



IO 2009 – 14º Congresso da Associação Portuguesa de  
Investigação Operacional

# Livro de Actas

7 a 9 de Setembro  
Faculdade de Ciências e Tecnologia - Caparica



## **Comissão de Programa**

Presidente:

Ana Barbosa-Póvoa, Instituto Superior Técnico, Universidade Técnica de Lisboa

Agostinho Agra, Universidade de Aveiro

Ana Paula Teixeira, Universidade de Trás-os-Montes e Alto Douro

Carlos Henggeler Antunes, Faculdade de Ciências e Tecnologia, Universidade de Coimbra

Filipe Alvelos, Universidade do Minho

Isabel Gomes-Salema, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

João Paulo Costa, Faculdade de Economia, Universidade de Coimbra

João Soares, Faculdade de Ciências e Tecnologia, Universidade de Coimbra

João Telhada, Faculdade de Ciências, Universidade de Lisboa

José Fernando Oliveira, Faculdade de Engenharia, Universidade do Porto

Margarida Vaz Pato, Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa

Maria Antónia Caravilha, Faculdade de Engenharia, Universidade do Porto

Maria Eugénia Captivo, Faculdade de Ciências, Universidade de Lisboa

Maria Sameiro Carvalho, Universidade do Minho

Marília Pires, Universidade do Algarve

Mónica Oliveira, Instituto Superior Técnico, Universidade Técnica de Lisboa

Pedro Oliveira, Universidade do Minho

José Manuel Valério de Carvalho, Universidade do Minho



## **Comissão Organizadora**

Presidente:

Ruy Costa, Faculdade de Ciência e Tecnologia, Universidade Nova de Lisboa

Carmo Brás, Faculdade de Ciência e Tecnologia, Universidade Nova de Lisboa

Isabel Correia, Faculdade de Ciência e Tecnologia, Universidade Nova de Lisboa

Isabel Gomes-Salema, Faculdade de Ciência e Tecnologia, Universidade Nova de Lisboa

Lídia Lourenço, Faculdade de Ciência e Tecnologia, Universidade Nova de Lisboa

Manuel V.C. Vieira, Faculdade de Ciência e Tecnologia, Universidade Nova de Lisboa

Nelson Chibeles Martins, Faculdade de Ciência e Tecnologia, Universidade Nova de  
Lisboa

Paula Amaral, Faculdade de Ciência e Tecnologia, Universidade Nova de Lisboa



## **Índice**

Editorial .....	1
Boschetto, Suelen N; Relvas, Susana; Barbosa-Póvoa, Ana P; Neves Jr, Flávio <i>Heuristic batch sequencing for multiproduct pipeline scheduling</i> .....	3
Carmo, José L; Rodrigues, António JL <i>Modelos de Previsão para Procura Irregular</i> .....	11
Carvalho, Leandro L; Sousa, Jorge F <i>Converting the BSC in a methodology to support management change in a public transport company: the STCP experience</i> .....	19
Cavique, Luís; Mendes, Armando B; Santos, Jorge M <i>Um Algoritmo para Encontrar a Cobertura de k-Cliques em Redes Sociais</i> .....	27
Cerqueira, Paula; Dias, Luís; Oliveira, José A <i>Furnace scheduling in a foundry</i> .....	35
Cerveira, Adelaide; Bastos, Fernando <i>Relaxations for the truss topology design problem</i> .....	43
Correia, Gonçalo HAR <i>Testing the Mixed Integer Programming approach for the vehicle relocation problem in one-way carsharing</i> .....	51
Fernandes, Leão J; Barbosa-Póvoa, Ana P; Relvas, Susana <i>Risk Management in Petroleum Supply Chain</i> .....	59
Geraldes, Carla AS; Carvalho, Maria S; Pereira, Guilherme A <i>Gestão de operações logísticas num armazém usando modelos de apoio à decisão</i> .....	67
Gomes, Margarida M; Ribeiro, Rita A; Amaral, Paula <i>Reducing the Number of Membership Functions in Linguistic Variables</i> .....	75
Gomes, Tiago; Alvelos, Filipe; Pimentel, Carina; Valério-Carvalho, José M <i>Uma heurística para a integração dos problemas de corte a duas dimensões e sequenciamento</i> .....	83
João, Isabel M; Bana-Costa, Carlos A <i>A técnica de bootstrap na comparação de métodos de regressão ordinal usados na medição da satisfação de clientes</i> .....	91
Margalho, Pedro; Nunes, Eusébio <i>Contributos para a melhoria da disponibilidade e para redução de riscos em unidades de transformação de subprodutos de origem animal</i> .....	99

Martins, Paulo T; Lobo, Victor; Vairinhos, Valter <i>Container Stowage Problem Solution for Short Sea Shipping .....</i>	107
Mendes, Armando B; Noncheva, Veska; Silva, Emiliana <i>Canonical Correlation Analysis in variable aggregation in DEA.....</i>	115
Morais, Rui M; Pavan, Claunir; Pinto, Armando N; Requejo, Cristina <i>Genetic Algorithm for the Design of Survivable Optical Transport Networks: Operators Comparison.....</i>	123
Mota, Alzira MT; Vaz, A I <i>Função de penalidade para programação semi-infinita não linear.....</i>	131
Neves, Sara J; Galvão, Teresa <i>The heterogeneous fleet multi depot vehicle scheduling problem.....</i>	139
Oliveira, Lia C; Figueiredo, Manuel C; Lima, Rui M <i>Optimização de Fluxos Inbound num TIER2 da Indústria Automóvel .....</i>	147
Pais, Tiago C; Amaral, Paula <i>A Tabu Search Based Approach For Exams Timetabling Problem .....</i>	155
Pereira, Teresa; Carvalho, Sameiro <i>Análise multicritério para a análise e selecção de um sistema de gestão de frota.....</i>	163
Pinto, Sofia V; Alvelos, Filipe; Silva, Elsa; Valério-Carvalho, José M <i>Heurísticas para empacotamento em placas de vários tamanhos.....</i>	171
Pinto-Varela, Tânia; Barbosa-Póvoa, Ana P; Novais, Augusto Q <i>A Multi-objective Optimization for the Design and Periodic Scheduling of Multipurpose Facilities under uncertainty .....</i>	179
Portela, M CAS; Camanho, Ana S; Fernandes, M; Borges, D <i>BESP – Benchmarking de escolas secundárias Portuguesas .....</i>	187
Ramos, Tânia RP; Gomes-Salema, Maria I; Barbosa-Povoa, Ana P <i>A multi-product, multi-depot vehicle routing problem in a reverse logistics system: comparative study of an exact formulation and a heuristic algorithm .....</i>	195
Rietz, Jürgen; Alves, Cláudio; Valério-Carvalho, José M <i>Exploring a New Family of Maximal Staircase Dual Feasible Functions .....</i>	203
Teixeira, Ana P; Monteiro, Helena <i>Estudo experimental relativo à utilização de diferentes softwares de Programação Linear no Ensino Secundário .....</i>	211
Varela, Maria LR; Ribeiro, Rita A; Silva, Silvio C <i>A Manufacturing Scheduling Web-based Decision Support System.....</i>	219
Vaz, Clara B; Camanho, Ana S; Guimarães, Rui C <i>Avaliação de desempenho de lojas de retalho utilizando um modelo de Data Envelopment Analysis multi-nível.....</i>	227
Índice de Autores.....	235

## **Editorial**

Dentro da série de congressos da Associação Portuguesa de Investigação Operacional (APDIO) o seu 14º Congresso IO2009 edita pela primeira vez um livro de actas, onde estão publicados alguns dos trabalhos a apresentar no IO 2009. Os 29 artigos que compõe este volume foram seleccionados de entre 35 submetidos após terem sido sujeitos a um processo de revisão garantido pela Comissão de Programa do congresso.

Com a edição deste livro de actas a direcção da APDIO pretende introduzir uma nova forma de publicação que permitirá, mais uma vez aos investigadores na área, divulgarem o seu trabalho. Com este documento a adicionar à comunicação oral, a realizar pelos diferentes autores no IO 2009, os trabalhos aqui publicados poderão ser alvo de uma análise mais profunda criando-se assim as condições necessárias para fomentar uma maior interacção entre quem desenvolveu o trabalho e quem o lê e estuda.

Os 29 artigos publicados inserem-se dentro das diferentes áreas da Investigação Operacional, abordando alguns deles a temática do próprio congresso “Vencer Novos Desafios nos Transportes e Mobilidade”.

Sem a disponibilidade manifestada pelos membros da Comissão de Programa e o apoio da Comissão Organizadora do IO 2009 este livro não poderia ter sido tornado realidade. Desde já o nosso obrigado a todos os envolvidos.

Agradecemos ainda aos diferentes autores a submissão dos seus artigos bem como a sua colaboração no cumprimento dos prazos. Fazemos votos para que nos próximos congressos mais trabalhos apareçam submetidos e publicados.

Para terminar, desejamos que este livro seja um documento científico importante para a comunidade científica de Investigação Operacional em Portugal, fornecendo a todos nós uma noção mais concreta dos diferentes trabalhos que se vêm realizando em Portugal, na área da Investigação Operacional

Ana Paula Barbosa Póvoa  
Maria Isabel Gomes Salema



# Heuristic batch sequencing for multiproduct pipeline scheduling

Suelen N. Boschetto <sup>†</sup>

Ana P. F. D. Barbosa-Povoa <sup>‡</sup>

Susana Relvas <sup>‡</sup>

Flávio Neves-Jr <sup>†</sup>

<sup>†</sup> CPGEI - Federal University of Technology - Paraná  
 Av. Sete de Setembro, 3165, 80230-901 – Curitiba, PR, Brazil  
 suelen@cpgei.ct.utfpr.edu.br  
 neves@utfpr.edu.br

<sup>‡</sup> CEG-IST, UTL  
 susanaicr@ist.utl.pt  
 apovoa@ist.utl.pt

## Abstract

This paper addresses the problem of developing a heuristic procedure to aid the operational decision-making of scheduling activities in a multiproduct pipeline. The pipeline connects a refinery and a tank farm, transporting six different oil products. The pipeline scheduling is obtained through the solution of a continuous time Mixed-Integer Linear Programming (MILP) model presented by Relvas et al. (2006). The sequence of products considered by the scheduling model is built considering typical batch volumes, capacity limits, daily products' demand, product quality constraints, initial inventory and the settling period constraints. Due to the complexity of the pipeline scheduling model a high computational burden is observed in solving the MILP model. In order to diminishing this computational burden a heuristic procedure is developed in this paper, which defines the sequence and volumes of products. These results are used as inputs by the MILP model. Such heuristic explores the characteristics of the problem and takes into account the product delivery time windows so as to create the appropriate pumping sequence. Through the use of the heuristic the impact on the CPU time is strongly reduced and the MILP model can find optimal solutions for the given sequence in few seconds. The presented solution strategy is applied to a real world system provided by a Portuguese oil products distribution company – Companhia Logística de Combustíveis (CLC).

**Keywords:** scheduling, heuristic, multiproduct pipeline

## 1 Introduction

The success of optimization strategies in process systems engineering constantly motivates the desire to formulate and solve larger problems over a wider set of domains (Grossmann and Biegler, 2004). In particular, oil supply chains are complex networks due to several facts such as extension, either geographical and/or on number of entities involved, variety of operations and strong dynamics that are highly influenced by the on

world economics. The logistics associated with the supply of oil and oil derivatives comprises several operations, transportation types, and constraints that can be reflected in the final performance, efficiency and coordination of the entire supply chain. In this way, the development of successful strategies to obtain the scheduling of each element in the chain will enable to operate a balanced and optimized supply chain. One of the most used systems within such chain are the pipelines that guarantee the connection between the different entities of the chain. The pipeline usage is characterized by an effective and economic method for the supply of products along the oil chain. However, the detailed scheduling of the operational activities in pipelines is a difficult task and efficient methods to solve such complex problem are still required. In general, models presented in the literature are not able to find optimal, fast and representative solutions for real pipelines scheduling problems due to the elevated complexity. Atamturk and Savelsbergh (2005) state that in most practical situations, one is usually satisfied with good feasible solutions, preferably within a tight margin from optimality.

From the published literature involving pipeline scheduling, it is possible to find different formulations, all of them attempting to find good solutions in reduced computational time. The system reported in Rejowski and Pinto (2003) is composed by an oil refinery, one multiproduct pipeline connecting several depots that receive large amounts of oil products. The formulation problem involves the use of disjunctive programming, and two MILP models. The corresponding large-scale MILP model was solved for a relaxation gap of less than 6% for a time horizon of over 3 days. Magatão *et al.* (2004) developed an optimisation structure based on a MILP model with uniform time discretization which contemplates a pipeline connecting an inland refinery to a harbour, conveying different types of oil derivatives. The computational burden is avoided by a decomposition strategy that made possible to obtain optimal solutions for approximately 20 days. Cafaro and Cerdá (2004) dealt with the scheduling of a multiproduct pipeline system receiving a number of liquid products from a single refinery source to distribute them among several depots. A MILP continuous mathematical formulation was presented. The model was compared with the model of Rejowski and Pinto (2003) and optimal solutions are obtained for a time horizon of 75 hours. A MILP approach is used to model the problem of oil derivatives pipeline transportation scheduling and inventory management in the work proposed by Relvas *et al.* (2006). A continuous representation of both time and pipeline volume was used. The pumping sequence was fixed or partially fixed. A pumping schedule, including the product sequence, batch volume, and timing issues, is obtained for a relaxation gap of less than 5% for a time horizon of one month. In order to avoid high computational effort and provide good solutions to be used in a real world environment, Relvas *et al.* (2009) proposed a sequencing heuristic that builds possible pumping sequences given the scenario data. This heuristic reduces the model complexity, enabling to obtain scheduling solutions for the required monthly horizon.

The approaches presented return, at least, a feasible solution for the problems. In some cases, optimal MILP solutions are found, but using supporting decomposition strategies or even heuristic procedures. In the end, these solutions are suboptimal of the original problem, but being found in a limited computational time. In this work a heuristic procedure is proposed for the sequencing problem proposed by Relvas *et al.* (2006). The results are compared with the ones obtained by Relvas *et al.* (2009).

## 2 Problem Description

The pipeline to be scheduled connects a refinery to a tank farm, transporting six different oil products. Figure 1 represents the system under study.

Given (a) the number of products to be transported; (b) The possible/forbidden sequences between two products; (c) The minimum/maximum storage capacity per product; (d) The pipeline capacity (e) the initial pipeline conditions; (f) the pumping rate limits; (g) the minimum settling period per product; (h) the time horizon; (i) the initial inventory per product; (j) the daily products' demand; (k) the planned pipeline stoppages.

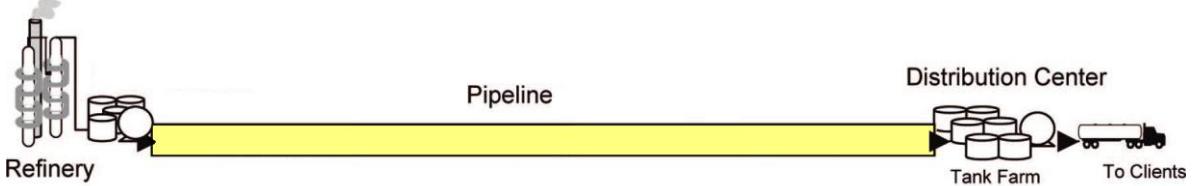


Figure 1: Single unidirectional multiproduct pipeline system

It is desired to obtain: (1) the pipeline schedule, with pumping sequence, batches' volumes, pumping rates and pumping and discharging timings, and (2) the inventory management at the distribution centre, including daily volume balances by product and monitoring of arrivals, settling and approving tasks as well as demands' fulfilment.

### 3 Heuristic Sequence Procedure

The proposed sequencing method was first proposed and used in Neves-Jr *et al.* (2007) and Boschetto *et al.* (2008) and is now adapted and enhanced for the presented problem that describes a real industrial situation. In both works a decomposition approach is proposed to address short-term scheduling in pipeline networks. One of the proposed blocks, the resource allocation, takes into account production and consumption functions and typical batch volumes (batch size) in order to determine a candidate sequence to pump. In addition, the resource allocation procedure indicates time-windows to the established sequences. Then, the pre-processed data are used by a continuous time MILP model, which determines the operational short-term scheduling for the entire pipeline network.

The proposed heuristic sequence is developed considering the following data: (a) The possible/forbidden sequences between two products; (b) The minimum/maximum storage limits per product; (c) the initial pipeline conditions; (d) the initial inventory per product; (e) the daily demand per product; (f) the settling period per new batch of a given product; and (g) the typical batch size, usually taken as a multiple of the tanks capacity for each product.

The time windows are obtained for each batch of each product to be received. The sequence can be created according to three priorities: minimum time window, maximum time window and product compatibility (stage 1 of the heuristic). As result, a fixed number of sequenced batches to be pumped are given, considering a fixed volume size for each product. After defining the number of sequence batches a second stage of the heuristics is performed where a batch correction is done in order to avoid negative inventory levels. The heuristic is detailed in the subsequent sections.

#### 3.1 Sets and Parameters

The indices/parameters used in this paper are the following:

Indexes/sets	
$p \in P$	Set of Products
$i \in I$	Set of Batches
$k \in K$	Set of days
Parameters	
$ID_{0,p}$	Initial inventory of product $p$ (without in transit batches) (vu)
$Dem_{p,k}$	Client's demands of each product $p$ at day $k$ (vu)
$Batch_p$	Typical batch volume of product $p$ (vu)
$Batch2_p$	Alternative batch volume of product $p$ (vu)
$ID_{min,p}$	Minimum storage for the product $p$
$ID_{max,p}$	Maximum storage capacity of the product $p$

$Stl_p$	Batch settling period of the product $p$
Parameters to be determined by sequencing procedure	
$Spc_{p,k}$	Available space to receive the product $p$ in the day $k$
$IDh_{p,k}$	Inventory of product $p$ at day $k$ (vu)
$BT_p$	Number of batches of product $p$
$Tmin_{j,p}$	Minimum time window of the batch $j$ of the product $p$ ( $j \in BT_p$ )
$Tmax_{j,p}$	Maximum time window of the batch $j$ of the product $p$ ( $j \in BT_p$ )

### 3.2 Time windows (Stage 1 – S1)

The time windows are determined for each batch to be sent of the product  $p$ . Considering this fact, it is necessary first to determine how many batches of each product ( $BT_p$ ) should be sent to supply the client's demand. The function  $round(x)$  returns the nearest integer to  $x$ .

$$BT_p = round\left(\frac{\sum_k Dem_{p,k}}{Batch_p}\right) \quad \forall p \in P \quad (1)$$

After determining the number of batches that should be sent for each product, it is possible to determine the time windows for each batch. This calculus is made through the following algorithms:

```

k=1
j=1
Spcp,0 = IDmaxp - ID0,p
while k ∈ K & j ≤ BTp
    Spcp,k = Spcp,k-1 + Demp,k
    while j ≤ BTp
        if Spcp,k ≥ j * Batchp
            Tminj,p = k
            j = j + 1
        else
            break
        k = k + 1
    
```

```

k=1
j=1
IDhp,0 = ID0,p - IDminp
while k ∈ K
    IDhp,k = IDhp,k-1 - Demp,k
    if IDhp,k ≤ 0
        Tmaxj,p = i - Stlp
        IDhp,k = IDhp,k + Batchp
        j = j + 1
    k = k + 1
for l = j.. BTp
    Tmaxl,p = h
    
```

(a) Minimum Time window ( $Tmin_{j,p}$ )(b) Maximum time window ( $Tmax_{j,p}$ )

Figure 2: Time windows algorithm

These algorithms are applied to each product  $p$  in order to determine the  $Tmin_{j,p}$  and  $Tmax_{j,p}$  for each batch  $j \in BT_p$ . A batch can be received after the time  $Tmin_{j,p}$  in the distribution center, if there is space available to allocate it ( $Spc_{p,k}$ ). Also, the same batch should be received before  $Tmax_{j,p}$ , when the inventory of product  $p$  ( $IDh_{p,k}$ ), after the lot settling period ( $Stl_p$ ), is near to minimum capacity.

The sequence can be established by considering the following criteria: (1) Maximum time window ( $Tmax_{j,p}$ ); (2) Minimum time window ( $Tmin_{j,p}$ ); (3) Lower compatibility between products. Firstly is compared the  $Tmax_{j,p}$  for every product. The product with smaller value is sequenced. If two or more products have the same value for  $Tmax_{j,p}$ , then the next step is to compare the value of the Minimum time window -  $Tmin_{j,p}$ . If the second item is the same, the product to be added in the sequence is chosen by the third item, the smaller products compatibility.

### 3.3 Batch Volume Correction (Stage 2 – S2)

Considering that the amount of product to be pumped, calculated in stage 1, in some cases is not exactly the amount requested by the product demand (see equation (1)), the difference was analysed in order to supply a possible lack of product. For some products, the batch size of the last sequenced product is increased. In other cases, an additional batch was considered with a volume size smaller than the typical value. The set of values to complement the demand are given by the parameter  $Batch2_p$ . The results of the both stages are presented in the next section.

## 4 Results

A case study taken from a Portuguese logistics company, CLC, is presented in this section. A medium term horizon of one month is considered for planning and six different oil products (P1–P6) are to be scheduled. Due to product quality, some sequences cannot be allowed inside the pipeline. Table 1 represents the allowed matches (X) or not allowed (–) to combine two consecutive products. Table 1 also contains values for typical batches' volumes, initial storage, and minimum and maximum capacities.

Table 1: Input data

$Batch2_p$ (vu)	$Batch_p$ (vu)	$ID_{0,p}$ (vu)	$ID_{min,p}$ (vu)	$ID_{max,p}$ (vu)	$Stl_p$ (days)	Possible sequences on two consecutive products						
						P1	P2	P3	P4	P5	P6	
P1	21800	18000	50677	0	81500	1	X	X	X	X	–	–
P2	8000	16000	16312	0	32000	2	X	X	–	–	–	–
P3	3800	8000	8069	0	24000	1	X	–	X	X	X	–
P4	3800	8000	17290	0	27800	1	X	–	X	X	X	–
P5	–	800	5895	0	10320	0,3	–	–	X	X	X	X
P6	4650	3100	5826	0	13120	0,3	–	–	–	–	X	X

For the Stage 1, the resulting time windows, determined by the algorithm presented in Figure 2, are presented in table 2.

Table 2: Time window for each batch (days)

$BT_p$	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>
<b>P1</b>	T <sub>min</sub>	1	2	4	6	10	12	16	18	20	24	26
	T <sub>max</sub>	8	10	12	16	18	22	24	26	30	31	31
<b>P2</b>	T <sub>min</sub>	2	9	16	23							
	T <sub>max</sub>	7	14	21	25							
<b>P3</b>	T <sub>min</sub>	1										
	T <sub>max</sub>	22										
<b>P4</b>	T <sub>min</sub>	1	3	6	11	13	18	20	25			
	T <sub>max</sub>	9	11	15	18	22	25	30	31			
<b>P5</b>	T <sub>min</sub>	1	1	1	1	1	2	4	6	10	12	16
	T <sub>max</sub>	19,7	23,7	24,7	26,7	30,7	31	31	31	31	31	31
<b>P6</b>	T <sub>min</sub>	1	1	4	10	17						
	T <sub>max</sub>	10,7	17,7	23,7	39,7	31						

The pumping sequences and the assigned batch volumes obtained with the proposed heuristic, from now on referred to as Heuristic 1 or H1, are presented in section 4.1. The results obtained in each stage will be reported as S1 for Stage 1 and S2 for Stage 2. In order to evaluate the quality of these results, they were fed to the MILP model proposed by Relvas *et al.* (2006). The operational results as well as the model statistics are compared with the results reported in Relvas *et al.* (2009), from now on referred to as Heuristic 2 or H2. Furthermore, both strategies are compared in terms of operational results with the real operation scheduled at CLC in the corresponding month. The comparison of results is presented in section 4.2.

#### 4.1 Heuristic 1: Proposed Sequence

In order to determine the sequence for the present case study, the criteria presented in subsection 3.2 were considered, that is: (1) Maximum time window ( $T_{max,j,p}$ ); (2) Minimum time window ( $T_{min,j,p}$ ); and (3) Lesser compatibility. Using the data in Table 1 and the results of table 2 the sequence of products shown in figure 3 was obtained for Stage 1.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
P1	P2	P1	P4	P5	P6	P5	P4	P1	P2	P1	P4	P5	P6	P5	P4	P1	P1	P2	P1	P3
22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
P4	P5	P6	P5	P4	P1	P2	P1	P4	P5	P6	P6	P5	P5	P5	P5	P4	P1	P1	P1	

Figure 3: Sequence obtained – H1-S1

Considering that the amount of product sent in some cases is less than the amount requested by the product demands (see equation (1)) the stage 2 of the proposed heuristics (H1) was applied). The resulting sequence is presented in figure 4.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
P1	P2	P1	P4	P5	P6	P5	P4	P1	P4	P1	P2	P1	P4	P5	P6	P5	P3	P1	P2	P1	P4	P5
24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
P6	P5	P4	P1	P1	P2	P1	P4	P5	P6	P5	P3	P5	P5	P5	P6	P5	P4	P1	P2	P1		

Figure 4: Sequence obtained – H1-S2

#### 4.2 Comparison of Results

Some notes should be made so as to establish a basis for comparison, regarding the results obtained and the number of batches. In Heuristic 2, the method proposes a set of different possible pumping sequences. After running the MILP model with each sequence, the model performance and operational results are used to select which schedule suits the requirements of the given scenario. In this way, only the selected schedule of H2 will be compared with H1 and it is represented in figure 5.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
P1	P2	P1	P4	P5	P6	P5	P4	P1	P2	P1	P3	P5	P6	P5	P4	P1	P2	P1	P4	P5
22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	
P6	P5	P4	P1	P2	P1	P4	P5	P6	P5	P4	P1	P2	P1	P4	P5	P6	P5	P4	P1	

Figure 5: Sequence determined through Heuristic 2

Regarding the number of batches, there is a slight difference between both methods. In H2 it is considered that two consequent batches do not contain the same product. This constraint is relaxed in H1. However, in practice, several subsequent batches of one product will correspond to one. Another difference is the lot size. While in H1 the lot size is determined in the heuristic procedure, in H2 the lot size of each batch is determined in the MILP model.

The MILP model statistics for the results obtained using the data provided by H1 and H2 are summarized in table 3. For H1 they are provided the results obtained for Stage 1 (H1-S1) and for Stage 2 (H1-S2). The stopping criteria are either the optimality or a maximum of computational effort corresponding to 2 hours. All model instances were run using GAMS 22.8/CPLEX 11.1 in an Intel Core 2 Duo P9400 2.40 GHz processor with 4

GB RAM. The objective function minimizes a dimensionless linear combination of operational objectives such as flowrate, pipeline usage or final inventory levels.

Table 3: MILP Model Statistics

	<b>H1-S1</b>	<b>H1-S2</b>	<b>H2</b>
Number of Batches	42	45	41
Number of continuous variables	21966	25089	21835
Number of binary variables	6058	6763	6485
Number of constraints	53157	60540	51089
Iterations	7181	8790	1760736
Nodes explored	221	273	7419
Solution Found	-2.171593	-2.120797	-2.231544
Computation Time (s)	5.064	6.789	561.7*
Relative Gap (%)	0.00	0.00	0.52

\*Within the maximum computational time span, optimality was not proven and/or found

Through the analysis of table 3, it is possible to conclude that, despite the differences on the number of batches (leading to different model sizes), the proposed heuristic (H1) pre-determines the batches and volumes to be pumped. The drawback is the reduction of the solution domain, which is proved by the objective function values. Nevertheless, obtaining the optimal (or near optimal) solution for the given scenario without the usage of the proposed heuristics leads to an intractable problem, as stated by Relvas *et al.* (2006). Additionally, the solutions found satisfy operational requirements and optimization goals when compared with trial-and-error and spreadsheet based solution approaches. Table 4 summarizes the operational results obtained either with H1 and H2 and also the real occurrences verified at CLC.

Table 4: Operational indicators

	<b>H1-S1</b>	<b>H1-S2</b>	<b>H2</b>	<b>CLC</b>
Number of Batches	42	45	41	38
1 Medium Flowrate (vu/h)	489.24	517.78	511.04	557.59
2 ΔInventory (vu)	-15320	+5630 <sup>1</sup>	+680 <sup>2</sup>	+19580
3 Pipeline Usage (%)	98.66	98.66	98.66	94.97
4 Final Inventory Level (%)	47.02	58.12	55.50	65.46
5 Minimum inventory (%, product)	2.95 (P3)	8.85 (P6)	23.58 (P3)	5.31 (P3)
6 Minimum final inventory (%, product)	20.69 (P3)	36.52 (P3)	52.53 (P1)	20.51 (P3)

<sup>1</sup>ΔInventory should be positive for each product; <sup>2</sup>ΔInventory should be positive for the sum of products

The number of batches pumped is within the regular amount for a medium term horizon of one month. Regarding the medium flowrate, it can be stated that with the results developed by CLC's schedulers the medium flowrate is considerably higher than the ones provided by the decomposition approach (heuristic combined with the MILP model). As a result, the amount of volume pumped through the pipeline is much higher than the requirements for the same period. On the one hand, the final inventory level is more comfortable, but it may originate scheduling problems for the following period. The pipeline usage is stretched to the maximum possible using the optimization approaches. The difference between H1-S1 and H1-S2 is emphasized by the difference on medium flowrate and ΔInventory. In the first situation, the results provided imply a major decrease in inventory level. A final note must be added regarding the information shared between heuristics and the MILP model. From the operational indicators presented in table 4, using the proposed approach, only indicators 1 and 3 are optimized at the MILP model, in contrast with the optimization of the complete set when using H2 with the MILP. The benefit from using the proposed approach is the reduction of the problem complexity and resource usage, which approximates the requirements of real world scenarios.

## 5 Conclusions

A heuristic procedure was proposed for the Relvas et al. (2006) sequencing problem in a multiproduct pipeline system. Through the proposed method, time windows were determined for each product batch, and pumping sequences were developed based on this information. The sequence was created according to three priorities: minimum time window, maximum time window and product compatibility. As result, a fixed number of sequenced batches to be pumped are obtained, considering a fixed batch size for each product.

The results show that the MILP model complexity is significantly reduced due to the elimination of a binary variable that assigns volumes to products, which are in turn assigned to a sequenced batch (three-index variable). In this way, the solution search is tighter and the computational effort is clearly reduced. This achievement allows schedulers to obtain fast solutions to evaluate/validate and apply for the current scenario.

In the H1 first stage (S1), only one volume size per product was only considered, implying that the amount of product supplied is not exactly as the amount demanded. This emphasized the inclusion of the second stage (S2). In future works, this procedure could be systematized for the complete sequence in order to pump the quantity of product that closely matches the demand. This is currently being explored by the authors.

## 6 Acknowledgements

The authors acknowledge financial support from CAPES/PDEE (grant 3262/08-1).

## 7 References

- Atamturk, A. and Savelsbergh, M. W. P. (2005) Integer-programming software systems, *Annals of Operations Research*, Vol 140, No 1, pp. 67-124.
- Boschetto, S. N. Felizari, L. C. Yamamoto, L. Magatao, L. Stebel, S. L. Neves-Jr, F. Arruda, L. V. R. Lüders, R. Ribas, P. C. and Bernardo, L. F. J. (2008) An Integrated Framework for Operational Scheduling of a Real-World Pipeline Network. In '18th European Symposium on Computer Aided Process Engineering – ESCAPE 18'. Lyon. (Eds Braunschweig, B. and Joulia, X.) pp. 259-264. (Elsevier Science)
- Cafaro, D. C. and Cerdà, J. (2004) Optimal scheduling of multiproduct pipeline systems using a non-discrete MILP formulation, *Computers & Chemical Engineering*, Vol 28, No 10, pp. 2053-2068.
- Grossmann, I. E. and Biegler, L. T. (2004) Part II. Future perspective on optimization, *Computers & Chemical Engineering*, Vol 28, No 8, pp. 1193-1218.
- Magatao, L. Arruda, L. V. R. and Neves, F. (2004) A mixed integer programming approach for scheduling commodities in a pipeline, *Computers & Chemical Engineering*, Vol 28, No 1-2, pp. 171-185.
- Neves-Jr, F. Magatao, L. Stebel, S. L. Boschetto, S. N. Felizari, L. C. Czaikowski, D. I. Rocha, R. and Ribas, P. C. (2007) An Efficient Approach to the Operational Scheduling of a Real-World Pipeline Network. In '17th European Symposium on Computer Aided Process Engineering – ESCAPE17'. Bucharest. (Eds V., P. and Agachi, P. S.) pp. 697-702. (Elsevier)
- Rejowski, R. and Pinto, J. M. (2003) Scheduling of a multiproduct pipeline system, *Computers & Chemical Engineering*, Vol 27, No 8-9, pp. 1229-1246.
- Relvas, S. Matos, H. A. Barbosa-Povoa, A. Fialho, J. and Pinheiro, A. S. (2006) Pipeline scheduling and inventory management of a multiproduct distribution oil system, *Industrial & Engineering Chemistry Research*, Vol 45, No 23, pp. 7841-7855.
- Relvas, S., Barbosa-Povoa, A., Matos, H. A. (2009) Heuristic batch sequencing on a multiproduct oil distribution system, *Computers & Chemical Engineering*, Vol 33, No 3, pp. 712-730.

# Modelos de Previsão para Procura Irregular

José Luís Carmo †

António J.L. Rodrigues ‡

† Faculdade de Ciências e Tecnologia, Universidade do Algarve  
and Centro de Investigação Operacional – FCUL  
jlcarmo@ualg.pt

‡ Faculdade de Ciências da Universidade de Lisboa  
and Centro de Investigação Operacional – FCUL  
ajrodrigues@fc.ul.pt

## Abstract

The problem of time series forecasting emerges in many different fields. Among them, we can find problems representing demand processes irregularly spaced in time, and with variable distribution. These processes are common in some areas of application of Operational Research methodologies — for instance, in inventory management or production planning problems.

The present work focuses on the study of problems related to forecasting irregularly demand processes. Several appropriate neural and exponential smoothing models are studied, and different approaches for the identification of those models are compared. In particular, adaptive methods for bivariate correlated processes (“when” and “how much”) are investigated and developed, through the application of gaussian basis function neural networks, among other models.

## Resumo

O problema da previsão de séries temporais surge nos mais diversos domínios. Entre eles, encontram-se os problemas representativos de processos de procura ocorrida ou registada em instantes de tempo não consecutivos, e com distribuição variável. Estes processos são comuns em diversos âmbitos de aplicação de metodologias de Investigação Operacional — por exemplo, na gestão e controlo de inventários, ou no planeamento de produção.

No presente trabalho, centra-se a atenção no estudo de problemas de previsão de processos de procura irregular. São estudados vários modelos neuronais e de alisamento exponencial apropriados, e são comparadas diversas abordagens para a identificação desses modelos. Em particular, são investigados e desenvolvidos métodos adaptativos para processos bivariados correlacionados (“quando” e “quanto”), através da utilização de, entre outros modelos, redes neuronais de funções de base gaussianas.

**Keywords:** Forecasting, neural networks

**Title:** Models for irregular demand forecasting

## 1 Introdução

Em problemas de gestão de inventários ou de planeamento de produção é relativamente comum observar processos de procura irregularmente espaçados no tempo. A Figura 1 mostra um fenómeno desse tipo: para um determinado artigo encontra-se uma colecção de pedidos de encomenda realizados de forma irregular no tempo. Pretende-se então modelar o conjunto dos dados de forma a responder a questões como as seguintes: qual é o valor esperado do próximo pedido e quando é que se espera que ocorra?; qual é o valor esperado da procura total num dado horizonte temporal?; etc.

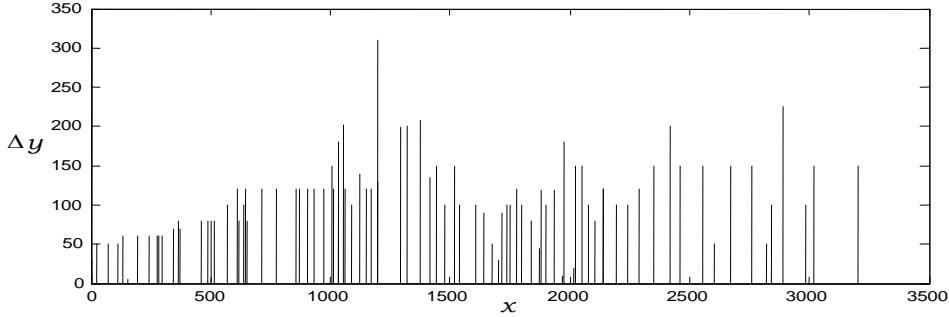


Figura 1: Ordens recebidas por uma fábrica de papel durante 9 anos (toneladas *versus* dias)

A grande maioria dos estudos conhecidos refere a aplicação de métodos de alisamento exponencial — Croston (1972), Wright (1985), Willemain *et al.* (1994), Cipra *et al.* (1995), Syntetos e Boylan (2001), Boylan e Syntetos (2007), Altay *et al.* (2008). Só mais recentemente têm sido considerados modelos neurais — Carmo e Rodrigues (2004), Gutierrez *et al.* (2008), Pour *et al.* (2008).

A partir de uma série irregularmente espaçada  $\{(x_i, \Delta y_i)\}_{i=1, \dots, N}$ , defina-se:

- $\Delta x_i = x_i - x_{i-1}$ : intervalo de tempo entre duas ocorrências consecutivas da procura;
- $y_k = \sum_{1, \dots, k} \Delta y_i$ : procura total até ao instante  $k$ ;
- $d_i = \Delta y_i / \Delta x_i$ : estimativa instantânea da taxa de procura.

Tendo em conta o facto de ser vantajoso modelar explicitamente o conjunto dos dados na forma de uma série bivariada, e considerando a dualidade dos papéis entre as variáveis “tempo” e “quantidade”, propomos considerar as duas alternativas seguintes:

- modelar a série bivariada dos dados diferençados,  $\{(\Delta x_i, \Delta y_i)\}$ ;
- modelar a série bivariada dos dados acumulados,  $\{(x_i, y_i)\}$ .

No primeiro caso, pode-se definir um esquema autoregressivo (linear ou não linear) em que se tente prever uma das variáveis supondo a outra conhecida, ou, mais geralmente, pode-se considerar a estimativa simultânea das duas variáveis:

- $\{(\Delta x_{k-p}, \Delta y_{k-p}), \dots, (\Delta x_{k-1}, \Delta y_{k-1}), \Delta x_k\} \longrightarrow \Delta \hat{y}_k$
- $\{(\Delta x_{k-p}, \Delta y_{k-p}), \dots, (\Delta x_{k-1}, \Delta y_{k-1})\} \longrightarrow (\Delta \hat{x}_k, \Delta \hat{y}_k)$

Como se irá mostrar, as redes neurais de funções de base gaussianas podem ser aplicadas à modelação de uma série bivariada na forma diferençada, utilizando qualquer dos esquemas anteriores. Também o método de alisamento exponencial de Croston utiliza as duas variáveis na forma diferençada; contudo fá-lo de forma independente, não tirando partido da possível correlação entre as variáveis. Para séries tratadas de forma acumulada, é aplicável o método de alisamento exponencial de Wright.

## 2 Previsão Através de Alisamento Exponencial

### 2.1 Método de Croston

Os métodos de alisamento exponencial são frequentemente utilizados para a previsão de séries temporais. São definidos através de fórmulas de actualização recursiva, onde são empregues pesos de forma a distinguir a importância entre as várias observações. Alguns desses métodos são adaptados para séries irregularmente espaçadas. O mais conhecido e usado para esse fim foi proposto por Croston (1972), estando inclusivamente disponível em algum software comercial de previsão. Consiste meramente na aplicação do método de alisamento exponencial simples, quer à série das encomendas, quer à série dos intervalos de tempo, mas com coeficiente de alisamento de valor comum às duas séries. Estas estimativas são então usadas para estimar a taxa de procura futura. O método de Croston pode apresentar estimativas enviesadas, pelo que Syntetos e Boylan (2001) propuseram uma fórmula correctiva, adoptada nas experiências realizadas no presente trabalho.

### 2.2 Método de Wright

Wright (1985) propôs uma extensão do método de alisamento exponencial de Holt para séries irregularmente espaçadas. A série da procura acumulada pode, então, ser objecto de aplicação deste método, por apresentar, em geral, um crescimento (tendência) localmente linear. Contudo, embora a série acumulada seja estritamente crescente, o método pode excepcionalmente produzir previsões negativas para as quantidades procuradas:

$$\hat{y}_{k+1|k} - y_k < 0$$

Para evitar essa possibilidade, consideramos correcções ao método original, descritas em Carmo e Rodrigues (2004). Para séries irregularmente espaçadas que apresentem efeitos sazonais, Cipra *et. al* (1995) propuseram a extensão do método de Wright, combinando na formulação o conhecido método de Holt-Winters.

## 3 Previsão Através de Redes Neuronais

### 3.1 Redes de Funções de Base Gaussianas

As redes neurais artificiais mais usadas em aplicações de previsão são modelos paramétricos de regressão não linear de mínimos quadrados. Em particular, as *redes de funções de base* apresentam algumas vantagens práticas relativamente a outros modelos mais complexos, como o *perceptrão multicamada* (PMC), nomeadamente, o facto de poderem ser heuristicamente interpretadas como modelos lineares relativamente aos parâmetros, permitindo a utilização de métodos eficientes de estimação.

A resposta de uma rede com  $n$  valores de entrada,  $m$  funções de base,  $\{\phi_i\}$ , e um único nó de saída é definida por uma função contínua parametrizada,  $f_{\mathbf{w}}: \mathbb{R}^n \rightarrow \mathbb{R}$ , não linear em relação às variáveis de entrada mas linear nos parâmetros  $\mathbf{w}$ :

$$f_{\mathbf{w}}(\mathbf{x}) = w_0 + \sum_{i=1}^m w_i \phi_i(\mathbf{x}); \text{ com } \phi_i(\mathbf{x}) = \phi(\|\mathbf{x} - \mathbf{c}_i\|)$$

onde  $\mathbf{x}$  é um padrão de entrada genérico e  $\mathbf{c}_i$  o *centro* para a função  $\phi_i$ . A optimização não linear das coordenadas dos centros pode ser computacionalmente proibitiva e, em geral, são consideradas estimativas obtidas heuristicamente por métodos de *clustering*.

Deste modo, dado um conjunto de padrões ou observações (*conjunto de treino*),  $\{(\mathbf{x}_k, y_k)\}_{k=1,\dots,N}$ , pretende-se estimar os parâmetros  $\{w_i\}$  segundo o modelo  $y_k = f_{\mathbf{w}}(\mathbf{x}_k) + \varepsilon_k$ , onde  $\{\varepsilon_k\}$  é geralmente assumido como sendo um processo de ruído branco. Razões de eficiência computacional e de eficácia do modelo recomendam que se considere  $m \ll N$ . Logo, como se está perante um modelo que é linear em relação aos parâmetros, e por se considerar como critério de optimalidade a minimização da soma dos quadrados dos erros de aproximação, pode-se obter eficientemente uma solução para o problema, seja por estimação directa, seja por estimação recursiva.

As redes de funções de base são habitualmente compostas por *funções de base radiais* (FBR), mas podem igualmente ser consideradas *funções de base elípticas* (FBE), ou ainda, *funções de base normalizadas* (FBN):

- FBR:  $\phi_i(\mathbf{x}) = \phi(\|\mathbf{x} - \mathbf{c}_i\|) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_i\|^2}{2\sigma_i^2}\right)$ ;  $\sigma_i$ : raio da f.b.  $\phi_i$
- FBE:  $\phi_i(\mathbf{x}) = \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{c}_i)^T \Sigma_i^{-1} (\mathbf{x} - \mathbf{c}_i)\right\}$ ;  $\Sigma_i = \text{diag}(\sigma_{1i}^2, \sigma_{2i}^2, \dots, \sigma_{ni}^2)$
- FBN:  $\tilde{\phi}_i(\mathbf{x}) = \frac{\phi_i(\mathbf{x})}{\sum_{j=0}^m \phi_j(\mathbf{x})}$

Habitualmente, são consideradas funções gaussianas para as FBR. Deste modo, o modelo é uma mistura linear de gaussianas multi-dimensionais, dependentes dos parâmetros  $\{w_i\}$  e dos hiperparâmetros de localização e dispersão  $\{(\mathbf{c}_i, \sigma_i)\}$ . O modelo FBE, com diferentes raios para as diferentes dimensões, embora mais versátil, é também mais exigente do ponto de vista computacional. Uma opção mais simples é a utilização de funções de base normalizadas (FBN): em cada ponto do espaço de entrada, a soma das saídas de todas as funções de base é igual à unidade, logo o espaço de entrada é coberto ao mesmo grau em todos os pontos.

### 3.2 Redes de Funções Normalizadas com Termo Independente

Não existem muitos estudos sobre as FBN, apesar de no trabalho seminal de Moody e Darken (1989) já se considerar este procedimento. Sarle (2002) apresentou um estudo comparativo entre vários modelos “clássicos” e normalizados, para o problema de aproximação de uma função. Para o exemplo aí exposto, mostra-se que a normalização origina uma redução acentuada do número de funções gaussianas — e, logo, de parâmetros —, para níveis de desempenho superiores.

Uma questão que não se encontra esclarecida na literatura é a que diz respeito à possível utilização de um termo independente,  $w_0$ , no modelo das FBN. Frequentemente, este termo ou não é considerado no modelo, ou, quando o é, não é incluído no processo de normalização. Essas opções não são justificadas, o que suscitou a realização de um estudo experimental (Carmo, 2007). Concluiu-se ser preferível manter esse termo independente no modelo, e a sua inclusão no processo de normalização:

$$f_{\mathbf{w}}(\mathbf{x}) = w_0 + \sum_{i=1}^m w_i \phi_i(\mathbf{x}) = \sum_{i=0}^m w_i \phi_i(\mathbf{x}) \xrightarrow{\text{normalização}} f_{\mathbf{w}}(\mathbf{x}) = \frac{\sum_{i=0}^m w_i \phi_i(\mathbf{x})}{\sum_{j=0}^m \phi_j(\mathbf{x})}$$

com  $\phi_0 = 1$ .

### 3.3 Identificação e Estimação

Havendo um número suficiente de observações, o conjunto de dados disponíveis deverá ser dividido em dois, formando o *conjunto de treino* e o *conjunto de teste*. É através do conjunto de teste que é aferida a capacidade de generalização da rede. Quando o conjunto de teste é igualmente utilizado para optimizar os hiperparâmetros do modelo, será antes designado por *conjunto de validação*. Neste caso, o conjunto de dados disponíveis poderá ser dividido não em dois, mas em três conjuntos, para que seja possível formar também o conjunto de teste.

Para a identificação e estimação do modelo, propomos os procedimentos sumariamente referidos na Tabela 1, e descritos em detalhe em (Carmo, 2007). Na opção de estimação “em bloco”, o mais recomendável é a optimização da ordem de autoregressão ( $p$ ) e do número de funções gaussianas ( $m$ ), através da utilização de um conjunto de validação. Os pesos são estimados em bloco empregando o método dos mínimos quadrados (OLS). Para a estimação recursiva, os centros, os raios e os pesos são continuamente adaptados através de, respectivamente, *k-means clustering* sequencial, *k*-vizinhos mais próximos, e a versão recursiva do método dos mínimos quadrados (RLS). Os valores para  $p$  e  $m$  podem ser obtidos, por exemplo, através de valores pré-definidos, ou, mais aconselhável, através de uma primeira fase de identificação do modelo com a aplicação da estimação em bloco.

Tabela 1: Esquema proposto para a estimação de redes de funções de base gaussianas

		Hiperparâmetros	Pesos
Em Bloco	$(p, m)$	min(REQM) no conj. de val.	OLS
	$\{\mathbf{c}_i\}$	<i>k-means clustering</i>	
	$\{\sigma_i\}$	<i>k</i> -vizinhos mais próximos	
	$\{\Sigma_i\}$	Covariância empírica	
Recursivamente	$\{\mathbf{c}_i\}$	<i>k-means</i> recursivo	RLS

## 4 Estudo Experimental

Pretende-se comparar o desempenho preditivo dos modelos antes referidos a um problema real de previsão da procura, englobado no problema mais geral de planeamento de produção e corte de papel numa empresa do sector. Para isso, foram consideradas 10 séries de procura irregularmente espaçadas com características e dimensões variadas. Cada uma foi pré-processada, por aplicação da transformação logarítmica, seguida de normalização estatística, excepto na aplicação do método de Wright. O procedimento inverso é aplicado às estimativas obtidas, de forma a permitir a apresentação dos resultados nas unidades originais.

Como medida de validação e avaliação dos modelos foi considerado o REQM (raiz do erro quadrático médio), baseado nos erros de previsão um passo à frente da série de valores da taxa de procura. Para os métodos recursivos, os 20 primeiros erros foram ignorados no cálculo da medida, para reduzir o risco de utilização de erros enviesados, devido às decisões de inicialização desses métodos. Os hiperparâmetros dos métodos de alisamento exponencial considerados serão optimizados por pesquisa em grelha.

#### 4.1 Estimação em Bloco — Modelos Neuronais

Pretende-se testar a aplicação das funções radiais (FBR), elípticas (FBE) e normalizadas (FBN) ao problema de previsão dos volumes de encomenda. Por outro lado, também se pretende comparar diferentes tipologias de redes neuronais, considerando, além das redes de funções de base gaussianas, o perceptrão multicamada (PMC).

Para a estimação das redes de funções de base considerou-se, para cada série, o processo de estimação em bloco referido na Tabela 1, com o conjunto de validação formado pelas últimas 33% observações. Para as redes PMC adoptou-se o mesmo esquema autoregressivo utilizado para as redes de funções de base, e o treino foi baseado na aplicação do método de optimização não linear de Levenberg-Marquardt.

Como mostra a Tabela 2, a rede FBN foi a que maioritariamente apresentou melhor desempenho no conjunto de validação. Para maior facilidade de interpretação dos resultados, apresenta-se para cada modelo um Índice de Desempenho Global (*IDG*), média de 10 quocientes entre o REQM obtido e o menor valor de REQM para a respectiva série.

Tabela 2: Previsão de encomendas: avaliação comparativa para diferentes modelos neuronais

Série ( <i>N</i> )	PMC	FBR	FBE	FBN
A (89)	3.35	3.05	4.38	<b>2.07</b>
B (212)	2.74	2.37	2.92	<b>2.19</b>
C (92)	22.26	17.49	22.59	<b>14.64</b>
D (167)	30.43	28.81	<b>28.18</b>	29.08
E (119)	5.86	5.64	6.50	<b>4.88</b>
F (210)	4.84	4.52	5.45	<b>4.46</b>
G (163)	0.88	0.78	0.94	<b>0.76</b>
H (210)	1.57	1.43	2.34	<b>1.23</b>
I (118)	0.86	<b>0.79</b>	1.37	0.95
J (263)	<b>19.25</b>	22.50	25.31	23.26
<i>IDG</i>	<b>1.23</b>	<b>1.13</b>	<b>1.47</b>	<b>1.04</b>

#### 4.2 Estimação Recursiva — Diferentes Abordagens

Pretende-se agora considerar a estimação recursiva, comparando o modelo FBN com métodos de alisamento exponencial. Dois diferentes casos foram analisados: a previsão das quantidades encomendadas, e a previsão simultânea das encomendas e dos respectivos instantes de ocorrência.

No primeiro caso, além da rede e do método de Wright, dois outros métodos foram igualmente aplicados, muito mais simples, cujos resultados servirão de referência para as abordagens em comparação. O mais simples, o método *naive*, determina cada previsão como o valor registado na ocorrência anterior. O outro método, frequentemente utilizado, consiste na aplicação de alisamento exponencial simples (AES) à sequência de valores da taxa de procura. Para a rede, como foi sugerido, optou-se por utilizar para *p* e *m* valores pré-definidos e comuns a todas as séries, respectivamente 3 e 15.

Mesmo não optimizando o modelo, a rede FBN mostrou uma clara supremacia relativamente aos outros métodos (Tabela 3). É de destacar o fraco desempenho do método de Wright, já que não conseguiu suplantar claramente os métodos mais simples. Relembre-se, no entanto, que as séries verdadeiramente estimadas por este método foram as séries de procura acumulada. Pode especular-se que seja um método mais indicado para previsões de longo prazo.

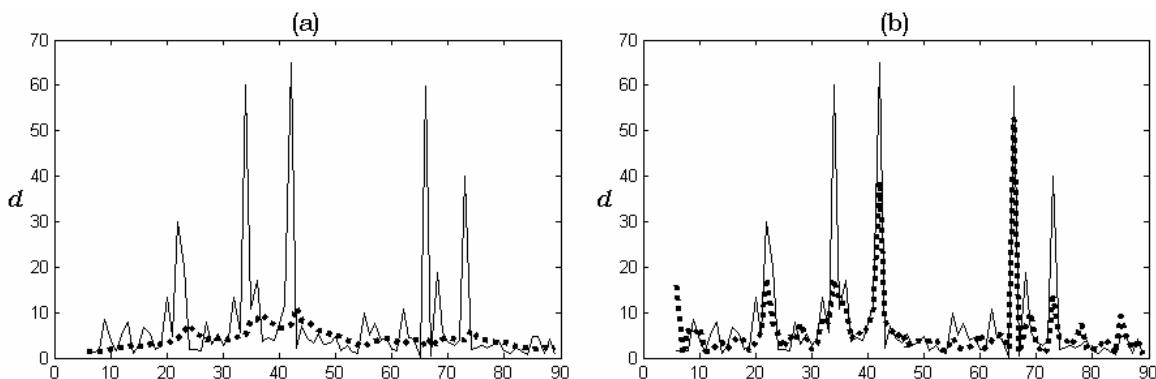
Tabela 3: Previsão de encomendas: avaliação comparativa de diferentes abordagens

Série	Naive	AES	Wright	FBN(3,15)
A	13.62	13.51	13.36	<b>3.73</b>
B	8.23	6.89	6.91	<b>5.03</b>
C	36.37	35.66	35.31	<b>29.41</b>
D	59.71	47.89	46.92	<b>39.38</b>
E	13.48	16.04	15.94	<b>10.22</b>
F	10.14	8.52	8.59	<b>6.23</b>
G	5.28	6.03	6.03	<b>4.24</b>
H	6.92	6.23	6.24	<b>3.87</b>
I	3.81	4.08	4.04	<b>2.17</b>
J	32.20	30.82	30.83	<b>25.90</b>
<i>IDG</i>	<b>1.70</b>	<b>1.65</b>	<b>1.64</b>	<b>1</b>

No segundo caso, considerou-se o método *naive*, o método de Croston e a sua versão revista (Croston-SB), e a rede FBN com duas saídas. Como se pode observar na Tabela 4, o modelo FBN(3,15) já não mostrou o predomínio revelado nos testes anteriores, apesar de globalmente continuar a apresentar os melhores resultados. Uma vantagem considerável só é atingida quando se procede à optimização dos valores de  $p$  e  $m$ . Por outro lado, a versão revista do método de Croston não produziu benefícios relativamente à versão original do método. Para ilustração gráfica dos resultados, apresentam-se na Figura 2 as estimativas obtidas para a série A para as duas principais abordagens em análise.

Tabela 4: Previsão bivariada: avaliação comparativa de diferentes abordagens

Série	Naive	Croston	Croston-SB	FBN(3,15)	FBN( $p^*, m^*$ )
A	18.86	13.60	13.69	11.68	<b>7.72</b>
B	9.72	7.28	7.30	7.21	<b>5.69</b>
C	53.25	37.48	37.52	37.50	<b>30.59</b>
D	65.61	47.77	47.78	42.44	<b>39.49</b>
E	22.64	16.23	16.30	15.20	<b>12.98</b>
F	12.14	8.93	8.93	9.05	<b>8.68</b>
G	8.60	6.13	6.14	4.59	<b>4.23</b>
H	8.98	6.51	6.53	6.47	<b>5.30</b>
I	5.79	4.04	4.05	4.18	<b>3.58</b>
J	45.56	32.18	32.24	32.26	<b>30.42</b>
<i>IDG</i>	<b>1.75</b>	<b>1.26</b>	<b>1.27</b>	<b>1.18</b>	<b>1</b>

Figura 2: Estimativas (a tracejado) para a série A: (a) Croston; (b) FBN( $p^*, m^*$ )

## 5 Conclusões

As redes de funções de base gaussianas, ao terem em conta as possíveis associações não lineares dos dados, demonstraram ser adequadas para a previsão de séries temporais irregularmente espaçadas, produzindo, em geral, melhores previsões que os modelos alternativos considerados, para os vários tipos de problemas estudados. Entre os vários tipos de funções de base, destacaram-se principalmente as redes FBN (normalizadas).

Para o problema da previsão bivariada, a opção pelo método de Croston poderá ser suficientemente adequada em muitos casos, tendo em consideração a maior complexidade da identificação e estimação exigida pelos modelos neuronais.

O presente trabalho centrou-se no estudo da *previsão pontual*, mas o valor obtido como previsão da próxima ocorrência, só por si, não contém informação relativa ao seu grau de incerteza. Em Carmo (2007) propõe-se a aplicação de modelos baseados em funções “kernel” gaussianas para a *previsão de densidades* que quantifiquem esse risco. Depois, numa segunda fase, face a um determinado critério de optimalidade, poder-se-á determinar a melhor decisão. O trabalho futuro será primordialmente dirigido para o aprofundamento desta nova abordagem.

## 6 Referências

- Altay, N., Rudisill, F. and Litteral, L.A. (2008) Adapting Wright's Modification of Holt's Method to Forecasting Intermittent Demand, *Int. J. Production Economics*, Vol. 111, pp. 389-408.
- Boylan, J.E. and Syntetos, A.A. (2007) The Accuracy of a Modified Croston Procedure, *Int. J. Production Economics*, Vol. 107, pp. 511-517.
- Carmo, J.L. (2007) *Previsão da Procura e Decisão Optimal: Modelos e Métodos Avançados*, Tese de Doutoramento, Universidade de Lisboa.
- Carmo, J.L. and Rodrigues, A.J. (2004) Adaptive Forecasting of Irregular Demand Processes, *Engineering Applications of Artificial Intelligence*, Vol. 17, pp. 137-143.
- Cipra, T., Trujillo, J. and Rubio, A. (1995) Holt-Winters Method with Missing Observations, *Management Science*, Vol. 41, pp. 174-178.
- Croston, J.D. (1972) Forecasting and Stock Control for Intermittent Demands, *Operational Research Quarterly*, Vol. 23, pp. 289-303.
- Gutierrez, R.S., Solis, A.O. and Mukhopadhyay, S. (2008) Lumpy Demand Forecasting Using Neural Networks, *Int. J. Production Economics*, Vol. 111, pp. 409-420.
- Moody, J. and Darken, C.J. (1989) Fast Learning in Networks of Locally-Tuned Processing Units, *Neural Computation*, Vol. 1, pp. 281-294.
- Pour, A.N., Tabar, B.R. and Rahimzadeh, A. (2008) A Hybrid Neural Network and Traditional Approach for Forecasting Lumpy Demand, *Proceedings of World Academy of Science, Engineering and Technology*, Vol. 30.
- Sarle, W.S. (ed.) (2002) Neural Network FAQ, URL: [ftp://ftp.sas.com/pub/neural/FAQ.html](http://ftp.sas.com/pub/neural/FAQ.html) (consultado a 15-Abril-2009).
- Syntetos, A.A. and Boylan, J.E. (2001) On the Bias of Intermittent Demand Estimates, *Int. J. Production Economics*, Vol. 71, pp. 457-466.
- Willemain, T.R., Smart, C.N., Shockor, J.H. and DeSautels, P.A. (1994) Forecasting Intermittent Demand in Manufacturing: A Comparative Evaluation of Croston's Method, *International Journal of Forecasting*, Vol. 10, pp. 529-538.
- Wright, D.J. (1985) Extensions of Forecasting Methods for Irregularly Spaced Data, in *Time Series Analysis: Theory and Practice*, Vol. 7, O.D. Anderson (ed.), Elsevier Science, pp. 169-181.

# Converting the BSC in a methodology to support management change in a public transport company: the STCP experience

Leandro Lopes de Carvalho †, Jorge Freire de Sousa †

† Faculdade de Engenharia da Universidade do Porto

leandromec@yahoo.com.br

jorgefreire@stcp.pt

---

## Abstract

In the last decades the public transport sector has been reevaluating its role in the society. Demands for quality of service improvement, cooperation for the sustainable development of community and environment, responsibility for the traffic and pollution reduction in the big cities and increased scrutiny in its financial and budgetary results are the main reasons. Therefore, to efficiently and effectively respond to the increasing necessities of the society, such companies have been modernizing their managerial and operational processes through the implementation of new techniques and methodologies.

The authors propose a new structure for the management control system of such companies based on the implementation of tools, like the Balanced ScoreCard (BSC), traditionally used by the private sector, and duly modified for this purpose.

This proposal is being applied in STCP, a public transport company in the city of Oporto, in order to implement a new strategic approach, more focused in the value creation for the stakeholder.

Finally, this paper will discuss the resulting benefits originated by the introduction of the methodology in the company.

**Keywords:** Public Transport, Strategic Alignment, Balanced Scorecard, Travel Behaviour.

---

## 1 Introduction

The companies of the public sector have the mission to, direct or indirectly, assure services which effectively contribute for the quality of life of the community and to make these services available to all citizens, as an obligatory complement for similar kind of services offered by the private sector. Therefore, people expect that they offer services with quality, but using the available resources with parsimony, efficiency and effectiveness. To correspond to these expectations from the society, it is necessary that such companies incorporate new concepts of management, able to drive their strategy to the value creation for the community.

In this context, this paper describes the adaptation of existing management control tools to the public transport sector and their application to STCP, the main public transport operator in the city of Oporto, Portugal. The Balanced Scorecard (BSC) and strategic maps

were used to support the introduction of a new management orientation, focused in the value creation for the stakeholders<sup>1</sup> in general and for the clients in particular.

The value creation for the clients is mainly based on the reinforcement of customer oriented principles in the culture of the company. With this purpose, this study integrates methods of increasing knowledge on customer expectations, using analysis of their behaviour and their attitudes towards transportation.

Finally, this paper will discuss the benefits obtained from the introduction of these tools in the strategic management control system of STCP, and the existent barriers to its implementation.

## **2 The STCP context**

Public transport companies must have equity and efficiency as one of their main principles as they are using resources from the society in order to satisfy the mobility needs of the community.

Nowadays, more than the mere accomplishment of these social responsibilities it is expected from these companies to contribute for improvements in other aspects of its activity which affect the society.

One of these aspects is the capacity to stimulate the travel mode switch from car to public transport, in order to contribute for the reduction of emissions, congestion, accidents and other issues related to the excess of traffic in the big cities. As quoted by Vuchic (2005), experience has shown that public transport has a great significance for reducing transit congestion, offering alternative means of travel, and contributing greatly for the quality of urban life.

In Portugal, these expectations from society are reflected in new demands from the government through the formalization of a new model of contract between the board of directors and the government, complemented by an annual sustainability report, beyond the traditional model of accountability report. These documents extend the evaluation of the company to non-financial performance measures such as efficiency in the operations, quality level of the service, environmental sustainability and social impact, complementing the budgetary and financial targets.

In the later years, STCP has been adapting its strategy and structure to the achievement of these challenges. This modernization focused on the customer orientation as a crucial point in the activity of the company. Moreover, the company aimed to adapt its network to integrate a new mode of transport in the city of Oporto, the light rail. STCP assumed itself as one of the components of a wider transport system, instead of keeping its traditional positioning as the dominant operator in the city.

This proposal seeks to use the BSC and strategic maps to translate this new strategic approach in an easier language, facilitating the discussion inside the company. Furthermore, the BSC will support the effective introduction of new acts in the company, such as new quality policy and security and environmental certification.

---

<sup>1</sup> In this study stakeholder means all the actors affected by the activities of the company, including the community, users and non-users of the service, the legislative bodies, the environment, the company's employees and the government.

### 3 Analysis of the management control system of STCP

Rather than acting as just an additional component of the performance management process, the BSC can act as a unifying force, tightening the links between the various dimensions of the system and making the whole significantly stronger than the sum of the parts. To accomplish this role, the BSC must find a place in the overall system allowing to maximize the considerable value it has proven to deliver (Niven, 2003).

Therefore, prior to the implementation of the BSC in STCP, an analysis of the management control system of the company was conducted based on the collection of internal written information. The objective is to determine how the company currently disseminates its strategy, organizes the initiatives and controls the performance indicators.

The central document in the strategic management system of the company is the management contract formalized between the board of directors and the government. It describes the 10 principles which drive its strategy, its vision and mission, its strategic objectives and the performance indicators which control them. Internally, this contract is complemented by departmental management contracts between the board of directors and the heads of the departments.

These contracts reflect the different departmental activities plans and are used as a basis to the boards of drivers. Each activities plan encloses the initiatives defined for a given department, also defining temporal targets for its implementation. The board of drivers entails the performance measures of the departments and the respective quantitative targets. The indicators included in these tools are controlled by the board of directors at each three months.

Complementing the board of drivers, for the day-to-day control management, there are indicators for the processes. The company has 18 documented processes, with their respective procedures and internal procedures, derived from a quality certification recently obtained by STCP. Similarly, there are additional indicators derived from the environmental and security certification. These indicators, however, are not yet entirely associated to the performance measures of the management contracts, activities plan and board of drivers.

In this study the adaptation of the BSC to the management control system of the company, in order to centralize all the management control information around a set of strategic objectives and measures, is proposed. It summarizes the system to three main documents, described in Figure 1. In this structure, the management contract formalizes the strategy of the company, which will be implemented and controlled by the BSC. Complementing these documents, the processes indicators will control the day-to-day activities of the company.

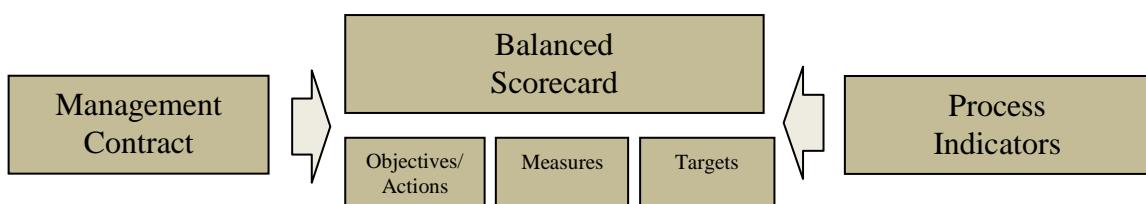


Figure 1: New proposal for STCP management control system

## **4 Adapting the BSC**

The BSC was originally developed by the researchers Robert Kaplan and David Norton as a performance measure tool aiming to be a set of measures capable to give top managers a fast but comprehensive view of the business. It includes financial measures that tell the results of actions already taken. And it complements the financial measures on customer satisfaction, internal processes, and learning and growth, operational measures that are the drivers of future financial performance (Kaplan & Norton, 1992). In this study the BSC will be used in a broader perspective, as a change management tool capable of clarifying the strategy and communicating it through the company, engaging people and creating synergies for the prosecution of a common goal.

The evolution of the BSC can be characterized in four steps – a system for performance management, a strategic management system, a tool for the communication of the strategy and finally a methodology to support the management change (Pinto, 2007). This evolution was possible due to the capacity of the BSC to link the companies' long-term strategies to short term actions, to align its management processes and to communicate the strategy, objectives and initiatives through the company allowing the alignment of people, processes and infrastructures to changes in the strategic direction. As quoted by Beckles (2006) this development can be chronicled through subsequent Kaplan and Norton articles and books (Kaplan & Norton, 1992; 1996; 2000; 2006).

A decisive point for this evolution was the creation of the strategic maps. The maps provide a visual representation of the company's critical objectives and the crucial relationships among them that drive organizational performance. Moreover, they show the cause-and-effect links by which specific improvements create desired outcomes (Kaplan & Norton, 2000). According to Pinto (2007) this tool had impact in all the BSC implementation process, but mainly in two aspects, the communication of the strategy and the translation of the strategy in operational terms.

The BSC and its first set of metrics were originally created to be used in the private sector. But, as Kaplan & Norton (2001) quote, in 2000 the BSC was already completely accepted and adopted by non-profit and public organizations all over the world. As a result several researchers described the introduction of the BSC in different types of public companies, either using the four original metrics or adapting them to the specificities of the service under analysis (Beckles, 2006; Greatbanks & Tapp, 2007; Lang, 2004; Niven, 2003; Radnor & Lovell, 2003).

According to Phillips (2004), the original metrics must be changed for the public organizations because they generally only pursue mission effectiveness and efficiency. That is, they want to accomplish their mission as efficiently as possible. Moreover, private organizations report mainly to their private shareholders, while public organizations respond to a broader range of stakeholder groups including citizens, public interest groups, and legislative bodies (Beckles, 2006).

Niven (2003) proposed a general structure of the BSC for public companies by changing its original metrics. He argues that if for profit-seeking enterprises to improve shareholder value is the endgame, the public and non-profit companies are completely oriented by its mission. In this sense he defined the mission as a new perspective, and all the other perspectives must converge in this main purpose.

#### **4.1 BSC for public transport companies**

In this section a new proposal of the BSC, adapted to the characteristics of the public transport companies, will be presented.

As a public transport company, the mission of STCP is the value creation for the stakeholder, the centre of its strategy. So, it is natural that this study proposes this perspective as central in its BSC structure. Other proposed change is to join the budgetary control to the financial perspective, in order to control the financial enablers and restrictors of the public sector. The other perspectives are similar to the original metrics: client, internal processes and infrastructure.

In order to use this proposal as a tool to management support, it is important to define the cause-effect relationship between these perspectives. In the proposed structure the client perspective is placed at the same level of the financial/budgetary one. This assembly assures that they will support and constraint each other for the achievement of the central target of the company, the value creation for the stakeholder. In the end, the challenge in a cause-effect relation is to assure synchronism between all the perspectives.

This proposal of BSC adapted for public transport companies was presented to the board of directors of STCP and approved for implementation. The resulting structure was the basis of the STCP strategic map presented in Figure 2.

#### **4.2 Implementation at STCP**

For people to act on the words given by the vision and strategy statements, those statements must be expressed as an integrated set of objectives and measures, agreed upon by all senior executives, that describe the long-term drivers of success (Kaplan & Norton, 1996). So, the first step for the BSC introduction was a discussion between the board of directors and the heads of departments in order to build a consensus about the organization's strategy. At this point a BSC workgroup responsible for further implementation was created.

The first author of this paper was integrated in this workgroup, for a participant observation data collection. In the case study strategy research, the participant observation is a special mode of observation in which the researcher is not merely a passive observer. Instead, he may assume a variety of roles within a case study situation and may actually participate in the events being studied (Yin, 1994).

The BSC and strategic map concepts were presented all over the company, through discussion meetings at each department. The objective, besides the clarification of the BSC language, was to discuss the STCP's strategy, to define the strategic objectives which better translate it and to describe how people interpret the role of their activities in its achievement. Finally, the workgroup met to consolidate the strategic map and BSC of STCP, based on the inputs of each department.

### **5 Results**

In the BSC all the measures should lead to improved bottom line performance (Niven, 2003). So, in the strategic map of the STCP, the three perspectives in the top (stakeholder, financial/budgetary and clients) entail the company's lag indicators, which show the result of the drivers of the strategy, and the bottom perspectives (internal processes and infrastructure) are controlled by lead indicators, which show the progress of the strategy

implementation in key areas. In this structure the strategy should be read by analyzing transversally the cause and effect relations between the objectives (see Figure 2).

The economic rationality, in the stakeholder perspective, encloses the company's main compromise with the government, its only shareholder. It seeks to reduce the impact of the activities of the STCP in the public accounts, by the cost structure optimization, a more efficient use of the resources and an increase of demand and income. The customer relationship management and the quality of service both aim to increase the demand and improve the profitability of the company. At an operational level many objectives will influence these top objectives, namely development/optimization of the service, maximization of business partnerships, internal efficiency and gain of new customers.

The other top objective is related to the other main stakeholders of the company, the society and the environment. Both aspects will be positively influenced by the increase of demand, based on the shift of car users to public transport users. The reduction of cars flow decrease emissions, congestion, accidents and other problems caused by the excess of traffic in big cities. Additionally, the social and environmental objectives are represented at the operational level.

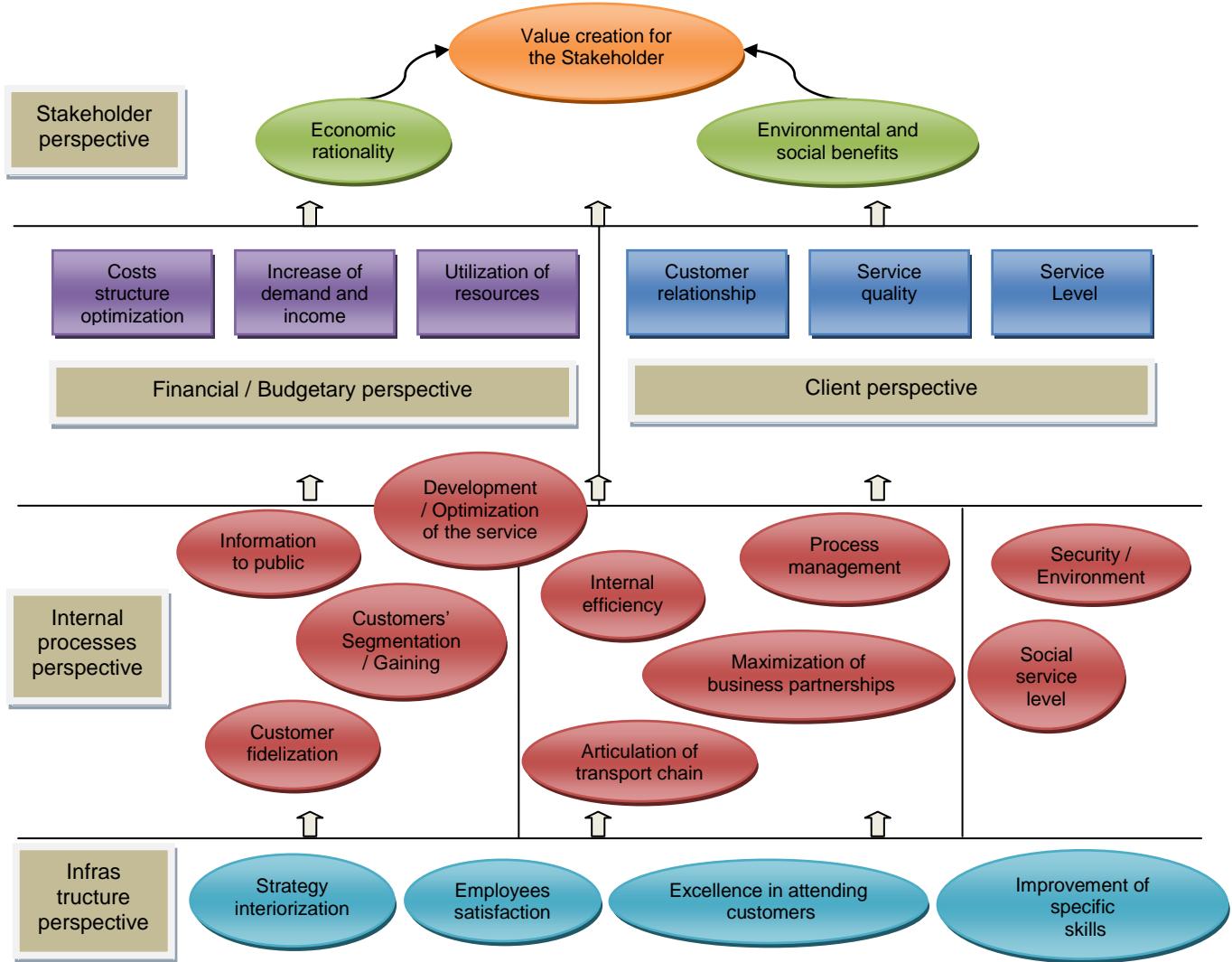


Figure 2: STCP Strategic Map

The customer orientation turned into one of the main objectives in the company's activities. However, despite its existence in the strategy in the later years, there is a general perception in the company that its principles are neither sufficiently translated in some processes nor interiorized by its employees, mainly in the lower levels. This occurs because STCP, as any other public company, historically considered the government as its prime client. This mentality should be definitively changed, once that the real STCP's client is the stakeholder, including the public transport users and potential users.

In this BSC proposal the customer orientation is reflected in the financial perspective through the increase of the demand. It reflects the company focus in increasing its share market both by augmenting the current clients use of the service and by gaining new customers. To achieve this objective is necessary to improve the quality of service and level of service offered in the customer perspective. Moreover, the company intends to continuously improve the relationship with the clients, namely through the image of being a reference public company in the city of Oporto, and a partner of the community.

However, the most important contribute of this proposal to the improvement of the customer orientation is in the operational level. It is reflected in a new approach for the market segmentation and target market definition, based on the behavioural segmentation.

New studies in transportation defend the use of behaviour segmentation in the basis of the transportation sector marketing strategy, complementing the traditional methods based on socio-demographic and geographical segmentation. This new approach seeks to use psychological factors including perceptions, identity, social norms and habit to understand travel mode choice (Anable, 2005).

In his study, Jensen (1999) developed six mobility types of passengers based on their attitudes towards transport - with the objective of understanding how transport became part of the everyday life -, on transport behaviour, on environmental consciousness and on transport as integrated in the modern culture.

Beirão and Cabral (2007) studied these mobility types in the city of Oporto, and suggest the need for segmentation taking into account travel attitudes and behaviour. They conclude that policies which aim to influence car usage should be targeted at the market segments that are most motivated to change and willing to reduce frequency of car use. In the internal processes perspective of the BSC, this proposal is reflected through the objective of gaining new customers in the targeted behavioural segments.

Nonetheless, other aspects considered strategical to the company are reflected in the BSC, such as the employees' satisfaction, the improvement of its process management systems and its role as main articulator of the transport chain of Oporto's metropolitan area.

## 6 Conclusions

This study has proposed a methodology to implement the BSC and strategic maps as management change tools in public transport companies, based on the case study of STCP. In order to introduce it, there was the need to propose a new model of management control system for the company. Additionally, a new structure for the BSC adapted to the public transport sector was proposed.

The development of the BSC in STCP allowed the development of a common language to describe its strategy. Moreover, it clarified the cause-and-effect relationships between the strategic objectives and actions, which was most of the time unclear for the employees. As a result, it eased the understanding of the strategy by the company.

Furthermore, the BSC simplified the management control system, clarified the indication of the responsible for the achievement of the performance goals and actions and highlighted the key performance indicators.

The main difficulty for a more efficient implementation of the BSC in STCP is to introduce it in the employees' evaluation system and beneficiation. The current evaluation system of the company is very rigid and it has been used since many years ago.

## **7 References**

- Anable, J. (2005). 'Complacent Car Addicts' or 'Aspiring Environmentalists'? Identifying travel behaviour segments using attitude theory. *Transport Policy*, Vol. 12, pp. 65-78.
- Beckles, G. (2006). Preferences for performance measures: A study of a federal agency. Unpublished PhD. Thesis, University of Central Florida, Florida.
- Beirão, Gabriela & Cabral, J.S. (2007). Understanding attitudes towards public transport and private car: A qualitative study. *Transport Policy*, Vol. 14(6), pp. 478-489.
- Greatbanks, R., & Tapp, D. (2007). The impact of balanced scorecards in a public sector environment - Empirical evidence from Dunedin City Council, New Zealand. *International Journal of Operations & Production Management*, Vol. 27(8), pp. 846-873.
- Jensen, M. (1999). Passion and heart in transport - a sociological analysis on transport behaviour. *Transport Policy*, Vol. 6(1), pp. 19-33.
- Kaplan, R. S., & Norton, D. P. (1992). The Balanced Scorecard: Measures That Drive Performance. *Harvard Business Review*, (Jan-Feb), pp. 172-180.
- Kaplan, R. S., & Norton, D. P. (1996). Using the Balanced Scorecard as Strategic Management System. *Harvard Business Review*, (Jan-Feb), pp. 150-161.
- Kaplan, R. S., & Norton, D. P. (2000). Having Trouble with Your Strategy? Then Map it. *Harvard Business Review*, (Sep-Oct), pp. 167-176.
- Kaplan, R. S., & Norton, D. P. (2001). *The Strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment*. Harvard Business School Press, Boston.
- Kaplan, R. S., & Norton, D. P. (2006). *Alinhamento: usando o Balanced Scorecard para criar sinergias corporativas* (1 ed.). Campus, São Paulo.
- Lang, S. (2004). Balanced Scorecard and Government Entities. *The CPA Journal*, (June), pp. 48-52.
- Niven, P. (2003). *Balanced Scorecard step-by-step for Government and Nonprofit agencies*, John Wiley & Sons, New Jersey.
- Philips, J. K. (2004). An application of the Balanced Scorecard to Public Transit System Performance Assessment. *Transportation Journal*, (Winter), pp. 26-55.
- Pinto, F. (2007). *Balanced Scorecard: Alinhar Mudança, Estratégia e Performance nos Serviços Públicos* (1<sup>a</sup> ed.), Edições Silabo, Lisboa.
- Radnor, Z., & Lovell, B. (2003). Defining, justifying and implementing the Balanced Scorecard in the National Health Service. *International Journal of Medical Marketing*, 3(3), pp. 174-188.
- Vuhic, V. R. (2005). *Urban transit: Operations, planning and economics* (1st edition). John Wiley & Sons, Inc, New Jersey.
- Yin, R. K. (1994). *Case Study Research - Design and Methods*, Vol 5, Sage Publications, California.

# Um Algoritmo para Encontrar a Cobertura por k-cliques de Redes Sociais

Luís Cavique †

Armando B. Mendes ‡

Jorge M.A. Santos §

† Universidade Aberta  
lcavique@univ-ab.pt

‡ Universidade dos Açores  
amendes@uac.pt

§ Universidade de Évora  
jmas@uevora.pt

---

## Resumo

Na análise de redes sociais, uma k-clique é a relaxação de uma clique, i.e., uma k-clique é um subgrafo quase completo. Um k-clique num grafo é um subgrafo onde a distância entre qualquer par de vértices não é maior do que k arestas. A visualização de um pequeno número de vértices é fácil de obter. Contudo, quando o número de vértices aumenta a visualização torna-se incompreensível. Nesta comunicação, propomos uma nova abordagem na extracção de conhecimento em grafos, utilizando k-cliques. O conceito de clique-relaxado é estendido a todo o grafo, quando este é coberto por k-cliques, obtendo assim uma visão geral das sub-componentes do grafo. Neste trabalho são apresentadas sequências de coberturas de k-cliques que combinam subgrupos coesos com o conceito de “pequenos mundos”. Resultados computacionais e exemplos são apresentados.

**Palavras-chave:** extracção de conhecimento em dados, extracção de conhecimento em grafos, redes sociais

**Title:** An Algorithm to Find the k-cliques Cover in Social Networks

---

## 1 Introdução

Depois da comunicação de Tim Berners-Lee [2] na conferência internacional de World Wide Web WWW2006 sobre as três idades da Web, as redes sociais associados à Web 2.0, suscitararam uma explosão de interesse, na tentativa de melhorar a socialização e na criação de novos modelos de gestão dos conhecimentos.

O termo "redes sociais" foi cunhado por Barnes em 1954, no entanto, a visualização através de grafos, chamados sociogramas, foi apresentado por Moreno [18]. Esta área científica da sociologia tenta explicar como emerge a liderança, como se criam alianças e

conflitos, como se difunde a inovação/epidemia e como é que a estrutura de um grupo afecta a sua eficácia.

Um importante desenvolvimento sobre a estrutura das redes sociais teve origem numa experiência realizada pelo psicólogo americano Stanley Milgram [16]. A experiência de Milgram consistiu em enviar cartas de pessoas no Nebraska, no Centro-Oeste, para serem recebidas em Boston, na costa leste dos EUA, com a seguinte particularidade - as pessoas eram instruídas a passar as cartas de mão em mão, a alguém que conheciam, até chegarem à origem. As cartas que chegaram ao destino foram passadas por cerca de seis pessoas. Milgram concluiu que, os americanos não estão a mais de seis passos entre si. Esta experiência deu origem ao conceito de “seis graus de separação” e ao conceito de “pequeno mundo”.

Um exemplo interessante dos “pequenos mundos” é o “Número de Erdos” [12]. Erdos é um dos mais prolíficos matemáticos de todos os tempos, sendo autor de mais de 1500 artigos publicados com mais de 500 co-autores. Erdos é o número zero e os investigadores que trabalharam com ele são chamados de Erdos número 1. Os co-autores de Erdos número 1 são conhecidos como Erdos número 2, e assim por diante, construindo um dos mais antigos “pequenos mundos”. O trabalho de Erdos e Renyi [8] apresenta interessantes propriedades dos grafos aleatórios. O interesse foi reavivado recentemente com o modelo de Watts e Strogatz, publicado na revista Nature [22], que apresenta novas propriedades dos pequenos mundos.

Os analistas de redes sociais precisam de realizar inquéritos de cada pessoa sobre os seus amigos, pedir a sua aprovação para publicar os dados e acompanhar a população durante anos. Por outro lado, as redes sociais como o LinkedIn, Facebook, Hi5 e o MySpace podem fornecer os dados necessários sem este tipo de esforço.

A extracção de conhecimento de grafos pode ser definida como a arte e a ciência de encontrar informação útil de grafos, como padrões e “outliers”, fornecidos respectivamente, por dados repetidos ou dados esporádicos existentes em grafos de grandes dimensões ou de redes complexas [9], [6].

Neste trabalho, propomos criar novas medidas com base na cobertura do grafo com  $k$ -cliques, com o objectivo de obter uma visão geral do grafo. Na secção 2, revemos os conceitos de redes sociais, detalhando as estruturas de subgrupos coesos. Na secção 3 apresentamos o algoritmo de duas fases que analisa em primeiro lugar para subgrupos coesos e, em segundo lugar descobre o conjunto mínimo de subgrupos coesos que cobrem todos os vértices. Na secção 4 são apresentados os resultados computacionais e exemplos. Finalmente, na secção 5, são apresentadas as conclusões.

## **2 Conceitos de Redes**

A análise de redes Sociais é uma abordagem muito que surgiu na moderna sociologia, e estuda a interacção entre indivíduos, organizações e outros tipos de entidades. Em [19] e [21] podemos encontrar as bases teóricas e as principais técnicas de redes sociais. A representação de uma rede social foi bastante influenciada pela teoria dos grafos. Nas redes sociais os vértices (ou nós) correspondem aos “actores” (i.e., pessoas, empresas, agentes sociais) e as arestas correspondem às “ligações” (isto é, relacionamentos, associações, links).

Uma das aplicações mais comuns da sociologia tem sido o tema da coesão social. Os subgrupos coesos podem incluir grupos como, grupos de trabalho, equipas desportivas, partidos políticos, cultos religiosos, ou estruturas secretas, como as organizações criminosas ou células terroristas. Nesta secção são apresentadas algumas técnicas de

subgrupos coesos como as cliques e cliques relaxadas, como o k-clique, k-club/k-clā e k-plex.

## 2.1 Clique

Dado um grafo não orientado  $G = (V, E)$  onde  $V$  denota o conjunto de vértices e  $E$  o conjunto de arestas, o grafo  $G_1 = (V_1, E_1)$  é chamado um subgrafo de  $G$  se  $V_1 \subseteq V$ ,  $E_1 \subseteq E$  e para cada aresta  $(v_i, v_j) \in E_1$  os vértices  $v_i, v_j \in V_1$ . Um subgrafo  $G_1$  é completo se existe uma aresta para cada par de vértices. Um subgrafo completo é também chamado de clique, ver figura 1. Uma clique é máxima, se não está contida numa outra clique. O número de clique de um grafo é igual à cardinalidade da maior clique de  $G$  e é obtido pela resolução do problema NP-difícil da máxima clique.



Figura 1: Cliques com 1, 2, 3, 4, 5 e 6 vértices

A estrutura de clique, onde deve existir uma aresta para cada par de vértices, apresenta muitas restrições na modelação da vida real. Deste modo, abordagens alternativas têm sido sugeridas, com o fim de relaxar o conceito de clique, como o k-clique, k-clā/k-club e k-plex.

## 2.2 k-clique

Dado um grafo não orientado  $G = (V, E)$ , um caminho em  $G$  é uma sequência de arestas de  $G$ . Luce [15] introduziu o modelo baseado na distância chamado k-clique, onde  $k$  é o comprimento máximo do caminho entre cada par de vértices. Uma k-clique é o subconjunto de  $C$  tal que para cada  $i, j \in C$ , a distância  $d(i, j) \leq k$ . Uma 1-clique é idêntica a uma clique, pois a distância entre os vértices é igual a um. Uma 2-clique é o subgrafo completo maximal com um caminho de uma ou duas arestas. O caminho de comprimentos dois pode ser exemplificado pelo "amigo do amigo" das relações sociais. Em sítios sociais como o LinkedIn, cada membro pode visualizar as suas ligações e também as ligações a dois e três graus de distância. O aumento do valor  $k$  corresponde a um relaxamento progressivo do conceito de clique. Um exemplo pode ser observado na figura 2.

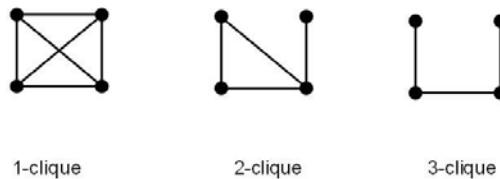


Figura 2: Exemplos de k-cliques

## 2.3 k-clā e k-club

Uma limitação do modelo k-clique é que alguns vértices podem ficar demasiado afastados do grupo. Para superar este inconveniente Alba [1] e Mokken [17] introduziram os modelos baseados no diâmetro, com o nome de k-club e k-clā, respectivamente. O comprimento do caminho mais curto entre os vértices  $u$  e  $v$  em  $G$  é denotado por  $d_G(u, v)$ . O diâmetro de  $G$  é dado por  $\text{diam}(G) = \max d_G(u, v)$  para todos  $u, v \in V$ . Para encontrar todas as k-clás, em primeiro lugar todas as k-cliques são encontradas e, em seguida, são removidas todas as k-cliques com diâmetro inferior a  $k$ . Na figura 3, à esquerda, a 2-

clique  $(1,2,3,4,5)$  foi removido porque a  $dG(4,5)=3$ . Outra abordagem possível, dos modelos baseados no diâmetro, é o  $k$ -clube que é definido como um subconjunto de vértices  $S$  tal que  $\text{diam}(G[S]) \leq k$ . Na figura seguinte, duas 2-cliques:  $(1,2,3,4,5)$  e  $(2,3,4,5,6)$ , um 2-clá:  $(2,3,4,5,6)$  e três 2-clubes:  $(1,2,3,4)$ ,  $(1,2,3,5)$  e  $(2,3,4,5,6)$  podem ser encontrados.

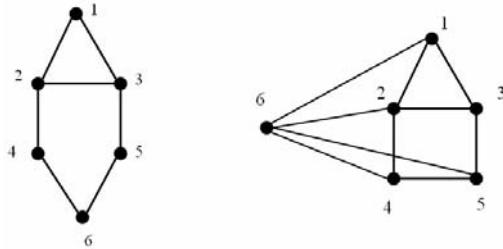


Figura 3: 2-clás, 2-clubes (esquerda) e 3-plex (direita)

## 2.4 k-plex

Uma outra forma de relaxar uma clique é através do conceito de  $k$ -plex, que tem em consideração o grau vértices. O grau de um vértice de um grafo é o número de arestas incidentes no vértice, e é denotado por  $\deg(v)$ . O grau máximo de um grafo  $G$  é o grau máximo dos seus vértices e é denotado por  $\Delta(G)$ . Por outro lado, o grau mínimo é o grau mínimo dos seus vértices e é denotado por  $\delta(G)$ . Um subconjunto de vértices  $S$  é o chamado de  $k$ -plex se o grau mínimo no subgrafo induzido for  $\delta(G[S]) \geq |S| - k$ . Na figura 3, à direita, o grafo com 6 vértices,  $|S| = 6$ , e do grau de vértices 1, 3, 4 e 5 não excede o valor 3. Assim, o grau mínimo, no subgrafo induzido  $\delta(G[S]) = 3$ .

## 3 O Algoritmo de Duas-Fases

As medidas utilizadas na análise de Redes Complexas e “Graph Mining” são baseadas em procedimentos de baixa complexidade computacional, como o diâmetro do grafo, o grau de distribuição dos nós e a verificação da conectividade, subestimando o conhecimento da estrutura das componentes do grafo. Neste trabalho, propomos uma medida baseada na cobertura do grafo com  $k$ -cliques, com vista a compreender melhor a estrutura do grafo, combinando a análise de subgrupos coesos com o conceito de “pequenos mundos”.

Com o fim de encontrar a sequência de coberturas mínimas por  $k$ -cliques é proposto um algoritmo de duas fases. Na primeira fase, introduzimos o conceito de grafo  $k$ - $G(V, E)$ . Dado um grafo  $G(V, E)$ , um grafo  $k$ - $G(V, E)$  é a transformação do grafo  $G(V, E)$ , tal que para cada  $i, j \in V$ , a distância  $d(i, j) \leq k$ . Para encontrar todas as  $k$ -cliques maximais do grafo, usamos uma simples transformação do grafo e, em seguida, utilizamos um algoritmo multi-partida da clique máxima. Esta fase tem como objectivo gerar todas as  $k$ -cliques maximais do grafo. Na segunda fase, cada subconjunto mínimo de  $k$ -cliques é escolhido para cobrir todos os vértices do grafo.

### Procedimento 1: Algoritmo de duas-fases para encontrar a cobertura por $k$ -cliques

Entrada: distância  $k$  e grafo  $G(V, E)$

Saída: cobertura por  $k$ -cliques

1. Encontrar todas as  $k$ -clique maximais do grafo  $G$ :

1.1. Transformar o grafo num  $k$ - $G(V, E)$  (Proc. 2)

1.2. Aplicar o algoritmo multi-partida da clique máxima (repetindo Proc. 3)

2. Aplicar o algoritmo da cobertura de conjuntos com  $k$ -cliques

### 3.1 Encontrar as k-cliques maximais no grafo k-G(V,E)

Esta secção detalha o processo de transformação do grafo e a geração das cliques maximais.

#### 3.1.1 Grafo Transformação

Para criar um grafo k-G(V,E), iremos utilizar o Algoritmo Floyd [10], que considera todos os caminhos mais curtos num grafo, e, em seguida, para cada par de vértices tal que a distância mínima seja inferior ou igual a k, é introduzida uma nova aresta caso não exista ainda.

#### Procedimento 2: Transformação do Grafo

Entrada: k, M[n,n] matriz de adjacência do grafo G

Saída: D[n,n] a matriz com k-distâncias do grafo G (ou k-Grafo)

1. D=M
2. Para cada  $(h,i,j)$   $D[i,j] = \min(D[i,j], D[i,h]+D[h,j])$
3. Para cada  $(i,j)$  se  $(D[i,j] \leq k)$   $D[i,j]=1$  senão  $D[i,j]=0$
4. Retornar D;

#### 3.1.2 Algoritmo da Clique Máxima

O problema da Clique Máxima é um problema NP-difícil que visa encontrar o maior subgrafo completo em um determinado grafo. Nesta abordagem, pretende-se encontrar um limite inferior para o problema de maximização, com base na heurística proposta por Johnson [13] e na meta-heurística, que utiliza a Procura Tabu, desenvolvida por Soriano e Gendreau [20]. Parte do trabalho descrito nesta secção também pode ser encontrado em [4] e [3]. Vamos definir  $A(S)$  como o conjunto de vértices que estão adjacentes aos vértices da solução corrente  $S$ . Seja  $n=|S|$  a cardinalidade da clique  $S$  e  $A^m(S)$  o subconjunto de vértices com  $m$  arestas incidentes em  $S$ .  $A(S)$  pode ser dividido em subconjuntos  $A(S)=\cup A^m(S)$ ,  $m=1,..,n$ . A cardinalidade do conjunto de vértices  $|V|$  é igual à soma dos vértices adjacentes  $A(S)$  mais os vértices não adjacentes  $A^0(S)$ , resultando em  $|V|=\sum |A^m(S)|+n$ ,  $m=0,..,n$ . Para uma determinada solução  $S$ , definimos uma estrutura de vizinhança  $N(S)$ , se esta gera uma solução admissível  $S'$ . Neste trabalho vamos utilizar três estruturas de vizinhança.

Vamos considerar a seguinte notação:

$$N^+(S) = \{S' : S' = S \cup \{v^i\}, v^i \in A^n(S)\}$$

$$N^-(S) = \{S' : S' = S \setminus \{v^i\}, v^i \in S\}$$

$$N^0(S) = \{S' : S' = S \cup \{v^i\} \setminus \{v^i\}, v^i \in A^{n-1}(S), v^i \in S\}$$

$S$  - a solução corrente

$S^*$  - a clique maximal corrente de maior cardinalidade

$T$  - a lista tabu

$N(S)$  - as estruturas de vizinhança

#### Procedimento 3: Heurística Tabu para o Problema da Clique Máxima

Entrada: k-Grafo,  $S$

Saída: clique  $S^*$

1. iniciar  $T$ ;  $S^*=S$ ;
2. enquanto não for condição de fim
  - 2.1. se  $(N^+(S) \setminus T \neq \emptyset)$  escolha a melhor  $S'$
  - 2.2. senão se  $(N^0(S) \setminus T \neq \emptyset)$  escolha a melhor  $S'$ ; atualizar  $T$ 
    - 2.2.1. senão escolha a melhor  $S'$  em  $N^-(S)$ ; atualizar  $T$
  - 2.3. atualizar  $S=S'$
  - 2.4. se  $(|S| > |S^*|)$   $S^*=S$ ;
3. fim ciclo enquanto;
4. retornar  $S^*$ ;

Encontrar uma clique maximal num k-grafo é o mesmo que encontrar uma k-clique maximal num grafo. Para gerar um grande conjunto de k-cliques maximais, um algoritmo multi-partida é utilizado, que chama a Heurística Tabu para resolver o Problema Clique Máxima. Cada solução inicial corresponde a afectar a solução inicial com um (e só um) vértice diferente. Cada iteração, correspondente à solução inicial diferente, pode fornecer um ou mais k-cliques maximais, sendo só guardados os k-cliques maximais diferentes dos anteriormente encontrados.

### 3.2 Cobertura Mínima com k-cliques

Os dados de entrada para o algoritmo de cobertura com k-cliques é uma matriz onde as linhas correspondem aos vértices do grafo e cada coluna é uma k-clique que abrange um determinado número de vértices. Uma heurística para a cobertura com cliques foi proposta por Kellerman [14] e melhorada por Chvatal [5].

Na heurística construtiva em cada iteração, é escolhida uma linha a ser coberta, e em seguida é escolhida uma coluna que cobre a referida linha, finalizando com a actualização da solução e dos restantes vértices cobertos. A linha escolhida é, normalmente, a linha que é mais difícil de cobrir, ou seja, a linha a que correspondem menos colunas. Depois de concluir a cobertura do grafo, o segundo passo é eliminar a redundância; começando por ordenar a cobertura por ordem decrescente de custo, e de seguida verificar se cada k-clique é realmente essencial.

Esta heurística construtiva pode ser melhorada utilizando uma heurística de Procura Tabu que remove as colunas mais caras e reconstrói uma nova solução, tal como apresentado em [11].

O problema de partição de um conjunto é muito semelhante ao problema de cobertura de um conjunto. A partição de um conjunto é uma divisão sem sobreposição de peças do conjunto. Por outro lado, a cobertura de um conjunto permite sobreposições ou sobre-coberturas. Em ambos os problemas sub-coberturas não são permitidos.

## 4 Resultados Computacionais

Para a implementação computacional do algoritmo algumas escolhas têm de ser feitas, como o ambiente computacional, as medidas de “graph mining” e as instâncias de teste.

Os programas foram escritos em linguagem C e o compilador Dev-C++ foi utilizado. Os resultados computacionais foram obtidos a partir de um processador 2.53GHz Intel Core 2Duo com 4,00 GB de memória principal, a funcionar sob o sistema operativo Windows Vista.

A medida de “graph mining” proposta é a sequência de cardinalidades das coberturas por k-cliques de um grafo, que se apresenta no procedimento seguinte. O algoritmo das Duas-Fases (geração de k-cliques e cobertura do grafo) é repetido para valores de  $k=1$  até  $k=\text{diam}(G)$ , com vista a encontrar a sequência de coberturas mínimas de  $k\text{-G(V, E)}$ . Quando  $k=\text{diam}(G)$ , ou a cardinalidade da cobertura por k-cliques é igual à unidade, estamos na presença de um “pequeno mundo”. Por isso dizemos que, a sequência de coberturas de  $k\text{-G(V, E)}$  combina a análise de subgrupos coesos com o conceito de “pequenos mundos”.

#### **Procedimento 0:** Geração da Sequência de Cardinalidades das Coberturas por k-cliques

Entrada: Grafo  $G(V, E)$ ,  $k$

Saída: sequência das cardinalidades das coberturas por k-cliques

1. Para  $k=1$  até  $k=\text{diam}(G)$ :

1.1. Aplicar o algoritmo de Duas-Fases (Proc. 1)

Nos resultados computacionais do Procedimento 0, geração da sequência de cardinalidades das coberturas por k-cliques foram utilizados dois grupos de conjuntos de grafos, os grafos Erdos e alguns ficheiros do concurso da clique DIMACS [7].

Tabela 1: Sequência das Cardinalidades das Coberturas por k-cliques

grafo	nr nós	diâmetro	cardinalidade da cobertura por k-cliques								
			k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=9	k=18
Erdos-97-1	472	6	9	8	7	7	4	1	--	--	--
Erdos-98-1	485	7	8	8	7	5	1	1	1	--	--
Erdos-99-1	492	7	8	8	7	7	1	1	1	--	--
brock200_1	200	2	24	1	--	--	--	--	--	--	--
brock200_2	200	2	26	1	--	--	--	--	--	--	--
brock400_1	400	2	26	1	--	--	--	--	--	--	--
brock400_2	400	2	23	1	--	--	--	--	--	--	--
c-fat200-1	200	18	16	11	9	8	7	7	6	5	1
c-fat200-2	200	9	9	7	5	4	3	3	3	1	--
c-fat500-1	500	40	16	12	9	7	7	6	6	4	3
											1

Nos grafos Erdos cada nó corresponde a um investigador, e dois nós são adjacentes se os investigadores publicaram juntos. Os grafos são denominadas "Erdos-xy", onde "x" representa os dois últimos dígitos do ano em que foi criado o grafo, e "y" é a distância máxima de Erdos para cada vértice no grafo. O segundo grupo de grafos inclui algumas instâncias de cliques DIMACS. Este grupo contém exemplos da família de grafos "brock" e da família de grafos "c-fat".

Para a análise de cada grafo, considerámos o número de vértices, o diâmetro e a cardinalidade do conjunto de k-cliques que cobrem todos os nós, variando k de 1 até ao valor do diâmetro. Os grafos Erdos-98-1 e Erdos-99-1, com diâmetro 7, são cobertos apenas com um 5-clique. Estes valores exemplificam a diferença entre o k-clique e o k-clã; estes grafos são de 5-clique, mas não 5-clãs porque o diâmetro é igual a sete. Os grafos "brock" têm diâmetro igual a 2 e uma cobertura por um 2-clique é suficiente. A maior parte dos casos DIMACS apresentam este perfil. Por outro lado, os grafos "c-fat", têm diâmetros maiores que 7, gerando longas sequências de coberturas de k-cliques. Na medida proposta, a sequência de cobertura por k-cliques, identifica famílias de grafos, e parece ser bastante promissora.

## 5 Conclusões

Dada a grande quantidade de dados, fornecidos pela Web 2.0, existe uma necessidade premente de criar novas medidas para melhor compreender a estrutura das redes, o modo como as suas componentes estão organizadas e o modo como evoluem ao longo do tempo.

As medidas na análise de Redes Complexas são essencialmente baseadas em procedimentos de baixa complexidade computacional, como o diâmetro do grafo, a distribuição do grau dos nós e a verificação da conectividade, subestimando o conhecimento da estrutura das componentes do grafo.

Neste trabalho o conceito de clique-relaxado é estendido a todo o grafo, quando este é coberto por k-cliques, obtendo assim uma visão geral das sub-componentes do grafo. A sequência de coberturas por k-cliques é apresentada, combinando a análise de subgrupos coesos com o conceito dos "pequenos mundos". A análise das sequências identifica diferentes tipos de grafos, mostrando que famílias de grafos têm estruturas diferentes.

Existe ainda um conjunto de características, não mencionadas no presente documento, mas que podem ser obtidas, como a sobre-coberturas dos nós, a cardinalidade das k-cliques e a composição das k-cliques.

Os trabalhos com redes sociais não excedem as dezenas de nós. Neste trabalho, a proposta do algoritmo de duas-fases trabalha com grafos com centenas de nós, com um tempo de execução de alguns segundos. Em trabalhos futuros, gostaríamos de alargar esta abordagem a um maior volume dados, com milhares de nós.

## 6 Referências

1. Alba, R. D.: A graph-theoretic definition of a sociometric clique, *Journal of Mathematical Sociology*, 3, 113-126 (1973)
2. Berners-Lee, T.: The Next Wave of the Web: Plenary Panel, 15th International World Wide Web Conference, WWW2006, Edinburgh, Scotland (2006)
3. Cavique L., C. Luz: A Heuristic for the Stability Number of a Graph based on Convex Quadratic Programming and Tabu Search, special issue of the *Journal of Mathematical Sciences*, Aveiro Seminar on Control Optimization and Graph Theory, Second Series (to appear 2009)
4. Cavique L., C. Rego and I. Themido: A Scatter Search Algorithm for the Maximum Clique Problem, In: Essays and Surveys in Metaheuristics, C. Ribeiro e P. Hansen (Eds), Kluwer Academic Publishers, 227-244 (2002),
5. Chvatal V.: A greedy heuristic for the set-covering problem, *Math. Oper. Res.* 4, 233-235 (1979)
6. Cook D.J., L.B. Holder, Editors: *Mining Graph Data*, John Wiley & Sons, New Jersey (2007)
7. DIMACS: Maximum clique, graph coloring, and satisfiability, Second DIMACS implementation challenge, URL <http://dimacs.rutgers.edu/Challenges/>, accessed April 2009 (1995)
8. Erdos, P., Renyi, A.: On Random Graphs. I., *Publicationes Mathematicae* 6, 290-297 (1959)
9. Faloutsos M., P. Faloutsos, C. Faloutsos: On power-law relationships of the Internet topology, In: SIGCOMM, 251-262 (1999)
10. Floyd, Robert W.: Algorithm 97: Shortest Path, *Communications of the ACM*, 5(6), 345 (1962)
11. Gomes M., L. Cavique, I. Themido: The Crew Time Tabling Problem: an extension of the Crew Scheduling Problem, *Annals of Operations Research*, volume Optimization in transportation 144(1), 111-132 (2006)
12. Grossman J., P. Ion, R. D. Castro: The Erdos number Project, URL <http://www.oakland.edu/enp/>, accessed April 2009 (2007)
13. Johnson D.S.: Approximation algorithms for combinatorial problems, *Journal of Computer and System Science*, 9, 256-278 (1974)
14. Kellerman E.: Determination of keyword conflict, *IBM Technical Disclosure Bulletin*, 16(2), 544-546 (1973)
15. Luce, R. D.: Connectivity and generalized cliques in sociometric group structure, *Psychometrika*, 15, 159-190 (1950)
16. Milgram, S.: The Small World Problem, *Psychology Today*, 1(1), 60-67 (1967)
17. Mokken, R. J.: Cliques, clubs and clans, *Quality and Quantity*, 13, 161-173 (1979)
18. Moreno, J. L.: *Who Shall Survive?*, Nervous and Mental Disease Publishing Company, Washington DC (1934)
19. Scott J.: *Social Network Analysis - A Handbook*, Sage Publications, London (2000)
20. Soriano P., Gendreau M.: Tabu search algorithms for the maximum clique, In: *Clique, Coloring and Satisfiability*, Second Implementation Challenge DIMACS, Johnson D.S., Trick M.A. (Eds.), 221-242 (1996)
21. Wasserman, S., K. Faust: *Social Network Analysis: Methods and Applications*, Cambridge University Press (1994)
22. Watts, D.J., Strogatz, S.H.: Collective dynamics of small-world networks, *Nature* 393(6684), 409-10 (1998)

# Furnace scheduling in a foundry

Paula Cerqueira<sup>‡</sup>

L.Dias<sup>‡</sup>

José A. Oliveira<sup>‡</sup>

<sup>‡</sup> Universidade do Minho  
 pcerqueira@fal.pt  
 lsd@dps.uminho.pt  
 zan@dps.uminho.pt

## Abstract

This paper presents a two-phase heuristic methodology to the production scheduling of a foundry. The first phase covers up a packing problem, in which is grouped together the set of orders, by the type of alloy in well-defined dimension groups. In the second phase is solved a knapsack problem in a defined time horizon, where the furnace set is processed in order to exhaust the daily capacity melting installed. A software tool was developed to implement the algorithm on Microsoft Office Excel 2007, using the VBA language. The results were very good and considerably better than those achieved in real production for the same list of orders.

**Keywords:** scheduling, heuristics, production management

## 1 Introduction

Nowadays because of competitiveness and efficiency, companies try constantly to provide mechanisms that frequently give improvements. The operational research plays an important role in achieving such goals, in the way that it provides organizations with efficient tools to solve their problems and optimize their processes. This article presents a production planning problem in an iron and steel foundry, as an operational research problem. Already other authors, Araujo (2003), Araujo and Arenales (2003, 2004), Araujo et al. (2004), Araujo et al. (2008), Fink (2007), Junior et al. (2006), Silva and Morabito (2004) and Tonaki (2006) studied this problem, however, with different approaches.

The production system consists in three sequential processes: molding, melting / casting and finishing. The planning of operations for the processing of orders is a problem of high complexity and its solution is not trivial. The orders must be divided by alloy types and grouped in sets of maximum size in order to minimize the production costs, which makes more difficult to accomplish delivery deadlines.

This article presents a heuristic methodology consisting in a two phases algorithm. In the first phase the problem is shaped as a *Bin Packing Problem* (Alvim et al. 2004; Coffman et al. 1996; Fleszar and Hindi 2002; Gupta and Ho 1999; Martello and Toth 1990; Osogami and Okano 2003) to group the orders in *bins*. In the second phase the problem is shaped as a *Knapsack Problem* (Kellerer et al. 2004; Martello and Toth 1990), to group the *bins*.

obtained in the previous phase. For both phases are presented two constructive algorithms, associated to a local search procedure (Michiels et al. 2007; Osogami and Okano 2003). The methods of exact solution are not suitable for the treatment of this problem due to the size of the instances of the real problem and to the need of make frequent executions of the algorithm. This methodology will be included in a decision support system and it is crucial to obtain solutions quickly, so that the coordinator can make a plan and respond to the permanent request of new orders.

A software tool was developed on Microsoft Office Excel 2007, using the Visual Basic for Applications (VBA) language. This option was chosen because the company information system allows moving data into this format, each order being presented in each line. The data reading (orders) will be made line by line and the information will be processed in the form of the desired output, also in an Excel sheet.

## 2 Problem Description

This section describes in detail the two processes of the production system identified as critical for scheduling, molding and melting, and the implications that generate on the production planning. The molding and melting processes were identified as the determinant points of the production and as the critical points of planning, once products arise to the finishing process, they are processed without major constraints. In the finishing process are included the final operations of dimensional correction and part cleaning. In molding, the molds are made of sand, with the format of the part to produce, in the necessary quantity to meet the production order. In the molding process there are three separate sectors, one automatic and two manuals, where the choice of molds to produce in each depends on its size and on its quantity. The automatic molding line is the one that most contributes to the total quantity of molds (72%) and metal weight (56%) to produce. The line of automatic molding has a special characteristic that distinguishes its planning from the conventional machines of metallurgy, which usually works only with one tool at a time. In the molding line four different tools (molds patterns) are used at the same time, with two half molds each. Figure 1 outlines the molding line process.

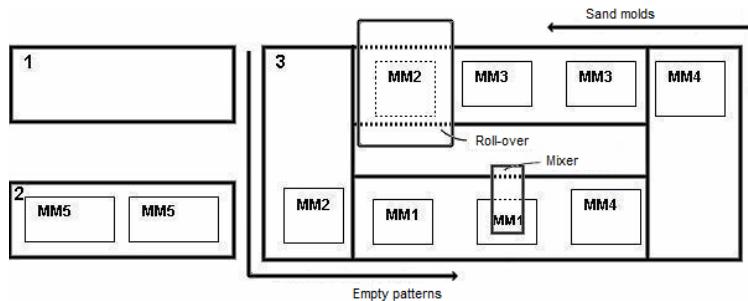


Figure 1: Diagram of the automatic molding line

To illustrate the operation, consider four different patterns (parts): 1, 2, 3 and 4. Knowing that each has two half patterns (MM) the sequence of processing in the carousel circuit is the one illustrated in Figure 1. Figure 1 shows three distinct areas of operation relating to the automatic molding machine. Areas 1 and 2, respectively, represent the exit and entry of patterns and zone 3 is the processing area. The patterns move in a circle, in zone 3, and are filled with sand under the mixer. The sand molds are removed from the patterns in the roll-over. When the amount of molds needed to meet the production orders is reached the corresponding two half patterns leave to zone 1 and two new half patterns (MM5) of a new part enter from zone 2. The selection of a new mold must comply with planning restrictions imposed by the subsequent operations (melting and casting).

In the second process, melting, the metal is melted into an alloy of a known chemical composition. For the melting there are two furnaces with melting capacity of 500 kg and

one with the capacity of 1500 kg. The melting timing of the 500 kg furnaces is about 1 hour and 10 minutes and on the 1500 kg furnaces is about 1 hour and 30 minutes. The scheduling of this study is focused only in the furnace of 1500 kg furnace where is produced 89% of the metal, constituting quite a reasonable and important simplification of the model. The casting is the operation where the molds are filled with the metallic alloy. The casting operation is a quite fast operation when compared with the others. However, this operation implies the existence and availability of all the molds before it begins.

The production planning represents an exercise of high degree of difficulty because involves a huge quantity of references, mainly in the automatic molding line considered crucial to the progress of all the other production process. Only when the metal weight from the molding process, by alloy, is enough, is that the melting process could be planned. Only parts of the same alloy can be considered in the same melting batch. This fact requires a great coordination between these two sequential processes.

Summarizing, it is important to obtain an adequate planning of molding and melting processes, to ensure enough weight of metal to compose meltings and to respect the daily capacities of molding and melting. The molding and melting capacities are known and established in terms of the number of daily molds and daily number of meltings in each furnace, respectively.

### 3 Resolution Methodology

The resolution methodology adopted is a two-phase algorithm, where the solutions obtained in a first phase determine the quality of solutions for the following phase. The first phase is approached as a packing problem and the second phase is solved as a knapsack problem. To shape the problem is considered the melting as early planning, where it is necessary to ensure enough weight of alloy to melt. Actually, it is the molding process that introduces the manufacturing orders in production, assuming a decisive role in processing order, and the melting is being scheduled as the molds are released from the first process. However, there may be profits with the anticipated scheduling of melting loads, with the balancing of metal weights per mold.

The aim of the first phase is to allocate orders, or part of them, to a certain melting (furnace) with some criterion of uniformity, in order to consider orders of the same alloy, with delivery deadlines the closest as possible. A melting, or a furnace, is a *bin* in this packing problem.

An intuitive approach is to assign to a furnace the orders in decreasing order of priority, providing a constructive and very simple heuristic. The criterion of priority for each order is determined by a weighted sum of four parcels: days in advance, order priority, order margin, which represents the company profit for each order, and the client priority.

Mathematically given a set of orders  $E = \{e_1, \dots, e_n\}$ , of weights  $w_j$ , alloys  $a_j$  and criteria  $d_j$  and a set  $m$  of furnaces of capacity  $c$ , each order is attached to each furnace according to the following terms:

$$d_{[1]} \geq d_{[2]} \geq \dots \geq d_{[n]} \quad (1)$$

$$\text{s.a. } \sum_{j=1}^{k_i < n} w_{[j]i} \leq c \quad , \quad i = 1, \dots, m \quad j \in E \quad (2)$$

$$a_{[1]i} = a_{[2]i} = \dots = a_{[j]i} \quad , \quad i = 1, \dots, m \quad j \in E \quad (3)$$

The orders fitting in a certain furnace are allocated up to its capacity before moving to the next furnace. At the end of this assignment there are  $m$  furnaces filled with orders of the same alloy, which constitutes the starting point for the next phase.

The constructive heuristic used for the definition of meltings is similar to that used in the *bin packing* problems. The purpose of heuristics is to group, in an appropriate manner, the set of orders  $E = \{e_1, \dots, e_n\}$  to compose the furnaces (meltings) required  $F = \{f_1, \dots, f_m\}$ .

This scheduling is exclusive to the furnace of the 1500 kg, because it is the one with the higher income of melting, and it contributes prominently to the total amount melted in the company. However, when the weight of a particular alloy is not enough to ensure a melting in this furnace, the orders are assigned to other existent furnaces, whose working decision belongs to the user, or, in case the minimum amounts of melting are not reached (2/3 of the melting weight of the smaller furnaces) a list with the information of these orders is emitted, and the scheduling decision of the orders in that list will also belong to the user.

The flowchart illustrated in Figure 2 schematizes the algorithm of furnaces definition (first phase), with the increases of the capacity restrictions adjusted to the reality of the company.

For the second phase of the methodology the resolution strategy for the scheduling is based on the resolution strategy of the knapsack problem. In this case, the furnaces to schedule will be the items to consider, and the days of planning will be the available knapsacks. The daily capacity of melting represents a knapsack of a well defined capacity. In this way, the knapsack problem could be used to shape the attribution of furnaces, built in the previous phase, to the knapsacks, which constitute the planning days. As the assumption, the number of items (furnaces) is superior to the available capacity in the several knapsacks (planning days) defined for the planning period.

Mathematically, given a set of furnaces  $F = \{f_1, \dots, f_m\}$ , of weights  $w_j = 1$  and criteria  $d_j$  and given a set of planning days /knapsacks  $M = \{M_1, \dots, M_k\}$  of capacity  $c$ , each furnace is attached to each knapsack according to the following terms:

$$d_{[1]} \geq d_{[2]} \geq \dots \geq d_{[m]} \quad (4)$$

$$\text{s.a. } \sum_{j=1}^{n_i < m} w_{[j]i} \leq c \quad , \quad i = 1, \dots, k \quad j \in F \quad (5)$$

Note that each furnace has an individual weight  $w_j = 1$ , as each furnace occupies a unitary space in a capacity translated in number of furnaces/meltings.

At this phase it must be taken into account the remaining restrictions of capacity. The molding capacities of each sector are known and must be kept. It is necessary to also consider the ability to process four tools simultaneously in the automatic molding line, constituting the usual processing situation.

The resolution will be done again with a constructive heuristic (filling of the knapsacks). Figure 3 shows a flowchart of the algorithm. It is considered that the set  $F = \{f_1, \dots, f_m\}$  generated in the previous phase is ordered by a well-defined criterion. For this purpose the criteria of the orders already determined are used, with a balance based on the weight that each order represents in the whole furnace weight in which is assigned to ( $\text{peso}_{\text{enc.}} / \text{peso}_{\text{forno}}$ ).

## 4 Local Search

In order to enhance the quality of the results obtained by the constructive algorithms we implement a first iteration of a local search procedure. In a local search algorithm, starting from an initial complete solution, repeatedly, the aim will be to replace the current solution for a better solution that exists in the neighborhood of the current solution (Michiels *et al.* 2007).

The original solution used, will be achieved through the use of constructive algorithms presented in the flowcharts of Figure 2 and Figure 3. From the current solution we start a process of transformation of this solution. The best solution of the neighborhood is assessed and then taken as the current solution if it is better than the actual current solution.

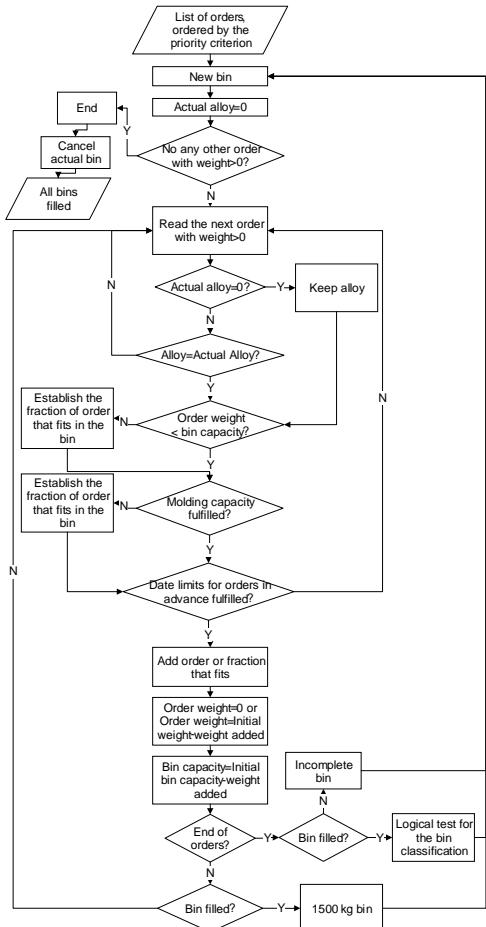


Figure 2:  
Flowchart of the algorithm of furnaces definition

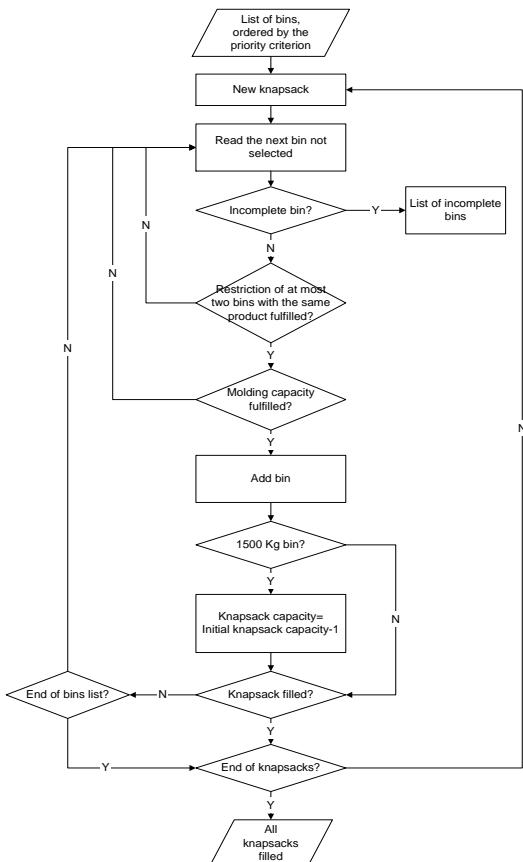


Figure 3:  
Flowchart of the algorithm of days definition

To test the viability of a local search as a quality improvement mechanism of the solution achieved through constructive algorithms (phase 1 and 2), was developed an algorithm that enables a blind order (for the first phase) or furnace (for the second phase) in a given iteration, preventing their assignment and making them available for review in the following iterations. These occultations are made randomly in each iteration, providing small changes regarding the original solution and giving rise to the possibility of a different solution generated randomly.

## 5 Computational Experiments

In order to validate the methodology designed, computational experiments were carried out with real data from the company and these were compared with the company procedure in two different situations. An additional study for the scheduling of two months of production in two different formats has also been carried, allowing to observe some important lessons.

Thus, the results were assessed in the following situations:

- Comparison of ten days of production (actual data obtained for comparable situations);
- Comparison of a month of production with the 2007 worst and best month of real production;
- Comparison of two months of planned production sequentially, deducting the orders planned in the first month in the list of orders of the second month;
- Comparison of two months of production planned simultaneously.

The fulfilling obtained for the 1500 kg meltings, in the first phase of planning, demonstrated to be sufficiently good. The average was 98% of fulfilling and 76% of furnaces had a fulfilling above 94%. In the second phase, for the purpose of comparison of the results obtained in the program with the real production in a ten working days, the real data was obtained between coating maintenances of the 1500 kg furnace (sinterizations), which prevents its normal work on two consecutive days, and in a date of molds stocks determined for this purpose.

For the same list of orders we found that the program results are rather better than the real ones. In the scheduling results the occupation of the melting capability of the 1500 kg furnace is total, unlike the actual production, and the consumption of molds generated in the three sectors of molding, on those ten days, is complete. In real production only took place 53 meltings (88%) of a total capacity of 60 meltings, all planned in the results of the program. In real production it took place, however, a greater use of one of the furnaces of 500 kg resulting in a higher number of total meltings ,than the one from the program (92 real versus 62 of the program) but with an inferior number of cast molds: in real production there were only casted 87% of the molds that the program results does allow to cast.

However, it was in the criterion of scheduling and in consideration of priority orders, that the program most stood out against the real scheduling. The program, within the possibilities determined by the restrictions of the model, considers the orders by their priority, established for the study. In real production, currently, there is not a well-defined scheduling criterion: orders with posterior delivery deadlines are processed before orders with a previous delivery deadline.

By comparison with the best and worst months of 2007 of the company, in terms of weight of metal cast, the result and conclusions are similar. The program ensures a better optimization of the melting, with a more uniform production than the real production. The main results, in terms of global indicators, are summarized in Table 1.

Table 1: Comparison per day between the overall results obtained in the program and the real production for a month of work (the best and the worst month):

	Program	Best mon.	Worst mon.
Nº of molds	107,7	106,8	84,8
Nº of parts	191,2	298,0	216,2
Casted metal (kg)	10119,9	7977,9	7100,0
Nº meltings 1500 kg	5,9	4,6	4,3
Nº meltings 500 kg	0,3	1,1	1,0
Nº total meltings	6,2	5,7	5,3

The results of the program have a greater number of cast molds, by the use of a better fulfilling of the capacity melting of the furnace of 1500 kg. As can be seen in Table 1 we can achieve a considerable improvement in average weight of cast metal a day using a greater number of meltings in the furnace of 1500 kg. In the two months scheduling, it was found that with the new list of orders, for the second month of the sequential scheduling, the results have conducted to a low occupation of the melting capacity. For this month, with a list of orders obtained by discounting the orders considered in the first month, only 68% of the available meltings were planned. With the initial list of orders, for the first month, the occupation of melting, in terms of number of meltings, was 98%. This low occupation of the second month is due to two reasons. A first reason due to the lack of molds from the larger manual molding sector, corresponding to higher weight molds and higher contribution to the melting. This sector is underused due to the lack of references pre-established to produce there. This may represent an improvement to the presented tool. Each product may correspond, for default, to a particular sector molding, however, it should be possible the assignment it to another sector, if necessary, and when it's possible. Apart from this identified factor there is another reason: the planning of

orders of metal weight per mold below the weights programmed for the first month. For a constant molding capacity, in terms of number of daily molds, the weight of melting obtained will be lower, because it results from the sum of the metal weights per molds, that are more reduced, and could constrain the ability to optimize the melting. This can also be an improvement for the developed tool.

Through the combination of orders by weight of metal molded we can ensure or not the melting weight necessary. The weight of the metal has been identified by a daily average molding, for an optimum, between molding and melting, of 94 kg. In this way, in the present study we observed that, for the actual molding capacity, through a daily average of metal weight per mold of 94 kg, we can obtain a full occupation of the molding capacity and the melting capacity and a balance between these two sequential processes.

In this study it is also verified that, for the molding and melting capacity studied, by the combination of weight of metal per mold, it may be the process of molding or the process of melting to strangle the production process. Therefore, through scheduling orders of reduced metal weight per mold, we can easily exhaust all the capacity of molding, without achieving the necessary melting weight. For orders of higher metal weight per mold happens exactly the opposite, where they exhaust the capacity of melting, even with uncast molds, not being able to avoid the stock accumulation.

The local search mechanism, through a first iteration, has led to some improvement in the value of objective functions defined for each of the phases. The objective function was established in order to minimize the number of bins for the first phase. For the second phase was introduced an equation to consider the parcels established for the criterion of priority of each order and, simultaneously, the weight of metal, appropriated with the reality of the company.

The tests were performed to 6 lists of different orders chosen randomly and for 2000 and 10000 iterations of transformation, in order to assess the influence of the size of the neighborhood set. It is not established that there is an improvement considering a larger neighborhood. For the first phase, the best improvement achieved was a 1,0% reduction in the number of bins (reduction of 4 bins in a total of 395). In the second phase, the improvement in the value of the objective function of the new solution was higher and reached 2,2% of the original solution. However, for the overall results, through the influence of the best solution of the first phase in the quality of the solution of the second phase, the results were not enough to conclude the existence of a pattern of improvement. Through the generation of 2000 and 10000 possible alternative solutions of the original solution, it is observed that the results achieved through constructive algorithms are already quite reasonable. This observation was achieved for the existence of quite worse alternative solutions, obtained with only a slight transformation of the original solution.

This study became of great relevance for the studied foundry because it gave a tool that, through the detail needed, program its orders, by a simple way, fast enough and efficient enough. This tool made a bridge between decisions taken by intuition and experience and well-programmed decisions. Today the organisation has a tool that provide a production planning, something that did not have before.

## 6 Conclusions

The problem of production scheduling of a foundry company is very complex and of hard solution. The use of exact solution methods may be compromised in the resolution of instances of large scale projects. The alternative involves the use of heuristic methodologies.

The complexity of the problem suggests the implementation of the strategy of dividing the original problem into two sub-problems and implement, in each case, a specific methodology: in an initial phase the bin packing problem, for the assignment of the orders to the furnaces, and in a second phase, the knapsack problem methodology, for the assignment of these furnaces to processing days. It is known that this methodology is an approach methodology, but it appears that the simplicity of the model does not question the achievement of good quality solutions. In order to enhance the quality of the

solutions of the problem we think that the implementation of procedures of local search is justified and will avoid the disadvantages arising from the use of two-phase strategy and constructive algorithms. The software tool developed through the application of algorithms identified under the form of VBA, in Microsoft Excel 2007 format, has proved to be very useful, with sufficiently efficiency and effectiveness for the uncertainties of real data that characterize this type of market. The results could be used for the early scheduling of production, according to well-defined criteria, and to achieve better results than the real production, for the same list of orders. There are many other studies and developments that could be made in the follow of this study, more specifically in the foundry industry, but also in other industries that are characterized by processes that depend on the production rate and release frequency of components from previous processes.

## **7 References**

- Alvim, A. C., Ribeiro C.C., Glover, F. and Aloise, D. J. (2004) A hybrid improvement heuristic for the one-dimensional bin packing problem, *Journal of Heuristics*, Vol. 10, pp. 205-229.
- Araujo, S. A.(2003) Modelos e métodos para o planejamento e programação da produção aplicados no setor de fundições. Universidade de São Paulo, Tese de Doutoramento.
- Araujo, S. A. and Arenales, M. N. (2003) Dimensionamento de lotes e programação do forno numa fundição automatizada de porte médio, *Pesquisa Operacional*, Vol. 23, pp. 403-420.
- Araujo, S. A., Arenales, M. N. and Clark, A. R. (2004) Dimensionamento de lotes e programação do forno numa fundição de pequeno porte, *Gestão Produção*, Vol. 11, 165-176.
- Araujo, S. A and Arenales, M. N. (2004) Planejamento e programação da produção numa fundição cativa automatizada de grande porte, *Investigação Operacional*, Vol. 24, pp. 197-210.
- Araujo, S. A., Arenales, M. N. and Clark, A. R. (2008) Lot sizing and furnace scheduling in small foundries, *Computers & Operations Research*, Vol. 35, No. 3, pp. 916-932.
- Coffman, E. G., Garey, M. R. and Johnson, D. S. (1996) Approximation algorithms for bin packing: a survey. In Hochbaum, D. (Ed.) *Approximation algorithms for NP-Hard problems*, Boston: PWS Publishing, p. 46-93.
- Fink, C. (2007) A programação da produção em fundições de pequeno porte: modelagem matemática e métodos de solução, Universidade de São Paulo, Dissertação de Mestrado.
- Fleszar, K. and Hindi, K. S. (2002) New heuristics for one-dimensional bin packing, *Computers & Operations Research*, Vol. 29, pp. 821-839.
- Gupta, J. N. D. and Ho, J. C. 1999) A new heuristic algorithm for the one-dimensional bin-packing problem, *Production Planning & Control*, Vol. 10, No. 6, pp. 598-603.
- Junior, R. F. T., Fernandes, F. C. F. and Pereira, N. A. (2006) Sistema de apoio à decisão para programação da produção em fundições de mercado, *Gestão & Produção*, Vol. 13, pp. 205-221.
- Kellerer, H., Pferschy, U. and Pisinger, D. (2004) *Knapsack Problems*, Springer, Berlin.
- Martello, S. and Toth, P. (1990) *Knapsack Problems: Algorithms and Computer Implementations*, John Wiley & Sons, Chichester.
- Michiels, W., Aarts, E. and Korst, J. (2007) *Theoretical aspects of local search*, EATCS monographs in Theoretical Computer Science, Springer, New York.
- Osogami, T. and Okano, H. (2003) Local search algorithms for the bin packing problem and their relationships to various construction heuristics, *Journal of Heuristics*, Vol. 9, pp. 29-49.
- Silva, R. J. and Morabito, R. (2004) Otimização da programação de cargas de forno em uma fábrica de fundição em aço-inox, *Gestão & Produção*, Vol. 11, No. 1, pp. 135-151.
- Tonaki, V. S. (2006) Uma heurística para o problema de dimensionamento de lotes em fundições de mercado, Universidade de São Paulo, Tese de Mestrado.

# Relaxations for the truss topology design problem

Adelaide Cerveira \*

Fernando Bastos †

\* Departamento de Matemática da UTAD, Vila Real  
 CIO- Centro de Investigação Operacional  
 cerveira@utad.pt

† Departamento de Estatística e Investigação Operacional da FC-UL, Lisboa  
 CIO- Centro de Investigação Operacional  
 fbastos@fc.ul.pt

## Abstract

This paper considers the classical problem of finding a truss design with minimal compliance subject to a given external force and a volume bound. This is a well-studied problem for continuous cross sectional areas. Generally, the optimal solutions obtained for this problem contain bars with many different cross sectional areas. In real life we have only a finite set of possible values, so it is important to consider discrete constraints (Cerveira (1997), Faustino et al. (2006), among others). In this paper we consider a single bar's cross sectional area case. Achtziger and Stolpe (2005) present a branch and bound method to solve this problem. We present semidefinite and lagrangean relaxations in order to get better lower bounds, to be used in a branch and bound method, than the one provided by linear relaxation. Some computational results are presented and the performance of the relaxations are compared. The best bound is provided by one of the lagrangean relaxations.

## Abstract

Neste artigo considera-se um problema clássico de optimização do desenho de estruturas que consiste em obter a estrutura mais resistente a um dado conjunto de forças exteriores e cujo volume não seja superior a um dado valor. No caso contínuo, que tem sido amplamente estudado, as estruturas óptimas contêm barras com muitos valores distintos para a área da secção transversal. No entanto, no mercado existe apenas um leque restrito de valores disponíveis. Surge então a necessidade de incluir restrições discretas (Cerveira (1997), Faustino et al. (2006), entre outros). Neste artigo considera-se que as barras da estrutura têm o mesmo valor para a área da secção transversal. Achtziger e Stolpe (2005) apresentam um método de *branch and bound* para resolver este problema. Neste artigo apresentam-se algumas relaxações semidefinidas e relaxações lagrangeanas para obter um limite inferior melhor que o fornecido pela relaxação linear, para se usar num método de *branch and bound*. São apresentados alguns resultados computacionais e comparam-se as várias relaxações. O melhor limite inferior é fornecido por uma das relaxações Lagrangeanas.

**Keywords:** truss topology design (TTD), stiffness, semidefinite programming (SDP), relaxation, lagrangean duality, cutting-plane method

## 1 Introduction

Truss topology design (TTD) deals with constructions like bridges, cantilevers and roof trusses supporting different load scenarios. The selection of an optimal configuration for the structure depends on the used criteria, see for instance Ben-Tal and Bendsøe (1993), Patriksson and Petersson (1995), Ben-Tall and Nemirovski (2001), Bastos et al. (2005), Faustino et al. (2006), Achtziger and Stolpe (2005). In this paper we examine the issue of the stiffness of the truss for a given amount of material for a single bar case. The design variables describe the cross-sectional area of the bars and these must be zero or a given value. This results in a large-scale non-convex problem with binary variables which can be formulated as a SDP Problem with binary variables. We derive semidefinite and Lagrangean relaxations for the binary problem.

The paper is organized as follows. In Section 2 we briefly present the formulation of the problem, in Section 3 we derive some relaxations of the problem and, in Section 4, we present computational results. The details of the proofs can be found in Bastos and Cerveira (2007).

## 2 Problem Formulation

A *truss* is a two or three dimensional structure composed of elastic *bars* linked at *nodes* which may be fixed, free or supported. We only consider two dimensional trusses. A fixed node has 0 degrees of freedom, a free node has 2 degrees of freedom and, a supported node has just 1. The number of degrees of freedom of the truss is the sum of the corresponding values on its nodes.

Initially we have a *ground structure* which is a previously chosen set of nodes and connecting bars. We can consider all possible links between the nodes, *rich topology*, or only the links between neighbouring nodes, *poor topology* (see Figure 1).

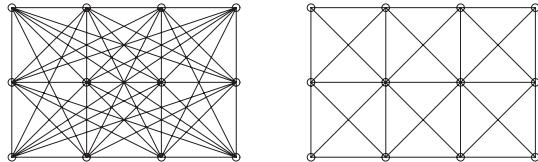


Figure 1: Rich and poor topologies.

The goal is to find the stiffest truss capable of withstand the given load acting on the nodes,  $f$ , with a volume not greater than a predefined value,  $v$ . The design variables are the cross-sectional area of the bars that can assume two values, 0 or  $a > 0$ . Only the bars with nonzero cross-sectional area are part of the final structure. When external loads are acting on the nodes the structure deforms. We may describe that deformation by the vector of nodal displacements,  $u$ , being the work done by external forces  $f^\top u$ . We call *compliance* to  $\frac{1}{2}f^\top u$ . This is a measure of the stiffness of the truss: the smaller the compliance the larger the stiffness of the truss with respect to the load. The problem can be formulated, in variables  $\alpha, u$ , as follows:

$$\begin{aligned} & \min f^\top u \\ \text{TTD} \quad & s.t. \quad K(\alpha)u = f, \\ & \alpha \in \mathcal{C}, \\ & u \in \mathbb{R}^n, \end{aligned} \tag{1}$$

where  $\mathcal{C} = \{\alpha \in \mathbb{R}^m : \sum_{i=1}^m \alpha_i a s_i \leq v, \alpha_i \in \{0, 1\}\}$ ,  $s_i$  is the length of bar  $i$ ,  $m$  is the number of bars in the ground structure,  $n$  is the number of its degrees of freedom,  $K(\alpha) = \sum_{i=1}^m \alpha_i t_i K_i$  is the stiffness matrix of the structure were  $t_i = a s_i$  and  $K_i$  is a dyadic matrix. This problem is equivalent to a min-max problem,  $\min_{\alpha \in \mathcal{C}} \max_{u \in \mathbb{R}^n} \{2f^\top u - u^\top K(\alpha)u\}$ , which can be written as

a SDP problem with binary variables,

$$\begin{aligned} & \min_{\tau, \alpha} \tau \\ \text{Bin} \quad & \text{s.t. } \begin{bmatrix} \tau & f^\top \\ f & \sum_{i=1}^m \alpha_i t_i K_i \end{bmatrix} \succeq 0, \\ & t^\top \alpha \leq v, \\ & \alpha \in \{0, 1\}^m. \end{aligned}$$

In the next section we will present several relaxations of problem *Bin*, the details can be found in Bastos and Cerveira (2007). The used semidefinite results can be found in Lemaréchal and Oustry (1999), Helmburg (2000), Klerk (2002).

### 3 Relaxations

#### 3.1 Linear relaxation - *LR*

The linear relaxation of problem *Bin*, denoted by *LR*, is obtained replacing  $\{0, 1\}^m$  by its convex hull,  $[0, 1]^m$ , which is equivalent to  $m$  linear matrix inequalities (LMI),

$$\begin{bmatrix} \alpha_i & 0 \\ 0 & 1 - \alpha_i \end{bmatrix} \succeq 0, \quad i = 1, \dots, m.$$

#### 3.2 Complete relaxation - *Rc*

If we express the 0-1 constraint,  $\alpha_i \in \{0, 1\}$ , as  $\alpha_i^2 - \alpha_i = 0$  and dualize all the constraints in problem *Bin* the lagrangean function is

$$\mathcal{L}_C(z, \gamma, b_0, b, B; \tau, \alpha) = \tau + \alpha^\top \text{Diag}(z)\alpha - z^\top \alpha + \gamma(t^\top \alpha - v) - \text{Tr} \left( \begin{bmatrix} \tau & f^\top \\ f & \sum_{i=1}^m t_i \alpha_i K_i \end{bmatrix} \begin{bmatrix} b_0 & b^\top \\ b & B \end{bmatrix} \right)$$

and the dual variables are  $z \in \mathbb{R}^m$ ,  $\gamma \in \mathbb{R}_+^+$  and  $\begin{bmatrix} b_0 & b^\top \\ b & B \end{bmatrix} \succeq 0$ . Letting  $w_i = t_i \text{Tr}(K_i B)$ ,

$$\mathcal{L}_C(z, \gamma, b_0, b, B; \tau, \alpha) = \begin{bmatrix} \tau & \alpha^\top \end{bmatrix} \begin{bmatrix} 0 & 0^\top \\ 0 & \text{Diag}(z) \end{bmatrix} \begin{bmatrix} \tau \\ \alpha \end{bmatrix} + \begin{bmatrix} 1 - b_0 & (-z + \gamma t - w)^\top \end{bmatrix} \begin{bmatrix} \tau \\ \alpha \end{bmatrix} - \gamma v - 2f^\top b. \quad (2)$$

The dual function is  $\theta_C(z, \gamma, b_0, b, B) = \min_{(\tau, \alpha)} \mathcal{L}_C(z, \gamma, b_0, b, B; \tau, \alpha)$  and the dual problem is

$$\begin{aligned} & \max_{r, z, \gamma, b_0, b, B} \theta_C(z, \gamma, b_0, b, B) \\ \text{s.t.} \quad & \begin{bmatrix} b_0 & b^\top \\ b & B \end{bmatrix} \succeq 0 \quad \Leftrightarrow \quad \begin{aligned} & \max_{r, z, \gamma, b_0, b, B} r \\ & \text{s.t.} \quad \theta_C(z, \gamma, b_0, b, B) \geq r, \\ & \begin{bmatrix} b_0 & b^\top \\ b & B \end{bmatrix} \succeq 0, \\ & \gamma \geq 0. \end{aligned} \end{aligned} \quad (3)$$

The lagrangean function, (2), has a finite lower bound if and only if:

$$\begin{bmatrix} 0 & 0^\top \\ 0 & \text{Diag}(z) \end{bmatrix} \succeq 0 \text{ and } \begin{bmatrix} 1 - b_0 \\ -z + \gamma t - w \end{bmatrix} \in \mathcal{R} \left( \begin{bmatrix} 0 & 0^\top \\ 0 & \text{Diag}(z) \end{bmatrix} \right).$$

When these conditions hold, we have

$$\theta_C(z, \gamma, b_0, b, B) = -\gamma v - 2f^\top b - \frac{1}{4} \left[ 1 - b_0 \quad (-z + \gamma t - w)^\top \right] \begin{bmatrix} 0 & 0 \\ 0 & \text{Diag}(z) \end{bmatrix}^\dagger \begin{bmatrix} 1 - b_0 \\ -z + \gamma t - w \end{bmatrix}.$$

So, using Schur complement and semidefinite results the dual problem, (3), can be written as a SDP problem,

$$Rc \quad \begin{aligned} & \max_{r,z,\gamma,b,B} r \\ & s.t. \left[ \begin{array}{cc} -\gamma v - 2f^\top b - r & \frac{1}{2}(-z + \gamma t - w)^\top \\ \frac{1}{2}(-z + \gamma t - w) & \text{Diag}(z) \\ & \begin{matrix} 1 & b^\top \\ b & B \\ & \gamma \end{matrix} \end{array} \right] \succeq 0. \end{aligned} \quad (4)$$

Problem  $Rc$  is strictly feasible and so if the optimal value of  $Rc$  is finite then an optimal solution for the dual of  $Rc$  is attained and the optimal values are equal. Unfortunately, the dual of  $Rc$  are equivalent to  $LR$  so, we have  $\text{val}(Rc) = \text{val}(DRc) = \text{val}(LR)$ <sup>1</sup>.

### 3.3 Linear matrix inequality - relaxation $Rlmi$

If we dualize only the LMI of problem  $Bin$ , the dual problem is:

$$\max_{b_0,b,B} \theta_{LMI}(b_0,b,B) \quad s.t. \quad \left[ \begin{array}{cc} b_0 & b^\top \\ b & B \end{array} \right] \succeq 0$$

with  $\theta_{LMI}(b_0,b,B) = \min_{\tau \in \mathbb{R}, \alpha \in \mathcal{C}} \left\{ (1-b_0)\tau - 2f^\top b - \sum_{i=1}^m \alpha_i t_i \text{Tr}(K_i B) \right\}$  where the dual variable is matrix  $\left[ \begin{array}{cc} b_0 & b^\top \\ b & B \end{array} \right] \succeq 0$ . As  $\theta_{LMI}(b_0,b,B) = -\infty$  if  $b_0 \neq 1$ , the dual problem can be written as

$$\max_{b,B} \left\{ -2f^\top b - \max_{\alpha \in \mathcal{C}} \sum_{i=1}^m \alpha_i t_i \text{Tr}(K_i B) \right\} \quad s.t. \quad \left[ \begin{array}{cc} 1 & b^\top \\ b & B \end{array} \right] \succeq 0. \quad (5)$$

Relaxation  $Rlmi$  can improve the lower bound, in fact  $\text{val}(Rlmi) \geq \text{val}(Rc)$ .

To solve the dual problem we apply a cutting-plane method, described below. The dual problem, formulation (5), can be written as

$$\begin{aligned} & \max_{z,b,B} z \\ & s.t. \left[ \begin{array}{cc} 1 & b^\top \\ b & B \end{array} \right] \succeq 0 \\ & z \leq -2f^\top b - \sum_{i=1}^m t_i \alpha_i \text{Tr}(K_i B), \quad \forall \alpha \in \mathcal{C}. \end{aligned} \quad (6)$$

Let us consider a relaxation,  $(SDP_k)$ , of (6) with only  $k$  linear constraints associated to  $k$  elements of  $\mathcal{C}$ ,  $\alpha^1, \dots, \alpha^k$ . When  $k$  grows, problem  $(SDP_k)$  will approach (6).

**Algorithm 1** Cutting-plane method to solve  $Rlmi$

0. Let  $e_1, \dots, e_m, \alpha^1 \in \mathcal{C}$  and  $k = 1$ <sup>2</sup>.
1. (iteration  $k$ )
  - (a) Solve  $(SDP_{m+k})$ . Let  $(z^k, b^k, B^k)$  be the optimal solution.
  - (b) Compute  $\theta'_{LMI}(b^k, B^k)$ . Let  $\alpha^{k+1}$  be the optimal solution.
2. If  $\theta'_{LMI}(b^k, B^k) \geq z^k$ , END: the optimal criterium is satisfied.  
Otherwise, add to problem  $(SDP_{m+k})$  another constraint, corresponding to  $\alpha^{k+1}$ . Replace  $k$  by  $k + 1$  and loop to step 1.

<sup>1</sup> $\text{val}(P)$  denotes the optimal value of problem  $(P)$ .

<sup>2</sup> $e_1, \dots, e_m$  are the canonical vectors and vector  $\alpha^1$  is obtained rounding the optimal solution of relaxation  $LR$ .

### 3.4 Lagrangean decomposition - $DL$

If we define  $\mathcal{Y} = \mathbb{R} \times \mathcal{C}$  and  $\mathcal{X} = \left\{ (\tau, \alpha) \in \mathbb{R}^{m+1} : \begin{bmatrix} \tau & f^\top \\ f & \sum_{i=1}^m t_i \alpha_i K_i \end{bmatrix} \succeq 0, t^\top \alpha \leq v, \alpha \in [0, 1]^m \right\}$ , problem  $Bin$  can be written as

$$\begin{array}{lll} \min_{\tau, \alpha} & \tau & \min_{\tau, \alpha, \tau', \alpha'} \tau \\ \text{s.t.} & (\tau, \alpha) \in \mathcal{X} \cap \mathcal{Y} & \iff \quad \text{s.t.} \quad (\tau, \alpha) \in \mathcal{X}, (\tau', \alpha') \in \mathcal{Y}, \\ & & (\tau, \alpha) = (\tau', \alpha'). \end{array}$$

If we dualize the equality  $(\tau, \alpha) = (\tau', \alpha')$ , the dual problem is  $\max_{(\lambda_0, \lambda) \in \mathbb{R}^{m+1}} \theta_{DL}(\lambda_0, \lambda)$  where vector  $(\lambda_0, \lambda) \in \mathbb{R} \times \mathbb{R}^m$  is the dual variable and

$$\begin{aligned} \theta_{DL}(\lambda_0, \lambda) = & \min_{\tau, \alpha, \tau', \alpha'} (1 + \lambda_0)\tau - \lambda_0\tau' + \lambda^\top \alpha - \lambda^\top \alpha' \\ \text{s.t.} & (\tau, \alpha) \in \mathcal{X}, \\ & (\tau', \alpha') \in \mathcal{Y}. \end{aligned}$$

If  $\lambda_0 \neq 0$  we have  $\theta_{DL}(\lambda_0, \lambda) = -\infty$ . Therefore, the dual problem can be written as

$$DL \quad \max_{\lambda \in \mathbb{R}^m} \left\{ \min_{(\tau, \alpha) \in \mathcal{X}} \{\tau + \lambda^\top \alpha\} + \min_{\alpha' \in \mathcal{C}} \{-\lambda^\top \alpha'\} \right\}. \quad (7)$$

Relaxation  $DL$  can improve the lower bound, in fact  $val(DL) \geq val(LR)$ .

To solve the dual problem, (7), we apply the cutting plane method similar to **Algorithm 1**. In the initialization step we consider one element in  $\mathcal{X}$  that is the optimal solution,  $(\tau^1, \alpha^1)$ , of  $LR$  and concerned to set  $\mathcal{C}$  we have considered two strategies:  $DL1$ - only one initial element in  $\mathcal{C}$ , the vector  $\alpha'^1$  obtained from  $\alpha^1$  by using a rounding scheme;  $DL2$ -  $m+1$  initial elements in  $\mathcal{C}$ : the vector  $\alpha'^1$  as previously and the  $m$  canonical basis vector of  $\mathbb{R}^m$ .

## 4 Computational Results

In this section, we present some computational results. To solve SDP problems we used the Brian Borchers's CSDP4.9 package. We adapted Pisinger's expknap.c program <sup>3</sup> to solve the Knapsack 0-1 problems with real coefficients. To solve the LP problems we used GLPK4.8 <sup>4</sup>.

Relaxation  $LR$  and  $Rc$ , being a SDP problem, are solved directly applying the CSDP package. Relaxation  $Rc$  is a very large problem compared to  $LR$ . As both relaxations give the same bound we only present results for relaxation  $LR$ . To solve relaxations  $Rlmi$  and  $DL$  we apply cutting plane methods described before. At each iteration, in  $Rlmi$  it is necessary to solve a knapsack problem and a SDP problem, and in  $DL$  it is necessary to solve a SDP problem, a knapsack problem and a LP problem. In Table 1 we compare the sizes of problems solved at each iteration: the number of variables, “ $nv$ ”; the number of blocks, “ $nb$ ”, and its dimensions; the number of constraints, “ $nc$ ”.

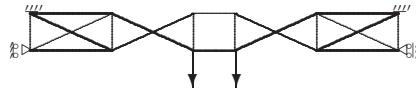
We only present results for two basic trusses ( $2 \times 6$ -truss and  $3 \times 3$ -truss) and some variants of each one. We consider that bars are all made of the same material with *Young's modulus*  $E = 69$  GPa. Figures 2,3 are graphical presentations of the obtained optimal solutions for  $2 \times 6$  nodes structure, considering both topologies. The ground structures has  $n = 18$  degrees of freedom and there are two external loads applied on two nodes as it is depicted by the arrows. In the bottom of each figure is written the optimal value, “ $val(.)$ ”; the volume of the final structure, “ $v$ ”; the cpu-time and the number of iterations, “# it.”. The cross-sectional area of the bars, given by  $a\alpha_i$ , are represented by the thickness of the bars.

<sup>3</sup>The program expknap.c is available in <http://www.diku.dk/~pisinger/codes.html>.

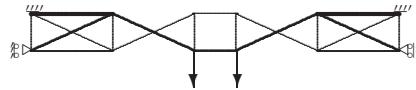
<sup>4</sup>The GLPK (GNU Linear Programming Kit) is available in <http://www.gnu.org/software/glpk/glpk.html>.

Table 1: Size of the solved problems at iteration  $k$  of the cutting plane method.

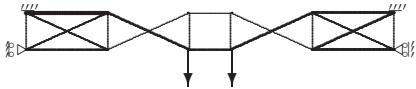
relaxation	SDP			Knap.		LP	
	$nv$	$nb$	block dimensions	$nv$	$nc$	$nv$	$nc$
$Rlmi$	$\frac{n^2}{2} + \frac{3n}{2} + 1$	$m + k + 2$	$\{n + 1, 1, \dots, 1\}$	$m$	1	—	—
$DL1$	$m + 1$	3	$\{n + 1, 1, 2m\}$	$m$	1	$m + 1$	$k + 1$
$DL2$	$m + 1$	3	$\{n + 1, 1, 2m\}$	$m$	1	$m + 1$	$k + 1 + m$



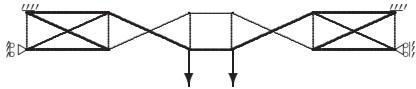
$val(LR) = 43.0412 \quad v = 31.0$   
cpu-time=0.078 sec.

(a) Relaxation  $LR$ .

$val(Rlmi) = 43.7294 \quad v = 30.6$   
cpu-time=17.391 sec. # it.=39

(b) Relaxation  $Rlmi$ .

$val(DL1) = 43.115 \quad v = 31.0$   
cpu-time=4.452 sec. # it.=62

(c) Relaxation  $DL1$ .

$val(DL2) = 43.2142 \quad v = 31.0$   
cpu-time=2.812 sec. # it.=45

(d) Relaxation  $DL2$ .Figure 2: A  $2 \times 6$  structure with  $a = 3$ ,  $v = 31$ ,  $m = 26$  (poor topology).

In Figure 2 the optimal designs are equal for all relaxations, only the cross-sectional area of the bars are different as we can see by its thickness or by  $\alpha$  values in Table 2. As we expected the optimal values for  $Rlmi$ ,  $DL1$  and  $DL2$  are greater than the optimal value of  $LR$ . The quickest relaxation is  $LR$  and the one that takes more time is  $Rlmi$ : using relaxation  $LR$  it is solved a SDP problem with  $72 \times 72$ -matrix and to solve  $Rlmi$  it is necessary to solve 39 SDP problems with dimensions between  $47 \times 47$  to  $85 \times 85$ .

As it was expected, considering rich topology, Figure 3, the optimal value is smaller. Being possible all the links between the nodes, we can get a “stronger” structure than in the poor topology case.

To shorten this section, in the next examples, we only present the optimal structure obtained by solving  $LR$ .

Table 2: Optimal solution  $\alpha^*$  for the structures in Figure 2.

solution	relaxation				solution	relaxation			
	$LR$	$Rlmi$	$DL1$	$DL2$		$LR$	$Rlmi$	$DL1$	$DL2$
$\alpha_1$	0.08	0.19	0.11	0.08	$\alpha_{14}$	0.19	0.10	0.14	0.18
$\alpha_2$	0.05	0.00	0.05	0.05	$\alpha_{15}$	0.00	0.00	0.00	0.00
$\alpha_3$	0.08	0.05	0.07	0.08	$\alpha_{16}$	0.46	0.74	0.53	0.44
$\alpha_4$	0.08	0.05	0.07	0.08	$\alpha_{17}$	0.19	0.39	0.26	0.20
$\alpha_5$	0.05	0.01	0.07	0.07	$\alpha_{18}$	0.21	0.11	0.16	0.20
$\alpha_6$	0.08	0.19	0.09	0.06	$\alpha_{19}$	0.00	0.00	0.00	0.00
$\alpha_7$	0.50	0.21	0.38	0.48	$\alpha_{20}$	0.32	0.40	0.37	0.32
$\alpha_8$	0.00	0.00	0.00	0.00	$\alpha_{21}$	0.34	0.12	0.32	0.37
$\alpha_9$	0.29	0.37	0.34	0.30	$\alpha_{22}$	0.34	0.12	0.26	0.33
$\alpha_{10}$	0.00	0.00	0.00	0.00	$\alpha_{23}$	0.32	0.40	0.37	0.32
$\alpha_{11}$	0.50	0.22	0.43	0.52	$\alpha_{24}$	0.00	0.00	0.00	0.00
$\alpha_{12}$	0.46	0.74	0.58	0.48	$\alpha_{25}$	0.21	0.11	0.16	0.20
$\alpha_{13}$	0.00	0.00	0.00	0.00	$\alpha_{26}$	0.19	0.38	0.21	0.15

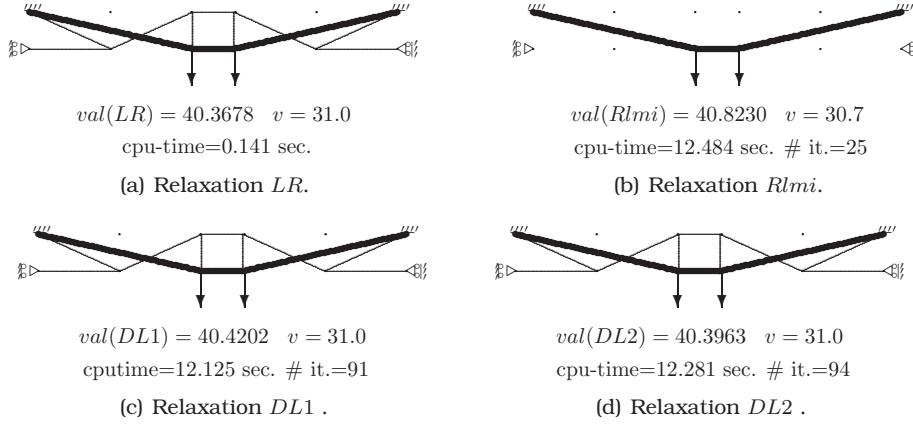
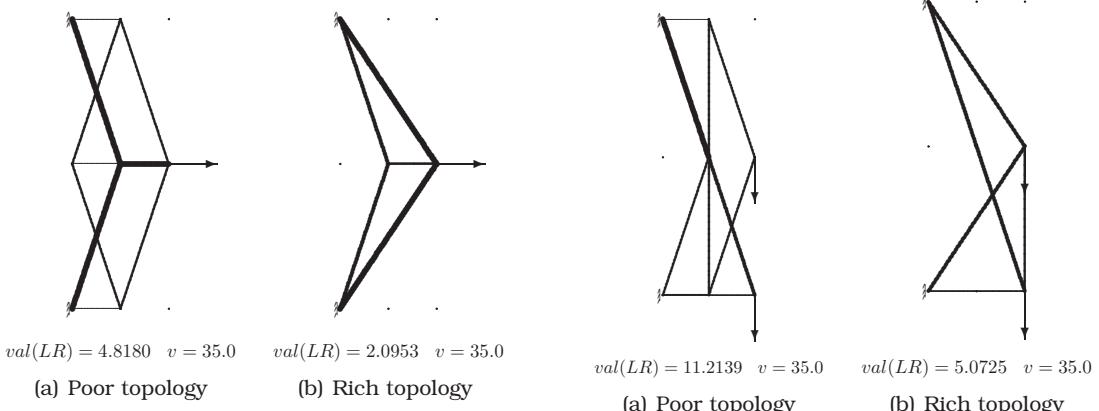
Figure 3: A  $2 \times 6$  structure with  $a = 3$  and  $v = 31$  (rich topology).Figure 4: A  $3 \times 3$ -truss with  $a = 3, v = 35$ .

Figure 5: Figure 4 in another load scenario.

In Figures 4 and 5 are presented the optimal  $3 \times 3$  structures obtained from poor and rich topology, considering two different load patterns. In Table 3 is presented the corresponding data. The structure has 2 fixed nodes and the remaining are free. In the structure of Figure 4 it is applied an external load with 300N and in Figure 5 are applied two external load with 300N each. In this last case, as we expected, the optimal value, *i.e.*, the deformation of the truss, is greater than in the other case.

Comparing the final design we conclude that the optimal structure for a given load scenario can be very different than the optimal structure for another load scenario. Analysing the optimal value we conclude that the rich topology is much more advantageous leading to a stronger structure. In the example of Figure 4  $DL2$  is the relaxation more time expensive.

## 5 Conclusions

We can conclude that concerned to the cpu-time the best relaxation is  $LR$ . Therefore this relaxation is not the best in terms of the obtained bound. In our experiments lagrangean relaxations gives greater bounds than  $LR$  and most of the time  $Rlmi$  gave the greatest one. However, it seems unlikely that those better bounds can overcome the problem of the cpu-time in the context of a branch-and-bound solving an integer problem.

Table 3: Data corresponding to Figure 4 and 5 .

relax.	n = 14; m = 20				n = 14; m = 28				Figure 4
	v	o.f./d.f.	# it.	cpu-time	v	o.f./d.f.	# it.	cpu-time	
LR	35.00	4.8180	—	0.070	35.00	2.0953	—	0.078	
Rlmi	33.68	5.0821	17	4.327	34.12	2.1595	21	5.359	
DL1	35.00	5.0249	139	8.832	33.78	2.1831	206	15.421	
DL2	35.00	5.0201	128	11.055	33.84	2.1817	218	16.203	
LR	35.00	11.2139	—	0.047	35.00	5.0725	—	0.094	Figure 5
Rlmi	34.19	12.5640	28	7.110	32.64	5.4891	15	3.719	
DL1	35.00	11.7167	65	3.156	35.00	5.3778	168	12.547	
DL2	35.00	11.7263	80	3.828	35.00	5.4086	135	9.891	

## 6 References

- Achtziger, W. and Stolpe, M. (2005) Global optimization of truss topology w.r.t. discrete bar areas - Part II: Implementation and numerical results. Report 309, Institute of Applied Mathematics, University of Dortmund, Germany.
- Achtziger, W. and Stolpe, M. (2008) Global optimization of truss topology w.r.t. discrete bar areas - Part I: Theory of relaxed problems. Comput. Optim. Appl., 40(2):247-280.
- Bastos, F. , Cerveira, A. and Gromicho, J. (2005) Using Optimization to Solve Truss Topology Design Problems, Investigaçāo Operacional, 22:123-156.
- Bastos, F. and Cerveira, A. (2007) Semidefinite and Lagrangean relaxations for the truss topology design problem, working paper 12/2008, CIO of Lisbon University, available at <http://cio.fc.ul.pt>
- Ben-Tal, A. and Bendsøe, M.P. (1993) A New Method for Optimal Truss Topology Design, SIAM Journal on Optimization, 3:322-358.
- Ben-Tal, A. and Nemirovski, A. (1995) Optimal design of engineering Structures, Optima 47:4-8.
- Ben-Tal, A. and Nemirovski, A. (2001) Lectures on modern convex optimization. Analysis, algorithms, and engineering applications. SIAM.
- Borchers, B. ( 1999) CSDP 3.2 User's Guide, Optimization Methods and Software, 11:597-611.
- Cerveira, A. (1997) Optimizaçāo do desenho de estruturas, master's thesis, Lisbon University, Portugal.
- Faustino, A., Júdice, J., Ribeiro, I. and Neves, A. (2006), An integer programming model for truss topology optimization, Investigaçāo Operacional, 26:111-127.
- Helmburg, C. (2000) Semidefinite programming for combinatorial optimization, tech. report, Konrad-Zuse-Zentrum für Informationstechnik Berlin.
- Klerk, E. de (2002) Aspects of Semidefinite Programming - Interior Point Algorithms and Selected Applications, Kluwer Academic Publishers, Norwell, Ma.
- Klerk, E. de, Roos, C. and Terlaky, T. (1995) Semi-definite problems in truss topology optimization, Tech. Report Nr. 95-128, Faculty of Technical Mathematics and Informatics, Delft University of Technology, November.
- Lemaréchal, M. and Oustry, F. (1999) Semidefinite relaxations and lagrangian duality with applications to combinatorial optimization, Rapport de Recherche 3710, INRIA.
- Patriksson, M. and Petersson, J. (1995) A subgradient method for contact structural optimization, LiTH-MAT-R-1995-25.
- Pisinger, D. (1995) An expanding-core algorithm for the exact 0-1 knapsack problem, European Journal of Operational Research, 87:175-187.

# Testing the Mixed Integer Programming approach for the vehicle relocation problem in one-way carsharing

Gonçalo H. A. Correia †

† Assistant Professor

Universidade de Coimbra  
Faculdade de Ciências e Tecnologia  
Departamento de Engenharia Civil  
Rua Luís dos Reis Santos  
3030-788 Coimbra  
gcorreia@dec.uc.pt

---

## Abstract

Carsharing is often wrongly mistaken as being the same as carpooling, when in reality these are two very different systems: carpooling denotes a system by which people share their own vehicles mainly in commuter trips in order for saving money on transport costs and reducing fuel emissions; whereas carsharing works as a near term car rental mainly suited for city centre trips avoiding car ownership and providing the comfort of using a private vehicle for trips between one station to another - one-way carsharing. These different characteristics have motivated two different problem formulations in Mixed Integer Programming (MIP). For carpooling the problem is to select optimum groups out of a list of clients, while for one-way carsharing, the problem has been to determine an optimum strategy for managing a vehicle relocation system which is needed for balancing vehicle stocks at each station during the day. For the carsharing problem it was only recently that a MIP formulation was proposed and used in an International project. In this paper this formulation is reviewed, tested and illustrated with results from its deployment in a highly used Optimization Software using the Boston ZipCar service as case study. We were able to conclude that the formulation is useful and interesting; however for real world instances it may be inefficient. It may be more productive to use it as part of a simulation tool for producing initial operational standards.

**Keywords:** Carsharing, transportation demand management, mixed integer programming

---

## 1 Introduction

The diversity of impacts caused by traffic congestion and parking demand increase in city centers has caused a problem too serious and complex for there to be any possibility of it being mitigated by unilateral, inconsistent or immediate measures. Theory and practice have concluded in the last years that if we want to reach more sustainable configurations of the mobility system there is the need for an interdisciplinary effort that besides including provision of adequate infra-structure (supply management) has necessarily to include Transport Demand Management (TDM) strategies.

The objective of TDM measures is to manage more efficiently the existing mobility requirements without building new infra-structure. It is opposite to the “predict and provide” approach which is not sustainable. The perspective of TDM includes various measures which increase travel choices and encourage consumers to use each option for what it does best. TDM helps create a more balanced, less automobile dependent transport system (Litman, 1999).

Carpooling and carsharing are part of these measures. These two are sometimes confused because the term carsharing is similar to the term “ridesharing”, however carsharing is very different from carpooling, it supposes the existence of a fleet of vehicles in several stations scattered in an urban geographical area, which people can use for near term trips thus avoiding the need to buy a second car. Carpooling on the other hand is a system by which people share their private vehicles with other who have the same, or aligned destinations, being especially suited for commuter trips.

The carsharing system has actually been very successful in several countries (Shaheen, et al., 1999), it allows people to avoid owning an automobile while providing the advantage of allowing to chose the vehicle that is right for a specific trip and saving money on maintenance costs.

These two transportation alternatives raise two different operational problems: For carpooling to be possible there has to be sufficient matching possibilities in order to put people together in the same vehicle (Maniezzo, et al., 2000, Varrentrapp, et al., 2002, Calvo, et al., 2004). Classic carsharing systems do not pose at the outset a particular problem of OR, however as the systems evolved toward more flexible operational models, practitioners understood that in order to increase the systems utilization they had to allow the drivers to chose the station for returning the vehicle - one-way carsharing - which was not being done due to the problem of vehicles stocks unbalance.

This is indeed an OR problem and it has motivated recent research on how to balance vehicle stocks according to the vehicles demand and the available staff for driving the vehicles between stations (Kek, et al., 2009). The question is not reduced only to determine the optimal number of employees needed for that activity, one is interested in knowing the priority of their activities and which replacement strategies to follow.

In the next section the MIP model proposed by Kek, et al. (2009) is reviewed and analyzed in its computational difficulty. In section 3 a numerical example is built and run in the Xpress-MP optimization software following a real word carsharing system configuration, the Bostonian ZipCar Service. The paper finishes with some conclusions about the model and its application in reality.

## **2 A Mixed Integer Programming model for the one-way carsharing vehicle relocation problem**

### **2.1 Defining the problem**

The definition for the one-way carsharing vehicle relocation problem is as follows: “given a set of geographically scattered stations, with each station having a capacity (number of parking stalls) and customer pick-up and return patterns, plus the

maintenance schedule for the vehicle fleet, the objective of the optimization problem is to allocate staff resource and staff activities so as to minimize the generalized cost associated with the vehicle relocation activities." (Kek, et al., 2009).

The activities of each staff member are: waiting (wait at a station for the next activity), maintenance (inspect or clean vehicles at a station), movement (travel between two stations without driving a vehicle) or relocation (drive a vehicle from one station to another).

The problem is defined in a two-dimensional time-space network as shown in Figure 1, with the x-axis representing time and y-axis representing space:

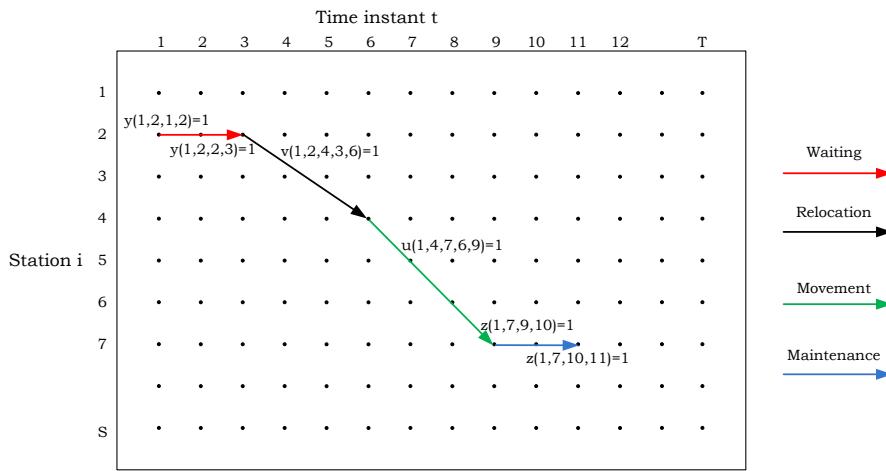


Figure 1: Optimization Space and binary variables example

Defining a set of available staff to carry out these activities as  $L = \{1, \dots, k, \dots, W\}$ , where  $W$  is the maximum number of staff available,  $V = \{1_1, \dots, i_{t-1}, i_t, i_{t+1}, \dots, S_T\}$  denoting all the  $S \times T$  nodes and  $A$  as any arc defined between any of these nodes. The MIP formulation for the problem proposed by Kek (2009) uses the following seven types of decision variables where a matrix of travel time between each station is given by  $t_{ij}$  and a time for a maintenance operation equal to  $t_m$ :

$x^k$ : Binary variable for staff usage, taking value 1 if staff  $k$  is ever used from  $t=1, \dots, T$ , and 0 otherwise,  $\forall k \in L$ ;

$y_{i_t i_{t+1}}^k$ : Binary variable associated with the waiting action, taking the value 1 if staff  $k$  waits at station  $i$  from time steps  $t$  to  $t + 1$ , and 0 otherwise ( $y(k, i, t, t + 1)$  in Figure 1);

$z_{i_t i_{t+t_m}}^k$ : Binary variable associated with the maintenance activity, taking the value 1 if staff  $k$  maintains a vehicle at station  $i$  from time steps  $t$  to  $t + t_m$  and 0 otherwise ( $z(k, i, t, t + t_m)$  in Figure 1);

$u_{i_t j_{t+t_{ij}}}^k$ : Binary variable associated with a staff movement, taking the value 1 if staff  $k$  moves from station  $i$  at time step  $t$  to station  $j$  at time step  $t + t_{ij}$  and 0 otherwise ( $u(k, i, j, t, t + t_{ij})$  in Figure 1);

$v_{i_t j_{t+t_{ij}}}^k$ : Binary variable associated with a vehicle relocation, taking the value 1 if staff  $k$  relocates a vehicle from station  $i$  at time step  $t$  to station  $j$  at time step  $t + t_{ij}$  and 0 otherwise ( $v(k, i, j, t, t + t_{ij})$  in Figure 1);

$d_{i_t}^r$ : Integer variable representing the number of rejected customer demand for vehicles at station  $i$  from time steps  $t - 1$  to  $t$  due to non availability of free parking stalls at the station.

$s_{it}^r$ : Integer variable, representing rejected customer return of vehicles at station  $i$  from time steps  $t - 1$  to  $t$ ;

Two additional dependent variables are:

$r_{it}$ : Number of available vehicles at station  $i$  at time step  $t$ ;

$\bar{r}_{it}$ : Number of unavailable vehicles at station  $i$  at time step  $t$ ;

The known constants are:

$C_{ij}$ : Fixed cost of a movement or relocation trip from stations  $i$  to  $j$ ;

$C_x$ : Fixed cost of utilizing one staff;

$C_d$ : Fixed cost of rejecting the demand of one customer-vehicle trip;

$C_s$ : Fixed cost of rejecting the return of one vehicle by a customer;

$r_{i_0}$ : Number of available vehicles at station  $i$  at time step  $t = 0$ ;

$\bar{r}_{i_0}$ : Number of unavailable vehicles at station  $i$  at time step  $t = 0$ ;

$d_{it}$ : Demand for vehicles at station  $i$  from time steps  $t - 1$  to  $t$ , from historical data;

$s_{it}$ : Number of vehicles returned by customers at station  $i$  from time steps  $t - 1$  to  $t$ , from historical data;

$m_{it}$ : Number of returned vehicles in need of maintenance at station  $i$  from time steps  $t - 1$  to  $t$ ;

$p_i$ : Number of parking stalls at station  $i$ .

$$\text{Min}(Z) = C_{ij} \left( \sum_{(i_t, j_t + t_{ij}) \in A} \sum_{k \in L} u_{i_t j_t + t_{ij}}^k + \sum_{(i_t, j_t + t_{ij}) \in A} \sum_{k \in L} v_{i_t j_t + t_{ij}}^k \right) + C_x \sum_{k \in L} x^k + C_d \sum_{i_t \in V} d_{i_t}^r + C_s \sum_{i_t \in V} s_{i_t}^r$$

This minimizes the total generalized cost, taking into consideration movement and relocation costs, staff cost and penalty costs of rejecting the demand or return of vehicles from customers.

Subject to,

$$\sum_{i \in N} y_{i_1 i_2}^k + \sum_{i \in N} z_{i_1 i_1 + t_m}^k + \sum_{\substack{i, j \in N \\ i \neq j}} u_{i_1 j_1 + t_{ij}}^k + \sum_{\substack{i, j \in N \\ i \neq j}} v_{i_1 j_1 + t_{ij}}^k = x^k \quad \forall k \in L$$

Serves the dual purpose of assigning a non-zero value to  $x^k$  when staff  $k$  is used and restricting staff to only performing one type of extra activity at  $t = 1$ .

$$\begin{aligned} & y_{i_{t-1} i_t}^k + z_{i_{t-t_m} i_t (i_t - t_m, i_t) \in A}^k + \sum_{(j_t - t_{ji}, i_t) \in A} u_{j_t - t_{ji} i_t}^k + \sum_{(j_t - t_{ji}, i_t) \in A} v_{j_t - t_{ji} i_t}^k - y_{i_t i_{t+1}}^k - z_{i_t i_{t+t_m}}^k \\ & - \sum_{(i_t, j_t + t_{ij}) \in A} u_{i_t j_t + t_{ij}}^k - \sum_{(i_t, j_t + t_{ij}) \in A} v_{i_t j_t + t_{ij}}^k = 0 \quad \forall i_t \in V, k \in L, t > 1 \end{aligned}$$

Ensures the conservation of a staff's activities at each node at  $i_t$ . It restricts each staff to only begin exactly one new activity after the previous one is completed.

$$\begin{aligned} r_{i_t} = & r_{i_{t-1}} + \sum_{(j_t - t_{ji}, i_t) \in A} \sum_{k \in L} v_{j_t - t_{ji} i_t}^k - \sum_{(i_t, j_t + t_{ij}) \in A} \sum_{k \in L} v_{i_t j_t + t_{ij}}^k + \sum_{\substack{(i_t - t_m, i_t) \in A \\ k \in L}} z_{i_t - t_m i_t}^k + (s_{i_t} - s_{i_t}^r) \\ & - (d_{i_t} - d_{i_t}^r) - m_{i_t} \quad \forall i_t \in V \end{aligned}$$

Update the number of available vehicles.

$$\bar{r}_{i_t} = \bar{r}_{i_{t-1}} - \sum_{k \in L} z_{i_t i_{t+t_m}}^k + m_{i_t} \quad \forall i_t \in V$$

Update the number of unavailable vehicles.

$$r_{i_t} + \bar{r}_{i_t} \leq p_i \quad \forall i_t \in V$$

Ensures that the sum of available and unavailable vehicles does not exceed the station's capacity at any time step.

$$d_{i_t}^r \leq d_{i_t} \quad \forall i_t \in V$$

Ensures that the rejected demand for vehicles does not exceed the requested demand for vehicles at a station at time step  $t$ .

$$s_{i_t}^r \leq s_{i_t} \quad \forall i_t \in V$$

Ensures that the number of rejected returns of vehicles does not exceed the number of requested returns of vehicles at time step  $t$ .

$$x^k = (0,1) \quad \forall k \in L$$

$$y_{i_t i_{t+1}}^k = (0,1) \quad \forall k \in L$$

$$z_{i_t i_{t+t_m}}^k = (0,1) \quad \forall k \in L$$

$$u_{i_t j_{t+t_{ij}}}^k = (0,1) \quad \forall k \in L$$

$$v_{i_t j_{t+t_{ij}}}^k = (0,1) \quad \forall k \in L$$

Impose binary conditions on these variables

$$d_{i_t}^r, s_{i_t}^r, r_{i_t}^r, \bar{r}_{i_t}^r \geq 0 \quad \forall i_t \in V$$

Impose non-negativity conditions on these variables.

## 2.1 Testing the problem complexity using the Bostonian ZipCar service as case-study

The four binary sets of variables:  $y_{i_t i_{t+1}}^k$ ,  $z_{i_t i_{t+t_m}}^k$ ,  $u_{i_t j_{t+t_{ij}}}^k$  and  $v_{i_t j_{t+t_{ij}}}^k$  represent the greatest burden for solving this MIP problem. The number of binary variables needed for solving a problem instance depends on the problem characteristics. Starting with variables  $y$  and  $z$  we may compute the total number of variables needed for the primal problem as:

$2 \times W \times S \times (T^2 - \sum_{t=1}^T t)$ , because there is no sense on having time steps  $t$  to  $t-1$  one is able to subtract  $\sum_{t=1}^T t$  to  $T^2$ .

As for variables  $u$  and  $v$  we have the following scale:

$2 \times W \times (S^2 - S) \times (T^2 - \sum_{t=1}^T t)$ , because there is no sense on having movements or relocations between the same station one is able to subtract  $S$  to  $S^2$ .

Thus, the total number of binary variables needed for solving the problem is:  $2 \times W \times S^2 \times (T^2 - \sum_{t=1}^T t)$ .

In order to test a typical real problem dimension we use the case of ZipCar, a Northern American carsharing company, for the city of Boston. This is a classical carsharing system (mandatory vehicle return on the same station) with 840 vehicles distributed over 250 stations, representing an average of about 3 vehicles per station.

If this system started working in one-way mode the request and return of vehicles considering 5 min intervals during one whole day, we have to define  $(24 \times 60)/5 = 288$  time instances for solving the relocation problem. Considering a maximum of 50 staff for balancing the system, the number of binary variables needed would be:

$$2 \times 50 \times 250^2 \times (288^2 - \sum_{t=1}^{288} t) = 2.583 \times 10^{11}$$

This is a great number of binary variables even for fast processors, thus showing the burden of solving this problem for a real word instance.

For simplifying the problem one may act on different dimensions: the day can be divided in 3 shifts of 8 hours each, allowing focusing on just one of those shifts. If this is divided in 15 min intervals rather 5 min, one obtains: T=32 time instants. One may act also on the number of stations, S, by selecting groups of the whole set, this entails that the staff in each optimization run will only perform activities between those stations, thus preventing relocations between groups of stations, which makes the choice of how to group them important. We will simplify and consider that one is able to divide the 250 stations of the Boston ZipCar Service in 25 groups of 10 stations. This would lead to an average of  $50/25 = 2$  staff, however one should allow the optimization to search for this best optimum number, hence a  $W = 5$  will be considered.

Thus the total number of binary variables is the following:

$$2 \times 5 \times 10^2 \times (32^2 - \sum_{t=1}^{32} t) = 496000$$

At the expense of these simplifications one gets a much reduced MIP problem. In the following section one problem of such dimension will be tested for the Bostonian ZipCar service.

### 3. Numerical Experiment

The relocation problem was modeled in the software Xpress-MP, allowing inputting any instance specified according to the formulation reviewed in Section 2. For testing it we take again the Boston Zipcar service as example and select 10 sets of stations for searching the best staff dimension and their activities accordingly to the problem reduction measures presented before. Some of the stations in Boston are actually very near to each other; hence it is more realistic to group some of them together in just one station with different parking areas, summing their capacities. The number of parking stalls was considered to be 40% higher than the number of available vehicles at the station, allowing some extra space for vehicle returning.

The time distances between the several groups are never greater than 1 quarter of hour, the time unit which will be used, thus simplifying the problem with a time-distance matrix just with 1's and 0's in the diagonal. The cost of each 15 min travel time will be considered as 2 euros, which produces a one hour trip of 8 euros (similar to the cost of ZipCar in dollars). The cost of utilizing one staff is considered equal to a minimum wage, 500 euros for simplification. The cost of rejecting a demand for a vehicle is considered to be the loss of one hour trip during 22 working days, about 170 euros, trying to translate the bad image that the client will have about the system during some time. The cost of not being able to return one vehicle is about 25% of the demand failure.

As for the demand and return historical data, having no better information, and considering the shift of 8 hours between 8:15 (t=1) and 16:00 (t=32), we consider a peak period in the morning and during lunch time, varying the demand for vehicles from 20% to 40% of the number of existing vehicles at each station. The number of returned vehicles is considered to be 50% of the corresponding demand, which implies that some of

the vehicles taken from the area of analysis will not return during this shift. For the number of returned vehicles in need of maintenance a 10% probability of that event is used, drawing a random number for each vehicle returned, there will be a 90% probability of it being operational and ready for being used on a new trip.

Running the problem, first for  $W = 5$  (maximum number of staff) in Xpress-MP, the Branch-and-Bound results were as follows:

	Solution#	Objective	BestBound	Gap	Time (s.) (Duo CPU @ 2.00GHz)	$\sum_{k=1}^W x^k$
W=5	7	27420	26474.67	3.45%	215	2
	8	27196	26474.67	2.65%	216	2
	9	27040	26475.6	2.09%	395	2
	10	26706	26476.23	0.86%	655	2
	11	26694	26476.31	0.82%	963	2
	12	26692	26476.31	0.81%	964	2
W=4	8	27430	26467.94	3.51%	144	2
	9	26868	26467.94	1.49%	151	2
	10	26698	26468.39	0.86%	405	2
	11	26696	26468.4	0.85%	559	2
W=3	1	27034	26466.87	2.10%	91	2
	2	27032	26466.87	2.09%	91	2
	3	26862	26466.92	1.47%	153	2
	4	26706	26467.01	0.89%	202	2
	5	26704	26467.65	0.89%	444	2
	6	26700	26467.65	0.87%	492	2
W=2	1	26866	26478.88	1.44%	34	2
	2	26700	26479.01	0.83%	79	2
	3	26694	26480.8	0.80%	145	2
	4	26692	26480.97	0.79%	232	2

We may observe that the number of staff does not vary much between the solutions and as we reduce the number of maximum staff  $W$  to match the results of the previous solutions, one gets good solutions much faster, as we may observe comparing solution 12 for  $W = 5$  and solution 4 for  $W = 2$ .

By reducing the maximum number of staff, one reduces the number of binary variables allowing Branch-and-Bound to search earlier for a better distribution of their activities. The resulting distribution of activities for the best solution can be seen in Figure 3.

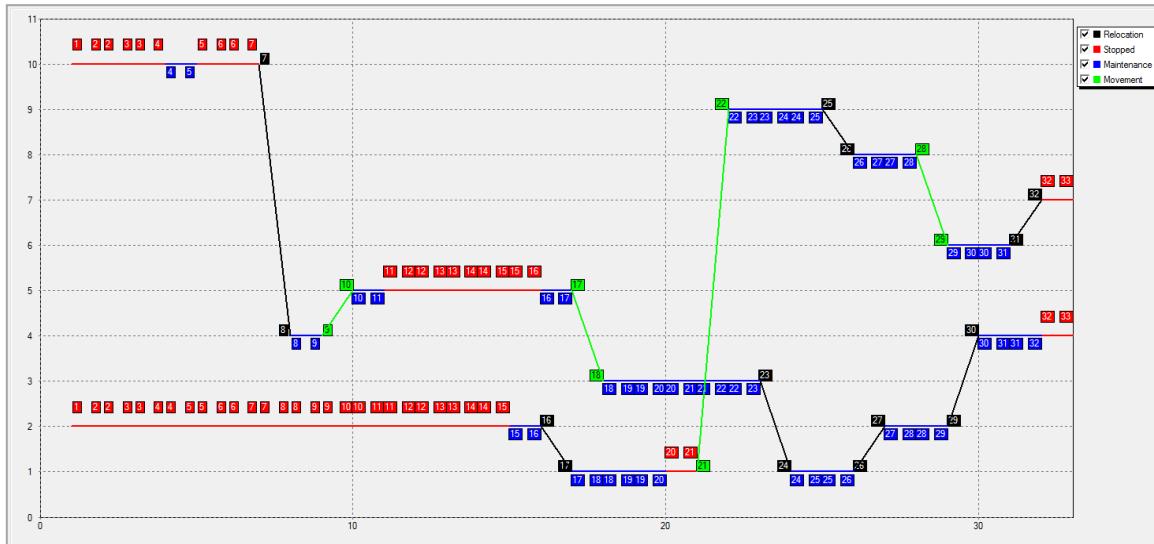


Figure 3: Xpress-MP graphical output of the optimization space

It is interesting to notice that for the initial instances the staff is mostly idled, until maintenance activities are needed and the number of available vehicles is not enough to meet the demand and the staff engages in some relocations.

## **4 Conclusions**

The MIP approach for the vehicle relocation problem in one-way carsharing was reviewed and tested in this paper. We started by observing the great need for a large set of binary variables for running the model, and then, for simplification purposes, some reduction measures were presented, allowing producing good results in acceptable time when applying the MIP problem to the Boston ZipCar service case-study.

The model produced an optimized staff number and the distributions of their activities, enabling to conclude on the periods of greater activity and the most performed activities, thus the simplifications allow getting interesting real time results. Nevertheless they may introduce some limitations to the adequacy of the model to the real system, since the need to divide the stations in groups subverts the reality. We may conclude that the MIP problem can be useful as a source for establishing operational standards which could then be tested using simulation techniques.

## **5 References**

- Calvo, R. W., Luigi, F. d., Haastrup, P. and Maniezzo, V. (2004) A distributed geographic information system for the daily car pooling problem, *Computers & Operations Research*, Vol. 31, pp. 2263 - 2278.
- Kek, A. G. H., Cheu, R. L., Meng, Q. and Fung, C. H. (2009) A decision support system for vehicle relocation operations in carsharing systems, *Transportation Research Part E-Logistics and Transportation Review*, Vol. 45, No.1, pp. 149-158.
- Litman, T. (1999) Guide to Calculating Transportation Demand Management Benefits, Victoria Transport Policy Institute.
- Maniezzo, V., Carbonaro, A. and Hildmann, H. (2000) An ANTS Heuristic for the Long - Term Car Pooling Problem, Paper presented at the ANTS'2000, 2nd International Conference Ant Colony Optimization, Brussels, Belgium, 8-9 September.
- Shaheen, S. A., Sperling, D. and Wagner, C. (1999) Carsharing and Partnership Management An International Perspective, *Transportation Research Record*, Vol. 1666, pp. 118-124.
- Varrentrapp, K., Maniezzo, V. and Stützle, T. (2002) The Long Term Car Pooling Problem - On the soundness of the problem formulation and proof of NP-completeness, Darmstadt, Technische Universität Darmstadt.

# Risk Management in Petroleum Supply Chain

Leão José Fernandes<sup>†\*</sup>

Ana Paula Barbosa-Póvoa <sup>†</sup>  
Susana Relvas <sup>†</sup>

<sup>†</sup> CEG-IST, Instituto Superior Técnico, \*CLC- Companhia Logística de Combustíveis  
leao.fernandes@clc.pt; apovoa@ist.utl.pt; susanaicr@ist.utl.pt

## Abstract

The petroleum supply chain (PSC) is a sequence of activities that source, transform and deliver petroleum products to end consumers. Due to intensive investment and accounting for its strategic nature, PSCs in general and risk management in particular, offer high potential for optimization and improving the interaction with related activities. In this paper we discuss risk management within PSC. We analyse some of the published literature, identifying aspects that should be accounted for when modelling risk within the PSC.

**Keywords:** Petroleum supply chain, risk management, uncertainty

## 1 Introduction

Starting 1990, the globalization boom has witnessed major developments in Supply Chain Management. Specific characteristics of certain problem types have motivated new strategies, definition of mathematical models whose primary aim is to support the strategic decision-making within the supply chain. However, problems related to operations' risks and that of infrastructures within the supply chains have been less developed, presenting an opportunity for development of new methods and algorithms.

Initial research in this area focussed on methodologies from economics, contributing to Value-Focused Process Engineering (Neiger et al, 2003, 2007), the Value-Focused Thinking (Keeney et al, 1992) and Supply Chain Risk Management (Peck et al, 2006). In recent years, the study of risk management gained a new impetus in the chemical process industry by the appearance of publications by Meel and Seider (2008), surging interest and a niche for research. The Conditional Value at Risk technique (Garcia-Gonzalez et al, 2007) and risks metrics in chemical plants (Applequist et al, 2000) are other examples in the chemical industry. Risk management now swings in as a hot area, where the need for development of optimization models has turned remarkably urgent.

Our research shows three levels of planning for supply chain risk management (SCRM): strategic or long-term above-year planning, tactical or mid-term up to a year planning, and operational or short-term day-to-day planning. Risk sources feature uncertainties in demand, price, supply yields, lead times, supply capacity, supply cost,

new technologies, breakdown of production capacity, uncertain information, terrorism, disaster and financial risks. These risk sources bring management issues requiring appropriate mitigation strategies. Management issues include demand shift across time, markets and products; supplier network design, relationships, selection process, order allocation, and contracts; product differentiation, postponement and process sequencing, logistics, environment health & safety, information, disruption and financial management.

Although we have extensively researched SCRM, we will limit our scope to the petroleum supply chain risk management (PSC-RM) in order to provide focus. The subsequent sections present our study on the PSC risk management literature concluding with directions for research in PSC-RM models and methodologies.

## 2 The Petroleum Supply Chain

The petroleum supply chain divides into two major areas: upstream and downstream. The upstream comprises of crude oil exploration, production and transportation. The downstream industry involves product refining, transport, storage, distribution and retail.

While a vast network of tankers, barges, pipelines, railways and trucks transport crude oil from the oil fields to the refineries, pipelines, trucks, vessels and trains transport refined products to distribution centres and finally to retailers which include oil and gas stations, factories, different business services (such as hotels or hospitals) that may directly consume the product or retail it to the end consumer. Storage facilities including above-ground and underground tanks, underground caverns and offshore storage, are located along transportation routes allowing upsizing or downsizing of batches to match the capacities of the next transportation mode (Figure 1).

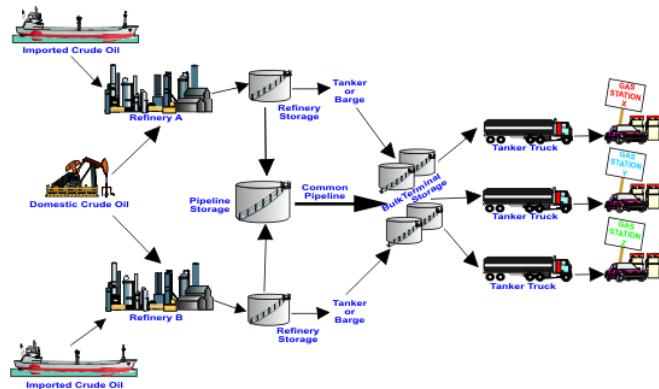


Figure 1: Petroleum distribution network (Source <http://www.eia.doe.gov>)

The complexity associated with such supply chain leads to the existence of different types of risks that need to be accounted for when designing, planning and operating such systems.

## 3 Risk Management literature from the Process Industry

In this section we present some of the risk management literature based on the process industry. Planning activities in these industries are complex and considerably developed. As PSC takes a strategic shift from long-term manufacture to short-term manufacture and from Make to Stock to the Make to Order paradigm, an excellent opportunity lies ahead in extending the existing planning models to incorporate RM, and build new methodologies for integrated PSC-RM. Research has evolved in two directions: early work has been mainly qualitative, while recent efforts have focused on quantitative approaches. Although literature is still scarce for risk management in PSC, we still find

some methodologies and models that could be loosely grouped into four major risk management processes. Consequently we present the literature organized as per the following structure: SC design and planning, risk assessment and measurement, simulation and early warning systems; risk control and management.

### **3.1 SC Design and Planning**

Neiro and Pinto (2004) developed an integrated refinery planning model that includes crude oil and production operations, crude oil unloading and blending, production unit operations and product blending and delivery. The downstream planning involves resource allocation of distribution centres, vehicles, distribution scheduling, pipeline and inventory management, and oil spill risk issues. Kumar and Viswanadham (2007) develop a Decision Support System (DSS) framework using a Case-Based Reasoning approach which assists decision makers in preventive as well as interceptive construction SCRM. They use fault tree analysis and case hierarchy diagrams.

Kuo and Chang (2008) present a mixed-integer linear program (MILP) that devises a planning model to coordinate the refinery production planning with the scheduling decisions. A two-stage stochastic programming (SP) approach is presented in Khor et al. (2008) with optimal midterm refinery planning addressing three sources of uncertainties: market demand, prices, and production yields of crude oil and products. They propose: i) the Markowitz's mean-variance model that minimizes the variance of the expected value; ii) the two-stage SP that minimizes the expected recourse costs; iii) incorporation of the MV model within the SP framework; iv) reformulation of the previous model by adopting mean-absolute deviation as the operational risk measure.

### **3.2 Risk assessment and measurement**

Cowan (1978) differentiates first degree hazards (sources of risk) from second degree hazards (resulting accidents). Later, Vatn (1997) discusses the philosophical aspects of risk acceptance criteria, and the role of statistical decision theory within safety management. Applequist et al. (2000) propose risk premium as a measure of return and risk. The risk premium objective achieves a trade-off of increasing return for increasing risk which seeks a higher return than an equally risky portfolio. The capital-asset pricing model developed in the mid-1960's (Sharpe, 1991) infer that risks, measured by the security to overall market volatility ratio, are rewarded by an increased rate of return. Christensen et al. (2003) salient the implications of REACH (Registration, Evaluation and Authorization of Chemicals), the European Union policy on the chemicals industry.

Coelho et al. (2005) propose risk assessment by way of a connectionist methodology using a feed-forward neural network as an alternative/complementary approach to the traditional Monte Carlo simulation to estimate the total drilling time of deep water wells. Pasman (2008) highlight the role of safety management databases identifying some examples: major hazard incident data system, major accident reporting system, failure and accidents technical information system, fault tree analysis, management oversight and risk tree, layer of protection analysis, hazard barrier and target diagrams and center for chemical process safety guidelines. Abrahamsen and Aven (2008) compare the risk acceptance criteria like fatalities acceptance ratio (FAR) and as low as reasonably practicable (ALARP), with the expected utility theory and the rank-dependent utility theory from decision theory. Fan et al. (2008) find that the generalized autoregressive conditional heteroskedasticity (GARCH) based Value-at-Risk approach is more effective than the autoregressive moving average (ARMA) forecasts for evaluation and forecasting of the risk situation in crude oil price and its spill over effect.

Reniers et al. (2008) alert that chemical industry cluster accidents are known for severity and domino effect. They recall the cluster PEMEX, Unigas and Gasomatico plants

explosions in Mexico in 1984 and the Buncefield oil storage terminal disaster in the United Kingdom in 2005. Cormier and Sutter (2008) integrate four assessments in their framework: i) condition assessments to detect chemical, physical, and biological impairments; ii) causal pathway assessments to determine causes and their sources; iii) predictive assessments to estimate risks and benefits of managerial actions; and iv) outcome assessments to evaluate the results of the decisions.

### **3.3 Simulation and Early warning systems**

Khan et al. (1998) use an accident simulation tool to forecast accidents, expected damage, and devise strategies to mitigate accidents in the chemical industry. Hazard analysis algorithms are proposed in Kang et al. (2001) for: i) deviation analysis, ii) malfunction analysis, and iii) accident analysis algorithm. They use process state variables to identify the fault propagation relationship.

Simulation models and DSS are discussed in Adhitya et al. (2007). They study PSC disruptions like crude oil delays, products out-of-spec, unexpected changes in product distribution, unavailable or constrained plant units, and demand fluctuations. Tian and Tianfield (2006) demonstrate the advantages of multi-agent modelling and simulation in SCM automation. This nested architecture encapsulates the complex system, and uses agents which are autonomous software programs, that perform given functions automatically or semi-automatically by communicating with other agents, programs or human agents. Later Tian et al. (2007) develop a multi-agent system to model the interaction among partners in a PSC. Meel and Seider (2008) present a plant-specific, real-time, risk assessment model using Bayesian analysis with copulas, that includes: i) a forecasting analyzer that predicts the occurrence of abnormal events, ii) a reliability analyzer that predicts the failure probabilities of safety systems, and iii) an accident closeness analyzer.

### **3.4 Risk control and management**

Cruz et al. (2001) suggest emergency response plans, mitigation measures, and design criteria to minimize hazmat release risk and natural disaster damages. Vinnem, (2004) present four consequence barrier functions from the BORA (Barrier and Operational Risk Analysis) project: avoid ignition; reduce release; avoid escalation and prevent fatalities. A two-stage method is presented in Lababidi et al. (2004) where a deterministic model minimizes the total production costs and raw material procurement. At the second stage, a scenario analysis technique controls risks like demands, market prices, raw material costs and production yields fluctuations.

Lavaja et al. (2006) propose a two-stage model using a deterministic MILP formulation and a stochastic model that generates 100 scenarios to consider uncertainty of the product price and demand in different markets, and resolve each scenario individually. Liu et al. (2007) propose risk assessment using Bayesian networks, fuzzy logic and hybrid networks. Risk management processes include process environment risk assessment, predictive risk index, rapid risk assessment, goal programming model, the analytic hierarchy process and the agent-based information flow. Six risks (supply, capacity, disruption, equipment failure, environment, delay risk) and five mitigation strategies (add capacity, redundant suppliers, safe management, increase flexibility, robust planning) are presented. Fernandes (2007) demonstrate the importance of integrated planning and develop a database for PSC contingency planning. Fernandes and Saldanha da Gama (2008) salient a shift from contingency planning to resilience building.

You et al. (2009) propose a two-stage linear stochastic programming approach within a multi-period planning horizon that takes into account production and inventory levels,

transport modes, shipment times and customer service levels under demand and freight rate uncertainty. They devise a simulation framework with a 12 month rolling horizon that uses a deterministic model and a stochastic model. They test four popular risk measures, namely the variance, the variability index, the probabilistic financial risk and the downside risk.

Although risk management has prime importance in the process industry, research in this area has been limited, especially in the development of quantitative models. This lack of models to control risks in the PSC, offers a tremendous opportunity for robust methodologies that could model and control the industry specific as well as overall risks.

## 4 Opportunities in PSC-RM modelling

The presented literature provides an overview of the modelling techniques in PSC-RM that address several unforeseen or uncertain situations along the chain. Company damage can come from negative impacts on reputation, operations, finance, safety and environmental sources. Based on our research in petroleum companies, we list typical PSC disruption scenarios and associated mitigation strategies in Table 1.

Table 1. Examples of Contingencies in the Petroleum Supply Chain

Disruption	Description	Mitigation strategy
Supply side	Product shortage can lead to lower profits due to unmatched orders	Use contingency plans including inter-company product exchanges. Optimize replenishment schedules to prioritize stocking of the product in shortage. Use strategic reserve capacity.
Supply side	Product out of specification can block over 25 million cubic meters of storage capacity per tank	Order immediate suspension of reception of contaminated batch. Expedite stock to industries with lower product specification needs. Supply demand from alternate facilities. Meet with clients and suppliers to initiate contingency measures.
Exceptional demand	Excessive demand due to increased air traffic or due to harsh winter	Optimize replenishment schedules. Optimize production shifts. Activate alternative supplies.
Transportation	Pipeline rupture can lead to the SC's prolonged inoperation	Supply demand from alternate facilities. Switch to alternate mode of transport. Design contingency plan with suppliers and clients
Facilities	Safeguard system out of order leading to unsafe operation	Switch to manual mode. Activate strict manual checks. Initiate immediate repair. Evaluate potential dangers and alternative revamping.
Facilities	Energy breakdown can disrupt production refining and pipelines.	Import refined product, transport by railway wagons, product exchange, postponement, shift production and insurance.
Information Systems	Failure of information systems leading to unavailability of customer orders	Activation of redundant systems. Identify problem area and action maintenance contract. Use sourcing information for alternate system in worst-case scenario.
Communications	Communication failure due to fiber optic interruption	Activate backup by telephone, mobile or satellite communications. Localize interruption and initiate immediate maintenance.

The ongoing study identifies various techniques and methodologies used for risk modelling. These include qualitative, case study, deterministic, stochastic, simulation, artificial intelligence and hybrid modelling. New realms are identified for further research. Several issues present research opportunities and hence should be studied in greater detail:

- Study of risk management issues affecting PSC modelling.
- Development of an appropriate metric for risk exposure and risk performance.
- Development of models and procedures to relate decision variables to performance measures.
- Linking downside risk measures with organizational strategy, structure, process, and environment variables.
- Study of disruption risks and their effect on the PSC partners' profitability.
- Incorporation of traditional risk analysis theory into PSC-RM.
- Development of realistic inventory, transportation and production cost functions.
- Simulating systems dynamics to analyze PSC configuration and inventory policies.
- Simulation of PSC vulnerability scenarios, assess potential for disruption, and strategies for response and recovery.
- Study of pricing decisions with RM.

We propose further investigation by projecting models and methodologies from related fields to construct or enhance risk management models for the PSC. Selected models will be case tested with the PSC in order to gain more insights. Specifically, we consider:

- The two-stage linear stochastic programming approach from You et al. (2009), based on a simulation framework with the deterministic and the stochastic model to manage production and inventory, transport modes, shipment times and customer service levels.
- The simulation models from Adhitya et al. (2007) for risk management strategies specially related to disruption management in PSCs.
- The risk assessment methods using Bayesian analysis with copulas from Meel and Seider (2008), that extend a real-time, dynamic model used to track the safety system's performance for plant-wide, real-time, risk assessment.
- Modelling approach from Grossman et al. (2005) for enterprise-wide optimization which involves optimizing of operations, supply, production and distribution activities for the process industry.
- The multi-objective SC design from Guillen et al. (2005) that study SC performance and financial risk, assessed by the profit over the time horizon, and the resulting demand satisfaction.
- The economic theory risks, capture and handling of the dynamics, design-planning and financial formulations of entire SCs as in Puigjaner et al. (2008) using model predictive control methodology and stochastic optimization.
- Methodologies from Srinivassan (2007) using artificial intelligence: software agents, pattern recognition and expert systems that react quickly and effectively to the changing markets and other exogenous factors.
- Research from Reklaitis et al (2004) using agent-based SC and plant design for discrete choices, dynamics, and parameter uncertainty, using process and logistical decision modelling and stochastic methods.

## **5 Conclusions**

This paper presents the problem of risk management in the PSC. The entire supply chain is summarily characterized in terms of equipment and processes. Current literature on risk management in the process industry is presented and analysed. Future research comprises of case testing the identified models, and developing quantitative methodologies from other supply chains for the PSC-RM.

Focusing on the Portuguese situation, the major petroleum companies, GalpEnergia, BP and Repsol manage storage depots linked by transportation facilities. The suppliers, companies, clients, systems and processes form a close fit PSC. A single member disruption can escalate to a crisis situation in a short time, due to the heavy influence of

PSC and petroleum dependence on its economy. In conclusion, it is extremely important to address risk modelling within the national PSC, by developing new methodologies with the main objective of reducing the risk associated with our infra-structures. This is the aim of the work in development by the authors.

## 6 Acknowledgments

The authors thank Fundação para Ciência e Tecnologia (FCT) and Companhia Logística de Combustíveis (CLC) for supporting this research.

## 7 References

- Abrahamsen, E.B., Aven, T., (2008). On the consistency of risk acceptance criteria with normative theories for decision-making. *Reliability Engineering and System Safety*, 93, 12, 1906-1910.
- Adhitya, A., Srinivasan, R., Karimi, I.A., (2007). Heuristic rescheduling of crude oil operations to manage abnormal supply chain events. *AIChE Journal*, 53, 2, 397-422.
- Applequist, G.E., Pekny, J.F., Reklaitis, G.V., (2000). Risk and uncertainty in managing chemical manufacturing supply chains. *Comput. Chem. Eng.*, 24, 9, 2211-2222.
- Christensen, F.M., Bruijn, J.H.M., Hansen, B.G., Munn, S.J., Sokull-Klüttgen, B., Pedersen, F., (2003). Assessment tools under the new European union chemicals policy. *Grenleaf Publishing*, GMI 41, Spring 2003, 5-19.
- Coelho, D.K., Roisenberg, M., Freitas Filho, P.J., Jacinto, C.M.C., (2005). Risk assessment of drilling and completion operations in petroleum wells using a Monte Carlo and a neural network approach. *IEEE Xplore, Proceedings 2005 Winter Simulation Conference*, 1892-1897.
- Cormier, S.M., Sutter II, G.W., (2008). A framework for fully integrating environmental assessment. *Environmental Management*, 42, 543-556.
- Cowan, J.M., (1978). Hazards in the chemical industry - Risk management and insurance. London Business School.
- Cruz, A.M., Steinberg, L.J., Luna, R., (2001). Identifying hurricane-induced hazardous material release scenarios in a petroleum refinery. *Natural Hazards Review*, 2, 4, 203-210.
- Fan, Y., Zhang, Y., Tsai, H., Wei, Y., (2008). Estimating 'Value at Risk' of crude oil price and its spillover effect using the GED-GARCH approach. *Energy Economics*, 30, 3156-3171.
- Fernandes, L.J. (2007). Um modelo de planeamento preventivo correctivo e de contingência em Sistemas de Informação – Aplicação a um parque de combustíveis, MSc Thesis, DEIO, FCUL, University of Lisbon.
- Fernandes, L.J., Saldanha da Gama, F. (2008). Contingency planning – a literature review, in Goyal, D.P., Varma, S., (Eds.), *Supply Chain Management and Competitiveness*, Macmillan, India, 457-467.
- Garcia-Gonzalez, J., Parrilla, E., Mateo, A., (2007). Risk-averse profit-based optimal scheduling of a hydro-chain in the day-ahead electricity market, *EJOR*, 181, 3, 1354-1369.
- Grossmann, I. E. (2005). Enterprise-wide optimization: A new frontier in process systems engineering. *AIChE Journal*, 51, 1846-1857.
- Guillen, G., Mele, E., Bagajewicz, M.J., Espuna, A., Puigjaner, L., (2005). Multiobjective supply chain design under uncertainty. *Chem. Eng. Sci.*, 60, 6, 1535-1553.
- Kang, B., Yoon, E.S., Suh, J.C., (2001). Application of automated hazard analysis by new multiple process-representation models to chemical plants. *Ind. Eng. Chem. Res.*, 40, 1891.
- Keeney, R.L., (1992). Value-focused thinking: a path to creative decision making. Harvard University Press, Cambridge.
- Khan, F.I., Rani, J.D., Abbasi, S.A., (1998). Accident simulation as a tool for assessing and controlling environmental risks in chemical process industries: A case study. *Korean Journal of Chemical Engineering*, 15, 2, 124-135.

- Khor, C.S., Elkamel, A., Ponnambalam, K., Douglas, P.L., (2008). Two-stage stochastic programming with fixed recourse via scenario planning with economic and operational risk management for petroleum refinery planning under uncertainty. *Chem. Eng. Proc.*, 47, 1744.
- Kumar, V., Viswanadham, N., (2007). A CBR-based decision support system framework for construction supply chain risk management. *Proceedings of the 3rd annual IEEE conference on automation science and engineering*, 980-985.
- Kuo, T., Chang, C., (2008). Application of a mathematic programming model for integrated planning and scheduling of petroleum supply networks. *Ind. Eng. Chem. Res.*, 47, 1935-1954.
- Lababidi, H.M.S., Ahmed, M.A., Alatiqi, I.M., Al-Enzi, A.F., (2004). Optimizing the supply chain of a petrochemical company under uncertain operating and economic conditions. *Ind. Eng. Chem. Res.*, 43, 63-73.
- Lavaja, J., Adler, A., Jones, J., Pham, T., Smart, K., Splinter, D., Steele, M., Bagajewicz, M.J., (2006). Financial risk management for investment planning of new commodities considering plant location and budgeting. *Ind. Eng. Chem. Res.*, 45, 7582-7591.
- Liu, L., Ji, J., Yu, H., Hu, J., Fan, T., (2007). Supply chain risk management in Chinese process industries. *IEEE Xplore, WiCom 2007 International Conference on Wireless Communications Networking and Mobile Computing*, 4923 - 4926.
- Meel, A., Seider, W.D., (2008). Real-time risk analysis of safety systems. *Comput. Chem. Eng.* 32, 827-840.
- Neiger, D., Rotaru, K., Churilov, L., (2007). Supply chain risk identification with Value-Focused Process Engineering. *J. Operations Management*, 27, 2, 154-168.
- Neiro, S.M.S., Pinto, J.M., (2004). A general modeling framework for the operational planning of petroleum supply chains. *Comput. Chem. Eng.*, 28, 871-896.
- Pasman, H.J., (2008). Learning from the past and knowledge management: Are we making progress?. *Journal of loss prevention in the process industries*, In press.
- Peck, H.H., (2006). Reconciling supply chain vulnerability, risk and supply chain management. *International Journal of Logistics: Research and Applications*, 9, 2, 127-142.
- Puigjaner, L., (2008). Capturing dynamics in integrated supply chain management. *Comput. Chem. Eng.*, 32, 11, 2582-2605.
- Reklaitis, G.V., (2004). Agent-based supply chain management parts 1 and 2. *Comput. Chem. Eng.*, 28, 9, 1547-1548.
- Reniers, G.L.L., Ale, B.J.M., Dullaert, W., Soudan, K., (2008). Designing continuous safety improvement within chemical industrial areas. *Safety Science*, 47, 5, 578-590.
- Sharpe, W.F., (1991). Capital asset prices with and without negative holdings. *Journal of Finance, American Finance Association*, 46, 2, 489-509.
- Srinivasan, R., (2007). Artificial intelligence methodologies for agile refining: an overview. *Knowledge and Information Systems*, 12, 2, 129-145.
- Tian, J., Tianfield, H., (2006). Multi-agent modelling and simulation for petroleum supply chain. *ICIC 2006, LNAI 4114*, 496-501.
- Tian, J., Tianfield, H., Chen, J., He, G., (2007). A Study upon the architectures of multi-agent systems for petroleum supply chain. Y.Luo (Ed.): *CDVE 2007, LNCS 4674*, 301-303.
- Vatn, J., (1997). A discussion of the acceptable risk problem. Department of production and quality engineering, Norwegian University of Science and Technology, Trondheim, Norway.
- Vinnem, J.E., Aven, T., Hauge, S., Seljelid, J., Veire, G., (2004). Integrated barrier analysis in operational risk assessment in offshore petroleum operations. *BORA (Barrier and Operational Risk Analysis) project*, Stavanger University College, Norway.
- You, F., Wassick, J.M., Grossmann, I.E., (2009). Risk Management for a global supply chain planning under uncertainty: Models and Algorithms. *AIChE Journal*, 55, 4, 931-946.

# Gestão de operações logísticas num armazém usando modelos de apoio à decisão

Carla Alexandra Soares Geraldes \* ‡      Maria do Sameiro Carvalho † ‡  
Guilherme A. B. Pereira † ‡

\* Instituto Politécnico de Bragança – Escola Superior de Tecnologia e Gestão  
Departamento de Gestão Industrial  
[carlag@ipb.pt](mailto:carlag@ipb.pt)

† Universidade do Minho – Escola de Engenharia  
Departamento de Produção e Sistemas  
[{sameiro|gui}@dps.uminho.pt](mailto:{sameiro|gui}@dps.uminho.pt)

‡ Centro Algoritmi – Universidade do Minho

## Abstract

Today's competitive and volatile market demands flexibility, quality and efficiency from the logistics operations. In this context, warehouses are an important link of the logistic chain and warehouse management plays a determinant role over customer's service. Throughout this work we support some warehouse management decisions with mathematical models. These models allowed us to redesign a particular warehouse and implement a new storage policy. The output of these models may also be taken as a starting point for further detailed warehouse operations decisions, especially as far as the optimization of the picking process is concerned.

## Resumo

Os armazéns são elos importantes da cadeia logística, pelo que, no contexto de um mercado global, altamente competitivo e exigente, a sua gestão cumpre um papel muito relevante no que diz respeito quer à obtenção dos níveis de serviço prestados ao cliente quer à minimização dos custos logísticos totais. Neste trabalho é realizada a reestruturação de operações logísticas num armazém. A reestruturação realizada teve por base a implementação dos resultados obtidos por resolução de dois modelos de programação matemática que permitiram redimensionar as áreas funcionais no armazém e localizar os produtos dentro dessas áreas, segundo a política de localização adoptada. As soluções obtidas permitiram não só reestruturar o armazém como abriram lugar ao estudo e implementação de novas estratégias, para posterior optimização do processo de *picking*.

**Keywords:** Logistics operations, Warehousing, Mathematical programming, Case study

**Title:** Decision support models on warehouse logistics operations

## 1 Introdução

O planeamento e concepção das operações num armazém é um problema complexo pois inclui um elevado número de decisões inter-relacionadas de *processos*, *recursos* e de *organização*. Nessas decisões, para além da definição da sua localização geográfica e da determinação da sua dimensão, inclui-se a definição do *layout*, a selecção de equipamentos e tecnologias, a definição de zonas funcionais, a localização dos produtos, a geração de listas de *picking*, a definição de rotas de *picking*, etc. O objectivo a atingir é maximizar a utilização os recursos disponíveis garantindo os requisitos dos clientes.

Rouwenhorst *et al.* (2000) classificou as decisões a tomar em *estratégicas*, *tácticas* e *operacionais*. As decisões de carácter estratégico são decisões consideradas de longo prazo. Neste grupo de decisões podemos incluir a determinação das principais políticas e planos para utilizar os recursos da empresa que melhor suportam a estratégia de competitividade de longo prazo adoptada. As decisões consideradas tácticas têm como objectivo atingir a máxima eficiência em termos de materiais e de mão-de-obra, de acordo com as restrições decorrentes das decisões estratégicas adoptadas. A este nível, as relações entre as várias decisões a tomar são menores se comparadas com as existentes ao nível estratégico. No entanto, as decisões tomadas a este nível têm um enorme impacto nos restantes problemas operacionais por solucionar. Ou seja, este conjunto de decisões traduz as limitações sobre as quais serão tomadas as decisões consideradas de nível operacional. Ao nível operacional as decisões envolvidas dizem respeito aos problemas de atribuição e controlo de pessoas e equipamentos. Atendendo a que as relações existentes entre os diferentes processos são, tipicamente, tratadas nos níveis estratégico e táctico podemos afirmar que, a este nível, as políticas adoptadas interagem pouco podendo, então, ser analisadas independentemente. Berg e Zijm (1999) debateram e apresentaram uma classificação hierárquica para os problemas operacionais nos armazéns. Adicionalmente, estes autores relembraram modelos de programação matemática que suportam algumas das decisões a tomar em cada um dos níveis considerados. Hassan (2002) apresentou uma estrutura de orientação para a concepção do *layout* de um armazém e para a definição dos processos que lhe estão associados. As linhas de orientação apresentadas têm em conta diversos factores e diversas operações sendo o objectivo principal a obtenção de um *layout* com as seguintes características: modular, adaptável, compacto, acessível, flexível, com utilização óptima do espaço disponível e onde os movimentos e os congestionamentos sejam reduzidos ao mínimo. Gray *et al.* (1992) descreveram e modelaram os problemas de concepção e de operação de um armazém típico de distribuição. Esta abordagem incluiu a definição do *layout*, a selecção de equipamentos e tecnologias, a localização dos produtos, a definição de zonas e rotas de *picking*, bem como a definição das listas de *picking* e do lote de encomendas a considerar. Dada a complexidade do problema considerado, os autores desenvolveram uma abordagem hierárquica multi-estágio que considera uma sequência de modelos matemáticos e as interacções entre eles. Essa abordagem caracteriza-se pela coordenação dos modelos matemáticos que incorporam as trocas económicas mais importantes, de forma a fazer recuar a decisão sobre um número reduzido de alternativas. A técnica de simulação foi também utilizada para que, de uma forma integrada, os resultados analíticos obtidos fossem validados. Heragu (2005) abordou a questão do dimensionamento das áreas do armazém juntamente com a atribuição dos produtos a essas áreas. Trata-se de um modelo de alto nível que, em simultâneo, permite determinar a dimensão de cada área funcional bem como o tipo de fluxo para cada produto minimizando os custos totais de movimentação e de armazenamento. Relativamente às políticas operacionais são particularmente relevantes as políticas de localização. Estas políticas são “um conjunto de regras que determinam a localização no armazém das unidades de carga dos diferentes produtos” (Goetschalckx *et al.*, 1990). Liu (1999) propôs um modelo matemático de programação binária que permite agrupar os produtos num armazém em *clusters*.

Considerando as implementações práticas, é bastante frequente a utilização de uma abordagem sequencial dos vários aspectos envolvidos aquando da concepção de um armazém. Tipicamente, há que decidir quais as necessidades de armazenamento e de *layout* para, posteriormente, determinar quais os sistemas de armazenamento e o equipamento de movimentação necessário para que se possam definir as políticas operacionais a adoptar. No entanto, na concepção de uma instalação óptima há que considerar as relações existentes entre todos os aspectos envolvidos.

Neste trabalho são abordados alguns dos problemas encontrados na reestruturação de um armazém de uma empresa Portuguesa (Geraldes, 2007). Em particular, serão analisados dois aspectos: o redimensionamento das áreas funcionais no armazém e a localização dos produtos dentro das áreas funcionais. Para o redimensionamento das áreas funcionais será construído um modelo de programação inteira mista, adaptado a partir modelo proposto por Heragu (2005), que permite, de forma integrada, encontrar a dimensão óptima de cada área funcional e, simultaneamente, distribuir os produtos pelas áreas consideradas. O dimensionamento e atribuição dos produtos é efectuado minimizando os custos totais de movimentação e de armazenamento. Posteriormente, e partindo dos resultados obtidos por resolução do modelo será também analisada uma nova política de localização baseada na técnica de *clusters*. Com este trabalho pretende-se não só contribuir para a resolução dos problemas com que a empresa se debate, como também analisar o desempenho dos modelos de programação usados, quando submetidos à resolução de problemas reais.

## 2 Caso de estudo

A *Fictícia*<sup>1</sup> é uma empresa Portuguesa localizada em Guimarães que produz e distribui cutelarias e utilidades domésticas. A empresa detém uma posição sólida no mercado contando com cerca de 1200 clientes distribuídos por todo o território nacional (hipermercados (85%), retalho (10%), hotéis (1%) e outros (1%)) e também em Espanha e África (3%). A par dos produtos adquiridos a fornecedores nacionais (25%) a empresa recebe cerca de 20-30% dos seus produtos de fornecedores da China e os restantes são oriundos de fornecedores Europeus. Em média, a *Fictícia* processa cerca de 40 encomendas diárias, com uma média de 20 itens por encomenda. Para o armazenamento dos seus produtos, a empresa dispõe de um edifício com um total de 4000 m<sup>2</sup> e tem arrendado um edifício com 1200 m<sup>2</sup> para armazenamento suplementar que dista 3 km do armazém principal. No armazém principal os produtos encontram-se armazenados em paletes em estantes com 4 níveis em altura. O armazém principal encontra-se dividido em duas áreas funcionais: armazenamento e *picking*. Diariamente, uma equipa de 15 funcionários efectua as operações de recepção, armazenamento, *picking*, consolidação e expedição das encomendas. As operações de recepção e armazenamento são, normalmente, realizadas durante o período da manhã. Durante o período da tarde, enquanto decorre a operação de *picking*, as encomendas processadas são verificadas, embaladas e expedidas. O transporte e entrega das encomendas estão a cargo de um operador logístico.

Consequência dos elevados custos operacionais e do baixo nível de produtividade verificada, a empresa enfrenta grandes dificuldades em se manter competitiva. Adicionalmente, a empresa depara-se com uma diminuição acentuada do nível de serviço aos clientes. Esta diminuição do nível de serviço prestado é resultado de i) processamento de encomendas com tempos de ciclo longos; ii) constantes alterações no padrão das encomendas; iii) erros no processamento das encomendas (p.e. resultado de rupturas de stock). Finalmente, a empresa depara-se com um problema de falta de espaço de armazenamento consequência da enorme variedade de produtos em armazém e dos elevados níveis de inventário necessários para fazer face aos longos tempos de entrega de alguns fornecedores (chegam a ser de 4 meses).

---

<sup>1</sup> Por questões de confidencialidade é omissa a denominação da empresa.

A escassez de área de armazenagem e a ineficiência das operações logísticas são, em parte, consequência da quantidade de produtos obsoletos e da existência de vários produtos de baixa rotação que vão sendo colocados ao longo dos corredores impedindo o fácil acesso às posições de stock. A realização de uma análise ABC permitiu identificar que do total de referências existentes apenas 43% haviam sido processadas e que 8 dos 1200 clientes são responsáveis por 70% do valor facturado.

Do tempo necessário para realizar as operações logísticas, 10% é gasto na procura dos produtos ao longo do armazém e cerca de 60% é ocupado com o processamento da encomenda. Os valores registados são consequência de um elevado nível de desorganização, da ausência de rotas de *picking*, da existência de um *layout* pouco eficiente e da falta de um sistema de informação integrado. Adicionalmente, o fraco planeamento das operações em armazém e a incapacidade de estabelecer previsões estão na origem da adopção de políticas erradas para a gestão dos inventários. Verificou-se, também, a ausência de codificação das posições em stock, a localização de produtos de alta rotação em locais de difícil acesso, existência de artigos em armazém não recepcionados e a inexistência de zonas funcionais bem definidas.

No caso em estudo foram considerados os dois modelos a seguir apresentados que, aplicados de forma sequencial, suportaram decisões tomadas aquando da reestruturação realizada.

### 3 Modelos de apoio à decisão

#### 3.1 Dimensionamento das áreas funcionais e identificação dos fluxos dos produtos no armazém

O primeiro modelo considerado (Geraldes, 2008b), adaptado de Heragu (2005), permitiu redimensionar as diferentes áreas funcionais no armazém, bem como identificar o fluxo de cada produto no armazém. Consideremos então as seguintes áreas funcionais: recepção, expedição, reserva e principal, e os possíveis fluxos em armazém (Figura 1).

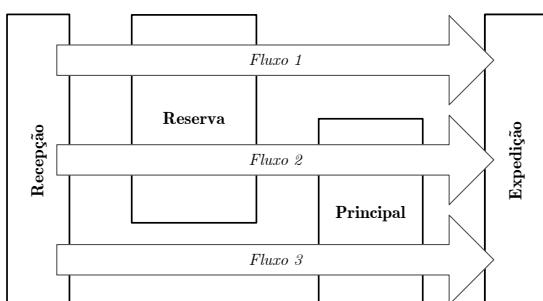


Figura 1: Fluxos de material no armazém.

Além do local de recepção e de expedição que se situa, respectivamente, junto ao cais de entrada e de saída das mercadorias, consideram-se duas áreas distintas de armazenamento – área de reserva e a área principal ou de *picking*. Na área de reserva serão localizados os artigos de menor rotação bem como o stock de reabastecimento dos artigos colocados na área principal. Na área principal encontraremos os produtos cuja rotação é maior. É nesta área que se realiza a operação de *picking* na sua quase totalidade.

O Fluxo 1 diz respeito aos produtos armazenados apenas na área de reserva. Como tal, a operação de *picking* é realizada nesta área conforme necessário. O Fluxo 2 representa o fluxo típico em armazém: os produtos são armazenados na área de reserva, sendo que uma parte da carga é movimentada para a área principal onde se realiza a

operação de *picking*. Finalmente, o Fluxo 3 representa os produtos que são enviados directamente para a área principal.

Considere-se então a seguinte notação:

#### Parâmetros

- $\lambda_i$  : procura anual do produto  $i$ , em unidades de carga
- $H_{ij}$  : custo anual de movimentar uma unidade do produto  $i$  no fluxo  $j$
- $C_{ij}$  : custo anual de armazenar uma unidade do produto  $i$  no fluxo  $j$
- $Q_i/2$  : stock médio em armazém do produto  $i$ , em unidades de carga
- $TS$  : espaço total disponível
- $p_i$  : percentagem média do tempo de permanência na área de reserva do produto  $i$  se este tem um fluxo do tipo 2
- $LL_R, UL_R$  : limites inferior e superior da área de reserva
- $LL_P, UL_P$  : limites inferior e superior da área de principal

#### Variáveis de decisão

$$X_{ij} = \begin{cases} 1 & \text{se o produto } i \text{ é atribuído ao fluxo } j \\ 0 & \text{caso contrário} \end{cases}$$

$\beta, \gamma$  : percentagem de espaço atribuída às áreas funcionais : reserva ( $\beta$ ) e principal ( $\gamma$ )

O modelo de programação matemática que permite determinar o fluxo ao qual cada produto deve ser afectado e, simultaneamente, a dimensão das áreas funcionais consideradas é:

$$\min \sum_{i=1}^n \sum_{j=1}^3 (H_{ij} \lambda_i X_{ij}) + \sum_{i=1}^n \sum_{j=1}^3 (C_{ij} Q_i / 2 X_{ij}) \quad (1)$$

$$\sum_{j=1}^3 X_{ij} = 1 \quad \forall_i \quad (2)$$

$$\sum_{i=1}^n (Q_i / 2 X_{i1}) + \sum_{i=1}^n (p_i Q_i / 2 X_{i2}) \leq \beta TS \quad (3)$$

$$\sum_{i=1}^n ((1 - p_i) (Q_i / 2 X_{i2})) + \sum_{i=1}^n (Q_i / 2 X_{i3}) \leq \gamma TS \quad (4)$$

$$\beta + \gamma \leq 1 \quad (5)$$

$$LL_R \leq \beta TS \leq UL_R \quad (6)$$

$$LL_P \leq \gamma TS \leq UL_P \quad (7)$$

$$\beta, \gamma \geq 0 \quad (8)$$

$$X_{ij} = 0 \text{ ou } X_{ij} = 1 \quad \forall_{i,j} \quad (9)$$

A função objectivo (1) considera os custos totais de movimentação e de armazenamento. A integridade do modelo é estabelecida garantindo que cada produto é atribuído a um e um só fluxo (2) e que o espaço total disponível não é ultrapassado (3)-(4), garantindo-se também que os limites inferior e superior impostos para cada uma das áreas funcionais são respeitados (6)-(7).

Os resultados obtidos, por resolução do modelo, têm um papel importante na reestruturação do armazém. No entanto, muitas outras decisões ficaram excluídas, por exemplo, definição da política de localização a adoptar. Vejamos de seguida o modelo que aborda a política de localização adoptada (Geraldes, 2008a). Trata-se de um modelo de programação matemática que considera as eventuais relações entre os produtos formando *clusters*.

### 3.2 Localização dos produtos na área principal

A necessidade de agrupar os produtos, formando *clusters*, para uma eficaz e eficiente localização do *stock* e posterior operação de *picking* tem sido bem recebida pelos gestores de armazéns. Um *cluster* diz respeito a um grupo de produtos similares. O objectivo é agrupar os produtos em *clusters* por forma a que os itens pertencentes ao mesmo *cluster* tenham entre si um elevado grau de associação (similaridade).

Seja  $x_{ij}$  uma variável binária que é 1 se o produto  $i$  é atribuído ao *cluster*  $j$  e 0 caso contrário. Considere-se também a variável binária  $y_j$  que representa o produto tipo de todos os produtos incluídos no *cluster* (*cluster* mediano). Se  $y_j = 1$  se produto  $j$  é seleccionado para *cluster* mediano e, caso contrário,  $y_j = 0$ . A formalização geral do modelo pode ser apresentada como:

$$\max \sum_{i=1}^P \sum_{j=1}^P s_{ij} x_{ij} \quad (10)$$

$$\sum_{j=1}^P x_{ij} = 1 \quad i = 1, \dots, P \quad (11)$$

$$x_{ij} \leq y_j \quad i, j = 1, \dots, P \quad (12)$$

$$\sum_{j=1}^P y_j = K \quad (13)$$

$$x_{ij} = 0 \text{ ou } x_{ij} = 1 \quad i, j = 1, \dots, P \quad (14)$$

$$y_j = 0 \text{ ou } y_j = 1 \quad j = 1, \dots, P \quad (15)$$

Onde  $s_{ij}$  representa a medida de similaridade adoptada e  $K$  é o número de *clusters* considerados. A função objectivo considerada (10) permite seleccionar  $K$  produtos tipo como medianas, para que a soma das similaridades de todos os produtos agrupados seja maximizada. As restrições garantem que cada produto pertence a um e um só *cluster* (11), que os produtos são atribuídos apenas a *clusters* medianos (12) e que não se ultrapassa o número máximo de *clusters* considerados (13).

### 4 Aplicação dos modelos ao caso de estudo

Neste estudo foram considerados os 1500 artigos considerados activos, um *stock* médio de 2257 unidades de carga (paletes) e o total de 2000 posições de *stock*. Os parâmetros envolvidos no modelo de dimensionamento (procura anual, *stock* médio e os custos de manuseamento e de armazenamento) foram calculados com base nos dados históricos referentes aos últimos dois anos e meio. A solução obtida define 880 posições de *stock* para a área de reserva (44% espaço total) e 300 posições de *stock* para a área principal (15% espaço total).

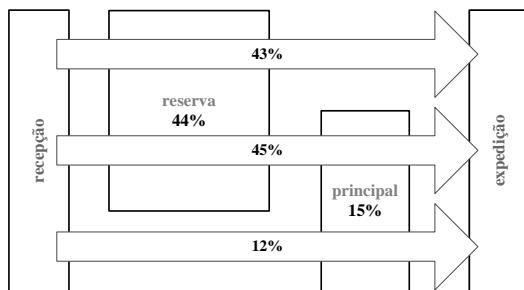


Figura 2: Solução do modelo de dimensionamento.

Cerca de 43% do total dos produtos ficam com um fluxo do tipo 1, 45% são atribuídos ao fluxo 2 e os restantes 12% têm um fluxo do tipo 3 (Figura 2).

Posteriormente, e considerando como medida de similaridade a probabilidade de um qualquer par de produtos constar na mesma encomenda (Tabela 1), foi determinada a estrutura de *clusters* para os produtos atribuídos à área principal. A título de exemplo foram considerados 30 *clusters*.

Quando considerados os parâmetros indicados a estrutura de *clusters* obtida é a apresentada na Tabela 2.

Tabela 1: Matriz de similaridade entre os produtos.

	P1	P2	P3	P4	P5	...	P184
P1	1.0						
P2	0.49	1.0					
P3	0.55	0.89	1.0				
P4	0.49	0.69	0.46	1.0			
P5	0.16	0.85	0.28	0.1	1.0		
...	...	...	...	...	...	...	
P184	0.26	0.09	0.43	0.38	0.81	...	1.0

Tabela 2: Exemplo de *clusters* obtidos.

Estrutura de Clusters
{P8, P37, P70, P90, P101}
{P103, P107}
{P11, P6, P42, P47, P87}
{P14, P51, P58, P81, P92, P112}
{P44, P60, P88}
...
{P52, P113, P120, P129}

A Tabela 3 apresenta alguns detalhes (N.º variáveis, N.º restrições, Tempo de CPU) das soluções obtidas por utilização da ferramenta de desenvolvimento e resolução de modelos de optimização – LINGO 9.0.

Tabela 3: Detalhes das soluções dos modelos.

	N.º variáveis total	N.º variáveis inteiros	N.º restrições	Tempo de CPU [hh:mm:ss]
Modelo 1	4502	4500	1505	00:01:44
Modelo 2	17030	17030	17031	00:00:27

Consequência deste estudo, a empresa decidiu reorganizar o armazém: foram identificadas as áreas funcionais e os produtos foram alocados às respectivas áreas. Relativamente aos produtos armazenados na área de *picking*, os de maior rotação foram posicionados perto do local de expedição e aos níveis 1 e 2 da estanteria. Os níveis 3 e 4 foram ocupados com *stock* de reabastecimento e com artigos de menor rotação. Adicionalmente foram propostas outras medidas com o objectivo de melhorar o desempenho do armazém, tais como: eliminação dos produtos sem rotação, codificação das posições de *stock*, o desenvolvimento de um sistema de informação e a introdução de medidas de desempenho.

## 5 Conclusões

O planeamento e concepção de um armazém é uma actividade complexa que deve ser minuciosamente examinada e adaptada para ir ao encontro e até antecipar futuras exigências. A complexidade associada ao planeamento e concepção destas instalações é originada pelo elevado conjunto de decisões inter-relacionadas, por envolver várias variáveis de decisão e, pelo facto de, por vezes, os objectivos a atingir serem contraditórios. Essa complexidade leva a que, e apesar das várias questões estarem bem identificadas, não as possamos reduzir a problemas isolados. Por tudo isto, utilizar um modelo único onde sejam incluídos todos os aspectos envolvidos torna-se uma tarefa árdua pois as alternativas existentes são inúmeras.

Este estudo abordou os problemas operacionais do armazém de uma empresa que produz e distribui cutelarias e utilidades domésticas. Um complexo e vasto conjunto de

problemas foi identificado e algumas soluções implementadas. Em particular, este estudo analisou o dimensionamento das áreas funcionais, a distribuição dos produtos pelas áreas funcionais e a adopção de uma política de localização para a área de *picking*. Para tal foram utilizados modelos de programação matemática de apoio à tomada destas decisões.

O modelo de programação inteira mista utilizado para o dimensionamento das áreas funcionais e para a distribuição dos produtos pelas diferentes áreas revelou ser bastante promissor na definição do novo *layout* tendo permitido uma redução substancial na área total de armazenamento. Consequentemente deixou de ser necessário o arrendamento do armazém auxiliar. A implementação dos resultados obtidos por resolução deste modelo abriu lugar ao estudo de uma nova política de localização. Apesar da política adoptada ter em consideração as correlações existentes entre os diferentes produtos, não são explicitamente considerados os custos operacionais envolvidos. Outra desvantagem prende-se com o facto de serem necessárias regras de sequenciamento para definir a localização dentro da área funcional.

A implementação dos resultados obtidos por resolução destes modelos mostrou que os modelos são eficientes, o que os torna promissores na resolução de problemas reais. No entanto, muitas decisões ficaram por tomar. Por exemplo, dado que a políticas de *picking* estão fortemente relacionadas com as políticas de localização torna-se imprescindível o estudo de políticas de *picking* para que seja possível avaliar o desempenho da política de localização adoptada.

## 6 Referências

- Berg, J. P. van den and Zijm, W. H. M. (1999) Models for warehouse management: Classification and examples, International Journal of Production Economics, Vol. 59, pp. 519-528.
- Geraldes, C. (2007) Modelos de programação linear aplicados à gestão das operações de um armazém – caso prático, Tese de Mestrado, Dep. Produção e Sistemas, Universidade do Minho, Braga, Portugal.
- Geraldes, C., Carvalho, S., Pereira, G. (2008a) A linear programming model for the storage location assignment problem – case study. Proceedings of the 15th European Concurrent Engineering Conference. Porto, Portugal. EUROSIS-ETI, ISBN: 978-90-77381-39-7.
- Geraldes, C., Carvalho, S., Pereira, G. (2008b) A warehouse design decision model – case study. Proceedings of the IEMC-Europe 2008 (International Engineering Management Conference). Estoril, Portugal. ISBN: 978-1-4244-2289-0, IEEE Catalog Number: CFP08EMS.
- Goetschalckx, M. and Ratlif, H. D. (1990) Shared Storage Policies Based on the Duration Stay of Unit Loads, Management Science, Vol. 36, No 9, pp. 1120-1132.
- Gray, A. E., Karmarkar, U. S. and Seidmann, A. (1992) Design and operation of an order-consolidation warehouse: models and application, European Journal of Operational Research, Vol. 58, No 1, pp. 14-36.
- Hassan, M. (2002) A framework for the design of warehouse layout, Facilities, Vol. 20, No 13/14, pp. 432-440.
- Heragu, S. S., Du, L., Mantel, R. J. and Schuur, P. C. (2005) Mathematical model for warehouse design and product allocation, International Journal of Production Research, Vol. 43, No 2, pp. 327-338.
- Liu, C. (1999) Clustering techniques for stock location and order-picking in a distribution center, Computers & Operations Research, Vol. 26, pp. 989-1002.
- Rouwenhorst, B. et al. (2000) Warehouse design and control: Framework and literature review, European Journal of Operational Research, Vol. 122, pp. 515-533.
- Salvendy, G. (eds.) (2001) Handbook of Industrial Engineering: Technology and Operations Management, 3rd ed., John Wiley & Sons, Inc., ch. 81, pp. 2083-2109.

# Reducing the Number of Membership Functions in Linguistic Variables

Margarida M. Gomes <sup>†</sup>

Paula Amaral <sup>‡</sup>

Rita A. Ribeiro <sup>‡</sup>

<sup>†</sup> FTC/UNL  
 mmg@uninova.pt  
 paca@fct.unl.pt

<sup>‡</sup> Uninova  
 rar@uninova.pt

## Abstract

In this paper clustering algorithms are used to reduce the number of membership functions in a fuzzy linguistic variable. A Scatter Search clustering algorithm is implemented and compared to a variation of the K-Means algorithm. A case study with linguistic variables belonging to a previously constructed inference system, where the task was to simplify the system without losing performance, is also studied. Using a hierarchical clustering algorithm relying on performance measures for the inference system, it was possible not only to simplify the inference system but also to improve its performance.

## Resumo

Neste artigo são utilizados algoritmos de agrupamento para reduzir o número de funções de pertença numa variável linguística. Foi implementado um algoritmo do tipo Scatter Search e este foi comparado com uma variante do algoritmo K-Means. É também apresentado um caso de estudo em que as variáveis linguísticas pertencem a um sistema de inferência construído previamente. O objectivo foi simplificar o sistema sem comprometer o seu desempenho. Usando um algoritmo de agrupamento hierárquico que tem em conta medidas de desempenho do sistema de inferência, foi possível não só simplificar o sistema, mas também melhorar o seu desempenho.

**Keywords:** Linguistic variable, fuzzy inference system, clustering

## 1 Introduction

The aim of this paper is to present algorithms to reduce the number of membership functions in a fuzzy linguistic variable (Zadeh 1975). The main idea when choosing a fuzzy model over a classical one is to obtain models that are less complex and easy to interpret. To achieve such interpretability, it is desirable that the linguistic variables in a fuzzy model are as intuitive as possible. This motivated the research of this paper. This problem

is approached as a clustering problem. We start by identifying groups of similar membership functions and merge each group into a new membership function.

Two main types of problems are presented. In the first, the purpose is to reduce the number of membership functions in linguistic variables that could be used in a fuzzy model. In the second, the linguistic variables belong to a previously constructed inference system and the aim is to achieve a compromise between performance and complexity.

A fuzzy set  $A$  in  $X$  is a set of ordered pairs  $A = \{(x, \mu_A(x)) : x \in X\}$ , where  $\mu_A(x)$  is called the membership function or grade of membership of  $x$  in  $A$  which maps  $X$  to the membership space  $M$ , usually the real interval  $[0,1]$  (Zimmermann, H. J. 1990)).

Some families of membership functions can be mapped to  $R^p$ , where  $p$  is the number of parameters of that family of functions. For instance, for  $p=3$  a triangular membership function can be represented by a triple  $(a, b, c)$ , where  $a$ ,  $b$  and  $c$  are the x-axis coordinates of its vertices. If the triangle is symmetrical, we can take  $p=2$  representing it by a pair  $(b, \varepsilon)$ , where  $\varepsilon = b - a = c - b$ . This representation will be used in section 2.

Given  $n$  trapezoidal membership functions represented by its vertices x-axis coordinates,  $T^i = (a_i, b_i, c_i, d_i)$ , the result of merging them using a generalization of the method proposed in (Setnes, Babuska et al. 1998) is a trapezoid  $T = (a, b, c, d)$ , where:

$$a = \min_{i=1,\dots,n} a_i \quad ; \quad b = \frac{1}{n} \sum_{i=1}^n b_i \quad ; \quad c = \frac{1}{n} \sum_{i=1}^n c_i \quad ; \quad d = \max_{i=1,\dots,n} d_i . \quad (1)$$

To merge triangular membership functions, it is considered  $b_i = c_i$ ,  $i = 1, \dots, n$ .

## 2 Heuristic Methods Based on Local Search

A Scatter Search based on the work presented in (Pacheco 2005; Abdule-Wahab, Monmarché et al. 2006) was implemented and compared with a variation of the K-Means algorithm (McQueen 1967), called K-Means++ (David and Sergei 2007).

### 2.1 Scatter Search

The Scatter Search (SS) algorithm operates on a small reference set (RS), incorporating good quality solutions and diverse solutions. An initial set of solutions is created by the Diversification Generation Method (DG). Each solution in this set is then improved by an Improvement Method (IM) before the creation of the RS by the Reference Set Update Method (RSU), which chooses for the RS the best quality solutions along with solutions with a high diversity level. The Subset Generation Method (SG) forms subsets of solutions from the RS to be combined by the Solution Combination Method (SC) into new solutions. The quality of such solutions is once again improved by the IM before the RS is updated. The algorithm continues until some stopping criterion is satisfied. The strategies adopted in this work for each of the main methods in the SS are based on the work of Pacheco (2005); and Abdule-Wahab et al (2006) and are summarized below.

#### 2.1.1 Fitness Function

The cluster validity index  $I$  discussed in Ujjwal and Sanghamitra (2002) was used as a fitness function to be maximized. This index is given by the following expressions:

$$I(K) = \left( \frac{1}{K} \times \frac{E_1}{E_K} \times D_K \right)^p ; \quad E_K = \sum_{k=1}^K \sum_{j=1}^n u_{kj} \|x_j - c_k\| ; \quad D_K = \max_{i,j=1,\dots,K} \|c_i - c_j\| , \quad (4)$$

where  $X = \{x_1, \dots, x_n\}$  is the data to be clustered,  $U = [u_{kj}]_{K \times n}$  represents a possible clustering of the data into  $K$  clusters and  $c_k$  is the centroid of cluster  $k$ .

### 2.1.2 Diversification Generation Method

The DG generates  $OS_{size}$  solutions that will serve as basis for constructing the RS. For each solution to be created, a random number of clusters,  $1 \leq K \leq K_{max}$ , where  $K_{max}$  is the maximum number of clusters allowed, is generated. Given the number of clusters  $K$  to be considered for the generation of a solution, the cluster centres  $S = \{c_1, \dots, c_K\}$  are randomly chosen from the data set  $X = \{x_1, \dots, x_n\}$  in the following way (Pacheco 2005):

1. Find  $x_{j^*}$ , the farthest point from the centroid of  $X$  and do  $c_1 = x_{j^*}$ ,  $S = \{c_1\}$ ,  $h = 2$ .
2. Fix  $0 \leq \alpha \leq 1$  and while  $|S| < K$  do:
  - a. Determine  $\Delta_j = \min \{\|x_j - c_i\| : c_i \in S\} \forall x_j \in X \setminus S$ .
  - b. Determine  $\Delta_{max} = \max \{\Delta_j : x_j \in X \setminus S\}$ .
  - c. Do  $L = \{x_j : \Delta_j \geq \alpha \Delta_{max}\}$ .
  - d. Choose  $x_{j^*} \in L$  according to an uniform distribution and do  $c_h = x_{j^*}$ ,  $S = S \cup \{c_h\}$ ,  $h = h + 1$ .

To avoid repetition of centres, the number of times that a point  $x_j$  is selected as a centre is stored in  $freq(j)$  and the values of  $\Delta_j$  in subsequent iterations are modified according to a parameter  $\beta$  that controls the importance given to the memory:

$$\Delta_j' = \Delta_j - \beta \Delta_{max} \frac{freq(j)}{freq_{max}} , \text{ where } freq_{max} = \max \{freq(j) : \forall j\} . \quad (5)$$

The remaining points are assigned to these clusters as in Pacheco (2005):

1. Let  $A$  be the set of unsigned points, i.e.,  $A = X \setminus S$ .
2. For each point  $x_j \in A$  and each cluster  $C_i, i = 1, \dots, K$  determine  $\Gamma_{ij} = \frac{|C_i|}{|C_i|+1} \|c_i - x_j\|^2$  where  $c_i$  is the centroid of  $C_i$ .
3. Calculate  $\Gamma_{ij^*} = \min \{\Gamma_{ij} : x_j \notin A, i = 1, \dots, K\}$ .
4. Assign  $x_{j^*}$  to  $C_i$  and set  $A = A \setminus \{x_{j^*}\}$ .
5. If  $A \neq 0$  return to 2, else stop.

### 2.1.3 Improvement Method

The improvement method is used to enhance the quality of the solutions generated during the diversification phase and after the SC method. The following steps are performed a number of times equal to  $MaxIterImp$  (Abdule-Wahab, Monmarché et al. 2006):

1. For each point  $x_j \in X$  do:
  - a. For each cluster  $C_i, i = 1, \dots, K$  determine  $v_{ij} = \frac{|C_i|}{|C_i|+1} \|c_i - x_j\|^2 - \frac{|C_i|}{|C_i|+1} \|c_i - x_j\|^2$   
where  $x_j \in C_i$  and  $c_i$  and  $c_i$  are the centroids of  $C_i$  and  $C_i$ , respectively.
  - b. Determine  $a = \arg \min_{j=1, \dots, K, j \neq i} \{v_{ij}\}$ .
  - c. If  $a < 0$  reassign  $x_j$  to  $C_a$ .
2. Compute the fitness of the new solution obtained.
3. If the fitness of the new solution is better than the original solution, replace the original solution by the new solution.

#### 2.1.4 Reference Set Update Method

To construct the initial RS, the first  $b_1$  best solutions are inserted in the RS. Next,  $b_2$  solutions are added one by one to the RS according to its diversity. In this work the diversity measure proposed by Pacheco (2005) is used. Let  $dif(\lambda, \lambda')$  be the number of assignments in solution  $\lambda$  that are different from the assignments in a solution  $\lambda'$ . Iteratively, it is chosen to enter the RS, the solution that maximizes:

$$\delta_{\min}(\lambda) = \min \{dif(\lambda, \lambda') : \lambda' \in RS\}. \quad (7)$$

In this implementation the RS is only updated based on the quality of solutions.

#### 2.1.5 Subset Generation Method and Solution Combination Method

The collection of subsets created through the SG method is composed of all pairs of solutions from the RS. To combine two solutions we implemented the path relinking strategy described in Pacheco (2005). A “path” is a series of simple movements from one solution to another. In this case we can start on one solution and move points from one cluster to another until the second solution is reached.

As in Pacheco (2005), given two solutions, the number of trial solutions that will be generated through the solution combination method varies. If the two solutions to be combined were chosen from the  $b_1$  high quality solutions in the RS, three trial solutions are generated. If the two solutions are chosen from the  $b_2$  diverse solutions in the reference set, only one solution is generated. Otherwise, two solutions are created. These solutions are randomly chosen from the solutions in the path.

## 2.2 Results

The algorithms were implemented in Matlab and run on an Intel Core2 Duo, CPU 2.2 GHz, 2 GB of RAM. The data used in this section is taken from Asuncion (2007). Wisconsin Diagnostic Breast Cancer (WDBC) data set contains 569 samples describing characteristics of the cell nuclei present in digitalized images of a fine needle aspirate (FNA) of a breast mass. Ten real-valued features were computed for each cell nucleus (Asuncion 2007): Radius, Texture, Perimeter, Area, Smoothness, Compactness, Concavity, Concave Points, Symmetry and Fractal dimension. For each of these features, the mean, standard error and mean of the three largest values (“worst”) of these features were computed for each image.

The mean ( $\mu$ ) radius can be used as a measure of the cell nuclei radius. The radius standard deviation ( $\sigma$ ) gives a measure of the error associated with this measure. To fuzzify the data we represented each sample's radius by a symmetric triangular membership function  $(\mu, 2\sigma)$ , as described in Section 1. The same fuzzification scheme was used for the rest of the features, resulting in 10 linguistic variables with 569 membership functions each. The objective is to reduce the number of membership functions in each linguistic variable using the algorithms previously described. The resulting linguistic variables could then be used in a fuzzy inference system or other fuzzy model to diagnose the type of cancer.

In the SS algorithm, we fixed  $K_{\max} = 100$ ,  $b_1 = b_2$  and  $OS_{size} = 10 \times (b_1 + b_2)$ . Several experiments were performed with different values for the other parameters of the SS. In average, the best results were obtained with  $\alpha = 0.5$ ,  $\alpha = 0.8$ ,  $b_1 = 5$  and  $MaxIterImp = 2$ . The results summarized in Table 1 correspond to these parameters. The SS algorithm was compared to the K-Means++ (David and Sergei 2007). The K-Means++ algorithm was run for  $1 \leq K \leq K_{\max}$  and the clustering that maximizes  $I(K)$  was chosen.

Table 1: K-Means++ vs Scatter Search

	K-Means++			Scatter Search		
	Time (sec.)	Nr. Clusters	I	Time (sec.)	Nr. Clusters	I
Radius	426,39	4	18,96	352,83	3	26,96
Texture	398,92	3	16,88	344,18	5	30,18
Perimeter	463,46	7	830,33	371,45	3	1286,17
Area	416,63	6	426500,7	394,15	3	668728,5
Smoothness	419,20	3	0,000143	413,96	4	0,000204
Compactness	407,77	3	0,0046	308,57	3	0,0045
Concativity	418,12	3	0,0802	305,72	3	0,0807
Concave P.	405,98	3	0,0023	471,0	3	0,0027
Symmetry	411,89	3	0,00092	672,41	3	0,00093
Fractal D.	413,90	4	0,0002	679,96	3	0,0003

The second part of the SS algorithm did not produce good solutions and the algorithm stopped after only a few iterations. Still, the first part of the SS algorithm achieved better results than the K-Means++ algorithm for most variables. With both algorithms, it was possible to significantly reduce the number of membership functions in the linguistic variables analysed. Figure 1 shows, for the linguistic variable Texture, the original membership functions, clusters found by the SS and the pruned linguistic variable.

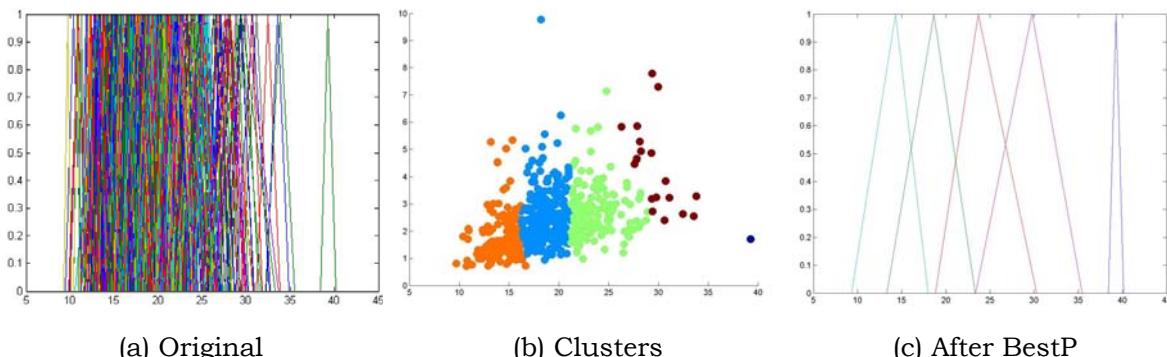


Figure 1: Linguistic Variable Texture

### 3 Case Study: a Fuzzy Inference System

In this chapter the goal is to reduce the number of membership functions in linguistic variables from a previously constructed inference system. The aim is to reduce the complexity of the system without losing too much performance.

This case study was developed at CA3 – UNINOVA in the scope of project “MODI-Simulation of a Knowledge Enabled Monitoring and Diagnosis Tool for ExoMars Pasteur Payloads” (CA3 2006; Gomes, Santos et al. 2008; Jameaux, Vitulli et al. 2006; Santos, Martins et al. 2008) for the European Space Agency. A terrain recognition inference system was automatically constructed from data collected by sensors while drilling different types of materials. The result of the inference system is the hardness type and the level of certainty in that classification.

#### 3.1 The algorithm

The algorithm adopted can be seen as a hierarchical clustering algorithm and is based in the algorithm proposed in Setnes, Babuska *et al.* (1998), where the most similar pairs of membership functions are iteratively merged into new membership functions until all membership functions are sufficiently dissimilar. Instead of defining that threshold heuristically, the algorithm was run until all membership functions were disjoint. The performance of the system, given by  $P(M)$ , was evaluated in each step of the algorithm. The inference system with best performance, from the ones generated during the algorithm, is returned (Figure 2).

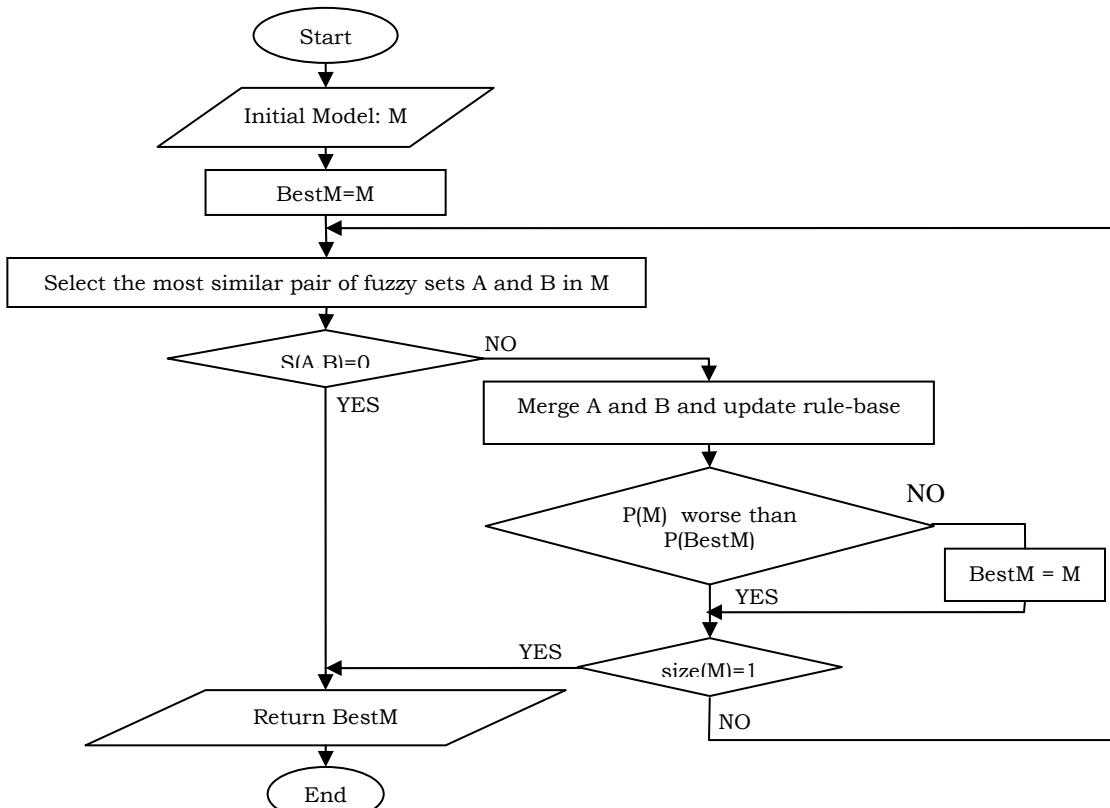


Figure 2: BestP algorithm

The algorithm is defined for any measure of the inference systems performance,  $P(\cdot)$ . In this case, the following performance measure, to be maximized, was used:

$$F = \frac{2 \cdot P \cdot MCL}{P + MCL} \quad (9)$$

where the Precision of the system ( $P$ ) is the ratio of well classified samples and the Mean Certainty Level ( $MCL$ ) is the average of the certainty levels for well classified samples.

If we want a compromise between the number of membership functions in the system and its performance we can combine these two objectives using appropriate weights and scales. In this case we used the Jaccard similarity measure (Miyamoto 1990) with the standard fuzzy intersection and union, but other similarity measures could be used.

### 3.2 Results

Drills were made in all combinations of 6 different types of terrain, 3 possible rotational velocities and 3 possible translational velocities of the drill, thus obtaining 54 membership functions per each of the linguistic variables. Applying the algorithm sequentially to each linguistic variable, it was possible not only to significantly simplify the system (Table 2) but also to improve its performance (Table 3). For instance, the Precision of the system ( $P$ ) increased from 72.33% to 85.47%. Figure 3 represents the linguistic variable Translational Voltage before and after the algorithm.

Table 2: Reduction of the membership functions

	Original	BestP
Rotation Current	54	3
Rotation Voltage	54	5
Rotation Speed	54	3
Thrust	54	17
Torque	54	46
Translational Voltage	54	3
Translational Current	54	3
Translational Speed	54	3
<b>TOTAL</b>	<b>432</b>	<b>45</b>

Table 3: Comparison of inference systems

	Original	BestP
P	72.33%	85.47%
MCL	34.49%	44.00%
F	46.71%	58.09%
N	423	45

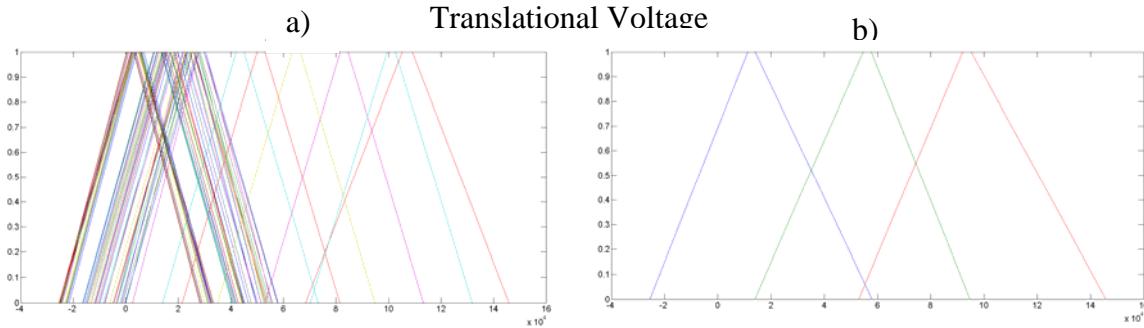


Figure 3: Translational Voltage before (a) and after (b) the algorithm

## 4 Conclusions

A SS algorithm was implemented to find good quality membership merging in reasonable time. This metaheuristic was compared to the K-Means++ algorithm. Although the second part of the algorithm was not able to produce good solutions, the first part was sufficient to obtain better results than those achieved by the K-Means++. Both methods were able to significantly reduce the number of membership functions in each linguistic variable of the case study presented. In the last chapter a case study where the linguistic variables to be pruned were part of a previously automatically constructed inference system was discussed. The results achieved were more than satisfying. Not only was it possible to significantly simplify the inference system, by reducing the number of membership functions in the system, but it was also possible to improve its performance.

## 5 References

- Abdule-Wahab, R. S., N. Monmarché, M. Slimane, M. A. Fahdil and H. H. Saleh (2006). A Scatter Search Algorithm for the Automatic Clustering Problem, Advances in Data Mining, pp. 350-364.
- Asuncion, A. N., D.J. (2007). UCI Machine Learning Repository , Retrieved March 2008, from [http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).
- CA3. (2006). MODI's homepage, Retrieved April 2007, from [http://www2.uninova.pt/ca3/en/project\\_MODI.htm](http://www2.uninova.pt/ca3/en/project_MODI.htm).
- David, A. and V. Sergei (2007). k-means++: the advantages of careful seeding, Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. New Orleans, Louisiana, Society for Industrial and Applied Mathematics.
- Gomes, M. M., B. R. Santos, T. Simas, P. Sousa and R. A. Ribeiro (2008). Reducing the Number of Membership Functions in Linguistic Variables: Application to a Fuzzy Monitoring System. Eight International Conference on Application of Fuzzy Systems and Soft Computing Helsinki, Finland, b- Quadrat Verlag.
- Jameaux, D., R. Vitulli, R. A. Ribeiro, T. Fonseca, B. R. Santos and M. Barata (2006). Monitoring & Diagnosis on-board software module for Mars driller, Proceedings of the 5th International Workshop on Planning and Scheduling for Space.
- McQueen, J. (1967). Some methods for classification and analysis of multivariate observations, In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281-297.
- Miyamoto, S. (1990). Fuzzy Sets in Information Retrieval and Clustering Analysis, Kluwer Academic Publishers.
- Pacheco, J. A. (2005). A scatter search approach for the minimum sum-of-squares clustering problem, Elsevier Science Ltd. Vol 32, pp. 1325-1335.
- Santos, B. R., G. Martins, M. Gomes and R. A. Ribeiro (2008). CCN for MODI - Simulation of Knowledge Enabled Monitoring and Diagnosis Tool for Mars Lander Payloads: Final Report, Uninova/CA3.
- Setnes, M., R. Babuska, U. Kaymak and H. R. van Nauta Lemke (1998). Similarity measures in fuzzy rule base simplification, IEEE Transactions on Systems, Man and Cybernetics, Part B, Vol 28, No 3, pp. 376-386.
- Ujjwal, M. and B. Sanghamitra (2002). Performance Evaluation of Some Clustering Algorithms and Validity Indices, IEEE Computer Society. Vol 24, pp. 1650-1654.
- Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-I, Information Sciences, Vol 8, No 3, pp. 199-249.
- Zimmermann, H. J. (1990). Fuzzy set theory and its applications. Boston, Kluwer Academic Publishers: 23-44

# **Uma heurística para a integração dos problemas de corte a duas dimensões e sequenciamento**

Tiago Gomes †

Filipe Alvelos ‡

Carina Pimentel †

J. M. Valério de Carvalho ‡

† Centro de Investigação Algoritmi, Universidade do Minho  
 tiago.gomes@dps.uminho.pt  
 carina@dps.uminho.pt

‡ Departamento de Produção e Sistemas e Centro de Investigação Algoritmi,  
 Universidade do Minho  
 falvelos@dps.uminho.pt  
 vc@dps.uminho.pt

## **Resumo**

Neste artigo propõe-se uma abordagem heurística para a integração dos problemas de corte a duas dimensões e sequenciamento. Tipicamente, um problema de corte consiste em cortar peças grandes (placas em problemas de duas dimensões, como o aqui abordado) em peças menores (itens), respeitando a procura dos itens e tendo como objectivo minimizar o desperdício. Um outro problema relacionado, usualmente tratado de forma independente, é a sequência pela qual as placas são cortadas. Na abordagem proposta, o problema de sequenciamento é integrado no problema de corte. O objectivo é a minimização do desperdício com a restrição do número máximo de pilhas abertas não poder exceder um determinado valor.

Propõe-se uma meta-heurística de descida em vizinhanças variáveis em que uma solução é codificada através de uma sequência de itens. Uma solução é construída, considerando um item de cada vez e posicionando-o, de acordo com critérios pré-definidos, num lugar disponível que respeita a restrição do máximo número de pilhas abertas. Se um item não pode ser posicionado em nenhum lugar nas placas até então consideradas, é adicionada uma nova placa à solução. As estruturas de vizinhança operam sobre a sequência de itens.

Foram realizados testes computacionais com instâncias “reais” fornecidas por uma empresa de mobiliário. Na maior parte das instâncias, as soluções obtidas com a abordagem proposta dominam as soluções actualmente implementadas na empresa.

## **Abstract**

In this paper a heuristic approach for the integration of the two-dimensional cutting stock and the sequencing problems is proposed. Generally, a cutting stock problem consists in cutting large pieces (boards in two-dimensional problems, as the one addressed) in smaller pieces (items), fulfilling the demand of items and with the objective of minimizing waste. A related problem, usually solved independently, is how the boards are sequenced. In the proposed approach, the sequencing problem is integrated in the cutting stock problem. The objective is to minimize the waste with a constraint on the maximum number of open stacks.

A variable neighborhood descent meta-heuristic is proposed. A solution is coded as a sequence of items and it is built by taking each item in turn and placing it in a best-fit location according to pre-defined criteria and taking into account the maximum number of open stacks constraint. If an item cannot be placed in any of the boards being used, a new board is added to the solution. Neighborhood structures are based on operations in the sequence.

Computational results are presented for “real-world” instances provided by a furniture company. In most instances, the solutions of the proposed approach dominate the ones in use at the company.

**Keywords:** Cutting and packing, pattern sequencing, heuristics, variable neighborhood descent

**Title:** Heuristic for the integration of the two-dimensional cutting stock and the sequencing problems

## 1. Introdução

Nas indústrias em que é necessário efectuar o corte da matéria-prima para se obterem os componentes ou os produtos desejados, um dos aspectos mais importantes consiste na diminuição de desperdício de material. O problema de corte abordado neste artigo é um problema com origem na indústria do mobiliário, em que se pretende cortar itens rectangulares a partir de placas rectangulares, com o objectivo de minimizar o desperdício (o que é equivalente a minimizar o número de placas usadas). Adicionalmente, considera-se que a sequência pela qual as placas são cortadas é relevante, já que esse aspecto permite diminuir as movimentações de material no chão da fábrica.

O transporte do material é uma operação que não acrescenta valor ao produto, sendo conveniente reduzi-lo ao máximo. Para além disso, existem muitas vezes limitações de espaço nas fábricas e para essas limitações não influenciarem muito o desempenho é conveniente planejar a produção de forma a reduzir os trabalhos em curso, assim como simplificar e reduzir o mais possível o transporte de material.

Em geral, à medida que as placas são cortadas em itens, estes são encaminhados por tapetes rolantes sendo agrupados por itens semelhantes em compartimentos (pilhas). À partida, define-se que a pilha agrupa os itens semelhantes ou de uma mesma ordem de fabrico, sendo aberta aquando do corte do primeiro item de um tipo e permanecendo aberta até ao corte do último item desse tipo. Nessa altura, a pilha é considerada fechada, podendo ser removida do compartimento e transportada para outro local, libertando assim espaço para criar outra pilha dum tipo de item diferente. Geralmente, nas linhas de produção existe um número limitado de compartimentos, o que origina uma solução de corte pior do que seria possível se não existisse essa limitação.

O problema em estudo pode ser modelizado através da decomposição em dois subproblemas: um relacionado com a minimização do desperdício do material cortado e outro relacionado com o sequenciamento das operações de corte. Na abordagem mais comum, estes dois subproblemas são considerados sequencialmente.

Primeiro é considerado o problema de corte, que tem sido estudo extensivamente nas últimas décadas. As referências (Dyckhoff *et al.*, 1997) e (Wäscher *et al.*, 2007) dão uma perspectiva geral destes problemas. Existe ainda uma base de dados de referências de corte e empacotamento classificadas disponíveis no sítio do "EURO Special Interest Group on Cutting and Packing" (<http://paginas.fe.up.pt/~esicup/>).

Num segundo momento é considerado o problema de sequenciamento dos padrões obtidos na fase anterior, de forma a minimizar, por exemplo, o número máximo de pilhas abertas. Este problema, designado por MOSP (*Minimization of Open Stack Problem*) ou problema de sequenciamento de padrões, tem também recebido um interesse significativo. Ver por exemplo, (Becceneri *et al.*, 2004; Yuen, 1995).

Em (Pileggi *et al.*, 2007) são propostas três abordagens para a integração dos dois problemas. Na primeira abordagem, em cada iteração, o problema de corte é resolvido heuristicamente com um conjunto de padrões proibido (inicialmente, este conjunto é vazio). O problema de sequenciamento é depois resolvido também heuristicamente e, se a solução não for admissível, alguns padrões são substituídos por padrões mais simples. Se um dado número máximo de iterações não for atingido, um conjunto de padrões é actualizado e é iniciada a iteração seguinte. Na segunda abordagem, a heurística que gera padrões e os repete até a procura dos itens ser esgotada é modificada para ter em conta a restrição relativa ao número máximo de pilhas abertas. Para evitar o aumento do desperdício nos últimos padrões, é utilizado um procedimento iterativo que actualiza o valor dos itens. Na terceira abordagem, baseada em geração de colunas, uma coluna é incluída na base apenas se o conjunto de padrões associado à nova base é admissível para o problema de sequenciamento (o que é verificado heuristicamente). De outra forma, é seleccionada a segunda coluna mais atractiva e o procedimento repetido. Em (Yanasse *et al.*, 2007), é apresentado um modelo de programação (linear) inteira também para o problema integrado. Um método de subgradiente modificado é utilizado para resolver o problema lagrangeano obtido por dualização das restrições que ligam os dois subproblemas (corte e sequenciamento). É também apresentada uma heurística.

Em geral, a obtenção duma boa solução num dos subproblemas faz com que o outro seja prejudicado. Ou seja, quanto menor for o maior número de pilhas abertas, maior

será a perda de material e para conseguir uma melhor solução em termos de desperdício será necessário utilizar mais pilhas. Torna-se necessário encontrar um ponto de equilíbrio de forma a obter soluções satisfatórias para ambos os subproblemas.

Na abordagem proposta neste artigo o problema de sequenciamento é integrado no problema de corte. O objectivo é a minimização do desperdício com a restrição do número máximo de pilhas abertas não poder exceder um determinado valor. Note-se que este é o problema enfrentado em muitas situações reais, já que o número de compartimentos para as pilhas é dado pelas máquinas utilizadas e/ou pelo *layout* fabril, sendo de facto uma restrição relevante.

Com a abordagem proposta é possível traçar uma curva do *trade-off* entre as perdas de material e o número máximo de pilhas abertas, bastando resolver várias vezes o problema de corte, fazendo variar o parâmetro relativo ao número máximo de pilhas abertas. Nesta abordagem é o agente de decisão quem escolhe qual a solução que mais lhe convém, podendo optar por uma solução com menor desperdício ou menor número máximo de pilhas abertas, mediante a sua percepção do problema. Um exemplo desta abordagem, proveniente da aplicação da abordagem proposta a uma instância “real” (AP-9\_8), é apresentado na Figura 1, em que cada ponto corresponde à resolução heurística de um problema de corte com uma restrição relativa ao número máximo de pilhas abertas.

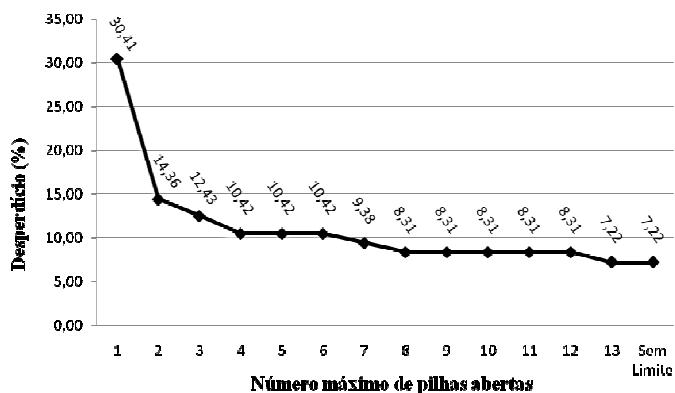


Figura 1 – Curva do trade-off entre as perdas de material e o número máximo de pilhas abertas para uma instância “real” (AP-9\_8).

A meta-heurística proposta baseia-se na colocação dos itens nas placas por uma determinada sequência. Sempre que se considera a colocação de um item, existe um conjunto de possíveis localizações (espaços livres). As possíveis localizações dependem também de informação relativa ao número de pilhas abertas em cada momento e relativa à posição de abertura e fecho da pilha do tipo de item que está a ser colocado. À medida que é construída uma solução para o problema de corte, a sequência das placas vai sendo alterada de forma a respeitar a restrição do máximo número de pilhas abertas.

Dada uma sequência de itens, o procedimento agora esboçado fornece uma solução admissível para o problema. Este procedimento permite a definição de estruturas de vizinhança baseadas na codificação de uma solução como uma sequência de itens. Estas estruturas de vizinhança são combinadas de forma sistemática num algoritmo de Descida em Vizinhanças Variáveis (*Variable Neighborhood Descent*) (Hansen *et al.*, 2008; Yagiura *et al.*, 2002), que permite a pesquisa do espaço de soluções.

A abordagem proposta pode ser vista como uma extensão, para lidar com o problema de sequenciamento, da abordagem descrita em (Alvelos *et al.*, 2009) para o problema de corte a duas dimensões.

Este artigo está organizado da seguinte forma: na secção 2 são apresentadas as características do problema de corte abordado e os conceitos mais relevantes relativos à sua integração com o problema de sequenciamento; na secção 3 é descrita a meta-heurística desenvolvida para o problema abordado; na secção 4 são apresentados e discutidos os resultados computacionais e, por fim, na secção 5, apresentam-se as conclusões deste trabalho.

## 2. Definição do problema

Segundo os cinco critérios da tipologia de Wäscher et al. (Wäscher et al., 2007) o problema de corte aqui abordado é um problema:

- a duas dimensões;
- de minimização de entrada (o valor das placas necessárias para acomodar todos os itens tem de ser minimizado);
- com uma grande variedade de itens (problema de empacotamento);
- em que as placas são todas iguais;
- em que a forma dos itens é regular (em particular, rectangular).

Utilizando estes cinco critérios é possível classificar o problema dumha forma básica e refinada. Em termos básicos classifica-se como sendo *Bin Packing Problem*, em termos refinados consiste num *Single Bin Size Bin Packing Problem* (SBSBPP). Ao problema de corte é adicionado o problema de sequenciamento. As placas de onde são cortados os itens têm de formar uma sequência de tal forma que em nenhuma ocasião o número máximo de pilhas abertas seja ultrapassado.

No problema em apreciação os seguintes aspectos são considerados: (i) os cortes são guilhotináveis, (ii) em três estágios, (iii) é permitida a rotação dos itens, (iv) existe uma restrição ao número máximo de pilhas abertas. Estes quatro aspectos são agora pormenorizados.

- (i) Os cortes guilhotináveis atravessam a placa dum lado ao outro.
- (ii) Um estágio corresponde a efectuar cortes guilhotináveis numa mesma direcção sobre o objecto ou sobre partes dele (resultantes do estágio anterior). Estágios sucessivos alternam a direcção do corte. Nos padrões de corte em três estágios é permitido efectuar três conjuntos de cortes em sentidos alternados. Por exemplo, efectuar cortes na placa no sentido horizontal, em seguida (nas partes resultante do primeiro estágio) efectuar cortes no sentido vertical, e por último efectuar um terceiro conjunto de cortes, de novo no sentido horizontal. Neste caso, um padrão de corte pode sempre ser representado como um conjunto de níveis (prateleiras) em que cada prateleira é formada por um conjunto de itens de igual largura empilhados verticalmente (pilhas do padrão). Na figura 2 é representada uma placa com um padrão de corte de três estágios exacto.

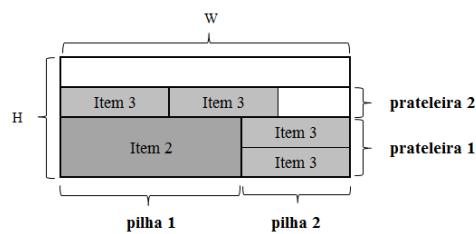


Figura 2 – Um padrão de corte em três estágios.

- (iii) A possibilidade de rodar de 90° os itens aumenta as combinações possíveis, logo poder-se-á obter soluções com menor desperdício.
- (iv) Uma pilha (relativa ao problema de sequenciamento e não a um padrão de corte) consiste num conjunto de itens do mesmo tipo. A pilha é aberta quando é cortado o primeiro item desse tipo e permanece aberta até o último item correspondente a essa pilha ser cortado, sendo nessa altura fechada. Após fechada, a pilha poderá ser removida, libertando espaço para a abertura de outra pilha.

A título exemplificativo, considere-se uma solução a que corresponde a matriz de corte da figura 3(a). Na matriz, o X significa que o tipo de item está presente na placa. O maior número de pilhas abertas é três, que ocorre quando é cortada a segunda placa, como é possível constatar na figura 3(b). A solução é admissível quando o número máximo de pilhas abertas é três, mas não quando o limite é inferior.

Sequência das placas				
	1	2	3	4
Tipo do item	1	X		
	2	X		X
	3		X	
	4		X	

(a) matriz de corte

Sequência das placas				
	1	2	3	4
Tipo do item	1	X		
	2	X	X	X
	3		X	
	4		X	X

(b) matriz das pilhas abertas

Figura 3 – Matrizes de corte e das pilhas abertas.

### 3. Uma meta-heurística de Descida em Vizinhanças Variáveis

A meta-heurística proposta baseia-se na codificação de uma solução como uma sequência de itens e numa heurística construtiva para transformar a sequência de itens numa solução admissível para o problema.

Nesta secção, descreve-se primeiro a heurística construtiva sem a restrição relativa ao sequenciamento, depois com a consideração da restrição relativa ao sequenciamento e, por fim, apresentam-se as estruturas de vizinhança utilizadas na descida em vizinhanças variáveis.

Enquanto todos os itens não estiverem posicionados em placas, o primeiro elemento da sequência é considerado (e dela removido). Em primeiro lugar, são avaliadas todas as possíveis localizações em pilhas (das placas) para as duas orientações possíveis do item e escolhida a melhor de acordo com um critério pré-definido (a apresentar posteriormente). Se o item não puder ser colocado em nenhuma pilha do padrão com nenhuma orientação, então são consideradas as localizações em prateleiras e o procedimento é repetido. Se o item não puder ser colocado em nenhuma prateleira com nenhuma orientação, então são consideradas as localizações no topo das prateleiras (placa). Caso também aí não seja possível posicionar o item com nenhuma orientação, é criada uma nova placa e o item é nela posicionado.

Este procedimento é aplicado numa fase preliminar da meta-heurística várias vezes, fazendo variar a sequência (ordenadas por altura, largura ou área) e os critérios para a colocação dos itens nas pilhas, prateleiras e placas. Por exemplo, a heurística construtiva pode ser executada com uma sequência de itens ordenados por altura decrescente e com os critérios de menor altura residual (altura que sobra após o posicionamento do item) para as pilhas; de menor largura residual para as prateleiras; de menor área residual para as placas. A sequência e os critérios que conduzem a uma solução de menor valor são utilizados como a sequência inicial e critérios para a descida em vizinhanças variáveis. O valor de uma solução  $s$  é dado por  $b_s a + a_s - n_s$ , em que  $b_s$  é o número de placas usadas na solução  $s$ ,  $a$  é a área de uma placa,  $a_s$  é a área ocupada na placa com menos área ocupada e  $n_s$  é o número de itens nessa placa. Desta forma privilegiam-se soluções que usem menos placas (o objectivo original do problema) e, em caso de empate, aquelas em que seja mais fácil esvaziar uma placa.

Considera-se agora a integração da restrição relativa ao número máximo de pilhas abertas.

Na heurística construtiva a gestão do estado da pilha de cada tipo de item é uma questão essencial. No inicio, nenhuma pilha está aberta e nenhuma está fechada. À medida que os itens são posicionados nas placas, e à medida que vão sendo usadas novas placas, os estados das pilhas e a sequência das placas são alterados de forma a respeitar o número máximo de pilhas abertas. A pilha de um tipo de item pode estar: (i) não inicializada, (ii) aberta, (iii) fechada. O estado (i) ocorre quando ainda não foi colocado nenhum item do tipo em questão. O estado (ii) implica que é conhecida a posição na sequência de placas em que pilha é aberta mas não aquela em que é fechada. No estado (iii) são conhecidas a posição da sequência de placas em que a pilha é aberta e aquela em que é fechada. No final da execução da heurística todas as pilhas de todos os itens estão no estado (iii).

Foram desenvolvidas duas versões da heurística construtiva. Na primeira, o estado (iii) só ocorre quando a colocação de outro tipo de item implica que a pilha do tipo em questão seja fechada. Na segunda, o estado (iii) também ocorre quando o item é o último do seu tipo. Assim, após a inserção de um item verifica-se se este é o último desse tipo, e caso seja, a pilha correspondente é fechada; caso contrário avança-se para o próximo item.

À medida que os itens são inseridos, dependendo do estado actual das pilhas, poderão ocorrer alterações no estado das pilhas e na ordem de corte das placas. A descrição desses momentos de decisão é efectuada de seguida.

Caso a pilha do tipo de item a inserir esteja no estado (i) é necessário abri-la. Para tal:

- Se o número de pilhas inicializadas é menor que o limite máximo de pilhas estabelecido à partida, a pilha desse tipo é aberta o mais cedo possível (na primeira posição da sequência de placas no caso da primeira versão da heurística);
- Caso contrário, é necessário fechar uma pilha de outro tipo de item para abrir a do tipo de item em consideração. Neste caso, é fechada a pilha que está aberta há mais tempo. Após a selecção da pilha a fechar, as placas que contêm esse tipo de item são ordenadas de modo a que essa pilha feche o mais cedo possível. De referir que a ordem estabelecida fica bloqueada até ao ponto de fecho da pilha. Em futuras ordenações, só será possível mexer nas placas posteriores ou inserir uma placa nova com apenas um tipo de item. Terminada a ordenação, a pilha seleccionada ficará fechada e a pilha do tipo de item pretendido ficará aberta, sendo inicializada na posição seguinte à pilha fechada.

Depois de decidido em que posição da sequência de placas é que a pilha do tipo de item em consideração abre, a pilha passa ao estado (ii) e todas as localizações a partir dessa posição são localizações possíveis para o item. Todas as localizações em placas anteriores à posição da sequência de placas em que pilha do tipo de item abre são não admissíveis.

A figura 4 contém um exemplo ilustrativo de quando é necessário fechar uma pilha para inicializar outra. Neste exemplo considerou-se que o limite de pilhas abertas é 2. Quando surge a necessidade de posicionar o primeiro item do tipo 3, existem duas pilhas abertas, logo é necessário fechar uma delas. Como a pilha que foi inicializada primeiro foi a do item 1, será essa a seleccionada para fechar. Nesta situação tenta-se fechar a pilha o mais cedo possível, trocando a ordem das placas. Neste caso é possível fazer uma troca entre a segunda e a terceira placa, permitindo dessa forma a pilha do item 1 fechar mais cedo, e a pilha do tipo de item 3 abrir na terceira placa.

Sequência de Itens: 1 2 2 2 2 2 2 1 1 ③...

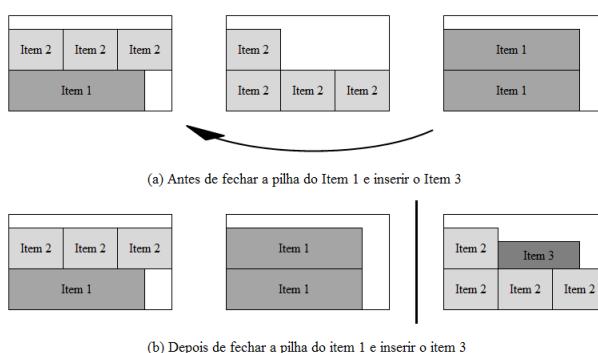


Figura 4 – Ilustração do fecho de uma pilha.

Caso a pilha do tipo de item a inserir esteja no estado (iii) é necessário verificar se o item pode ser colocado entre a posição de abertura e fecho da pilha. Caso seja possível, é seleccionada a melhor localização e não é operada nenhuma alteração no estado das pilhas. Caso tal não seja possível, é necessário colocar o item numa nova placa e essa placa será colocada na posição seguinte à posição em que estava a fechar a pilha. Após estas trocas é necessário actualizar o estado das pilhas que sofreram alteração.

Como já foi referido, na 2<sup>a</sup> versão da heurística, sempre que é inserido um item, é verificado se esse item é o último do seu tipo. Caso seja, a pilha é fechada, mas antes é verificado se é possível ordenar as placas que contêm esse tipo de item, de forma a fechar o mais cedo possível. De referir ainda que, caso haja troca na ordem das placas, a ordem estabelecida fica bloqueada até ao ponto de fecho da pilha e em futuras ordenações, só será possível mexer nas placas posteriores ou inserir uma placa nova com apenas um tipo de item.

Foram implementadas três estruturas de vizinhança e combinadas num algoritmo de descida em vizinhanças variáveis. A primeira baseia-se na troca de tipos de itens que aparecem em posições consecutivas na sequência. A segunda estrutura de vizinhança na troca de posições entre subsequências consecutivas e a terceira na inversão de subsequências. As duas últimas estruturas de vizinhança são pesquisadas com diferentes valores de parâmetros que determinam o tamanho das subsequências.

#### **4. Resultados computacionais**

As duas versões da heurística foram codificadas em C++, no ambiente de desenvolvimento Microsoft Visual Studio 2008. Os testes computacionais foram efectuados num computador pessoal, com processador Intel Core 2 Duo CPU, de 2.00GHz e com 2.00GB de memória RAM. Foram testadas 47 instâncias “reais” fornecidas por uma empresa do sector mobiliário. Para as 536 execuções das duas versões (obtidas fazendo variar o limite ao número de pilhas abertas), a segunda versão obteve melhores resultados 131 vezes e a primeira 34 vezes. Em termos de número total de placas usadas, com a segunda versão obtiveram-se 17338 placas e com a primeira 17447, logo com segunda versão seriam necessárias menos 109 placas.

Pela observação dos resultados, verificou-se que a restrição do número máximo de pilhas abertas tem um grande impacto na qualidade da solução do problema de corte, principalmente quando o número máximo de pilhas abertas é pequeno. A redução média do número de placas é: da primeira para a segunda pilha cerca de 30%; da segunda para a terceira cerca de 12%; da terceira para a quarta cerca de 6%; da quarta para a quinta cerca de 3%; da quinta para a sexta cerca de 1%; para as seguintes a redução é ainda menor. Como já foi referido, cabe ao agente de decisão optar por uma solução com menor desperdício ou menor número máximo de pilhas abertas, mediante a sua percepção do problema.

Na Tabela 1 são comparadas as soluções actualmente implementadas na empresa com as soluções obtidas com a meta-heurística proposta, para as 27 instâncias para as quais a empresa forneceu a solução que adopta. As três primeiras colunas caracterizam as instâncias testadas em que cada uma tem um determinado número de tipos de itens (a), que por sua vez têm uma procura. A procura somada dá o número total de itens (b). A procura média (c) é obtida através da divisão do número total de itens pelo número de tipos de itens. As três colunas seguintes dizem respeito às soluções da empresa: o maior número de pilhas abertas (d), o número de placas (e) e o desperdício de material que lhe está associado (f). Note-se que dado que não é permitida produção em excesso, para o mesmo número de placas o desperdício é igual. Para as soluções com a abordagem proposta, os resultados estão divididos em dois grupos: para o mesmo número de pilhas abertas da solução usada na empresa, são apresentados o número de placas utilizadas (g) e o desperdício correspondente (h); para um desperdício igual ou para o valor de desperdício inferior mais próximo ao valor da empresa (k) são apresentados o número máximo de pilhas abertas (i) e o número de placas (j). A última coluna corresponde ao tempo médio de resolução (em segundos) para a obtenção das soluções dadas pela abordagem proposta (l).

Tomando a perspectiva de o problema ter dois objectivos, uma solução de uma abordagem domina a solução da outra quando não é pior nem em termos de desperdício nem em termos do número do número máximo de pilhas abertas. As soluções da abordagem proposta dominam as fornecidas pela empresa em 22 das 27 instâncias para as quais a empresa forneceu resultados. Apenas em duas instâncias (TRAS-BC\_2 e TRAS-BC\_5) as soluções da empresa dominam as soluções da meta-heurística proposta. Em três instâncias os resultados são iguais (AP-9\_2, AP-9\_4 e TRAS-BC\_4).

#### **5. Conclusões**

Neste artigo apresentou-se uma abordagem heurística para o problema de corte a duas dimensões integrado com o problema de sequenciamento. O objectivo consiste na minimização do desperdício (ou, de forma equivalente, do número de placas utilizadas), sendo imposta uma restrição no número máximo de pilhas abertas.

A abordagem baseia-se na colocação sequencial dos itens nas placas respeitando a restrição do número máximo de pilhas abertas e, em certas situações, reordenando as placas. Foram utilizadas estruturas de vizinhança que operam sobre a sequência dos itens num algoritmo de descida em vizinhanças variáveis.

Os testes computacionais realizados em instâncias fornecidas por uma empresa de mobiliário indicaram que as soluções obtidas com a meta-heurística proposta dominam as soluções actualmente implementadas na empresa. Em quase todas as instâncias o maior número de pilhas abertas e o desperdício foram reduzidos, ou o maior número de pilhas abertas foi reduzido sem piorar o desperdício, ou o desperdício foi reduzido sem piorar o maior número de pilhas abertas. Para o mesmo número de pilhas abertas, as soluções dadas pela meta-heurística proposta utilizam menos 109 placas. Para um desperdício igual ou inferior, as soluções dadas pela meta-heurística proposta reduzem o valor total do maior número de pilhas abertas em 59.

Tabela 1 – Comparaçao dos resultados obtidos com os da empresa.

Instância	Caracterização das instâncias			Solução da Empresa			DVV para o mesmo número de pilhas abertas		DVV para o desperdício inferior mais próximo			Tempo médio da DVV (segs)
	Número de tipos	Número de itens	Procura média	Maior número de pilhas abertas	Número de placas	Desperdi-cio	Número de placas	Desperdi-cio	Maior número de pilhas abertas	Número de placas	Desperdi-cio	
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
AP-9_1	30	770	25.7	8	66	17.1%	61	10.3%	3	63	13.1%	90.1
AP-9_10	20	559	28.0	6	73	19.1%	69	14.4%	2	73	19.1%	61.5
AP-9_11	27	507	18.8	6	71	28.3%	57	10.7%	2	60	15.1%	34.4
AP-9_12	10	215	21.5	4	29	22.7%	25	10.3%	2	26	13.8%	2.2
AP-9_13	21	450	21.4	5	36	29.6%	29	12.7%	2	29	15.6%	17.8
AP-9_15	8	248	31.0	3	16	26.2%	14	15.7%	2	14	15.7%	2.5
AP-9_2	3	44	14.7	2	3	38.5%	3	38.5%	2	3	38.5%	0.1
AP-9_3	20	724	36.2	6	54	19.8%	49	11.6%	2	52	16.7%	50.3
AP-9_4	3	44	14.7	2	14	22.4%	14	22.4%	2	14	22.4%	0.1
AP-9_5	8	182	22.8	4	16	25.9%	14	15.3%	2	15	20.9%	3.6
AP-9_6	31	809	26.1	6	80	21.2%	70	9.9%	2	72	12.4%	91.5
AP-9_7	12	339	28.3	4	42	16.5%	39	10.1%	2	40	12.4%	12.2
AP-9_8	27	774	28.7	6	96	18.8%	87	10.4%	2	91	14.4%	168.9
AP-9_9	3	135	45.0	3	6	25.7%	5	10.8%	2	5	10.8%	0.6
FA+AA-9_10	36	64	1.8	11	8	15.0%	8	15.0%	10	8	15.0%	0.5
FA+AA-9_14	26	46	1.8	9	6	23.4%	6	23.4%	6	6	23.4%	0.2
FA+AA-9_4	34	61	1.8	10	8	20.2%	8	20.1%	6	8	20.1%	0.4
FA+AA-9_7	54	106	2.0	14	12	17.5%	12	17.4%	10	12	17.4%	1.8
FA+AA-9_9	24	39	1.6	11	4	17.9%	4	17.8%	7	4	17.8%	0.1
FA+AP-9-10MM_1	16	52	3.3	5	10	35.1%	8	18.9%	3	9	27.9%	0.2
FA+AP-9-10MM_2	8	78	9.8	4	5	17.2%	5	17.2%	3	5	17.2%	0.2
FA+AP-9-10MM_3	42	160	3.8	7	24	16.0%	24	16.0%	4	24	16.0%	3.6
FA+AP-9-10MM_4	11	22	2.0	5	4	31.7%	4	31.7%	4	4	31.7%	0.1
TRAS-BC_2	40	163	4.1	8	18	20.4%	19	24.6%	9	18	20.4%	4.0
TRAS-BC_3	32	71	2.2	8	20	29.5%	20	29.4%	6	20	29.4%	0.7
TRAS-BC_4	8	13	1.6	2	8	44.8%	8	44.8%	2	8	44.8%	0.0
TRAS-BC_5	11	22	2.0	3	7	32.3%	8	40.8%	4	8	32.3%	0.0

## Agradecimento

Este trabalho foi financiado pelo projecto “SCOOP - Sheet cutting and process optimization for furniture enterprises” (Contract COOP-CT-2006-032998), financiado pela Comissão Europeia, no âmbito do Sexto Programa-Quadro de Investigação e Desenvolvimento Tecnológico.

## Referências

- Alvelos, F., Chan, T.M., Vilaça, P., Gomes, T., Silva, E. and Carvalho, J.M.V.d. (2009) Sequence based heuristics for two-dimensional bin packing problems, *accepted for publication in Engineering Optimization*.
- Becceneri, J.C., Yanasse, H.H. and Soma, N.Y. (2004) A method for solving the minimization of the maximum number of open stacks problem within a cutting process, *Computers & Operations Research*, Vol. 31, No. 14, pp. 2315-2332.
- Dyckhoff, H., Scheithauer, G. and Terno, J. (1997) Cutting and Packing. In MD Amico, F Maffioli, S Martello (Eds.), *Annotated Bibliographies in Combinatorial Optimization*, pp. 393-412. John Wiley and Sons.
- Hansen, P., Mladenovic, N. and Pérez, J.A.M. (2008) Variable neighborhood search: methods and applications, HEC Montreal and GERAD, Montreal.
- Pileggi, G.C.F., Morabito, R. and Arenales, M.N. (2007) Heurísticas para os problemas de geração e sequenciamento de padrões de corte bidimensionais, *Pesquisa Operacional*, Vol. 27, No. 3, pp. 549-568.
- Wäscher, G., Haubner, H. and Schumann, H. (2007) An improved typology of cutting and packing problems, *European Journal of Operational Research*, Vol. 183, pp. 1109-1130
- Yagiura, M. and Ibaraki, T. (2002) Local Search, IN Pardalos, P. M. and Resende, M. G. C. (Eds.) *Handbook of Applied Optimization*, Oxford University Press.
- Yanasse, H.H., Lamosa, M. J. P. (2007) An integrated cutting stock and sequencing problem, *European Journal of Operational Research*, Vol.183, No. 3, pp. 1353-1370
- Yuen BJ (1995) Improved heuristics for sequencing cutting patterns, *European Journal of Operational Research*, Vol. 87, No. 1, pp. 57-64.

# A técnica de *bootstrap* na comparação de métodos de regressão ordinal usados na medição da satisfação de clientes

Isabel M. João <sup>†,‡</sup>

Carlos A. Bana e Costa <sup>‡</sup>

<sup>†</sup> Departamento de Engenharia Química, Instituto Superior de Engenharia de Lisboa,  
Rua Conselheiro Emídio Navarro, 1959-007 Lisboa.  
[ijoaq@deq.isel.ipl.pt](mailto:ijoaq@deq.isel.ipl.pt) ou [isabel.joao@tagus.ist.utl.pt](mailto:isabel.joao@tagus.ist.utl.pt)

<sup>‡</sup> Centro de Estudos de Gestão do Instituto Superior Técnico, Universidade Técnica de Lisboa  
Av. Rovisco Pais, 1049-001 Lisboa.  
[carlosbana@ist.utl.pt](mailto:carlosbana@ist.utl.pt)

## Abstract

This paper compares the results of the application of two ordinal regression methods in customer satisfaction measurement programs. The evaluation of customer's satisfaction can be regarded as a multicriteria problem in which overall satisfaction depends on a set of service quality dimensions. The additive mathematical modelling of the problem is not the same in the two methods, with different objective functions, however both of them estimate overall and partial satisfaction functions from qualitative judgments of the customers about the quality of service with respect to each one of the criteria, as well as their weights. Bootstrap analysis was used to compare the results of the two methods: 10,000 bootstrap samples were generated from the initial data set and the criteria weights obtained by the bootstrap samples for both methods were compared. Analysis of the histograms consisting of 10,000 bootstrap replications shows that the stability of the results obtained is not the same in the two methods.

## Resumo

Este artigo compara os resultados obtidos por dois métodos de regressão ordinal aplicados em estudos de medição da satisfação de clientes. A avaliação da satisfação de clientes pode ser entendida como um problema de análise multicritério em que a satisfação global depende de várias dimensões da qualidade do serviço. A modelação matemática aditiva do problema é diferente nos dois métodos, com distintas funções objectivo, mas ambos estimam funções de satisfação global e parciais baseadas em julgamentos qualitativos dos clientes sobre a qualidade de serviço nos vários critérios, bem como os pesos destes. Para comparar os resultados obtidos utilizou-se a análise de *bootstrap* e gerou-se um total de 10000 amostras *bootstrap* a partir do conjunto de dados inicial. A análise dos histogramas dos pesos dos critérios para o conjunto das 10000 amostras mostra que a estabilidade dos resultados não é a mesma nos dois métodos.

**Keywords:** ordinal regression, multicriteria analysis, *Bootstrap* analysis, customer satisfaction

**Title:** Using bootstrap analysis to compare the results of two ordinal regression methods applied to customer satisfaction appraisal

## 1 Introdução

Diferentes métodos têm sido usados para definir, medir e analisar a satisfação de clientes. Alguns deles não consideram ou não respeitam a natureza qualitativa dos julgamentos dos clientes, apesar desse tipo de informação constituir a base de dados de entrada. Noutros, as medidas que oferecem não são suficientes para analisar em detalhe a satisfação de clientes uma vez que decorrem de uma análise descritiva simples (Grigoroudis e Siskos, 2002).

Os métodos de regressão ordinal que serão comparados neste artigo fazem uso da abordagem multicritério de agregação-desagregação de preferências. As preferências individuais dos clientes são agregadas numa função de valor global, assumindo que a satisfação global do cliente pode ser explicada pela sua satisfação parcial em cada um de um conjunto de múltiplos critérios de avaliação. Enquanto que numa abordagem multicritério tradicional o modelo de agregação de preferências parciais é definido à priori e usado para ajudar a construir preferências globais desconhecidas, na filosofia de desagregação verifica-se o oposto, conhecendo-se as preferências globais e inferindo-se os modelos de preferências parciais (Jacquèt-Lagrèze e Siskos, 2001).

Nos métodos de desagregação-agregação usa-se um procedimento que se repete até se atingir um nível de consistência considerado satisfatório entre as preferências globais do decisor e as funções de valor parciais (Jacquèt-Lagrèze e Siskos, 2001; Siskos *et al.*, 2005). Os dois métodos de regressão ordinal a comparar assentam neste paradigma de desagregação-agregação, mas têm diferentes modelações matemáticas fazendo uso de diferentes critérios de optimização.

Para comparar a estabilidade dos resultados produzidos pelos métodos utilizou-se a técnica de Bootstrap. Esta técnica foi introduzida na literatura por Efron no final dos anos 70 (Efron, 1979) e pode-se dizer que é uma técnica que procura substituir a análise estatística teórica pela “força bruta” da computação, cada vez mais acessível e menos dispendiosa. Na terminologia usada por Efron e Tibshirani (1993) a técnica de bootstrap permite aproximar a distribuição de uma função das observações pela distribuição empírica dos dados baseada numa amostra de dimensão finita.

## 2 Métodos de regressão ordinal na análise da satisfação de clientes

A análise da satisfação de clientes pode ser encarada como um problema de análise multicritério no qual a satisfação global do cliente depende de um conjunto de critérios ( $i$ ) em avaliação. Considere-se que  $v_i(g_i)$ , com  $i = 1, \dots, n$  são as funções de satisfação parciais correspondentes aos vários critérios ( $n$  critérios), as quais se assume serem funções monótonas crescentes, e  $V(g)$  representa a função de satisfação global, também ela uma função monótona crescente.

O modelo aditivo de agregação na sua forma não normalizada pode ser simplesmente representado por:

$$V(g) = \sum_{i=1}^n v_i(g_i)$$

O problema de avaliação da satisfação de clientes consiste em estimar uma função de satisfação global  $V(g)$ , bem como as funções de satisfação parciais  $v_i(g_i)$ , correspondentes aos vários critérios.

Considera-se  $A = \{a_j, j = 1, 2, \dots, m\}$  o conjunto de  $m$  clientes a utilizar para estimar as funções de satisfação. Pede-se a cada cliente,  $(a_j, j = 1, 2, \dots, m)$ , para expressar o seu grau de satisfação respeitante a cada critério em avaliação, bem como o seu grau de satisfação global.

O método pela primeira vez apresentado por Siskos *et al.* (1998) e mais tarde designado de MUSA (Grigoroudis e Siskos, 2002) baseia-se nos princípios da análise

multicritério, usando um método de regressão ordinal com restrições e recorrendo a técnicas de programação linear para sua resolução. A implementação do método requer a identificação de uma família consistente de critérios, permite a utilização de escalas ordinais de satisfação e faz uso dos habituais questionários de satisfação de clientes. Os resultados que se obtêm com este método (funções de valor e pesos dos critérios) constituem uma fonte de informação para a gestão desenvolver acções de melhoria da qualidade de serviço. O método segue os princípios da análise de regressão ordinal com restrições e a função objectivo corresponde à minimização da soma dos erros absolutos (LAD - least absolute deviation):

$$\text{Min} \sum_{a_j \in A} |\varepsilon(a_j)|$$

A regressão LAD não tem uma resolução analítica e como tal é necessário usar um método iterativo. Neste artigo designa-se por método A o método de regressão ordinal que corresponde ao programa linear utilizado pelo MUSA com o critério de optimização linear usado. Para uma melhor compreensão do método ver Grigoroudis e Siskos (2002).

Em alternativa, o método que designaremos por B neste artigo é também um método de regressão ordinal com restrições devido à natureza qualitativa dos dados (João, 2009). De modo a estimar os parâmetros de regressão, o método B faz uso do método dos mínimos quadrados (Least squares – LS) cujo objectivo consiste na minimização da soma do quadrado dos erros. A modelação matemática do método B baseia-se na regressão com variáveis *dummy* ( $X_{ik}$  e  $Y_k$ ). Uma variável *dummy* do tipo  $X_{ik}$  representa um nível  $k$  da escala de satisfação do critério  $i$ . Uma variável *dummy* do tipo  $Y_k$  representa um nível  $k$  da escala de satisfação global. A escala a utilizar para a função de satisfação global é definida por  $[g^0, g^L]$  e para cada critério ( $i$ ) é definida por  $[g_i^0, g_i^{L_i}]$ , com  $i = 1, 2, \dots, n$ .

Podem utilizar-se vários esquemas de codificação tal como descrito em João *et al.* (2007).

O problema de avaliação da satisfação dos clientes consiste assim em estimar uma função de satisfação global  $V(g)$ , bem como as funções de satisfação parciais  $v_i(g_i)$  correspondentes aos vários critérios  $i$  em avaliação.

Representa-se o modelo de regressão com variáveis *dummy* para um cliente “ $a$ ” por:

$$\sum_{i=1}^n \sum_{k=0}^{L_i-1} D_{ik} \times X_{ik} - \sum_{k=0}^{L-1} Z_k \times Y_k + \varepsilon(a) = 0$$

Onde  $D_{ik}$  e  $Z_k$  são os parâmetros da regressão com variáveis *dummy*. Cada parâmetro de regressão representa a diferença de valor de um nível  $k$  em relação ao nível considerado como nível de referência  $L_i$  e  $L$ , respectivamente:

$$D_{ik} = v_i(g_i^k) - v_i(g_i^{L_i}), \quad k = 0, \dots, L_i - 1, \quad i = 1, 2, \dots, n$$

$$Z_k = V(g^k) - V(g^L), \quad k = 0, \dots, L - 1$$

De modo a estimar os parâmetros de regressão utiliza-se o método dos mínimos quadrados (Least squares – LS).

A forma final da modelação matemática é:

$$\text{Função objectivo: } \text{Min} \sum_{a_j \in A} [\varepsilon(a_j)]^2$$

Sujeito a:

$$\left\{ \begin{array}{l} \sum_{i=1}^n \sum_{k=0}^{L_i-1} D_{ik} \times X_{ikj} - \sum_{k=0}^{L-1} Z_k \times Y_{kj} + \varepsilon(a_j) = 0, \text{ para } j = 1, \dots, m \\ \sum_{i=1}^n D_{i0} = \text{constante } (<0) \\ D_{i(L_i-1)} \leq 0 \\ D_{ik} \leq D_{i(k+1)}, \text{ para } \forall i, \text{ com } k = 0, 1, \dots, L_i - 2 \\ Z_0 = \text{constante } (<0) \\ Z_{L-1} \leq 0 \\ Z_k \leq Z_{k+1}, \text{ com } k = 0, 1, \dots, L - 2 \end{array} \right.$$

Os pesos relativos ( $w_i$ ) dos critérios definem-se da seguinte forma:

$$w_i = \frac{\nu_i(g_i^{L_i}) - \nu_i(g_i^0)}{\sum_{j=1}^n [\nu_j(g_j^{L_i}) - \nu_j(g_j^0)]}$$

É procedimento usual normalizar as funções de valor no intervalo [0,1] de modo a proceder à sua comparação. Os valores normalizados,  $\bar{\nu}_i(g_i^{\alpha_i})$  para  $\alpha_i = 0, 1, \dots, L_i$  e  $\bar{V}(g^\alpha)$  para  $\alpha = 0, 1, \dots, L$  são obtidos por:

$$\begin{aligned} \bar{\nu}_i(g_i^{\alpha_i}) &= \frac{\nu_i(g_i^{\alpha_i}) - \nu_i(g_i^0)}{\nu_i(g_i^{L_i}) - \nu_i(g_i^0)} \\ \bar{V}(g^\alpha) &= \frac{V(g^\alpha) - V(g^0)}{V(g^L) - V(g^0)} \end{aligned}$$

De acordo com o procedimento de normalização efectuado, o valor 0 corresponde ao valor do menor grau de satisfação,  $\bar{\nu}(g_i^0) = 0$  e o valor 1 ao valor do maior grau de satisfação,  $\bar{\nu}_i(g_i^{L_i}) = 1$ .

Uma explicação detalhada deste método é apresentada em João *et al.* (2007).

### **3 Estudo de um caso: Satisfação de clientes com o serviço prestado por uma agência de viagens**

Para comparar os dois métodos de regressão ordinal analisou-se a satisfação de clientes com o serviço prestado por uma agência de viagens, conforme descrito em João *et al.* (2007), em que a satisfação do cliente foi avaliada por três critérios: comunicação (1), competência (2) e capacidade de resposta (3). Utilizou-se uma escala de satisfação de cinco níveis para os critérios comunicação e capacidade de resposta, bem como para avaliar a satisfação global do cliente com o serviço prestado pela agência de viagens. Os cinco níveis da escala utilizada são descritos por: totalmente insatisfeito, insatisfeito, satisfeito, muito satisfeito e totalmente satisfeito. Para o critério competência foi utilizada uma escala com os três níveis seguintes: insatisfeito, satisfeito e muito satisfeito. Em João *et al.* (2007) encontram-se as respostas para um conjunto de 100 clientes. Por utilização da técnica de *bootstrap* e partindo de uma amostra inicial de 100 observações (respostas dos clientes), procede-se à comparação dos métodos A e B.

#### **3.1 A técnica de *bootstrap* na comparação de métodos**

A técnica de *bootstrap* trata a amostra como se esta representasse exactamente toda a população. Constrói-se  $\mathcal{B}$  amostras  $X^{*(1)}, \dots, X^{*(\mathcal{B})}$  independentes e identicamente distribuídas (i.i.d.) cada qual com dimensão igual a 100, a partir da população finita  $(x_1, x_2, \dots, x_{100})$ . Tal corresponde a amostras com substituição a partir do conjunto inicial  $X$ .

A utilização da técnica de *bootstrap* será de grande utilidade pois a partir de uma amostra de 100 observações pode-se construir um grande conjunto de amostras e calcular as funções de valor e pesos pelo método A ou pelo método B. Pode-se representar a distribuição dos pesos obtidos para ambos os métodos e também determinar os intervalos de confiança para os pesos (o mesmo procedimento pode ser desencadeado para as funções de valor obtidas).

Na figura 1 encontra-se representado de uma forma esquemática o processo de *bootstrap* usado. 10000 amostras *bootstrap* foram geradas a partir do conjunto de dados original (dados do inquérito). Cada uma das amostras *bootstrap* é constituída por 100 elementos tal como o conjunto de origem.

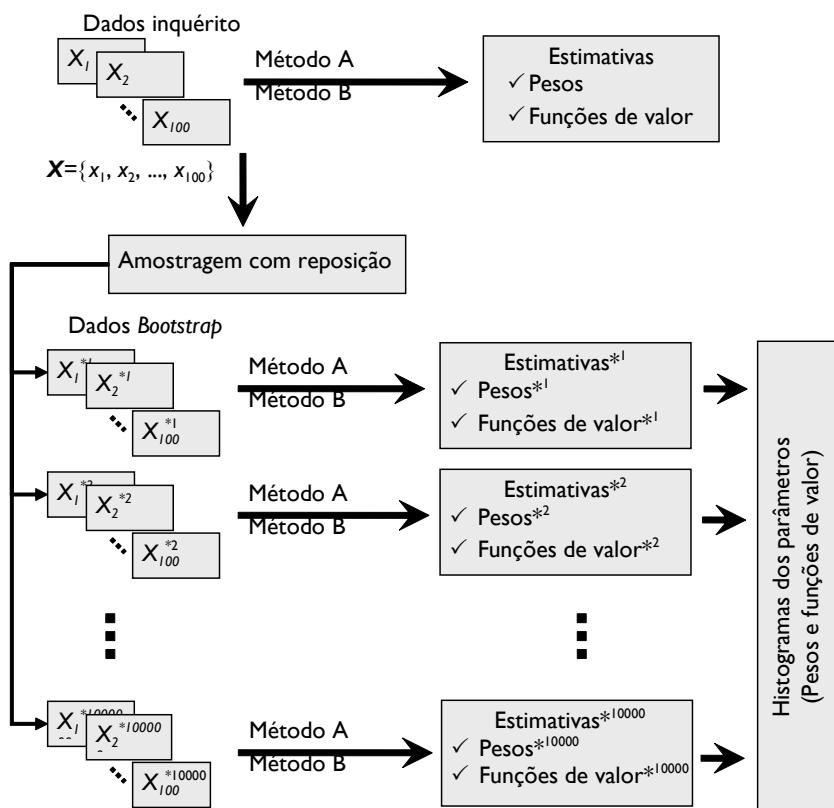


Figura 1: Esquema de implementação da análise por *bootstrap*

Para cada amostra *bootstrap* calcularam-se os pesos dos critérios e as funções de valor utilizando cada um dos dois métodos A e B. Nas Figuras 2 e 3 apresentam-se as distribuições de frequência dos pesos dos critérios obtidos por cada um dos métodos.

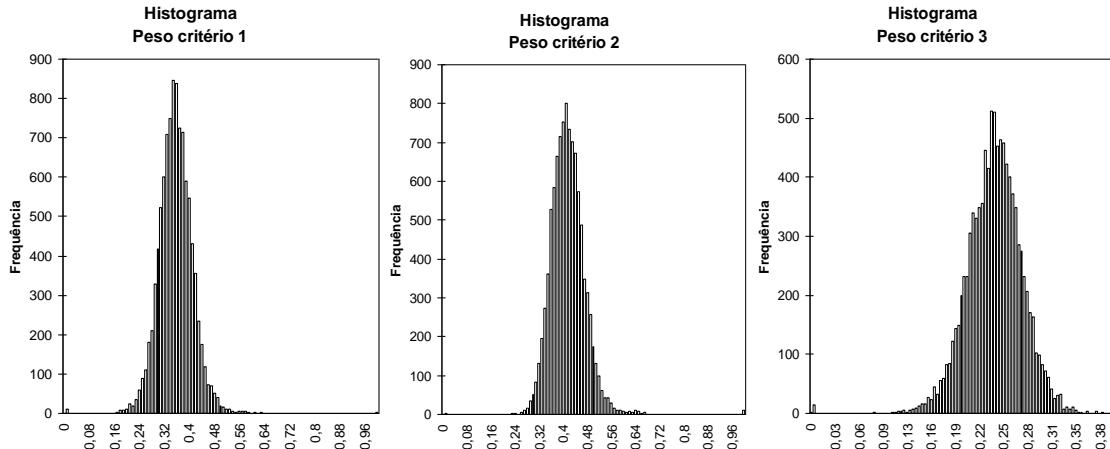


Figura 2: Histogramas dos pesos dos critérios obtidos com o Método B – agência de viagens (10000 amostras *bootstrap*)

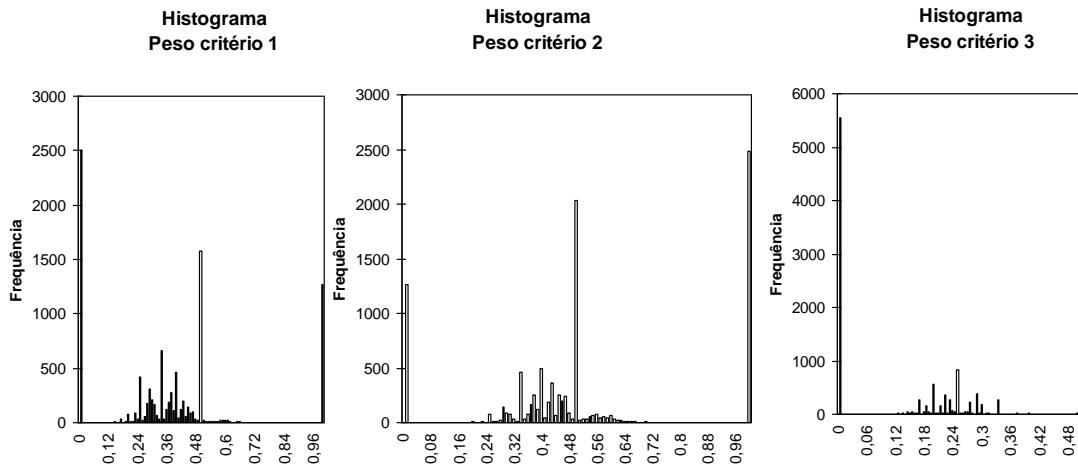


Figura 3: Histogramas dos pesos dos critérios obtidos com o Método A – agência de viagens (10000 amostras *bootstrap*)

Por observação dos histogramas conclui-se que as funções de distribuição empíricas obtidas pelo método B apresentam uma variabilidade muito inferior à das funções obtidas com o método A (e um comportamento aproximadamente Normal). No método A a variabilidade é muito elevada e verifica-se tipicamente uma elevada frequência de amostras com peso 1 ou peso 0 para um mesmo critério.

## 4 Conclusões

Efectuou-se a comparação de dois métodos de regressão ordinal com restrições, com modelações matemáticas distintas. A utilização da técnica de *bootstrap* revelou-se útil para comparar os métodos. No caso estudado, as funções de distribuições empíricas obtidas para os pesos dos critérios com o método B apresentam uma variabilidade muito inferior às registadas com o método A; verifica-se existir instabilidade das soluções obtidas com o método A, comprovada pela elevada frequência de pesos com valor 1 e 0 para um mesmo critério. O método B é de simples implementação, e os resultados obtidos são estáveis, o que, ao ser confirmado por novos estudos, será uma grande vantagem face ao método A, em que os problemas referidos são fruto das técnicas de programação linear utilizadas e do método de minimização do erro escolhido.

## 5 Agradecimentos

Isabel M. João agradece o apoio financeiro concedido pelo Turismo de Portugal, I.P. para a realização do trabalho de investigação que deu origem a este artigo.

## 6 Referências

- Efron, B. (1979) Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics*, Vol. 7, No. 1, pp.1-26.
- Efron, B. e Tibshirani, R.J. (1993) An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability, Chapman & Hall, New York.
- Grigoroudis, E. e Siskos, Y. (2002) Preference Disaggregation for Measuring and Analysing Customer Satisfaction: The MUSA Method, *European Journal of Operational Research*, Vol. 143, pp. 148-170.
- Jacquèt-Lagrèze, E. e Siskos, Y. (2001) Preference Disaggregation: 20 years of MCDA Experience, *European Journal of Operational Research*, Vol.130, pp. 233-245.
- João, I.M. (2009) Um Método Multicritério para Avaliação da Satisfação de Clientes na Indústria Hoteleira, Tese de Doutoramento, IST-UTL.
- João, I.M., Bana e Costa, C.A. e Figueira, J.R. (2007) An Alternative to the MUSA Method for Customer Satisfaction Analysis. Working Paper CEG-IST, 20. <https://fenix.ist.utl.pt/investigacao/ceg-ist/lateral/working-papers>
- Siskos, Y., Grigoroudis, E., e Matsatsinis, N.F. (2005) "UTA Methods", em Figueira, J., Greco, S., Ehrgott, M. (Editors), "Multiple Criteria Decision Analysis: State of the Art Surveys", Springer's International Series in Operation Research Management Science, Springer.
- Siskos, Y., Grigoroudis, E., Zopounidis, C. e Sauris, O. (1998) Measuring Customer Satisfaction Using a Collective Preference Disaggregation Model, *Journal of Global Optimization*, Vol.12, pp.175-195.

