um ficheiro fonte em LaTeX com ficheiros EPS para as figuras, juntamente com um ficheiro PDF ou Postscript. Em alternativa, ficheiros fonte em Word são também aceites. Para garantir uma boa qualidade gráfica, as figuras devem ser em formato vectorial; formatos raster como JPG, BMP, GIF, etc. devem ser evitados.
12. *Provas dos artigos* serão enviadas por e-mail como ficheiros PDF para o autor correspondente. Corrigir as provas cuidadosamente, e restringir as correcções apenas aos pontos em que as provas diferem do manuscrito. Desvios à versão aceite pelo editor são apenas possíveis com a autorização prévia e explícita do editor. Trinta separatas de cada artigo são enviados gratuitamente ao autor correspondente.

### **Informação sobre a Publicação**

Investigação Operacional (ISSN 0874-5161) está registada na Secretaria de Estado da Comunicação Social sob o número 108335. Os volumes da Revista são publicados em papel, e distribuídos a todos os associados da APDIO, e em formato electrónico na rede SciELO - Scientific Electronic Library Online. O preço da assinatura anual é de 25 euros. Os volumes são enviados por correio normal. Informação adicional sobre assinaturas pode ser solicitada ao Secretariado da APDIO- CESUR, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 LISBOA, Portugal. Tel. +351 218 407 455 - [www.apdio.pt](http://www.apdio.pt) - [apdio@cesur.civil.ist.utl.pt](mailto:apdio@cesur.civil.ist.utl.pt)

## **JOURNAL INVESTIGAÇÃO OPERACIONAL**

### **Editorial Policy**

*Investigação Operacional (IO)* is the scientific journal of APDIO - Associação Portuguesa de Investigação Operacional (the Portuguese Operational Research Association). The editorial policy of IO is to publish high quality and original articles that contribute to theory, methodology, techniques and software of Operational Research (OR) and its application to different fields. It also publishes articles with relevant reviews of OR subjects. Cases of successful application of OR to practical problems are specially welcome.

### **Acceptance Process**

All manuscripts submitted for publication are refereed and accepted only on the basis of its quality, importance and adequacy to the editorial policy. It will be the responsibility of the Editor to interpret the referee's assessment. The contribution of each paper should be clearly stated in the introduction. Criteria such as relationship with existing literature, length and style are taken into account. A clear indication on the suitability of a manuscript is usually provided after the first round of refereeing. The authors of an accepted paper will be asked to transfer its copyright to the publisher, which will ensure the widest possible dissemination of information. The volumes of the journal are published in hardcopies, which are distributed to all APDIO associates, and in electronic format in SciELO - Scientific Electronic Library Online.

### **Articles are abstracted/indexed in**

IAOR - International Abstracts in Operations Research

### **Instructions to Authors**

1. *Submit papers* for publication to the main editor, preferably by e-mail in Microsoft Word or "Portable Document Format"(PDF) to jfo@fe.up.pt, or by ordinary mail (four copies) to the following address: Prof. José Fernando Oliveira, Departamento de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 PORTO, Portugal.
2. *Language.* Papers must be written in Portuguese, English or Spanish.
3. *Manuscripts* should be typewritten or typeset. Number the pages consecutively.
4. *The first page* of the manuscript written in English should contain the following information: (a) Title; (b) names, e-mails and institutional affiliations of the authors; (c) an abstract; (d) keywords (f) identification of the corresponding author.
5. *Acknowledgements*, including support information, should be placed prior to the references section.
6. *Footnotes* should be avoided.

7. *Formulas* that are referred to should be numbered consecutively throughout the manuscript as (1), (2), etc. on the right.
8. *Figures*, including graphs and diagrams, should be numbered consecutively in Arabic numbers.
9. *Tables* should be numbered consecutively in Arabic numbers.
10. *References*. Cite only the most relevant references and list only those cited in the text. Indicate citations in the text by bracketed numbers, e.g., [4]. At the end of the paper list the references alphabetically by the surname of the first author and number them consecutively, according to the following formats: Articles: author(s), title, name and number of the journal (or book, but in this case include the editors names), year, pages. Books: Author(s), title, publisher, year.
11. *Accepted papers* are to be sent by the author to the editor, preferably in the form of a source file in LaTeX and EPS files for the figures together with a PDF or postscript file. Alternatively, source files in Word are also accepted. To ensure good publishing quality the figures should be in vector formats; raster formats like JPG, BMP, GIF, etc. should be avoided.
12. *Page proofs* will be e-mailed as a PDF file to the corresponding author. Correct proofs carefully, and restrict corrections to points at which the proof is at variance with the manuscript. Deviations from the version accepted by the editor are only possible with the prior and explicit approval of the editor. Thirty offprints of each paper are supplied free of charge to the corresponding author.

### **Publication information**

*Investigação Operacional* (ISSN 0874-5161) is registered in the *Secretaria de Estado da Comunicação Social* under number 108335. The volumes of the journal are published in hardcopies, which are distributed free of charge to all APDIO associates, and in electronic format in SciELO - Scientific Electronic Library Online. Subscription price is 25 euros. Issues are sent by standard mail. Additional subscription information is available upon request from APDIO Secretariat - CESUR, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 LISBOA, Portugal. Tel. +351 218 407 455 - [www.apdio.pt](http://www.apdio.pt) - [apdio@cesur.civil.ist.utl.pt](mailto:apdio@cesur.civil.ist.utl.pt)



## ÍNDICE

M. Matos, C. Lopes, S. Nunes, I. Venâncio <i>Reflexões sobre os rankings do Secundário</i> .....	1
J. Moreira, J. Sousa <i>Applying Data Warehousing Technology to Support Planning and Control on Mass Transit Companies</i> .....	23
P. Infante, J. Rodrigues Dias <i>Carta X-bar com Amostras de Tamanho Variável: Um Novo Procedimento Dinâmico</i>	45
E. Gomes, J.C. Soares de Mello, J.A. Mangabeira <i>Fronteira DEA Difusa na Avaliação de Eficiência em Agricultura</i> .....	65
R. Marques, D. Silva <i>Inferência Estatística dos Estimadores de Eficiência obtidos com a Técnica Fronteira Não Paramétrica de DEA. Uma Metodologia de Bootstrap</i> .....	89
A.M. Faustino, J.J. Júdice, I.M. Ribeiro, A. Neves <i>An integer programming model for truss topology optimization</i> .....	111