

INVESTIGAÇÃO OPERACIONAL

Volume 28 — nº 1 — Junho 2008

Publicação Semestral

Editor Principal: José F. Oliveira
Universidade do Porto

Comissão Editorial

M. Teresa Almeida Inst. Sup. Econ. e Gestão	J. Rodrigues Dias Univ. de Évora	N. Maculan Univ. Fed., Rio Janeiro
C. Henggeler Antunes Univ. de Coimbra	Laureano Escudero IBM, Espanha	Rui Oliveira Inst. Superior Técnico
Marcos Arenales Univ. de São Paulo	Edite Fernandes Univ. do Minho	J. Pinto Paixão Univ. de Lisboa
Jaime Barceló Univ. de Barcelona	J. Soeiro Ferreira Univ. do Porto	M. Vaz Pato Inst. Sup. Econ. e Gestão
Eberhard E. Bischoff Univ. of Wales, Swansea	J. Fernando Gonçalves Univ. do Porto	Mauricio G. Resende AT&T Labs Research
C. Bana e Costa Inst. Superior Técnico	Luís Gouveia Univ. de Lisboa	A. Guimarães Rodrigues Univ. do Minho
M. Eugénia Captivo Univ. de Lisboa	Rui C. Guimarães Univ. do Porto	António J. L. Rodrigues Univ. de Lisboa
Domingos M. Cardoso Univ. de Aveiro	Joaquim J. Júdice Univ. de Coimbra	J. Pinho de Sousa Univ. do Porto
João Clímaco Univ. de Coimbra	J. Assis Lopes Inst. Superior Técnico	Reinaldo Sousa Univ. Católica, Rio Janeiro
J. Dias Coelho Univ. Nova de Lisboa	Carlos J. Luz Inst. Polit. Setúbal	L. Valadares Tavares Inst. Superior Técnico
João P. Costa Univ. de Coimbra	Virgílio P. Machado Univ. Nova de Lisboa	B. Calafate Vasconcelos Univ. do Porto
Ruy Costa Univ. Nova de Lisboa	Manuel Matos Univ. do Porto	L. Nunes Vicente Univ. de Coimbra
		Victor V. Vidal Techn. Univ. of Denmark

A PROGRAMAÇÃO MATEMÁTICA POSITIVA COMO INSTRUMENTO DE CALIBRAÇÃO E PRESCRIÇÃO DOS MODELOS DE OFERTA AGRÍCOLA

Rui Manuel de Sousa Fragoso[†]
Maria Leonor da Silva Carvalho [‡]
Pedro Damião de Sousa Henriques [‡]

[†] Universidade de Évora, Departamento de Gestão, Instituto de Ciências Agrárias Mediterrânicas (ICAM), Centro de Estudos e Formação Avançada em Gestão e Economia (CEFAGE)
rfragoso@uevora.pt

[‡] Universidade de Évora, Departamento de Economia, Instituto de Ciências Agrárias Mediterrânicas (ICAM), Centro de Estudos e Formação Avançada em Gestão e Economia (CEFAGE)
leonor@uevora.pt

[‡] Universidade de Évora, Departamento de Economia, Centro de Estudos e Formação Avançada em Gestão e Economia (CEFAGE)
pdamiao@uevora.pt

Abstract

In this paper, calibration and prescription capacity of different types of positive mathematical programming models applied to the Alentejo agricultural sector is analysed. Model results are compared in the 2000 and 2004 agricultural price and subsidies scenarios, regarding optimal combination of activities. Results allow concluding that positive mathematical programming is an efficient instrument on calibration of agricultural supply models, as well as on prescription of their results.

Resumo

Neste artigo avalia-se a capacidade de calibração e de prescrição de resultados de um modelo de oferta agrícola da Região Alentejo. A capacidade de calibração é analisada para o regime de preços e de ajudas agrícolas em vigor no ano 2000, comparando os resultados de diferentes formas de especificação da função dos custos variáveis totais do modelo de programação matemática positiva com os resultados do modelo tradicional de programação linear e com os dados estatísticos observados. Depois de calibrado, o modelo de programação matemática positiva é utilizado na prescrição dos resultados relativos ao cenário de preços e ajudas em vigor no ano de 2004. Conclui-se que a programação matemática positiva para além de ser um eficaz instrumento de calibração dos modelos de oferta agrícola, constitui também uma forma de prescrição de resultados futuros.

Key Words: positive mathematical programming, agricultural supply, calibration, prescription.

Title: Positive Mathematical Programming: an instrument for calibration and prescription of agricultural supply models.

1 Introdução

As primeiras aplicações da programação linear (PL) à economia agrícola realizaram-se no contexto da empresa agrícola (Throsby, 1974; Martin, 1977). Esses modelos, fáceis de construir, revelaram-se muito úteis para compreender a realidade. A sua ampla utilização no estudo de problemas económicos aplicados à agricultura deve-se principalmente à facilidade com que incorporam na sua estrutura os princípios da teoria económica do produtor e ao facto das necessidades de informação serem substancialmente inferiores às dos métodos econométricos (Howitt, 1995).

O problema económico do produtor agrícola é geralmente formulado sob a forma primal da PL, em que o objectivo é determinar a combinação das actividades agrícolas que maximizam o lucro e que são admissíveis relativamente à disponibilidade dos recursos fixos. Frequentemente, a solução deste problema afasta-se da realidade observada e é sobre-especializada nas actividades que mais contribuem para a formação do lucro do produtor.

Segundo Howitt (1995) a origem do problema de sobre-especialização da solução do modelo de PL, principalmente nos modelos agregados, está no reduzido número de restrições empíricas comparado com o número observado de actividades agrícolas na situação de referência, na falta de especificação da não linearidade das tecnologias agregadas e no facto de ser difícil considerar os preços endógenos dos produtos e o risco no comportamento dos agentes económicos.

No modelo de PL, as variáveis não zero, que representam as actividades agrícolas realizadas, estão limitadas pelo número das variáveis básicas do problema. Cada variável básica está associada a uma restrição do problema que, por sua vez, representa a disponibilidade dos recursos e o seu valor dual. Para cada restrição, o valor dual corresponde ao valor da produtividade marginal do recurso. Deste modo, se o número de restrições empíricas for inferior ao número de variáveis básicas, a solução do modelo será necessariamente sobre-especializada, verificando-se que várias actividades agrícolas observadas na situação de referência não constam da solução do modelo.

Para aproximar os resultados dos modelos de PL à situação de referência, podem sempre acrescentar-se restrições que permitem aumentar o número das variáveis básicas e condicionar os valores das variáveis de decisão. Apesar da complexidade dos factores que estão na origem da discrepância dos resultados, os esforços dos investigadores neste sentido têm sido significativos.

Day (1961) defende que o realismo da solução dos modelos de PL pode ser aumentado, através da introdução de limites superiores e de limites inferiores ao valor das variáveis que traduzem o nível das actividades agrícolas no modelo. Para McCarl (1980), a solução está em aproximar o custo de produção do verificado na situação de referência através da decomposição das tecnologias (variáveis) e dos recursos (restrições) de acordo com a sua heterogeneidade. No entanto, é necessário dispor de dados detalhados e os resultados são válidos apenas para a situação de referência, sendo, na

maior parte dos casos, inapropriados para estudar os efeitos decorrentes de alterações técnicas, económicas e institucionais.

As decisões de afectação dos recursos fixos resultam geralmente da maximização dos rendimentos e dos *trade-offs* entre rendimentos, custos de produção e externalidades entre culturas, sendo raramente tomadas com base em critérios agronómicos. O estabelecimento de rotações de culturas é, geralmente, o reflexo da escassez dos recursos agrícolas e dos preços relativos dos produtos e dos factores de produção.

De acordo com Heckelei e Britz (2005) uma abordagem alternativa para a calibração dos modelos de PL é a utilização de funções objectivo não lineares para explicitar no modelo o comportamento de risco dos agentes económicos ou os preços endógenos dos produtos. A função objectivo não linear permite a existência de soluções interiores independentemente das restrições definidas, atenuando, deste modo, a sobre-especialização da solução. No entanto, mesmo nestas condições não é possível calibrar completamente a solução do modelo (Meister et al., 1978).

A necessidade de informação substancial para descrever adequadamente as tecnologias de produção, os problemas de enviesamento decorrentes do processo de agregação e a fraca aderência dos resultados aos comportamentos observados são as principais limitações dos modelos de PL no apoio à decisão e na avaliação de medidas de política agrícola e de desenvolvimento rural. A Programação Matemática Positiva (PMP) constitui uma alternativa viável, na medida em que permite calibrar automaticamente os modelos sem necessidade de recorrer a restrições adicionais de flexibilização e as respostas às alterações dos parâmetros reflectem comportamentos mais regulares dos agentes económicos (Howitt, 1995). São várias as aplicações recentes da PMP ao estudo de problemas sistémicos e multidisciplinares de sustentabilidade económica, social e ambiental, como é o caso das últimas reformas da Política Agrícola Comum (Heckelei e Britz, 2005).

O objectivo deste artigo é avaliar a capacidade de calibração e de prescrição de resultados de um modelo de PMP de oferta agrícola da Região Alentejo.

O artigo comprehende mais quatro secções que dizem respeito a uma apresentação do modelo geral de PMP e de diferentes formas de especificação da função dos custos variáveis totais, ao desenvolvimento de um modelo de oferta agrícola para o Alentejo, aos resultados e por último à conclusão.

2 Programação Matemática Positiva

A PMP permite obter de forma automática a calibração exacta do modelo em termos dos níveis das actividades, da produção e dos preços. Mesmo antes de ser formalmente apresentada por Howitt (1995), a PMP já tinha sido utilizada na modelação de problemas económicos aplicados ao sector agrícola (House, 1987; Kasnakoglu e Bauer, 1988; Bauer e Kanakoglu, 1990; Horner et al., 1992). Depois do artigo de Howitt sucederam-se várias utilizações do método, tendo inclusivamente surgido novos desenvolvimentos que vieram reforçar o seu interesse (Arfini e Paris, 1995; Cypris, 2000; Gohn e Chantreuil, 1999; Barkaqui e Butault, 1999; Barkaqui et al., 2001; Graindorge et al., 2001; Helming et al., 2001).

No modelo de PMP, a informação contida nas variáveis duais das restrições, que limitam as actividades de um problema de programação linear de maximização do lucro aos níveis observados, é utilizada para especificar uma função objectivo não linear, de tal modo que a solução óptima reproduzirá os níveis observados dessas actividades.

A PMP pode ser entendida como um compromisso entre os modelos econométricos e os modelos de programação matemática, na medida em que a estimação dos parâmetros é feita, tal como nos primeiros, com base nos comportamentos observados e a solução primal exibe uma especificação explícita da tecnologia, como em qualquer modelo de programação matemática.

Os procedimentos empíricos do problema de PMP têm lugar em duas fases, que compreendem a estimação dos parâmetros de calibração (fase I) e a especificação de uma função objectivo não linear com base nesses parâmetros (fase II).

Na fase I deste processo usam-se restrições de calibração que forçam a solução do modelo de PL aos níveis observados das actividades:

$$\begin{aligned}
 & \max_x Z = p'x - c'x \\
 & \text{sujeito a} \\
 & Ax \leq b [\lambda] \\
 & x \leq (x^0 + \varepsilon)[\rho] \\
 & x \geq 0
 \end{aligned} \tag{1}$$

em que:

- Z = valor da função objectivo
- p = vector (nx1) de preços dos produtos
- c = vector (nx1) de custos variáveis por unidade de actividade
- x = variáveis de decisão
- A = matriz (mxn) de coeficientes técnicos
- b = vector (mx1) de disponibilidades dos recursos
- λ = vector (mx1) de variáveis duais associadas às restrições nos recursos
- x^0 = vector (nx1) de níveis observados das actividades
- ε = vector (nx1) de números positivos muito pequenos
- ρ = variáveis duais associadas às restrições de calibração

Para garantir a independência linear das restrições do problema, e deste modo evitar a degenerescência da solução dual, é necessário adicionar um pequena perturbação positiva ε aos termos independentes das restrições de calibração x^0 . Este procedimento leva a que o valor dual da restrição de calibração da última actividade a fazer parte da base da solução do modelo de PL seja nulo. Por conseguinte, o parâmetro de especificação desta variável na função objectivo não linear também será nulo.

Nessas condições, o nível de pelo menos uma das actividades no modelo de PL não é limitado pela respectiva restrição de calibração, mas por uma das restrições dos recursos fixos, o que levará a que essa variável exiba rendimentos constantes à escala. Deste modo, o vector x pode ser dividido num vector ((n-m)x1) de actividades ditas preferenciais (x^p), que são limitadas pelas restrições de calibração, e num vector (mx1) de actividades marginais, x^m , limitadas pelas restrições dos recursos fixos.

As condições de Kuhn-Tucker demonstram que o valor dual das restrições de calibração das actividades preferenciais (ρ^p) é dado pela diferença entre a receita marginal e os custos marginais, que o valor dual das restrições de calibração das actividades marginais (ρ^m) é nulo e que o valor dual das restrições dos custos fixos depende dos coeficientes das variáveis marginais na função objectivo:

$$\rho^p = p^p - c^p - A^p \lambda \tag{2}$$

$$\rho^m = [0] \quad (3)$$

$$\lambda = (\rho^m - c^m)(A^m)^{-1} \quad (4)$$

Na fase II, os valores duais das restrições de calibração, ρ^p , que acomodam erros de agregação, má qualidade da informação disponível e comportamentos de aversão ao risco, são utilizados para especificar uma função objectivo não linear, tal que o custo marginal das actividades preferenciais seja igual ao respectivo preço para os níveis das actividades observadas na situação de referência, x^0 . Nestas condições, o modelo deverá reproduzir exactamente o vector x^0 .

Para traduzir correctamente o lucro do produtor na função objectivo, é necessário introduzir termos não lineares do lado da oferta, consistentes com as tecnologias agrícolas, com a teoria micro-económica e com os dados disponíveis.

A não linearidade do problema primal do produtor é explicada pela heterogeneidade da qualidade da terra e pelos rendimentos decrescentes à escala em função do aumento da área das culturas (Peach, 1993).

Qualquer tipo de função não linear pode ser utilizado para representar o lucro do produtor, desde que a função dos custos variáveis das actividades seja convexa. Pelo facto de se adaptar bem à hipótese dos rendimentos decrescentes da produção agrícola e também por falta de argumentos mais fortes para outro tipo de funções, a função quadrática é frequentemente utilizada. O lado da oferta representado pela função dos custos variáveis de produção pode ser especificado como:

$$c' = d'x + \frac{1}{2}x'Qx \quad (5)$$

em que:

d = vector ($nx1$) de parâmetros associados ao termo linear, e

Q = matriz ($n \times n$) simétrica, definida ou semi-definida positiva, dos parâmetros associados ao termo quadrático.

A função dos custos marginais das actividades resulta da soma dos custos lineares c e dos custos marginais ρ . Os custos lineares são explicitados pelo vector d com base na informação contabilística disponível e os custos marginais das actividades são dados pela matriz Q , que corresponde aos termos quadráticos da função custo:

$$Cm^v = \frac{\partial C^v(x^0)}{\partial x} = d + Qx^0 = c + \rho \quad (6)$$

Uma vez especificados d e Q , o problema de programação não linear que reproduz os níveis observados das actividades é:

$$\begin{aligned} \max_x Z &= p'x - d'x - \frac{1}{2}x'Qx \\ \text{sujeito a} \\ Ax &\leq b [\lambda] \\ x &\geq 0 \end{aligned} \quad (7)$$

A condição $Cm = c + \rho$, definida em (6), constitui um sistema indeterminado de $n+n(n+1)/2$ parâmetros e n equações, que está associado à existência de uma infinidade de padrões de resposta. Na tentativa de evitar simulações arbitrárias do comportamento

dos agentes económicos, têm sido desenvolvidas diferentes formas de especificação dos parâmetros d e Q da função do custo variável (Heckelei e Britz, 2005).

Nas primeiras utilizações da PMP, o problema da especificação quadrática da função custo foi resolvido fazendo d=c, e os n elementos da diagonal de Q, q_{jj} , eram então calculados como:

$$q_{jj} = \frac{\rho_j}{x_j^0} \quad j = 1, 2, \dots, n \quad (8)$$

Esta especificação da função custo, designada de *standard*, é linear nas actividades marginais em que $\rho^m=0$. Por conseguinte, o valor dual do recurso fixo permanece constante, o que implica que uma alteração no preço de uma actividade preferencial conduz a uma substituição da actividade marginal, sem afectar os níveis das outras actividades preferenciais.

Este método é de simples especificação, especialmente quando a informação disponível é reduzida. No entanto, Cypris (2000) verificou que os resultados obtidos evidenciavam padrões de comportamento pouco coerentes com as alterações introduzidas nos coeficientes do modelo, nomeadamente, respostas a incentivos que traduzem elasticidades implícitas demasiado elevadas.

Paris (1988) usou uma especificação alternativa à anterior, em que o parâmetro d da função custo é igual a zero e os elementos da matriz Q são calculados em função dos custos explícitos observados na situação de referência, c, e dos valores duais das restrições de calibração, ρ :

$$\mathbf{d} = \mathbf{0} \quad (9)$$

$$q_{jj} = \frac{c_j + \rho_j}{x_j^0} \quad j = 1, 2, \dots, n$$

Neste caso, os elementos da diagonal da matriz Q, correspondentes às actividades marginais, são todos positivos.

Apesar deste método representar uma melhoria relativamente ao anterior, o padrão de resposta continua a ser arbitrário e está condicionado pela informação contida numa única observação.

Outra forma de especificação da função custo consiste em partir do pressuposto que o vector observado dos custos unitários por actividade, c, é igual ao *custo médio* da função quadrática de custos variáveis:

$$q_{jj} = \frac{2\rho_j}{x_j^0} \quad j = 1, 2, \dots, n \quad (10)$$

$$e$$

$$d_j = c_j - \rho_j \quad j = 1, 2, \dots, n$$

Neste método, os valores obtidos para a diagonal de Q são superiores aos obtidos com a regra *standard* em (8), o que implica a existência de elasticidades implícitas menores. Persiste o problema da linearidade da função custo nas actividades marginais para as quais não são definidos custos médios.

A utilização de *elasticidades exógenas* na especificação da função dos custos variáveis possibilita a introdução de informação relativa a várias observações e reduz a PMP ao seu papel de calibração no contexto de uma única observação dos níveis das actividades. A elasticidade preço da actividade j é calculada por:

$$\varepsilon_{jj} = \frac{1}{q_{jj}} \frac{p_j^0}{x_j^0} \quad j = 1, 2, \dots, n$$

Os parâmetros q_{jj} e d_j da função custo são obtidos do seguinte modo:

$$q_{jj} = \frac{1}{\varepsilon_{jj}} \frac{p_j^0}{x_j^0} \quad j = 1, 2, \dots, n$$

e

$$d_j = c_j + p_j - q_{jj} x_j^0 \quad j = 1, 2, \dots, n \quad (11)$$

Nesta especificação os elementos da diagonal principal da matriz Q são nulos e o efeito marginal do valor dual dos recursos fixos (λ) é ignorado, dependendo as alterações do nível das actividades, x, dos seus preços relativos e de q_{jj}^{-1} , i.e., da derivada parcial de x em ordem ao preço p ($\partial x / \partial p$).

3 O modelo da oferta agrícola da região alentejo

Para avaliar a capacidade de calibração e de prescrição da PMP, nomeadamente das regras de especificação da função custo de produção, foi desenvolvido um modelo de PMP adaptado às características regionais da produção agrícola da Região Alentejo. A sua formulação simplificada é a seguinte:

$$\begin{aligned} \text{Max } Z = & \sum_j p_j X_j + \sum_j a_j X_j + \sum_i p_i Y_i + \sum_i a_i Y_i \\ & - \sum_j c_j X_j - \frac{1}{2} \sum_j q_{jj} X_j^2 - \sum_i c_i Y_i - \frac{1}{2} \sum_i q_{ii} Y_i^2 - phT - piE \end{aligned} \quad (12)$$

sujeito a

$$\sum_i e_{jf} Y_i \leq X_{jf} \quad (13)$$

$$\sum_{js} X_{js} * 0,1 \leq X_{set} \quad (14)$$

$$\sum_j X_j \leq b_s \quad (15)$$

$$\sum_j h_j X_j + \sum_i h_i Y_i \leq b_t + T \quad (16)$$

$$\sum_j c_j X_j + \sum_i c_i Y_i \leq b_c + E \quad (17)$$

em que X_j e Y_i são variáveis de decisão que traduzem, respectivamente, a área das actividades vegetais j em hectares (ha) e a dimensão dos efectivos pecuários reprodutores i em Cabeças Normais (CN); T e E são variáveis de contratação adicional de trabalho e capital no curto prazo; p , a , c e h são, respectivamente, o valor da produção, as ajudas directas, os custos variáveis e as necessidades unitárias de trabalho das actividades j e i ; ph e pi são preços unitários da contratação de unidades adicionais de trabalho e de capital; $e_{j|i}$ são os encabeçamentos das actividades pecuárias i ; e b_s , b_t e b_c são os termos independentes das restrições dos recursos fixos terra, trabalho e capital.

A função objectivo (12) traduz a maximização do VAB no sector agrícola do Alentejo. Este indicador económico corresponde ao lucro do produtor no curto prazo. O seu valor, em euros, é obtido retirando, ao total dos proveitos, os custos variáveis totais. Os proveitos incluem a venda dos produtos nos mercados agrícolas e as ajudas directas em função das áreas cultivadas e do número de animais elegíveis. Nos custos variáveis consideram-se os custos dos factores de produção proporcionais às quantidades produzidas (c_j e c_i), os custos de contratação de recursos fixos adicionais no curto prazo, como o trabalho (ph) e o capital (pi) e os coeficientes dos custos marginais das actividades (q_{jj} e q_{ii}). De acordo com a regra de especificação utilizada, a função dos custos variáveis totais pode assumir as formulações que constam no Anexo 1.

As variáveis de decisão incluem dezoito actividades de produção agrícola na Região Alentejo. Essas actividades estão divididas em actividades de produção vegetal e actividades de produção pecuária. As actividades vegetais abrangem a produção de culturas arvenses (trigo, milho, arroz e girassol), horto-frutícolas, frutícolas, vinha, olival, pastagens permanentes, culturas forrageiras, pousio obrigatório e pousio de rotação e ainda uma actividade destinada às ocupações de matas e florestas. A produção pecuária inclui: a produção de bovinos de carne, com uma tecnologia de engorda de novilhos e com uma outra de venda de vitelos ao desmame; a produção de ovinos de carne e a produção de suínos alentejanos em regime extensivo.

No Anexo 2 são apresentados os coeficientes que descrevem as actividades no modelo, para os preços e as ajudas agrícolas em vigor em 2000 e em 2004.

A produção de pastagens permanentes e de culturas forrageiras constituem actividades intermédias, reutilizadas nas actividades pecuárias. Por isso, estas actividades não apresentam proveitos, mas apenas custos, uma vez que os seus proveitos são os que se obtêm indirectamente das actividades pecuárias. Essa transferência de rendimento entre actividades é realizada através da equação (13) que estabelece o balanço entre as áreas forrageiras que são disponibilizadas pela produção vegetal e o número total de animais.

A equação (14) modela a retirada de terras de cultivo imposta pela PAC após a reforma de 1992. Trata-se de um instrumento de controle da oferta agrícola, que obriga a que uma percentagem da área de culturas arvenses seja retirada da produção agrícola e posta em pousio. De acordo com a regulamentação em vigor, essa percentagem foi fixada em 10% da área de culturas arvenses e do próprio pousio obrigatório.

As equações (15), (16) e (17) modelam, respectivamente, a utilização dos recursos terra, trabalho e capital. Estas equações garantem que a procura destes recursos é inferior à sua disponibilidade. Para a terra considerou-se uma disponibilidade de aproximadamente 2 milhões de ha, que correspondem à superfície agro-florestal no Alentejo, incluindo a superfície agrícola utilizada e a superfície de matas e florestas. A disponibilidade anual de trabalho foi estimada em cerca de 24 milhões de horas, o que equivale a perto de 12,5 mil Unidades de Trabalho Anual (UTA). No caso do capital operacional, considerou-se uma disponibilidade inicial de 350,8 milhões de €.

Apesar da função objectivo representar a retribuição dos recursos primários da produção agrícola (terra, trabalho e capital), a solução do modelo está limitada apenas pela disponibilidade do recurso terra na equação (15). A procura de trabalho na equação (16) e de capital na equação (17) pode exceder as suas disponibilidades através da contracção, no mercado local, de unidades adicionais de trabalho por 3,5 euros por hora no primeiro caso e de capital à taxa de juro anual de 7% no segundo caso.

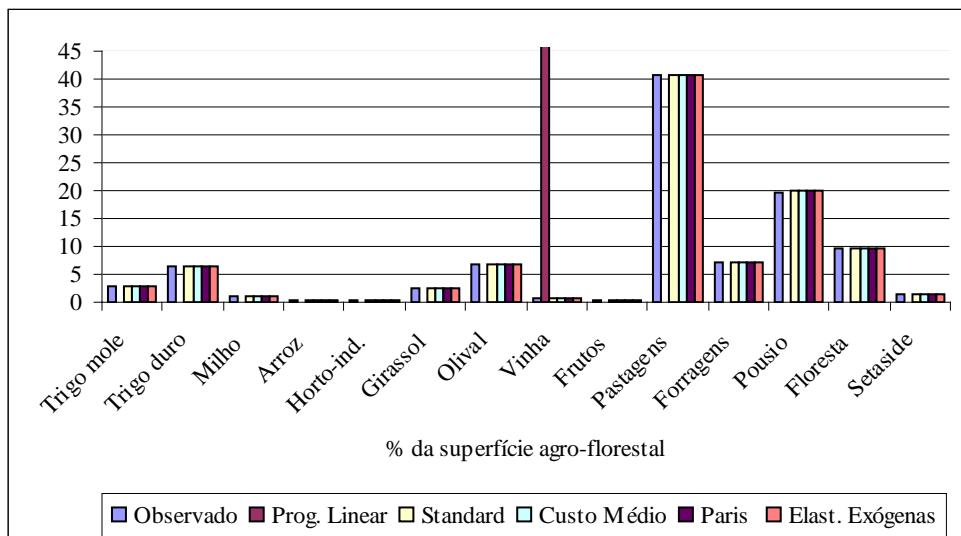
4 Resultados

A capacidade de calibração do modelo de PMP é avaliada através da comparação dos resultados com os dados observados e com os resultados do modelo tradicional de PL de oferta agrícola para a Região Alentejo para o cenário de preços e ajudas de 2000. Nas Figuras 1 e 2 apresenta-se a comparação dos resultados com os dados observados em 2000, em termos da percentagem das actividades na superfície agro-florestal total e dos efectivos pecuários reprodutores expressos em milhares de cabeças normais (CN), respectivamente.

Os resultados obtidos para o modelo de PL indicam que a totalidade da superfície da agro-florestal deverá ser utilizada na produção de vinha para vinho e que a produção pecuária deverá ser abandonada. Estes resultados são inaceitáveis na medida em os dados estatísticos de 2000 mostraram que apenas 0,7% da superfície agro-florestal da Região Alentejo é ocupada com vinha e que o efectivo pecuário reproduutor é composto por 27,1 mil CN de suínos alentejanos, 176,3 mil CN de ovinos de carne e por 183,5 mil CN de bovinos de carne, dos quais 64,2 mil CN no sistema de engorda de novilhos e 119,3 mil CN no sistema de venda de vitelos ao desmame.

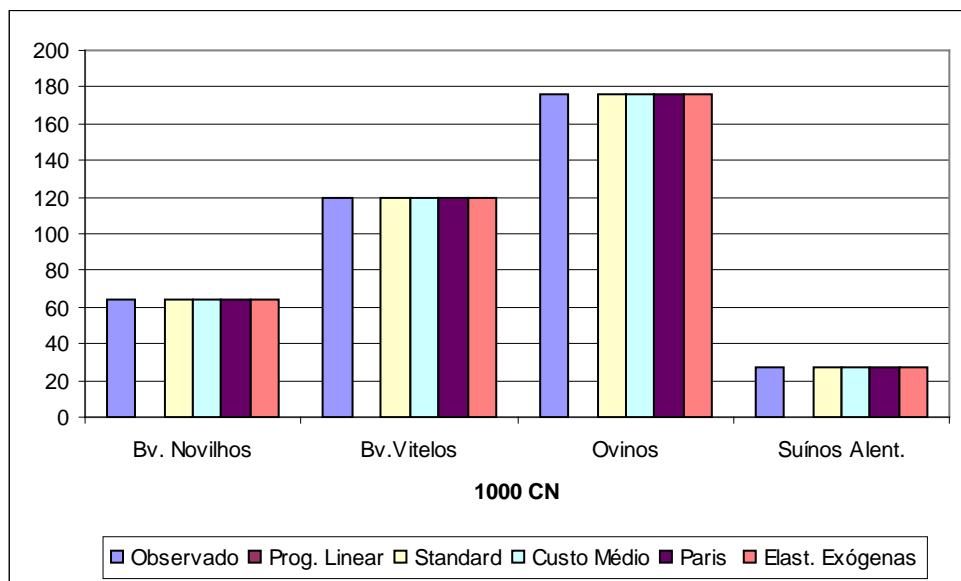
Contrariamente, os resultados, obtidos com os modelos de PMP para as várias especificações consideradas para a função dos custos variáveis, reproduzem com precisão o nível das actividades observadas nos dados estatísticos de 2000.

O facto de nesta fase se obterem os mesmos resultados para as diferentes formas de especificação da função custo de produção, deve-se à condição $C_m = c + \rho$ definida em (6), que constitui um sistema indeterminado de $n+n(n+1)/2$ parâmetros e n equações. Como para este sistema existe uma infinidade de valores para os parâmetros q_{ij} e q_{ii} que satisfazem as condições do problema de PMP, são de esperar padrões de resposta distintos para o modelo, consoante a forma de especificação utilizada para a função dos custos variáveis.



Fonte: INE, 2000 e Resultados do modelo de PL e dos modelos de PMP

Figura 1: Níveis observados e estimados das actividades agro-florestais para 2000



Fonte: INE, 2000 e Resultados do modelo de PL e dos modelos de PMP

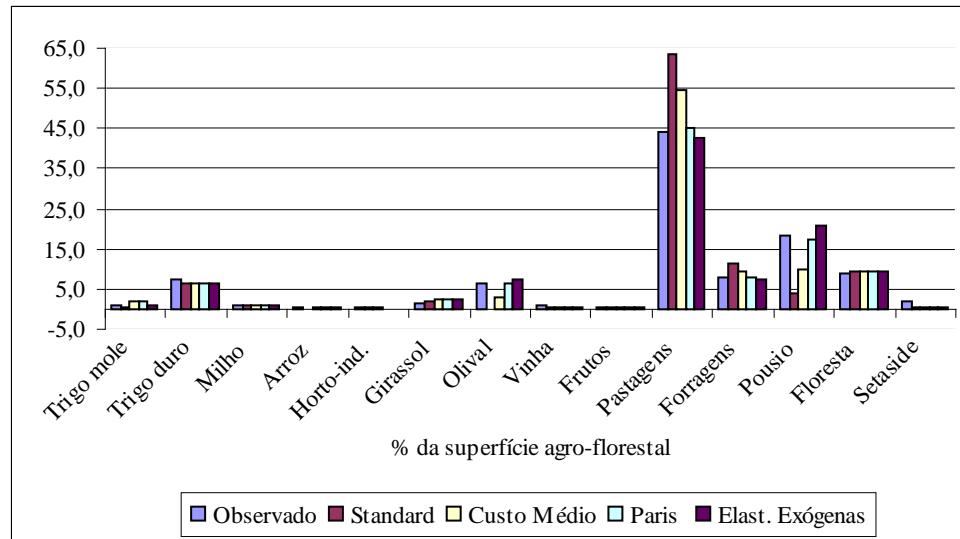
Figura 2: Níveis observados e estimados dos efectivos pecuários para 2000

O padrão de resposta do modelo relativamente a alterações dos seus coeficientes do ponto de vista da análise económica das políticas agrícolas e das condições de mercado traduz a sua maior ou menor capacidade de prescrição.

Para avaliar a capacidade de prescrição do modelo de PMP e das diferentes formas de especificação das funções dos custos variáveis alteraram-se os coeficientes do modelo introduzindo os preços e as ajudas em vigor no ano de 2004.

Nas Figuras 3 e 4 apresenta-se a comparação dos resultados dos modelos de PMP com os dados observados em 2004 em termos da percentagem das actividades na

superfície agro-florestal total e dos efectivos pecuários reprodutores expressos em milhares de cabeças normais (CN).

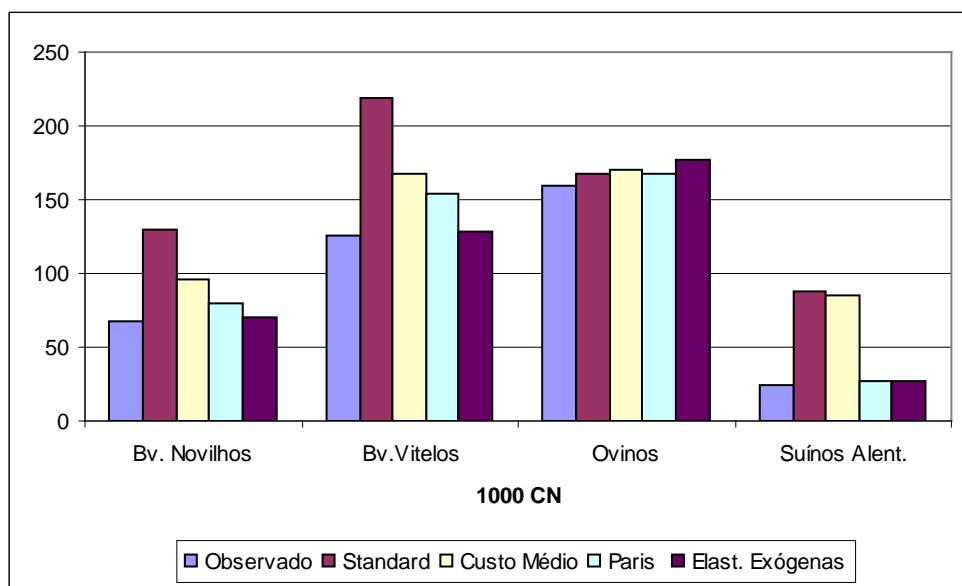


Fonte: INE, 2004 e Resultados dos modelos de PMP

Figura 3: Níveis observados e estimados das actividades agro-florestais para 2004

Estes resultados mostram que a especificação da função dos custos variáveis pela regra das *elasticidades exógenas* é superior às restantes, na medida em que a sua solução é a que se afasta menos dos valores observados nos dados estatísticos de 2004. Neste caso, o desvio absoluto ponderado na afectação das actividades à superfície agro-florestal é de 7,3% e o desvio absoluto ponderado na composição dos efectivos pecuários reprodutores é de 7,8% (Quadro 1). Segue-se a regra de *Paris* com desvios absolutos ponderados de 8,3% e de 14,4%, respectivamente.

Os resultados obtidos para as regras especificação da função dos custos variáveis *standard* e do *custo médio* são claramente inferiores, apresentando, respectivamente, desvios absolutos ponderados de 39,2% e 26,2% na afectação das actividades à superfície agro-florestal e de 88% e 63,4% na composição dos efectivos pecuários reprodutores.



Fonte: INE, 2000 e Resultados do modelo de PL e dos modelos de PMP

Figura 4: Níveis observados e estimados dos efectivos pecuários para 2004

Tabela 1: Desvios absolutos ponderados dos níveis das actividades para 2004 (%)

	Standard	Paris Standard	Custo Médio	Elasticidades Exógenas
Actividades Vegetais	39,2	8,3	26,2	7,3
Actividades Animais	88,0	14,4	63,4	7,8

Na especificação pelo método das *elasticidades exógenas* apenas três actividades agrícolas apresentam um desvio absoluto superior aos 15% indicados por Hazell (1986) como limiar mínimo de calibração desejável. Essas actividades são o arroz (-23,5%), o girassol (76,4%) e a vinha (-48,6%).

No caso da regra de *Paris*, existem seis actividades que apresentam desvios absolutos acima dos 15%, quatro actividades vegetais e duas actividades pecuárias. Particularmente elevados são os desvios absolutos registados nas superfícies de trigo mole (109,9%) e de girassol (63,7%). No que diz respeito às actividades pecuárias, os desvios absolutos são da ordem dos 20% nas actividades de produção de carne.

Para as regras de especificação *standard* e do *custo médio*, das dezoito actividades agro-florestais previstas no modelo, dez apresentam desvios absolutos acima dos 15%, sendo particularmente elevados para a composição dos efectivos pecuários reprodutores.

5 Conclusão

Este artigo avalia a capacidade de calibração e de prescrição futura dos resultados de um modelo de Programação Matemática Positiva, desenvolvido para as condições de oferta agrícola da Região Alentejo. Foram consideradas em alternativa as regras de especificação

da função dos custos variáveis totais *standard*, *Paris*, *custo médio* e *elasticidades exógenas*.

Os resultados obtidos demonstram que o modelo de Programação Matemática Positiva reproduz com precisão o nível das actividades observadas na situação de referência, independentemente da regra utilizada para especificar a função dos custos variáveis de produção.

Relativamente à prescrição de resultados futuros, a Programação Matemática Positiva revelou ser também uma boa opção metodológica, especialmente se forem utilizadas na especificação da função dos custos variáveis de produção a regra das *elasticidades exógenas* ou a regra de *Paris*.

As regras de especificação da função dos custos variáveis baseadas no método *standard* e no método do *custo médio* revelaram uma menor capacidade de prescrição dos resultados futuros.

Pode concluir-se que as propriedades da Programação Matemática Positiva não se esgotam apenas na calibração exacta dos modelos de oferta agrícola, mas também se estendem à sua capacidade de prescrição de resultados futuros.

6 Referências

- Arfini, F., Donati, M., Zuppiroli, M. and Paris, Q. (2005) Ex-post evaluation of set-aside using Symmetric Positive Equilibrium Problem. Selected paper presented at the 89th EAAE symposium on "Modelling agricultural policies: state of the art and new challenges", Parma.
- Barkaoui, A. and Butault, J. (1999) Positive Mathematical Programming and Cereals and Oilseeds Supply with EU under Agenda 2000. Paper presented at the 9th European Congress of Agricultural Economists, Warsaw.
- Baskaoui, A., Butault, J. and Rousselle, J. (2001). Positive Mathematical Programming and Agricultural Supply within EU under Agenda 2000. In: Heckelei T., Witzke and W. Henrichsmeyer (Eds.): Agricultural Sector Modelling and Policy Information System, Vauk Verlag Kiel.
- Bauer, S. and Kanakoglu, H. (1990) Non Linear Programming Models for Sector Policy Analysis, Economic Modelling 7, pp. 272-90.
- Cypris, C. (2000) Positiv Mathematische Programmierung (PMP) im Agrarsektormodells RAUMIS. Dissertation, University of Bonn.
- Day, R. H. (1961) Recursive programming and the production of supply. In: Heady et al. (eds): Agricultural Supply Functions, Iowa State University Press, Iowa, USA.
- Gohn, A. and Chantreuil, F. (1999) La programmation mathématique positive dans les modèles d'exploitation agricole. Principes et l'importance du calibrage, Cahiers d'Economie et de Sociologie Rurales, 5, pp. 59-79
- Graindorge, C., Henry de Frahan, B. and Howitt, R. (2001) Analysing the effects of Agenda 2000 using a CES Calibrated Model of Belgian Agriculture. In: Heckelei T., Witzke and W. Henrichsmeyer (Eds.): Agricultural Sector Modelling and Policy Information System, Vauk Verlag Kiel, pp. 176-186.
- Hazell, P. and Norton, R. (1986) Mathematical Programming for Economic Analysis in Agriculture, Mac Millan Publishing Company, New York.
- Heckelei, T. and Britz, W. (2005) Models Based on Positive Mathematical Programming: State of the Art and Further Extensions, In Arfini, F. (Ed.): Modelling Agricultural Policies: State of the Art and New Challenges, Parma, Italy, pp. 48-73.

- Helming, J., Peeters, L. and Veendendaal, P. (2001) Assessing the Consequences of Environmental Policy Scenarios in Flemish Agriculture. In: Heckelei T., Witzke and W. Henrichsmeyer (Eds.): Agricultural Sector Modelling and Policy Information System, Vauk Verlag Kiel, pp. 237-245.
- Horner, G., Corman, J., Howitt, R. Carter, C. and Macgregor, R. (1992) The Canadian Regional Agriculture Model: Structure, Operation and Development. Agriculture, Canada. Technical Report 1/92, Ottawa.
- House, R. (1987) USMP Regional Agricultural Model. Washington DC: USDA. National Economics Division Report, ERS, 30.
- Howitt, R. (1995) Positive Mathematical Programming. American Journal of Agricultural Economics, 77 (2), pp. 329-342.
- Kasnakoglu, H. and Bauer, S. (1988) Concept and Application of an Agricultural Sector Model for Policy Analysis in Turkey. In: Agricultural Sector Modelling. S. Bauer and W. Henrichsmeyer (Eds.), Vauk Verlag: Kiel.
- Martin, L.R. (1977) A Survey of Agricultural Economics Literature. vol 2, University of Minnesota Press, Minneapolis.
- McCarl, B. And Spreen, T. (1980) Price Endogenous Mathematical Programming as a Tool for Sector Analysis, American Journal of Agricultural Economics, pp. 87-102.
- Meister, A., Chen,C. and Heady, E. (1978) Quadratic Programming Models Applied to Agricultural Policies, Iowa State University Press, Ames.
- Paris, Q. (1988) PQP, PMP, Parametric Programming and Comparative Static. Chapter 11 in Notes for AE 253. Department of Agricultural Economics, University of California, Davis.
- Peach, T. (1993) Interpreting Ricardo. Cambridge University Press, Cambridge.
- Throsby, C. (1974) New methodologies in agricultural production economics: A review. In The Future of Agriculture. Papers and reports, 15 th international Conference of Agricultural Economics, S. Paulo, Brazil, pp. 150-169.

Anexo 1 Formas de especificação da função dos custos variáveis totais no modelo de PMP

Standard

$$\sum_j c_j X_j - \frac{1}{2} \sum_j \frac{\rho_j}{X_j^0} X_j^2 + \sum_i c_i Y_i + \frac{1}{2} \sum_i \frac{\rho_i}{Y_i^0} Y_i^2 + phT + piE$$

Paris

$$\frac{1}{2} \sum_j \frac{c_j + \rho_j}{X_j^0} X_j^2 + \frac{1}{2} \sum_i \frac{c_i + \rho_i}{Y_i^0} Y_i^2 + phT + piE$$

Custo Médio

$$\sum_j (c_j - \rho_j) X_j + \frac{1}{2} \sum_j \frac{\rho_j}{X_j^0} X_j^2 + \sum_i (c_i - \rho_i) Y_i + \frac{1}{2} \sum_i \frac{\rho_i}{Y_i^0} Y_i^2 + phT + piE$$

Elasticidades Exógenas

$$\begin{aligned} & \sum_j (c_j + \rho_j - \frac{1}{\epsilon_j} \frac{p_j^0}{X_j^0}) X_j + \frac{1}{2} \sum_j \frac{1}{\epsilon_j} \frac{p_j^0}{X_j^0} X_j^2 \\ & + \sum_i (c_i + \rho_i - \frac{1}{\epsilon_j} \frac{p_i^0}{Y_i^0}) Y_i + \frac{1}{2} \sum_i \frac{1}{\epsilon_i} \frac{p_i^0}{Y_i^0} Y_i^2 + phT + piE \end{aligned}$$

Anexo 2 Coeficientes técnicos das actividades agroflorestais no modelo de PMP

	Valor da produção		Ajudas		Custos variáveis	Mão-de-obra
	2000 €/ha	2004 €/ha	2000 €/ha	2004 €/ha		
Trigo mole	340	181	65	129	260	17,9
Trigo duro	310	240	415	482	260	17,9
Milho	1550	1480	411	479	1154	19,1
Arroz	1545	1350	319	319	1490	15,8
Horto-frutícolas	3673	4052	-	-	1837	186,2
Girassol	158	123	90	129	152	17,4
Olival	375	390	-	-	362	29,0
Vinha	4160	4000	-	-	925	153,0
Fruticultura	3988	3910	-	-	615	537,1
Pastagens permanentes	-	-	-	-	43	1,2
Culturas forrageiras	-	-	-	-	246	14,4
Pousio obrigatório	-	-	75	129	-	-
Pousio	-	-	-	-	-	-
Matas e florestas	750	750	-	-	-	-
	€/CN	€/CN	€/CN	€/CN	€/CN	h/CN
Bovinos engorda novilhos	313	1512	205	317	228	29
Bovinos venda de vitelos	287	330	227	327	190	19
Ovinos de carne	293	280	167	200	203	4,8
Suínos alentejanos	669	669	-	-	475	4,8

Fonte: Contas de actividade agrícola, Staff do Departamento de Gestão da Universidade de Évora

Métodos de Penalidade Exacta para Resolução de Problemas de Optimização não Linear

Aldina Correia *

João Matias †

Carlos Serôdio ‡

* Escola Superior de Tecnologia e Gestão de Felgueiras
Instituto Politécnico do Porto
e Centro de Matemática da UTAD (CM-UTAD)
aldinacorreia@eu.ipp.pt

† Centro de Matemática da UTAD (CM-UTAD)
Universidade de Trás-os-Montes e Alto Douro
j_matias@utad.pt

‡ Centro de Investigação e de Tecnologias Agro-Ambientais e Biológicas (CITAB)
Universidade de Trás-os-Montes e Alto Douro
cserodio@utad.pt

Abstract

In this work we present a classification of some of the existing Penalty Methods (denominated the Exact Penalty Methods) and describe some of its limitations and estimated.

With these methods we can solve problems of optimization with continuous, discrete and mixing constrains, without requiring continuity, differentiability or convexity.

The boarding consists of transforming the original problem, in a sequence of problems without constrains, derivate of the initial, making possible its resolution for the methods known for this type of problems.

Thus, the Penalty Methods can be used as the first step for the resolution of constrained problems for methods typically used in by unconstrained problems.

The work finishes discussing a new class of Penalty Methods, for nonlinear optimization, that adjust the penalty parameter dynamically.

Resumo

Neste trabalho pretende apresentar-se uma classificação dos Métodos de Penalidade existentes (salientando os Métodos de Penalidade Exacta) e descrever algumas das suas limitações e pressupostos.

Esses métodos permitem resolver problemas de optimização com restrições contínuas, discretas e mistas, sem requerer continuidade, diferenciabilidade ou convexidade.

A abordagem consiste em transformar o problema original, numa sequência de problemas sem restrições, derivados do inicial, possibilitando a sua resolução pelos métodos conhecidos para este tipo de problemas.

Assim, os Métodos de Penalidade podem ser usados como o primeiro passo para a resolução de problemas de optimização permitindo a resolução de problemas com restrições por métodos tipicamente utilizados em problemas sem restrições.

O trabalho termina com a discussão de uma nova classe de Métodos de Penalidade, para optimização não linear, que ajustam o parâmetro de penalidade dinamicamente.

Keywords: Optimização não linear com restrições, Métodos de Penalidade, Métodos de Penalidade Exacta, Métodos de Penalidade Dinâmica.

Title: Exact Penalty Methods for Nonlinear Optimization Problems

1 Introdução

A modelação matemática tem sido utilizada para o estudo e compreensão de muitos problemas e fenómenos reais, na engenharia, economia, medicina, entre outras. Os problemas de optimização são abundantes, havendo a necessidade de determinar soluções que correspondam o melhor possível à realidade. Podem encontrar-se alguns desses problemas, por exemplo, em Ferris (1997).

As técnicas ou estratégias utilizadas em cada método dependem da quantidade de informação disponível, e passível de ser utilizada, da maior ou menor eficiência e robustez do algoritmo a utilizar, da facilidade de implementação, e, obviamente, das especificidades do próprio problema.

Quando o problema tem restrições (NLPs, do inglês *NonLinear Programming Problems*) são, muitas vezes, usadas as funções de penalidade, que vêm permitir efectuar uma transformação ao problema original, para que a solução seja obtida pela resolução, não de um, mas de uma sequência de outros problemas, derivados do inicial, todos eles sem restrições. Esta abordagem, possibilita, assim, a utilização de outro tipo de métodos ou algoritmos, nomeadamente a classe dos Métodos dos Gradientes ou a classe dos Métodos de Pesquisa Directa.

Aliás, as estratégias existentes para a resolução de problemas não lineares com restrições, consistem, em geral, em os transformar em problemas sem restrições, de resolução mais fácil, cuja solução é igual ou se relaciona de alguma forma com a solução do problema original.

Os métodos disponíveis para efectuar este procedimento são diversos, por exemplo:

1. Métodos de Penalidade;
2. Métodos Barreira;
3. Método dos gradientes reduzidos;
4. Método de projecção do gradiente;
5. Métodos baseados em funções Lagrangeanas aumentadas;
6. Métodos de Lagrangeanas projectadas;
7. Método dos filtros.

Neste trabalho começam por abordar-se, genericamente, os Métodos de Penalidade, para seguidamente se apresentar uma classificação de alguns dos Métodos de Penalidade existentes, apresentando algumas das suas limitações e pressupostos.

Nos últimos anos surgiu um grande interesse nos Métodos de Penalidade (em 2006, Byrd (2006), em 2005 Chen (2005), em 2002 (2002), em 2003 Gould (2003), Leyffer (2006) em 2006 , Klatte (2002) em 2002, em 1995 Mongeau (1995) e em 2005 Zaslavski (2005),

entre outros), por causa da sua capacidade de resolução de problemas degenerados e com restrições não lineares.

Os Métodos de Penalidade Exacta foram usados com sucesso para resolver problemas com restrições de Complementaridade (MPCCs - do inglês *Mathematical programs with complementary constrains*), por Benson (2003) e por Leyffer (2006), e, por Byrd (2006) e por Chen (2005), foram usados em programação não linear com restrições (CNLP, do inglês *Constrained NonLinear Programming*) para assegurar a admissibilidade de subproblemas e melhorar a robustez da iteração.

Assim, os Métodos de Penalidade podem ser usados como o primeiro passo para a resolução de problemas de optimização, pois permitem a resolução de problemas com restrições por métodos tipicamente utilizados em problemas sem restrições.

2 Formulação Matemática do Problema

De uma forma geral, a formulação de um problema, P , de optimização não linear com restrições, não lineares de igualdade, desigualdade e limites simples, pode ser posta na seguinte forma, Matias (2003):

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimizar}} & f(x) \\ \text{sujeito a} & e_i(x) \geq 0, \quad i = 1, 2, \dots, s \\ & d_j(x) = 0, \quad j = 1, 2, \dots, t \\ & a_k \leq x_k, \quad k = 1, 2, \dots, n \\ & x_l \leq b_l, \quad l = 1, 2, \dots, n \end{array} \quad (2.1)$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é a função objectivo do problema, $e_i : \mathbb{R}^n \rightarrow \mathbb{R}$, com $i = 1, 2, \dots, s$, são as s restrições de desigualdade, $d_j : \mathbb{R}^n \rightarrow \mathbb{R}$, com $j = 1, 2, \dots, t$, representam as t funções das restrições de igualdade, e as duas últimas condições representam os limites simples impostos às variáveis.

3 Métodos de Penalidade

Os Métodos de Penalidade foram criados para resolver um problema, P , resolvendo uma sequência de problemas sem restrições especialmente escolhidos. Ou seja, é efectuada uma transformação ao problema original e é feita a resolução de uma sequência de outros problemas, sem restrições, derivados do inicial, pelos métodos conhecidos para este tipo de problemas.

Assim, é construída uma nova função objectivo, Φ , que contém informação relativa à função objectivo inicial, f , e, simultaneamente, às restrições do problema. São, desta forma, construídos problemas sucessivos sem restrições, que dependem de um parâmetro positivo, r , cujas correspondentes soluções, $x^*(r)$, convergirão para a solução do problema inicial, x^* .

A **região admissível** de P é o conjunto R definido pelas condições:

$$\begin{array}{ll} e_i(x) \geq 0, & i = 1, 2, \dots, s \\ d_j(x) = 0, & j = 1, 2, \dots, t \\ a_k \leq x_k, & k = 1, 2, \dots, n \\ x_l \leq b_l, & l = 1, 2, \dots, n. \end{array} \quad (3.2)$$

Nos Métodos de Penalidade, a região admissível, R , é expandida a todo o \mathbb{R}^n , mas é aplicada uma grande penalização à função objectivo nos pontos que estão fora da região admissível original, R .

Considere-se o problema P , (2.1). Convertendo as restrições, $e_i(x) \geq 0$ em $-e_i(x) \leq 0$, $d_j(x) = 0$ em $d_j(x) \leq 0$ e $-d_j(x) \leq 0$ e $a_k \leq x_k$ e $x_l \leq b_l$ em $a_k - x_k \leq 0$ e $x_l - b_l \leq 0$, podemos assumir P na forma:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimizar}} & f(x) \\ \text{sujeito a} & C_i(x) \leq 0, \end{array} \quad (3.3)$$

onde $C_i(x)$ representa o conjunto de todas as $m = s + t + 2n$ restrições, $c_i(x) \leq 0$, $i = 1, 2, \dots, m$, do problema.

Assim, o problema original, com restrições de todo o tipo, é agora um problema com restrições apenas de desigualdade da forma menor ou igual.

Então pode construir-se uma nova função objectivo, Φ , que contém informação relativa à função objectivo inicial, f , e simultaneamente, às restrições do problema:

$$\Phi(x, r) = f(x) + \Theta(x, r), \quad (3.4)$$

onde $\Theta : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$ é uma função de r , parâmetro real positivo, denominado **parâmetro de penalidade** e da função de penalidade, sendo $\Theta(x, r) = rp(x)$.

Definição 1 - Freund (2004)

Uma função $p : \mathbb{R}^n \rightarrow \mathbb{R}$ diz-se **função de penalidade** para P , se verifica:

- $p(x) = 0$ se $c_i(x) \leq 0$
- $p(x) > 0$ se $c_i(x) > 0$

Assim, o problema a resolver, que substitui o problema P , é um problema, P_m , com uma nova função objectivo e sem restrições:

$$P_m : \underset{x \in \mathbb{R}^n}{\text{minimizar}} f(x) + r_m p(x), \quad (3.5)$$

representando r_m uma sucessão crescente de constantes tal que $r_m \rightarrow +\infty$.

Seja $r_k \geq 0$, $k = 1, 2, \dots, \infty$ a sucessão de parâmetros de penalidade, tal que, $r_{k+1} > r_k, \forall k$ e $\lim_{k \rightarrow +\infty} r_k = +\infty$.

Sejam $q(x, r) = f(x) + rp(x)$, x^k a solução exacta do problema $P(r_k)$ e x^* a solução óptima de P .

Nestas condições prova-se o **Teorema da Convergência dos Métodos de Penalidade**, com f a função objectivo, c_i as funções restrição e p a função de penalidade:

Teorema 1 - Freund (2004)

Suponha-se que f , c_i e p são funções contínuas. Seja $\{x^k\}, k = 1, 2, \dots, \infty$ uma sucessão de soluções de $P(r_m)$, então qualquer ponto limite \bar{x} de $\{x^k\}$ é solução óptima de P .

3.1 Funções de Penalidade usadas frequentemente

Uma classe de Funções de Penalidade usada frequentemente é:

$$p(x) = \sum_{i=1}^m [\max \{0, c_i(x)\}]^\rho, \rho \geq 1, \quad (3.6)$$

- se $\rho = 1$, $p(x)$ em (3.6) diz-se **função de penalidade linear**.

Nota: Esta função pode não ser diferenciável nos pontos onde $c_i(x) = 0$, para algum i .

- (3.6) com $\rho = 2$, é a função de penalidade mais usada na prática e, diz-se **função de penalidade quadrática**.

3.2 Métodos de Penalidade Exacta - Definição

A característica principal dos Métodos de Penalidade Exacta é que, ao contrário dos Métodos de Penalidade Inexacta, apresentados anteriormente, que transformavam o problema original numa sequência infinita de problemas, a sequência de problemas é finita, ou seja, a solução do problema NLP é determinada resolvendo um número finito de problema sem restrições.

Definição 2 - Zaslavski (2005)

Um Método de Penalidade Exacta é um método que escolhe a função de penalidade $p(x)$ e o parâmetro r de tal forma a que a solução óptima \tilde{x} do problema sem restrições, P_m , seja também a solução óptima, x^* , do correspondente problema com restrições, P .

4 Classificação de alguns dos Métodos de Penalidade existentes

Nesta secção faz-se uma revisão e apresenta-se uma classificação para alguns dos Métodos de Penalidade existentes. São ainda apresentados alguns dos seus pressupostos e limitações.

Na tabela seguinte apresenta-se resumidamente essa classificação.

Tabela 1: Uma Classificação de alguns dos Métodos de Penalidade existentes

Optimização Global	Penalidade Exacta	Penalidade Estática
		Penalidade Dinâmica
	Penalidade Inexacta	Penalidade de Recusa
		Penalidade Discreta
Optimização Local	Penalidade Exacta	Lagrangeana
		Penalidade l_1

Uma das dificuldades dos Métodos de Penalidade é encontrar parâmetros de penalidade convenientes de forma a que a solução \hat{x} , que minimiza o problema sem restrições corresponda, de alguma forma, ao mínimo do respectivo problema com restrições.

Se o mínimo do problema sem restrições é admissível e correspondente ao mínimo global do problema com restrições (CGM, do inglês *Constrained Global Minimum*), ou seja, se é o ponto da região admissível com melhor valor da função objectivo, diz-se que o método utilizado é um *Método de Optimização Global*. Se esse mínimo, correspondente a um mínimo local do problema com restrições (CLM, do inglês *Constrained Local Minimum*), ou seja, se é o ponto de uma vizinhança admissível pré-definida, com melhor valor da função objectivo, diz-se que o método utilizado é um *Método de Optimização Local*.

Assim, os Métodos de Penalidade podem ser classificados como:

- MÉTODOS DE PENALIDADE DE OPTIMIZAÇÃO GLOBAL, (GOPM, do inglês *Global Optimal Penalty Methods*), se permitem obter soluções globais, CGM, de P
- MÉTODOS DE PENALIDADE DE OPTIMIZAÇÃO LOCAL, (LOPM, do inglês *Local Optimal Penalty Methods*), se permitem obter soluções locais, CLM, de P .

Por outro lado, podem ser (Bertsekas (1999)):

- MÉTODOS DE PENALIDADE INEXACTA, nos quais a minimização da nova função objectivo, Φ , não encontra os pontos CGM e CLM exactos, apenas caminha para pontos infinitamente próximos destes, pela minimização sucessiva dos problemas obtidos;
- MÉTODOS DE PENALIDADE EXACTA, se permitem encontrar os CGM e CLM exactos, através de uma sequência finita de problemas de penalidade.

4.1 Métodos de Penalidade de Optimização Global - GOPM

Os GOPM podem ser Inexactos ou Exactos. Nesta classe de métodos estão Métodos de Penalidade Estática e Métodos de Penalidade Dinâmica.

4.1.1 Métodos de Penalidade Estática

Os MÉTODOS DE PENALIDADE ESTÁTICA foram propostos inicialmente por Homaifar (1994). Nestes métodos é considerada uma família de níveis da violação para cada tipo de restrição. Cada nível da violação impõe um nível diferente da penalidade. A desvantagem deste método é o número de parâmetros a ser seleccionado. O número de parâmetros aumenta rapidamente com o número de restrição e de níveis da violação.

Nestes métodos é fixado um parâmetro de penalidade, r para todo o processo. Existem duas dificuldades associadas a esta abordagem, Deb (2001), Godfrey (2004):

- A solução da função objectivo depende do parâmetro de penalidade, r . Diversos autores procuraram encontrar o melhor valor para r , nestes métodos, para conduzir a procura dentro da região admissível. Isto requereu uma extensa experimentação para encontrar uma solução razoável. Este problema é tão indomável que alguns investigadores usaram diferentes valores de r , dependendo do nível de violação das restrições, enquanto outros actualizaram o parâmetro de penalidade a cada iteração a partir de parâmetros que fixavam o raio de evolução (métodos de penalidade dinâmica)
- A inclusão do termo de penalidade distorce a função objectivo, Deb (2001). Para valores pequenos de r , a distorção é pequena, mas o óptimo de $\Phi(x, r)$ pode não estar perto do verdadeiro óptimo. Por outro lado, se é usado um r grande, o óptimo de $\Phi(x, r)$ é próximo do mínimo, mas a distorção pode ser tão grande que $\Phi(x, r)$ pode ter mínimos fictícios.

Considerando, as restrições $a_i(x) \geq 0$ como $-a_i(x) \leq 0$, $i = 1, 2, \dots, s$; $a_k - x_k \leq 0$, $k = 1, 2, \dots, n$ e $x_l - b_l \leq 0$, $l = 1, 2, \dots, n$, como restrições do tipo menor ou igual, $g_i(x) \leq 0$, $i = 1, 2, \dots, s + 2n$, (2.1), pode ser escrito como:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimizar}} & f(x) \\ \text{sujeito a} & d_j(x) = 0, \quad j = 1, 2, \dots, t \\ & g_i(x) \leq 0, \quad i = 1, 2, \dots, s + 2n \end{array} \quad (4.7)$$

Com os vectores de penalidade α e β , um exemplo de Problema de Penalidade Estática, para (4.7), $\rho \geq 1$, é:

$$\underset{x \in \mathbb{R}^n}{\text{minimizar}} L_s(x, \alpha, \beta) \quad (4.8)$$

com

$$L_s(x, \alpha, \beta) = f(x) + \sum_{j=1}^t \alpha_j |d_j(x)|^\rho + \sum_{i=1}^{s+2n} \beta_i (\max(0, g_i(x)))^\rho. \quad (4.9)$$

Um Método de Penalidade Estática pode ser Exacto ou Inexacto. Por exemplo, no Problema (4.8), se $\rho = 1$ em (4.9), estamos na presença de um Método de Penalidade Estática Exacto, se $\rho > 1$ é um Método de Penalidade Estática Inexacto, Bertsekas (1999).

Isto é, quando $\rho = 1$ existem valores de penalidade α e β tais que o ponto que minimiza a função de penalidade é exactamente o CGM de P .

Contudo, quando $\rho > 1$, o Método de Penalidade Estática é um Método Inexacto e convergirá para CGM como aproximação infinita dos valores de penalidade.

O Método de Penalidade Estática de Homaifar, Homaifar (1994), resolve um problema semelhante a (4.8), mas requer a escolha de um número muito grande de parâmetros e além disso é um Método de Penalidade Inexacta.

Assim, a limitação comum de todos os métodos de penalidade estática é que geralmente é muito difícil escolher os valores apropriados de penalidade estaticamente. Além disso, estes métodos foram desenvolvidos para encontrar CGM e não permitem encontrar um CLM de P . Portanto, são computacionalmente dispendiosos porque envolvem a procura de um mínimo global de uma função de penalidade não linear.

Como alternativa à procura de parâmetros de penalidade por tentativa em erro existem os Métodos de Penalidade Dinâmica.

4.1.2 Métodos de Penalidade Dinâmica

Os MÉTODOS DE PENALIDADE DINÂMICA, Wang (2006), discutidos na secção 5, incrementam as penalidades em (4.8) gradualmente encontrando o mínimo global \tilde{x} de (4.8), para cada combinação de penalidades, terminando quando \tilde{x} é uma solução admissível de P .

Tal como os Métodos de Penalidade Estáticos, os Métodos de Penalidade Dinâmica podem ser métodos exactos ou inexactos, dependendo do valor de ρ . Além disso, têm uma limitação comum com os anteriores, uma vez que também só permitem encontrar um mínimo global de funções não lineares.

Existem muitas variantes dos Métodos de Penalidade Dinâmica.

Uma variante amplamente conhecida é o *Método de não estacionaridade*, que resolve uma sequência de problemas do tipo de (4.8), verificando em cada iteração de cada subproblema k , as condições que se seguem, com $C > 0$ e $\rho > 1$ parâmetros constantes:

$$\alpha_j(k+1) = \alpha_j(k) + C \cdot |d_j(x)| \quad (4.10)$$

$$\beta_i(k+1) = \beta_i(k) + C \cdot (\max(0, g_i(x))) \quad (4.11)$$

Outro Métodos de Penalidade Dinâmica é o *Método de Penalidade Adaptativa*, que faz uso do *feedback* do processo de pesquisa. Este método resolve o seguinte problema na iteração k :

$$\underset{x \in \mathbb{R}^n}{\text{minimizar}} \ L_k(x, \alpha, \beta) \quad (4.12)$$

com

$$L_k(x, \alpha, \beta) = f(x) + \sum_{j=1}^t \alpha_j(k) (d_j(x))^2 + \sum_{i=1}^{s+2n} \beta_i(k) (\max(0, g_i(x)))^2, \quad (4.13)$$

onde $\alpha_j(k)$ é, respectivamente, incrementado, decrementado ou deixado inalterado quando a restrição $d_j(z) = 0$ é, respectivamente, não admissível ou admissível ou nenhuma delas, nas últimas l iterações. Isto é:

- $\alpha_j(k+1) = \alpha_j(k)/\lambda_1$
 - se $d_j(x(i)) = 0$ é uma iteração admissível $k-l+1, \dots, k$

- $\alpha_j(k+1) = \lambda_2 \cdot \alpha_j(k)$
 - se $d_j(x(i)) = 0$ é uma iteração não admissível $k-l+1, \dots, k$
- $\alpha_j(k+1) = \alpha_j(k)$
 - caso contrário.

onde l é um inteiro positivo, $\lambda_1, \lambda_2 > 1$ e $\lambda_1 \neq \lambda_2$ (para evitar ciclos nas actualizações). O parâmetro β é actualizado de forma similar.

Existem outras duas variações de penalidade global, ambas abordagens exactas, são os Métodos de Penalidade de Recusa e os Métodos de Penalidade Discreta.

4.1.3 Métodos de Penalidade de Recusa e Métodos de Penalidade Discreta

Um MÉTODO DE PENALIDADE DE RECUSA (*Death penalty Methods*), Hu (2002), Zhang (2005), é um método de penalidade que rejeita simplesmente todos os pontos não admissíveis. Começa com um ou mais pontos admissíveis e procura novos pontos, se o novo ponto não for admissível é rejeitado.

Nesta categoria, a dificuldade reside em gerar os pontos admissíveis iniciais e computacionalmente é ainda mais complicado, nomeadamente quando a região admissível é muito pequena.

Assim, este método resolve o Problema (4.7) considerando:

$$L_p(x, \alpha, \beta) = f(x) + \alpha^T P(d(x)) + \beta^T Q(g(x)),$$

onde

$$\begin{aligned} P(d(x)) &= +\infty \text{ se } d(x) \neq 0 \\ &\quad \text{e} \\ P(d(x)) &= 0 \text{ se } d(x) = 0 \\ &\quad \text{e} \\ Q(g(x)) &= +\infty \text{ se } g(x) > 0 \\ &\quad \text{e} \\ Q(d(x)) &= 0 \text{ se } g(x) \leq 0. \end{aligned}$$

Este método é um Método de Penalidade Exacta. Para quaisquer valores finitos dos parâmetros de penalidade α e β , o ponto mínimo da função penalidade será admissível e terá o menor valor da função objectivo, sendo dessa forma exactamente o CGM de P.

Outro Método de Penalidade Exacta é o MÉTODO DE PENALIDADE DISCRETA que usa o número de restrições que foram violadas em vez do grau de violações na função penalidade. Este tipo de métodos é usada muitas vezes em Métodos de Elementos Finitos, Dai (2007).

Conclui-se assim, que os Métodos de Optimização Global de (4.7) têm aplicação prática limitada, porque a procura do mínimo global é computacionalmente dispendiosa e as técnicas de optimização global, como o método de não estacionaridade, também são lentas pois só alcançam óptimos globais com convergência assimptótica, Kirkpatrick (1983).

4.2 Métodos de Penalidade de Optimização Local - LOPM

Para evitar a dispendiosa optimização global têm sido desenvolvidos métodos de optimização local. Estes incluem, por exemplo, os Métodos dos Multiplicadores de Lagrange e os Métodos de Penalidade l_1 , que são ambos Métodos de Penalidade Exacta. Estes métodos foram criados para resolver problemas de optimização não linear contínuos, isto é problemas do tipo (4.7), onde f é contínua e diferenciável e g e d podem ser descontínuas, não-diferenciáveis.

Nestes métodos o objectivo é encontrar um mínimo local \check{x} , relativamente à vizinhança $N(\check{x}) = \{x^* : \|x^* - \check{x}\| \leq \epsilon \text{ e } \epsilon \rightarrow 0\}$ de x^* .

Definição 3 - Gould (2003)

Um ponto \check{x} diz-se um mínimo local de P_m relativamente à vizinhança $N(\check{x})$, se \check{x} é admissível e $f(\check{x}) \leq f(x)$ para todo o x admissível dessa vizinhança.

4.2.1 Métodos dos Multiplicadores de Lagrange

A Teoria Lagrangeana tradicional funciona para (4.7) com funções restrição g e h contínuas e diferenciáveis.

A Função Lagrangeana de (4.7) com multiplicadores de Lagrange λ e μ é definida por:

$$L(x, \lambda, \mu) = f(x) + \lambda^T d(x) + \mu^T g(x).$$

Assumindo a continuidade e diferenciabilidade, o CLM satisfaz a condição necessária de Karush-Kuhn-Tucker (KKT) e a condição suficiente de ponto sela, Bertsekas (1999).

Esta abordagem é, portanto, limitada à resolução de CNLPs com funções contínuas e diferenciáveis e não pode ser aplicada a problemas discretos ou a problemas onde não se conheçam as derivadas das funções. Esta limitação deve-se ao facto da existência dos multiplicadores de Lagrange depender da existência dos gradientes das restrições e da função objectivo e da regularidade das restrições (independência linear dos gradientes das restrições) nos pontos solução.

Além disso, é difícil encontrar pontos e multiplicadores que satisfaçam a condição suficiente de ponto sela, porque esta é expressa como sistema de desigualdades não lineares, nem sempre fácil de resolver. Assim, esta condição é usada simplesmente para verificar se as soluções encontradas pela condição KKT são, efectivamente, soluções óptimas.

4.2.2 Métodos de Penalidade l_1

O MÉTODOS DE PENALIDADE l_1 é outro método de penalidade exacta de minimização local e que permite, portanto, resolver CNLPs. Este método resolve problemas de minimização com a seguinte função, Gould (2003):

$$l_1(x, \mu) = f(x) + \mu \sum_{j=1}^t |d_j(x)| + \mu \sum_{i=1}^{s+2n} \max[g_i(x), 0] \quad (4.14)$$

A teoria desenvolvida à volta desta expressão mostra que existe uma correspondência bijectiva entre o CLMs e o mínimo global da função l_1 (4.14), quando μ é suficientemente grande, Bertsekas (1999).

Como é sabido, para valores apropriados do parâmetro de penalidade μ , os pontos estacionários de $l_1(x, \mu)$ são também pontos KKT do problema não linear (4.7) ou pontos estacionários não admissíveis, ver por exemplo, Byrd (2003). Esta propriedade é a mais apelativa destes métodos de penalização exacta porque uma escolha adequada de μ deverá ser adequada para todo o processo de minimização.

Este como os outros Métodos de penalização exacta são, assim, menos dependentes no parâmetro de penalidade do que os métodos de penalização quadrática para os quais é necessário resolver uma sucessão de subproblemas com séries divergentes de parâmetros de penalidade.

A função (4.14) serviu de base a muitos métodos de penalidade propostos na literatura.

Apresenta-se de seguida o correspondente Algoritmo, Byrd (2006):

Algoritmo: Método de penalidade l_1 clássico

- Dados:
 - $\mu_0 > 0$
 - a tolerância $\tau > 0$
 - um ponto inicial x_0^s
- **Para** $k = 0, 1, 2, \dots$
 - Encontrar um minimizante aproximado x_k de $l_1(x, \mu)$, começando em x_k^s ;
 - Se** $\sum_{j=1}^t |d_j(x)| + \sum_{i=1}^{s+2n} \max[-g_i(x), 0] \leq \tau$
 - Parar e considerar a solução aproximada x_k ;
 - Senão**
 - Escolher um novo parâmetro de penalidade $\mu_{k+1} > \mu_k$;
 - Escolher um novo ponto inicial x_{k+1}^s ;

A minimização da função de penalidade l_1 , $l_1(x, \mu)$, é difícil porque é não diferenciável. Como resultado destes obstáculos, esta aproximação sem restrições, não é infelizmente viável como técnica para a programação não linear.

A limitação maior deste método, tal como nos Métodos de Multiplicadores de Lagrange, é que só se aplicam a problemas contínuos e diferenciáveis e é difícil encontrar um valor de μ consistente para todos os subproblemas.

As alternativas a estes métodos têm vindo a aparecer, nomeadamente no que diz respeito à procura de soluções para o problema da escolha de parâmetros de penalidade.

5 Uma nova classe de Métodos de Penalidade Dinâmica

Os métodos de penalidade cresceram em três etapas de desenvolvimento desde que foram introduzidas em 1950. Primeiro foram vistos como meio de resolução de problemas de opti-

mização com restrições através de técnicas de optimização sem restrições. Esta abordagem não provou ser efectiva, excepto para classes especiais de aplicações.

Na segunda etapa, os problemas de penalidade foram substituídos por uma sucessão de subproblemas com restrições lineares. Estas técnicas, descritas nas abordagens de programação quadrática sequencial, são muito mais eficientes do que as aproximações sem restrições, mas deixam em aberto a questão de como escolher o parâmetro de penalidade.

Infelizmente a escolha dos parâmetros de penalidade era, frequentemente, muito difícil porque a maioria das técnicas utilizadas são estratégias heurísticas.

Como alternativa a estas estratégias apareceu o método dos filtros, introduzidos por Fletcher and Leyffer, Fletcher (2002). Desde aí, as técnicas de filtros foram aplicadas a métodos SLP (Sequential Linear Programming) e SQP (Sequential Quadratic Programming), porque constituíam métodos menos dependentes de parâmetros do que os métodos de Penalidade.

Na etapa mais recente do desenvolvimento, os métodos de penalidade ajustam o parâmetro de penalidade em cada iteração, com o intuito de atingir um nível estipulado de admissibilidade linear. A escolha do parâmetro de penalidade deixa de ser heurística e torna-se parte integral do processo de cálculo.

Um exemplo deste tipo de abordagem é a apresentada em Byrd (2006), no contexto do algoritmo da progressão linear quadrática sequencial (SLQP). Neste artigo é feita uma análise e generalização desta estratégia para a sua aplicação em outros métodos.

A estratégia consiste em ajustar o parâmetro de penalidade dinamicamente; por controlo do grau de linearidade admissível considerado em cada iteração, elas promovem um progresso balanceado de optimabilidade e admissibilidade. Para escolher um parâmetro de penalidade que permita cumprir o objectivo deve resolver-se um subproblema adicional em cada iteração.

Em contraste com as aproximações clássicas, a escolha do parâmetro de penalidade deixa de ser uma heurística e é determinado, através de um subproblema com objectivos claramente definidos. A nova estratégia de actualização da penalidade é apresentada no contexto de métodos de programação quadrática sequencial (SQP) e programação linear quadrática sequencial (SLQP), métodos que usam regiões admissíveis para promover a convergência.

A abordagem apresentada em Byrd (2006), consiste em fazer uma reformulação do problema linearizando as restrições, depois, exigindo que em cada passo haja progresso na admissibilidade flexível (*linear feasibility*) que é proporcional ao possível progresso óptimo. Esta nova estratégia, incrementa automaticamente o parâmetro de penalidade e supera o comportamento indesejável da dificuldade de escolher valores apropriados para μ_k nos métodos de penalidade. Esta dificuldade, causou que os métodos de penalidade não diferenciáveis perdessem o interesse durante os anos 90 e originaram o desenvolvimento do método dos filtros, Gonzaga (2003), Wachter (2004), que não exigem um parâmetro de penalidade.

Outras estratégias de actualização de penalidade foram propostas recentemente. Chen e Goldfarb, Chen (2005), propuseram regras que actualizam o parâmetro de penalidade baseadas na admissibilidade e no tamanho dos multiplicadores. Leyffer, Leyffer (2006), considera métodos de penalidade descrevendo critérios dinâmicos para actualizar parâmetros de penalidade baseado no decrescimento médio das restrições penalizadas.

6 Conclusão

Neste trabalho apresenta-se uma Classificação dos Métodos de Penalidade existentes e descrevem-se algumas das suas limitações e pressupostos.

Esta Caracterização baseia-se essencialmente na distinção dos métodos de acordo com o tipo de óptimo que permitem encontrar (global ou local) e na distinção no que diz respeito à exactidão das soluções encontradas (Métodos exactos ou inexactos).

Se o mínimo do problema sem restrições é admissível e correspondente ao mínimo global do problema com restrições (CGM) diz-se que o método utilizado é um Método de Optimização Global (GOPM). Se esse mínimo, correspondente a um mínimo local do problema com restrições (CLM), diz-se que o método utilizado é um Método de Optimização Local (LOPM).

Se a minimização da nova função objectivo, Φ , não encontra os pontos CGM e CLM exactos, apenas caminha para pontos infinitamente próximos destes, pela minimização sucessiva dos problemas obtidos diz-se que o método utilizado é um Métodos de Penalidade Inexacta. Se o método permite encontrar os CGM e CLM exactos, através de uma sequência finita de problemas de penalidade diz-se que o método utilizado é um Métodos de Penalidade Exacta.

O trabalho termina com a discussão de uma nova classe de Métodos de Penalidade, para optimização não linear, que ajustam o parâmetro de penalidade dinamicamente.

7 Referências

- Benson, H.Y, Sen, A., Shanno, D.F and Vanderbei, R. J. (2003) Interior-Point Algorithms, Penalty Methods and Equilibrium Problems, Technical Report ORFE-03-02, Department of Operations Research and Financial Engineering, Princeton University, Princeton NJ, 08544.
- Bertsekas, D. P. (1999) Nonlinear Programming, Athena Scientific, Belmont, Massachusetts.
- Byrd, R. H., Nocedal, J. e Waltz, R. A. (2006) Steering Exact Penalty Methods for Optimization, Technical Report, Optimization Technology Center, Northwestern University, Evanston, IL 60208, USA.
- Byrd, R. H., Gould, N. I., Nocedal, J. e Waltz, R. A. (2002) On the convergence of successive linear-quadratic programming algorithms, Technical Report OTC 2002/5, Optimization Technology Center, Northwestern University, Evanston, IL, USA.
- Chen, L. and Goldfarb, D. (2005) Interior-point l2-penalty methods for nonlinear programming with strong global convergence properties, Mathematical Programming, Technical report, IEOR Dept, Columbia University, New York.
- Dai, X. (2007) Finite element approximation of the pure Neumann problem using the iterative penalty method, *Applied Mathematics and Computation*, 186(2):1367-1373.
- Deb, K. (2001) Multi-Objective Optimization Using Evolutionary Algorithms, John Wiley and Sons.
- Ferris, M. C. and Pang, J. S. (1997) Engineering and Economic Applications of Complementarity Problems, SIAM Review, 39-4: 669-713.
- Fletcher, R. and Leyffer, S. (2002) Nonlinear Programming Without a Penalty Function, *Mathematical Programming*, 91(2):239-270.
- Freund, Robert M. (2004) Penalty and Barrier Methods for Constrained Optimization, Massachusetts, Institute of Technology.
- Godfrey C. Onwubolu, and Babu, B. V. (2004) New Optimization Techniques in Engineering, Springer.
- Gonzaga, C.C., Karas, E. and Vanti, M. (2003) A Globally Convergent Filter Method for Nonlinear Programming, SIAM J. Optimization, 14(3):646-669.

- Gould, N. I., Orban, D. e Toint, P. L. (2003) An interior-point l_1 -penalty method for nonlinear optimization, Technical Report RAL-TR-2003-022 Rutherford Appleton Laboratory Chilton, Oxfordshire, UK.
- Homaifar, A., Lai, S. H. V. and Qi, X. (1994) Constrained Optimizatin via generic algorithms; Simulation 62(4), 242-254.
- Hu, X. and Eberhart, R. (2002) Solving constrained nonlinear optimization problems with particle swarm optimization,Proceedings of the Sixth World Multiconference on Systemics, Cybernetics and Informatics 2002 (SCI 2002), Orlando, USA.
- Kirkpatrick, S., Gelatt, C. D. Jr. and Vecchi, M.P. (1983) Optimization by simulated annealing, Science, 220(4598):671-680.
- Klatte, D. and Kummer (2002) Constrained Minima and Lipschitzian Penalties in Metric Spaces, SIAM J. on Optimization, 13(2):619-633.
- Leyffer, S., López-Calva, G., and Nocedal (2006) Interior Methods for Mathematical Programs with Complementarity Constraints, SIAM J. on Optimization 17(1):52-77.
- Matias, J. L. H. (2003) Técnicas de Penalidade e Barreira Baseadas em Métodos de Pesquisa Directa e a Ferramenta PNL-Pesdir, Tese de Doutoramento, UTAD.
- Mongeau, M. and Sartenaer, A. (1995) Automatic decrease of the penalty parameter in exact penalty function methods, European Journal of Operational Research,83(3):686-699.
- Wächter, A. and Biegler, L.T. (2004) On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming, Tech. Rep. RC 23149, IBM T. J. Watson Research Center, Yorktown - USA.
- Wang, F.Y.and Liu, D. (2006) Advances in Computational Intelligence: Theory and Applications, World Scientific, ISBN 9812567348.
- Zaslavski,A. J. (2005) A Sufficient Condition for Exact Penalty in Constrained Optimization, SIAM Journal on Optimization, 16(1):250-262.
- Zhang, S. (2005) Constrained Optimization by ϵ Constrained Hybrid Algorithm of Particle Swarm Optimization and Genetic Algoritm, Proceedings of AI 2005: Advances in Artificial Intelligence, Springer.

Proposal of a heuristic model using genetic algorithms to solve an operational port problem

Vanina Macowski Durski Silva †

Antônio Sérgio Coelho ‡

Sérgio Fernando Mayerle ‡

† Universidade Federal de Santa Catarina

Brasil

vaninadurski@gmail.com

mayerle@deps.ufsc.br

‡ Universidade Federal de Santa Catarina

Brasil

coelho@deps.ufsc.br

Abstract

This article is characterized by the presentation of some problems found in ports that emphasize the Berth Allocation Problem, for which a heuristic model has been proposed to resolve this problem. This heuristic proposal is based on Genetic Algorithms and its objective is to allow the learning of this subject, besides urging the reader to implement a computer-based tool that uses these heuristics to solve this problem, practically and efficiently.

Keywords: Genetic algorithm; berth allocation problem; Competitiveness.

Resumo

O presente artigo caracteriza-se pela apresentação de alguns dos problemas portuários, enfatizando o Problema de Alocação de Navios em Berços, para o qual se propõe um modelo heurístico de resolução. A heurística proposta baseia-se nos Algoritmos Genéticos e tem por objetivo tornar possível o aprendizado deste assunto, além de instigar o leitor a implementar uma ferramenta computacional baseada nesta heurística para resolver o problema de maneira prática e eficiente.

Palavras-chave: Algoritmo genético; Problema de alocação de berços; Competitividade.

1 Introduction

The administration of a maritime port complex involves many decision-making problems, which can happen at: strategic, tactical and operational levels. There are many operational problems and methods addressed in technical literature; however, they are aimed at the sizes of the mooring berths, which are compatible with an expected demand

of vessels. Other studies seek to simulate operations that consider costs, investments and responsibilities, in other words, directed towards a system of economic operational sizing of container terminals (FERNANDES, 2001).

According to Silva and Coelho (2007), a gap still remains to be explored regarding the research and methods of relevant operational problems that are found in port systems: the berth allocation problem, BAP.

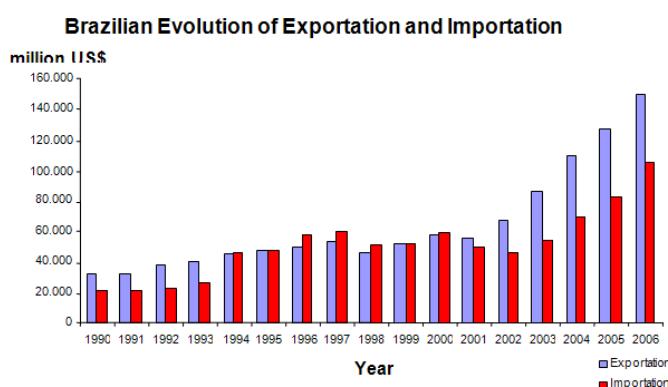
In this case, the article is divided into 4 chapters, including this introduction. In chapter 2, transport will be presented as an economic developer, specifically related to maritime transport and, also, information relating to the port operation, as well as the problem under study in this article. In Chapter 3, a heuristic method is proposed that solves the BAP, for which the main variables involved in problem have been identified. The model that has been developed endeavors to determine the sequence of vessels allocation to mooring berths, therefore minimizing the total time spent servicing the vessels, as well as the operational cost. Finally, in Chapter 4, the final considerations, conclusions and recommendations are shown for future studies

2 Contextualization and Systemic View of the Maritime Transport System

According to Novaes (2004) the distribution of products from factories to wholesale or retail outlets can be done using several transportation methods: highway, rail, waterway transports, by air, pipe work systems for special cases (gas, gasoline, diesel, ethanol). Like this, the investments in transports such as in any project, should be preceded by technical-economical viability studies. For this, it is necessary to carry out a discerning comparative analysis of the costs and benefits of the project (PEIXOTO, 1977).

2.1 The Importance of the Maritime Transport and Costs

Waterway transport includes all types of transport made on water. It includes the fluvial and lake transport (inland waterways) and maritime transport (NOVAES, 2004). According to Peixoto (1977) the technological progress and the increase in the world's population will cause a great increase in the demand for some items and therefore, as can be seen in Graph 1 both the exported and imported volume show a significant growth.



Graph 1 –Brazilian export and import volumes (1990-2006)

Source: Site of the Ministry of the Development, Industry and Overseas Trade

Therefore, to meet the whole demand that may occur, marine transport should concentrate its efforts on the recovering the transportation deficiency that has become worse over recent decades, this may done by the construction of new ports or the refurbishing of existing ones, the construction of shipyards, the upgrading of the

merchant fleet, the training of labor or even the search for the optimization of the current control systems for the ports operation, which have been obstacles to the country's economic expansion.

Adapting old ports to new cargo operational methods, requires large investments, in an attempt to reduce the length of time that the vessel remains in the port, thereby maximizing its performance for the owner and reducing the operational cost with the use of mechanization. Where investments do not occur, the cargo and discharge operations of the vessels are slow, the costs and the losses are high.

2.2 Port Operations

2.2.1 Some Problems Found in Ports

Several problems are encountered in ports, some of them are shown below:

- Problems with the acquisition or rental of equipment;
- Problems with the rental of equipment to be used in port services;
- Problems with the sizes of the berths;
- Problems with the port layout;
- Problem with the allocation of vessels to berths.

According to Nishimura et al. (2001), the majority of berths in large ports are rented, in other words, rented by the ship's operators for the containers processing to obtain a greater productivity.

In this context, it is interesting to limit the number of effective berths. The allocation of vessels in this system, that is, the attribution of berths to vessels that dock for loading/unloading, the reduction of time required to carry out this task becomes important; because the handling time for a specific vessel will not necessarily be the same for each berth.

2.2.2 Definition of the Allocation Problem

According to Guan et al. (2004), the problem of allocating space in the berths for vessels in port terminals, is considered as being the berth allocation problem. In the definition of Moon (2000) the problem consists of determining the moment of the mooring and the positions for each vessel in the port terminal. A more complete definition is that the problem consists of determining the allocation plan for the vessel that moor in the port in the berths, so that each vessel is allocated a berth for a length of time to do the loading/unloading activities with the intention of reducing operational costs.

Because of berth space being very limited and thousands of containers are handled daily, an effective allocation of berths becomes critical for the efficient administration of the flow of containers. When there is no available space at the berths, the vessel needs to wait before mooring and, this time should be the least possible, because a stopped vessel generates unnecessary arrears in delivery schedules, as well as extra costs (Waterway Survey CNT, 2006).

However, the mooring position is also a very important variable in the decision for the following reasons. The containers that are to be loaded onto the vessels arrive at the port several days before the vessels arrive at the port. Therefore, if a vessel is moored near to the container storage area, of the containers to be loaded on the vessel, the transportation cost of the containers by trucks or other internal transport equipment that is used in the port, can be minimized because the distance traveled will be less than if the storage area of the containers is far from the mooring berth.

2.2.3 BAP Mathematical Formulation

Amongst the formulations found in the bibliographical survey, the work done by Imai et al. (2001) stands out, initially developed for a static model, in other words, when is considered that all vessels arrive in the port, for later allocate each one to an available berths, always respecting the restrictions.

Later the authors solved this problem with the use of a dynamic model, in which the information that is known prior to the arrival of each vessel is taken into consideration and that, vessels don't arrive in the port before the berths which have been attributed to them, become available.

In the dynamic BAP formula the defined binary variables x_{ijk} specify if vessel j should be supported as the k_{th} vessel in the berth i as follows:

$$\text{Minimize} \sum_{i \in B} \sum_{j \in V} \sum_{k \in O} \left\{ (T - k + 1) \cdot C_{ijk} + S_i - A_j \right\} \cdot x_{ijk} + \sum_{i \in B} \sum_{j \in W_i} \sum_{k \in O} (T - k + 1) \cdot y_{ijk} \quad (1)$$

Subject to:

$$\sum_{i \in B} \sum_{k \in O} x_{ijk} = 1 \quad \forall j \in V, \quad (2)$$

$$\sum_{j \in V} x_{ijk} \leq 1 \quad \forall i \in B, k \in O, \quad (3)$$

$$\sum_{l \in V} \sum_{m \in P_k} (C_{il} x_{ilm} + y_{ilm}) + y_{ijk} - (A_j - S_i) \cdot x_{ijk} \geq 0 \quad \forall i \in B, j \in W_i, k \in O, \quad (4)$$

$$x_{ijk} \in \{0,1\} \quad \forall i \in B, j \in V, k \in O, \quad (5)$$

$$y_{ijk} \geq 0 \quad \forall i \in B, j \in V, k \in O \quad (6)$$

Where:

$i (=1, \dots, I) \in B$	group of berths,
$j (=1, \dots, T) \in V$	group of vessels,
$k (=1, \dots, T) \in O$	group of service orders,
S_i	the moment in which the berth i becomes available for the planning of berth allocation,
A_j	the arrival of vessel j ,
C_{ij}	the handling time of the vessel j in berth i ,
x_{ijk}	1 if the vessel j is handled as the k_{th} vessel in the berth i 0 if not.
P_k	subgroup of O such as $P_k = \{p / p < k \in O\}$,
W_i	subgroup of vessels as $A_j \geq S_i$,
y_{ijk}	time available in berth i between the departure of the k_{th} vessel and the arrival of the k_{th} vessel when the vessel j is being served as the k_{th} vessel.

The objective function (1) is to minimize the total waiting and handling time for each vessel. The group of restrictions (2) guarantees that each vessel should be handled in some berth in a service order (the order of being attended to). The group of restrictions (3) guarantees that each berth attends to, at the most, one vessel at a time.

Bearing in mind that the system is static, $S_i \geq A_j$, in other words, the vessels will always arrive before the time of the berth becoming available. The restrictions (4) guarantee that the vessels should be handled after arriving.

Later a heuristics solution was proposed for this problem, because this could not

be solved in polynomial time, using the method of the optimization of the sub-gradient, based on the Lagrangian relaxation of the original problem - dynamic BAP.

2.2.3.1 Proposed Techniques to Solve the BAP

The majority of the studies related to ports, focus their attention on strategic and tactical problems. Because most berths are operated by private shipping companies, few studies have ever been carried out into the allocation of vessels in berths (Imai et al., 2001). According to Dai et al. (2004) there is a large amount of studies on operational research applications about container operations. In Table 1, the relationship between research that was found in the literature for solving the BAP is shown:

Table 1 – Bibliography found on BAP

Pesquisador	Ano	Trabalho
Brown et. al	1994, 1997	Allocation problem in naval ports
Lim	1998	Bidimensional problem of packing
Moon	2000	Integer linear programming
Imai, Nishimura e Papadimitriou	2001	Lagrangean Relaxation
Nishimura, Imai e Papadimitriou	2001	Genetic algorithms
Park e Kim	2003	Subgradient optimization, dynamic programming
Kim e Moon	2003	Simulated annealing

There is additional research that seeks a solution for the BAP, is available from Park and Kim (2003), Dai et al. (2004), Guan and Cheung (2004) and, Mulato and Oliveira (2006).

3 Proposal to solve the BAP

3.1 General Idea of the Proposed Method

Similar to the series of studies carried out by Imai, Nishimura and Papadimitriou (2001, 2003), the intention of this survey is to propose an alternative tool for solving the BAP, which differs from other studies that were found that deal with the insertion of variables related to operational costs, that are carried out in practice, making the model the closest to the activities carried out in the ports. An example is, the inclusion of the berthing fees for a vessel in the port, which is of great relevance in the decision making process.

In practice, we know that a port receives several sizes of vessels, therefore, they charge different rates for the length of time spent in the port (termed as berthing fees). In a survey held in the port of Itajaí (SC) was obtained the information that the daily berthing rate for a ship in the port is close to US\$ 15,000.00. Thus, in this way, probably, it will be more viable to service a large vessel first, leaving the smaller vessels for later, in order to avoid high costs for staying in the port.

Therefore, in this survey some parameters are considered as being of great relevance for solving the BAP, proposing a solving methodology that is easily implemented and that can offer good results.

3.2 Parameters for solving the BAP

In order to demonstrate the port operations with respect to its mooring plan it is necessary to determine some parameters that are pertinent to the problem, so that these reflect to the maximum, the real problems found in ports.

3.2.1 Costs (Vessel and Port)

The port system, because of its responsibilities and activities, involves several taxes that are charged for its operations. Firstly, the construction cost of the terminal,

then, costs related to the acquisition of trucks and equipment, labor expenses, fuel, amongst others. As the focus of this survey is the berth allocation problem, only the port operational costs are considered.

For the preparation of this article the mooring tariffs and transport of cargo were considered, these were considered to have the greatest impact on making up the operational cost of the port.

The mooring rates are charged per linear meter of the wharf that is occupied by the moored vessel and for a certain period, or fraction, in hours. Some ports adopt the nomenclature 'period' to designate a certain number of elapsed hours from the moment of the vessels mooring in a berth, until its removal from that berth. For instance, the port of Santos (SP) uses a period of six continuous hours. In this way, if a ship is moored in certain berth for seven hours, it will have to pay the mooring rate for two periods, because it went over the limit of six hours in a single period.

The rates charged for cargo handling is charged for the handling that is done with the vessel. This rate may be based on tons moved or number of containers; however, the charging criteria should be defined at the time of the negotiations between the owner and the port authority.

These two charges are considered as being the main tariffs that are charged for using the port infrastructure, there are also charges for the use of the terrestrial infrastructure and general services.

3.2.2 Restrictions

Amongst the several restrictions that influence the mooring decision process of a certain vessel, the following should be highlighted: time, depth, length of equipment and the vessels schedules.

With regard to the time restriction, it is necessary to consider the arrival time of the vessel in the port, as well as the time of the availability of the berth. Berth availability, is understood to be moment when a vessel leaves a berth, thereby freeing it to receive another vessel. In practice, ports consider a time interval between the release time of the berth and a new mooring at that berth. This interval is used to prepare the berth for the next mooring, by removing or the repositioning of equipment, relocation of labor, etc. The Itajai Port allows two hours for this interval.

The draft restriction is responsible for ensuring that the berth depth is greater than the draft of the vessel so that the vessel can be moved without any problems. Usually the vessels draft represents 70% of the depth of the water at the berth, this for greater security in the mooring operation.

The length restriction should verify that the length of the berth is greater than the length of the vessel so that it can adequately come alongside the berth.

Regarding the equipment, it is necessary to evaluate the probable berths (which already complied with previous restrictions) that will receive a certain vessel, ensure that it has suitable equipment to undertake the loading and unloading operation of the merchandise. Assume that a certain berth complies with the length and depth restrictions, and it is also unoccupied, already taken into account the period of time required between the release of a vessel and the mooring of another; nevertheless, it is necessary to verify that this berth has equipment capable of meeting the needs of the vessel which is a candidate for the mooring, because depending on the type of merchandise, different equipment will be required. For this study, it has been considered that all berths have suitable equipment available for any type of vessel to be handled.

The restriction scheduling to be considered in this study is to ensure that the staging of the vessels occurs in two possible ways: the earliest date of liberation of the vessel or by the least allocation cost. In some cases, and depending on the merchandise type to be handled, it may be better to carry out the load/unloading operation in the shortest period of time; mainly in the case of perishable products, where the time factor plays a fundamental role in the shelf life of the product, even if that allocation has a greater cost.

In other cases, usually for nonperishable products, such as ores, iron,

aluminum, it is more convenient to guarantee that the allocation has the smallest possible cost, not to increase the cost of the merchandise. Like this, probably, the ship will wait for a mooring for a longer period, however, with a lower allocation cost if the criterion of the earliest release date was used.

3.2.3 Decision Making Variables (When and How to Allocate Each Vessel)

The port administration, that has a list of vessels to be moored within a certain period of time, a week, a fortnight or a month, should prepare a mooring plan, in other words, create a system of scheduling vessels, to guarantee that all vessels are dealt with in the shortest period of time or at the lowest possible cost. For this, it is necessary to verify if all the restrictions are satisfied for the staging to begin.

The mooring plan should be capable of allocating the vessels to berths in the best possible way, in other words, to determine in 'which' berth should each vessel be moored, as well as 'the time' of the mooring, and analyze if there was any waiting time on the part of each vessel, considering costs that will optimize the ports operating system, at the end of the staging, to determine a group of berths that are handling a group of vessels moored in different periods, with different capacities, tariffs and productivity rates.

Port activities involve the daily arrival and departure of several vessels (the port of Rotterdam, has 60 vessels on average), as well as the use of several berths (Rotterdam has five mooring areas and another two areas for emergency mooring), the scheduling of the vessels can be slow on many occasions, requiring the use of Information Technology (IT).

3.3 Heuristic Allocation Algorithm

In an attempt to offer an easily implementable method that can provide good results for solving the BAP, a heuristic algorithm has been suggested for the allocation of berths for vessels.

Initially a simplified algorithm has been proposed that is capable of analyzing a list of vessels (each one containing information on the time of arrival, draft, load, length and differential tariffs), and a list of berths (with information relating to the release time, the considered interval, depth, length, productivity, tariffs), to verify that the restrictions have been satisfied and, later, allocate each vessel to a berth.

The list of vessels that are to be moored, is not necessarily informed according to the arrival schedule, and therefore the algorithm has the allocation criteria to verify the final time of release of a certain berth, add the interval period to be considered (two hours, in the case of the Itajai port), then allocate a ship to this berth.

Supposing there is a list with two ships to be moored, according to Table 2:

Ship ID	Length	Depth	Cargo	Arriving time	Cost to be waiting
n1	220	14	150	18/8/2007 17:00	2.000,00
n2	210	14	125	18/8/2007 12:30	1.700,00

Table 2 – Example of a list of vessels to be moored

Source: the author

And a list of two existent berths, according to Table 3:

Berth ID	Length	Depth	Productivity (container/hour)	Release Time	Anchorage tariff	Movement tariff (per container)
b1	260	16	50		2,50	45
b2	250	13	45		2,30	45

Table 3 – Example of a list of berths to be occupied

Source: the author

After scheduling the vessels, it is noted that the berth $b2$ does not comply with draft restriction, because both vessels need 14 meters for mooring and berth $b2$ only has 13 meters of water. This way, both will have to be handled in berth $b1$, which meets the length and depth restrictions.

As the list of vessel starts with vessel $n1$ with an arrival time of 17:00 h, this should be the first ship to be allocated. It is worth mentioning that this criteria is used in this survey, and may be altered between authors, in other words, initially the list of vessels may be arranged in accordance with the time of arrival, to later to carry out the staging.

Also considered in this example is the time of liberating both berths as being free to make for staging at any moment, in other words, a vessel can be moored in any one of the berths from when it arrives (if the restrictions are complied with).

We can then go to vessel $n1$ that has its arrival time set 17:00 h. As already mentioned, it will be handled by berth $b1$, from the time of its arrival. As this vessel has 150 tons of merchandise to unloaded and berth $b1$ has a capacity of 50 tons per hour, this vessel $b1$ will take 3 hours to be handled, therefore, terminating its port activities at 20:00 h. From this moment forward, a period of two hours will be added (using this example) so that the berth can undergo the necessary treatment to receive the next vessels at about 22:00 h.

Using this time schedule, vessel $n2$ moors into berth $b1$, it has 125 tons of cargo to be handled, and it will finish it handling at 00:30 h, freeing the berth at that time. Like this, it can be concluded that vessel $n1$ didn't have any waiting time and vessel $n2$ had a waiting time of 9:30 h, according to Figure 1 (a).

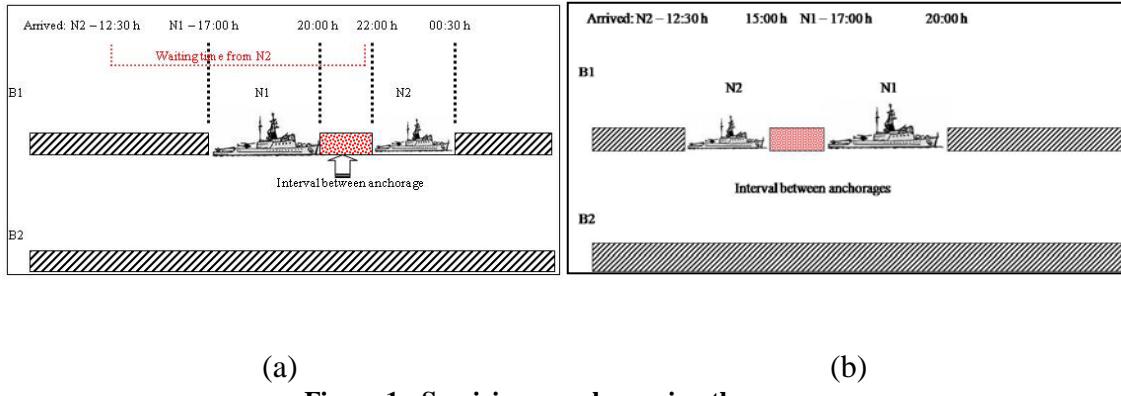


Figure 1 - Servicing vessels varying the sequence
Source: the author

To calculate the total cost of the allocation (CA), consider:

$$\text{MinCA} = \sum_{i \in N} \sum_{j \in B} \left\{ \left[(E_i + A_{ij}) \cdot CP_i \right] + \left[C_i \cdot TMov_j \right] + \left[(L_i \cdot TAttrac_j) \cdot P \right] \right\} \cdot x_{ij} \quad (7)$$

Subject to:

$$\sum_{i \in N} (L_j - L_i) \cdot x_{ij} \geq 0 \quad \forall j \in B \quad (8)$$

$$\sum_{i \in N} (D_j - D_i) \cdot x_{ij} \geq 0 \quad \forall j \in B \quad (9)$$

$$Mcheg_i + \sum_{i \in N} A_{ij} \cdot x_{ij} - Mlib_j \geq 0 \quad (10)$$

Where:

CA: is the objective function of the allocation cost;
 E_i: waiting time for the vessel *i*;
 A_{ij}: time servicing the vessel *i* in the berth *j*;
 CP_i: Cost of the vessel *i* stopped;
 Mchegi : arrival time of the vessel *i*;
 Mlibj: time of liberating the berth *j*;
 L_i: length of vessel *i*;
 L_j: length of berth *j*;
 Di: draft of vessel *i*;
 Dj: depth of berth *j*;
 TAtracj: mooring fees charged for the berth *j* of vessel *i*;
 Ci: vessels load *i*;
 TMovj: movement fees charged for the berth *j* for vessel *i*;
 P: number of periods charged for;
 x_{ij}: binary variable of the type 0-1, where it is assumed that the value 1 if the vessel *i* has been serviced at berth *j* and assume 0 otherwise;
 N = {n₁, n₂, ..., n_k}, represents the group of vessels;
 B = {b₁, b₂, ..., b_m}, represents the group of berths.

The equation (8) represents the restriction on the length of the berth *j* which should be greater than the length of the vessel *i*; restriction (9) indicates that the draft of the ship *i* should be less than depth at berth *j* and, restriction (10) indicates that the time of mooring and servicing of the vessel *i* should be greater than the time of freeing of berth *j*, in other words, a vessel cannot be moored and serviced by a given berth before being liberated from the handling of the previous vessel.

Therefore, the total cost of this allocation would be \$29,500.00 obtained using equation (7). If the service list began with vessel *n*₂, which would be handled first, the total allocation cost would be \$23,700.00, because the vessels would stay in the port for less time, there would be no waiting time for any vessel, according to Figure 1 (b).

In consequence, this results in a vessel downtime cost less than the first allocation, in other words, in this case, the cost reduction would be approximately 20%.

Because the cost varies a lot from one allocation sequence to another, it becomes necessary to evaluate the several possibilities of sequences that can exist, and for this search not to be exhaustive, a genetic search algorithm has been proposed that analyzes the possible allocation sequences.

3.4 Genetic Search Algorithm

According to Soares (1997), a procedure for the selection of the method is to carry out an exhaustive study on available optimization algorithms, verifying the characteristic of reaching the global solution more frequently for a certain number of executions. That is a factor for measuring the potentiality of the algorithms and, amongst the most effective methods, are the Genetic Algorithms - GA. GA's are applied as a random search technique that have been used by a great number of followers (DÁVALOS & STANGE, 1994).

In accordance with Tcholakian and Stange (1994), in the GA's each point of the space solution is considered as a chromosome or an individual. The group of chromosomes is called a population. The individuals in the population grow generation by generation through operations between the chromosomes, such as the selection, crossover and mutation operations.

3.4.1 Characterization of the Chromosome

For Mayerle (1996) it is in the chromosome where the characteristics of the researched solutions are stored. According to Goldbarg and Luna (2000), a chromosome is usually defined as a vector of components.

Like this, for the BAP, a chromosome may be represented by a sequence of

vessels, which carry in themselves the values of the variables:

$$\text{Chromosome} = \{N3_1, N2_2, N1_3, \dots, Nn_k\}$$

where Nnk represents the k_{th} vessels in the list, and each vessel occupies a position in this list (for instance: the vessel $n3$ occupies position #1 in the list).

3.4.2 Generation of an Initial Population

An initial population is formed, in principal, through some evaluation performance mechanism (GOLDBARG & LUNA, 2000). The individuals are coded in a finite sequence of vessels and in this way, each component of the sequence is called a gene, which is associated to a variable of the problem.

This evaluation performance, better known as fitness, represents the individual's capacity to adapt to the environment. In the case of GA, when applied to the optimization of combinatory problems, its size relates to the value of the objective function, which can be calculated according to the expression shown in (7).

Considering that for a BAP we want to find the minimum value for the objective function, the fitness measure should be considered as; the smaller the value is, the bigger the adaptation capacity will be.

3.4.3 Selection Technique

To arrange the individuals of the population during the search process of the fitness size is (represented by f_n), calculated by (7), as already mentioned. The order is given in the following way: $f_1 \leq f_2 \leq \dots \leq f_n$. Moreover, this way, the first individual of the population presents the best performance, while the last individual has the worst performance of all the population. This way the method always chooses individuals with best performances causing them to reproduce.

3.4.5 Crossover

The crossover operator, is considered as the fundamental characteristic of the GA. Pairs of genitors are chosen, randomly, from the population and, based on the aptitude, new individuals are formed from the exchange of genetic material. This way, the descendants will be different from their parents, but with genetic characteristics of both genitors (PACHECO, year unknown).

There are several rules that exist to make up the individuals crossing and like this, for the problem related to this survey, is was opted to innovate and to differentiate the other rules, considering the average of the occupied positions for the individuals in the chromosomes

Considering two chromosomes:

$$C1 - \{N3_1, N2_2, N4_3, N1_4\}$$

$$C2 - \{N2_1, N4_2, N1_3, N3_4\}$$

The averages of the positions are calculated occupied by each individual in the two chromosomes:

N1: occupies positions 4 and 3, therefore the average is 3.5;

N2: occupies positions 2 and 1, therefore the average is 1.5;

N3: occupies positions 1 and 4, therefore the average is 2.5;

N4: occupies positions 3 and 2, therefore the average is 2.5;

Next, list them in ascending order and, when there is parity, randomly select any one of these to occupy the position:

$$\begin{array}{cccc}
 1,5 & \leq & 2,5 & \leq 3,5 \\
 \downarrow & & \downarrow & \downarrow \\
 N2_1 - N3_2 - N4_3 - N1_4
 \end{array}$$

As there was parity between the N3 and N4 individual averages, it was decided to insert in the list, in the second position, vessel N3 and, followed by vessel N4.

3.4.6 Mutation Operation

After making the crossover, the mutation takes space. This is a genetic operator which is introducing new characteristics to the individual or even to restore characteristics that were lost in operations, for instance, in the crossover.

In this way, the mutation assures that the probability of arriving at any point in the search space is never zero, as well as solving any problem relating to maximums (or minimums) places, because with this mechanism, the direction of the search may be altered slightly. In the GA system, a mutation is an event that has a probability p_m for each bit of the chain of characters of all the individuals in the population, which could become an expensive operator.

In the BAP, the mutation should be done randomly after the crossover, using a mutation rate, in other words, a probability that this procedure will occur.

The mutation operator has an important role in the growth of the population (introduction and restoration of characteristics), however a secondary genetic operator is considered, which should happen with a low probability, considering the high computing cost.

3.5 Implementation of the Proposed Algorithm

For the implementation of the suggested method, it is suggested that this be based on the Imai et. al (1994, 2001, 2003, 2005) and Nishimura et. al (2001) studies.

In this way, the proposal for solving the BAP is the construction of a program, which can be developed using Delphi®, where the behavior of the proposed variables method should be implemented, tested and analyzed. The suggestion to use Delphi® is because this includes several simulation methods and it allows for the possible representation of dynamic aspects, where some variables can be implemented – such as berth productivity - with the objective of being closest to the reality that is needed.

The suggestion of the heuristic method for solving the BAP is idealized for generic use, so that it can be adapted to any marine port terminal by simply modifying a few parameters.

4 Final Considerations

At the end of this article it is hoped that some implications that are present in a port administration have been conveyed concisely, more specifically the problem relating to the allocation of ships to berths and the respective implications.

It is hoped that something has been learnt about genetic algorithms, an easily understood tool and application, which is proposed for the solving of the BAP.

To take this survey a step further, it is intended to implement a program to use the suggestions in this article, the use of the genetic algorithm and the inclusion of pertinent variables relating to the real problem, allowing for testing of the suggested methodology and the evaluation of the results, validating them with the intention of helping in the administration of the port system.

Special thanks

I would like to thank to the CNPq – National Council for Scientific and Technological Development for the financial support in the execution of this research through a scholarship.

References

- Anuário Estatístico dos Transportes. Disponível em:
[< http://www.geipot.gov.br/anuario2001/complementar/Tabels/722.xls>](http://www.geipot.gov.br/anuario2001/complementar/Tabels/722.xls)
- BROWN, G. G., LAWPHONGPANICH, S., THURMAN, K. P. (1994): Optimizing ship berthing. *Naval Research Logistics*, v. 41, 1-15.
- BROWN, G. G., CORMICAN, K. J., LAWPHONGPANICH, S., WIDDIS, D. B. (1997). Optimizing submarine berthing with a persistence incentive. *Naval Research Logistics*, v. 44.
- DAI, J., LIN, W., MOORTHY, R., TEO, C.-P. Berth allocation planning optimization in container terminal. Disponível em:
[<http://www.bschool.nus.edu.sg/staff/bizteocp/berthplanningjuly2004>](http://www.bschool.nus.edu.sg/staff/bizteocp/berthplanningjuly2004)
- DÁVALOS, R. V., STANGE, P. (1994). Uma comparação entre otimização global via algoritmos genéticos e via Gams. Anais do congresso XXVI Simpósio Brasileiro de Pesquisa Operacional, p. 696-701. Florianópolis.
- FERNANDES, M. G. (2001). Modelo econômico-operacional para análise e dimensionamento de terminais de contêineres e veículos. Dissertação de Mestrado. São Paulo: Universidade de São Paulo, Departamento de Engenharia Naval e Oceânica da Escola Politécnica da Universidade de São Paulo.
- GOLDBARG, M. C. LUNA, H. P. L. (2000). Otimização combinatório e programação linear: modelos e algoritmos. Rio de Janeiro: Campus.
- GUAN, Y.; CHEUNG, R. K. (2004). The berth allocation problem: models and solutions methods. *OR Spectrum*, v. 26, 75-92.
- IMAI, A., NAGAIWA, K., TAT, C. W. (1994). Efficient planning of berth allocation for container terminals in Asia. *Journal of Advanced Transportation*, v. 31, n. 1, 75-94.
- IMAI, A., NISHIMURA, E., PAPADIMITRIOU, S. (2001). The dynamic berth allocation problem for a container port. *Transportation Research Part B*, v. 37, 401-417.
- IMAI, A., NISHIMURA, E., PAPADIMITRIOU, S. (2003). Berth allocation with service priority. *Transportation Research Part B*, v. 37, 437-457.
- IMAI, A., SUN, X., NISHIMURA, E., PAPADIMITRIOU, S. (2005). Berth allocation in a container port: using a continuous location space approach. *Transportation Research Part B*, v. 39, 199-221.
- KIM, K. H., MOON, K. C. (2003). Berth scheduling by simulated annealing. *Transportation Research Part B*, v. 37, 541-560.
- LIM, A. The berth planning problem. (1998). *Operations Research Letters*. v. 22, 105-110.
- MAYERLE, S. F. (1996). Um sistema de apoio à decisão para o planejamento operacional de empresas de transporte rodoviário urbano de passageiros. Tese de Doutorado-UFSC. Florianópolis.
- MOON, K. C. (2000) A mathematical model and a heuristic algorithm for berth planning. Brain Korea 21 Logistics Team, July.
- MULATO, F. M., OLIVEIRA, M. M. B. de. (2006). O impacto de um sistema de agendamento antecipado de docas para carga e descarga na gestão da cadeia de suprimentos. *Revista Produção Online*. v. 6, n. 3, p. 96, set./dez.
- NISHIMURA, E., IMAI, A., PAPADIMITRIOU, S. (2001). Berth allocation planning in the public berth system by genetic algorithms. *European Journal of Operational Research*, v. 131, 282-292.
- NOVAES, A. G. (2004). Logística e gerenciamento da cadeia de distribuição. 2^a ed. Rio de Janeiro: Campus.
- OBITKO, M. Website tutorial. (1998). Genetic algorithms with Java examples.

- PACHECO, M. A. C. Algoritmos genéticos: princípios e aplicações. ICA: Laboratório de Inteligência Computacional Aplicada. Departamento de Engenharia Elétrica. Pontifícia Universidade Católica do Rio de Janeiro. Fonte desconhecida.
- PARK, K. T., KIM, K. H. (2002). Berth scheduling for container terminals by using a sub-gradient optimization technique. *Journal of the Operational Research Society*, v. 53, 1054-1062.
- PARK, Y.-M., KIM, K. H. (2003). A scheduling method for berth and quay cranes. *OR Spectrum*, v. 25, 1-23.
- PEIXOTO, J. B. (1977). Os transportes no atual desenvolvimento do Brasil. Rio de Janeiro: Biblioteca do Exército.
- Pesquisa Aquaviária CNT: Portos Marítimos: Longo curso e cabotagem. Brasília: Confederação Nacional do Transporte: 2006.
- SILVA, V. M. D., COELHO, A. S. (2007). Uma visão sobre o problema de alocação de berços. *Revista Produção Online*, v.7, n. 2. ISSN 1676-1901.
- SOARES, G. L. (1997). Algoritmos genéticos: estudo, novas técnicas e aplicações. Dissertação de Mestrado. Belo Horizonte: Escola de Engenharia da Universidade Federal de Minas Gerais.
- TCHOLAKIAN, A. B., STANGE, P. (1994). Um algoritmo para a construção de funções de pertinência via algoritmos genéticos. *Anais do congresso XXVI Simpósio Brasileiro de Pesquisa Operacional*, p. 111-116. Florianópolis.

Site:

Ministério do Desenvolvimento, Indústria e Comércio Exterior:

<http://www.aprendendoaexportar.gov.br>

Initial Electronic Spare Parts Stock and Consumption Forecasting

Guilherme NEVES ‡
Madiagne DIALLO ‡†
Leonardo Junqueira LUSTOSA ‡

† Research Productivity Grant of CNPq—Brazilian National Research and Development Centre

‡ Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro,
Rua Marquês de São Vicente, 225 , sala 950L, CEP: 22453-900, Gávea, Rio de Janeiro - RJ
Brazil

gneves01@gmail.com
<http://www.ind.puc-rio.br/~diallo>
diallo@puc-rio.br
<http://www.ind.puc-rio.br/~leonardol>
ljl@puc-rio.br

Abstract

There is a consensus that conventional continuous demand distribution methods are not appropriate for forecasting replacement parts. However, many forecasting tools available in market still use them. This work presents an application of the Poisson distribution to forecast the needs of electronic spare parts. Using basic stock management notions and usual concepts of reliability, availability and the Poisson process, an alternative method is proposed for sizing the initial stock of replacement parts to be purchased along with a electronic equipment. The results from the application of the proposed method and its comparison to the SAGA method, which is based on time series and normal distribution, are presented. The analyses of results have shown that it is possible to reduce the forecast errors; hence the stock costs, and the number of stockouts, thus enhancing the operational availability.

Keywords: Replacement parts, spares, forecast, stock management, supply, time series, availability, reliability, Poisson.

1 Introduction

The adequacy of the stock control type depends on the purpose for which the material is intended. Thus, when a spare item is seen from the manufacturer or the seller's point of view, it should be analyzed through an appropriate method for production or sales. However, if the same material is seen under the user's point of view, the appropriate method must serve the purpose of a "Maintenance Support Stock" (MSS).

Often, replacement parts are expensive, with long re-supply time, the stockout cost very high and the patterns of demand are not the same found in the retail sales and production. There is a consensus among researchers that the spare parts consumption forecast should not be made by traditional methods such as moving average and exponential smoothing. This is mainly due to the low consumption rate, the long time between demand occurrences, characteristic of spare parts. Moreover, the demand generating process depends on the number of equipment units in use and in some cases, on the intensity of use, which are usually well known data. When the reliability of the system is critical, manufacturers and users focus their attention on the failure probability, which is the time elapsed until it fails (MTBF - Mean Time Between Failure); for repairable parts, the average time to fail (MTTF - Mean Time To Failure); for disposable parts, the time between repairs (TBO - Time Between Overhaul), informed by the manufacturer. For these reasons, the management of parts must be based on the process that generates failures, which is endogenous to the system. However, the software most commonly used in the management of spare part stocks disregards these characteristics and use the exponential smoothing forecasting methods.

Another important feature related to spare parts in a MSS is the desired operational availability, Ao, of the associated equipment. The availability can be understood as the parameter that measures the proportion of the time the equipment is up and ready for use. Thus, the shorter the time off is, the higher the availability of the equipment and the better the level of service will be. However, for the availability be high, there should be spare parts ready for the replacement of the defective items in case of failures.

The literature on inventory management is rich in methods for controlling stocks in wholesale, retail sales and manufacture, but in what the supply of spare parts is concerned, it does not have the same coverage and is relatively scarce. In spite of its actual importance to many industries, engineering or business academic courses seldom address subjects approaching the management of replacement part inventories.

The purpose of this study is to present a real case of forecasting the consumption of electronic spare parts adequate for sizing stocks to cover the needs between acquisition epochs. The method is mainly aimed at the application in companies that provide services that need to maintain high operational availability of their equipment and thus to have stock of spare parts for the ready replacement of the defective items.

The next section will present a brief review of the literature on spare parts consumption forecasting. Following that, Section 3 will present details of the exponential smoothing forecasting method adopted by the SAGA System (Management and Support Automated System) used by the studied organization managing spare parts. Section 4, presents the proposal of an alternate methodology for forecasting the needs of electronic spare parts, as well as the present situation, and compares the expected results from both methodologies. Finally, in the last section, the benefits of the proposed methodology will be discussed.

2 Literature Review

Methods for forecasting and determining the re-supply of wholesale, retail and manufacture stocks are easily found in books and inventory management courses. In such cases, the demand and the response time frequently have good adherence to the normal distribution, and time series methods are adequate for forecasting the demand. However, in the case of spare parts for maintenance, the problem is clearly different. When compared with retail items, spare parts are usually more expensive, of sporadic need and low consumption rate; the availability is critical (high stockout costs). This intermittent demand rules out the normal distribution as a reasonable representation, and the time series methods, designed for continuous distribution, becomes inappropriate in face of high probabilities of zero demand.

Frequently, time intervals between failures are completely random, and many studies found in literature employ mainly Gamma and Poisson distributions to represent the demand for spare parts (CROSTON, 1972).

These distributions are associated with the known Poisson process characteristic of phenomena in which age or wear of the component does not affect the likelihood that it will fail, and also the fact that given that a failure has just occurred has no influence on the time elapsed until the next failure. A characteristic of the Poisson distribution (i.e. the probability distribution that, in a Poisson process, x failures will occur in a given time interval t), which makes it easy to use, is that its average is equal to its variance and is a parameter that completely characterizes the distribution. Therefore, if the failure process has the characteristics of a Poisson process, it is enough to use the average demand of historical data to estimate the probability of any given number of failures to occur in any time interval. Strijbosch et al. (2000) discuss the selection of adequate distribution for methoding demand distribution for spare parts for an (s, Q) control system, and use a compound Bernoulli distribution.

Studying the case of a manufacturer of electronic products in Taiwan, Yeh (1997) adopted the premise of the Gamma distribution for demand to determine the spare parts stock policy. As usual, the normal distribution proved to be inadequate by the fact that most of the items studied presented annual demand less than ten units. The Poisson distribution also showed little adherence to data since the variances and average historical demands were significantly different.

Wanke (2005) studied the case of a Brazilian manufacturer of agricultural equipments and used the Gamma distribution for methoding the consumption of spare parts adjusting the actual data to test alternative methods based on the key characteristics of the spare parts. In another study, Wanke (2006) noted that the demand for spare parts had good adherence to the Poisson distribution. He also pointed out that the properties of this distribution become particularly interesting when examining how different safety stock levels would affect the likelihood of lack of material, especially in environments where the annual consumption is between 1 and 300 units.

From the logistic point of view, spare parts systems can become extremely complex, especially when the company provides maintenance service for its customers. Some factors that may exacerbate the complexity are the geographic location and the existence of repairable items. Some interesting case studies are presented by Cohen, and Zheng Wang (1999) and by Fleischmann et al. (2003) for companies that offer service to their customers.

After checking the forecasting methods used by a central audio manufacturer, S.I.T.T.I.

(1999), by the France Telecommunications Committee, C.C.T. (1983), by a radar manufacturer, FIAR (1988) and by the authority responsible for the logistic development of equipments used for flight protection in Brazil, CISCEA (1994), it was found that all use the Poisson distribution in the methodology for estimating the initial purchase of spare parts.

The methodologies developed by Yeh (1997), and by Wanke (2005) were aimed at the stock of suppliers or manufacturers of spare parts. However, the methodologies employed by manufacturers of equipment S.I.T.T.I. (1999), FIAR (1988), CCT (1983), and mentioned by Wanke (2006) and CISCEA (1994), were developed for users of equipment maintainers. This latter seemed to be more adherent to the reality proposed in this work and was chosen to be the basis of its development.

3 The SAGA Stock Management Method

The SAGA System (Management and Support Automated System) is a software system containing a popular time series method (exponential smoothing) that allows forecasting demands of items based on its consumption records. The method is based on two main variables: the consumption average level (MI) and the growth trend (TI). For the calculation of future consumption, the past MI and TI values will be analyzed and used to calculate future values.

The procedures presented below have been taken from the technical documentation ISDS013 (PAME, 1986).

3.1 Presentation of the Method

The exponential smoothing method with trend can be described in two steps:

- a) startup , and
- b) (continuous) re-estimation of the parameters of the method.

3.2 Startup

In the startup step, the system provides starting values for α (smoothing coefficient for the average level estimate) and β (smoothing coefficient for the trend estimate) chosen from a set of pre-defined values that best represents the series. In other words, the pair that gives the smallest mean absolute deviation (MAD) value for a starting sample of the series of observed demands of the item (or, perhaps of some similar item, if the item has no history).

α	0.10	0.10	0.10	0.15	0.15	0.15	0.20	0.20	0.20	0.30	0.30	0.30
β	0.40	0.20	0.10	0.40	0.20	0.10	0.40	0.20	0.10	0.40	0.20	0.10

Adapted from PAME, 1986.

Table 1 – Initial values of α and β for the selection of the best pair

Different values of α and β can be entered manually.

Two situations are considered:

- a) There is a historical series of consumption; in this case, the system methods the series and computes the mean absolute deviation (MAD) of the forecasts produced by each pair α and β . The pair selected will be the one presenting the lowest MAD; and
- b) There is no historical series of consumption; in this case, the system estimates the future consumption considering $\alpha = 1$ for the first forecast, $\alpha = 0.8$ for the second forecast, $\alpha = 0.6$ for the third forecast, $\alpha = 0.4$ for the fourth forecast, $\alpha = 0.2$ for the fifth forecast and $\alpha = 0.15$ for the sixth forecast. In the seventh forecast, the system considers that there is already a historical series of consumption, and reevaluates the parameters accordingly.

3.3 Estimation of Model Parameters

The calculation of the consumption forecasts is done using the formulae described below:

$$\hat{L}_t = \alpha Q_t + (1 - \alpha) \hat{M}_t \quad \rightarrow \text{level in month } t$$

$$\hat{T}_t = \beta (\hat{L}_t - \hat{L}_{t-1}) + (1 - \beta) \hat{T}_{t-1} \quad \rightarrow \text{trend in month } t$$

$$\hat{M}_{t+1} = \hat{L}_t + \hat{T}_t \quad \rightarrow \text{consumption forecast for month } t+1$$

$$Q_t \quad \rightarrow \text{actual consumption in month } t$$

3.4 Calculation of the Mean Absolute Deviation (MAD)

The MAD is calculated using the formula below:

$$EMA_{t_0} = \alpha |Q_{t_0} - \hat{M}_{t_0}| + (1 - \alpha) EMA_{t_0-1}$$

$$\hat{M}_{t_0} \quad \rightarrow \text{Average consumption forecast for the current month made in previous month.}$$

$$Q_{t_0} \quad \rightarrow \text{Actual consumption in the current month.}$$

$$EMA_{t_0} \quad \rightarrow \text{Mean absolute deviation estimate of the level estimates.}$$

3.5 Estimation of Future Values

The estimation of future values is performed taking the optimum pair α e β , the last forecast and actual consumption as reference.

$$\hat{L}_t \quad \rightarrow \text{Level estimate in month } t.$$

\hat{T}_t

→ Trend estimate in month t.

$$\hat{M}_{t+n} = \hat{L}_t + n \hat{T}_t$$

→ Consumption forecast for month t+n (made in month t).

The total consumption forecast for the next n months will be given by:

$$\hat{C}_n = n \times \hat{L}_t + n \left(\frac{n+1}{2} \right) \hat{T}_t$$

→ Consumption forecast for n future months.

3.6 Considerations on the SAGA

The system was developed in the 80's of the last century to manage around 100,000 different items. At that time, the computing resources in terms of memory, capacity and processing speed were quite limited when compared to today. Thus, programmers used tables and simplifications to avoid operations such as exponential, which consumed large processing time.

If the same system would be developed today, it would not be so worried about processing speed and could use procedures such as linear regression for the initialization of coefficients α and β , as well as the mean quadratic error instead of the MAD, making the calculations more accurate.

4 The Forecast Methodology Proposed

Modern electronic equipments present very low failure rates, something around one failure every 100,000 hours. This characteristic would lead us to think that electronic equipments would have long lifetime. However, one must consider another aspect: the obsolescence. There are two types of obsolescence: the operational obsolescence due to the emergence of new equipment providing greater convenience and operational performance, and the technical obsolescence due to the evolution of components, leading manufacturers to seek for components with better performance and lower cost. Obsolescence, even of reasonably large and expensive electronic equipment, has made the time to replacement that, not long ago was of 20 years, to drop to some 5 to 10 years.

It is then assumed that an electronic component does not wear during the lifetime of 5 to 10 years, and in this case, ruling out the phase of "infant mortality," the failures follow a stable stationary stochastic process. This assumption is fundamental for understanding the proposal made in this work, to use the Poisson distribution as an alternative to the use of time series to achieve better results in forecasting consumption of electronic replacement parts.

Initially, a brief review of the concepts that will support the methodology proposed will be presented followed by a methodology for calculating the initial purchase of spare parts used by the majority of suppliers of equipments.

4.1 Failure Rate λ

The failure rate of a given part can be defined as "the proportion of entities that having survived a certain time t failed within the time interval $[t, t+T]$ ", (VILLEMEUR, 1992). The failure rate can also be defined for a single item, by:

$$\lambda = \frac{f}{T}$$

λ → Failure rate

f → number of failures occurred in the time interval T

T → observation time interval,

and for N items installed in K equipments, by:

$$\lambda_{eq} = \frac{f}{KNT}$$

K → number of identical equipments containing the item,
N → quantity of the item installed in each equipment,

A variant of the failure rate, also widely used, is the MTBF (Mean Time Between Failures) and MTTF (Mean Time To Failure). For obvious reasons, the MTBF is customarily used to repairable items and the MTTF for disposable items.

$$MTBF = MTTF \frac{1}{\lambda}$$

4.2 Availability and Level of Service

The operational availability is related to the time to repair the equipment. It can also be seen as level of service and depends basically on the following:

- a) Frequency and duration of the preventive maintenance, and
- b) Quantity and duration of corrective maintenances.

Mathematically, the operational availability is the relationship between the operational time and the sum of the operation time and the time off (or, down time). A value of around one means maximum availability.

$$A_o = \frac{T_{ON}}{T_{ON} + T_{OFF}}$$

A_o → Operational availability

T_{ON} → Operational time

T_{OFF} → Time off

Time off is a function of several factors, of which, the most important are (C.C.T., 1983):

- a) Constructive characteristics of the equipment:
 - Devices for testing and locate failures
 - Facilities for the removal of the defective item and its replacement;
- b) Maintenance personnel:
 - Quantity;
 - Classification;

- Experience;

- c) Technical documentation:

- Scope;
- Update;

- d) Testing tools and equipment:

- Adequacy;
- State;
- Calibration;

- e) Replacement parts availability:

- Adequacy;
- Quantity.

This work is focused on the issue of sizing spare parts inventory. The target stock should be such that the probability of no stockout (replacement parts availability) provides an operational availability index (level of service) specified for the equipment. Modeling the probability of no stockout as a function of available stock is, therefore, key issue on controlling operation availability. The model here proposed is based on the Poisson distribution.

4.3 Poisson Distribution

The Poisson distribution, whose concept was presented in Section 2, is used to method the occurrences of an event considered "rare" in a given period of time, (SHAPIRO, 1967).

Probability Density Function

$$f(k) = P(K = k) = (\lambda t)^k \frac{e^{-\lambda t}}{k!}, \quad t > 0$$

λ → Failure Rate

λt → Mean failures in time period t

k → Number of failures in time period t

Cumulative Probability (probability of k or less failures)

$$P(K \leq k) = \sum_{i=0}^k (\lambda t)^i \frac{e^{-\lambda t}}{i!}, \quad t > 0$$

4.4 Initial Purchase Forecast

The purpose of the initial purchase of spare parts is to minimize the probability that particular equipment becomes inoperable due to the lack of parts in the initial operation period. This is important because the cost of the additional parts at the time of acquisition is far lower than when it will be years latter when the part is already out of regular production.

Usually, the initial quantity of spare parts is purchased along with the equipment and is calculated according to the buyer's criteria and suggestion from the supplier. It is essential to maintain the availability of the equipment within the operation period in which the maintainability is estimated based on data from the manufacturer.

The problem is, essentially, to determine the number of spare component units to buy along with new equipment so that the probability of running out of stock is some given value.

4.4.1 Assumptions

The following assumptions are used in the inventory sizing method:

- a) the defects due to the infant mortality were eliminated in the burn in process;
- b) the components follow an exponential failures law;
- c) there is no degradation of parts during the storage period (bottom of the "bath-tub" curve);
- d) only the items damaged in the equipment are replaced by spare parts during the repair;
- e) the failure rate is considered constant over the period forecasted for the stock, and the failures are resulting from stochastic and independent events, by intrinsic mechanisms of the own item (related to assumption (c) above);
- f) the equipment is operating according to normal specification conditions, receiving all the recommended preventive maintenance, as well as all the corrective maintenance needed to restore it to original specification conditions and level of quality, and
- g) the spare quantities of the item necessary for repairing failures, other than those resulting from the processes listed above, are supplied by additional stocks.

4.4.2 Mathematical Method

The cumulative Poisson distribution defined above gives the probability of not running out of stock during the maintainability period. With a notation and terminology more adequate to the application (CISCEA, 1994) it is presented below:

$$Ps(s) = \sum_{x=0}^s \frac{F^x}{x!} \times e^{-F},$$

- | | |
|-------|---|
| s | → number of units of the spare item bought with the equipment, |
| Ps(s) | → also called the "probability of no stock disruption," is the cumulative probability that the number of failures, x, of the item is less than or equal to s. |

Using the notation defined earlier:

$F = \lambda_{eq} t$ → mean number of failures of the item under consideration during a given a time interval t, assuming that all equipment units are identical:

$$F = K N M L T_d;$$

- K → total number of units of equipment that contain the item,
- N → quantity of the item installed in each unit of equipment,
- M → ($M < 1$) average utilization rate of the KN items installed ($M = 1$ if under 24h a day, year-round operation),
- L → failure rate of the item (failure per time unit), usually designated as λ ,
- T_d → time to decommissionement (or support, or maintainability period) forecasted as of the initial purchase.

4.5 Methodology Proposed

An item is considered critical when its failure puts the equipment, which contains it, out of operation. Hence, the first step in the stock-replacement methodology proposed is the identification of critical items. These items should have priority treatment in relation to others, regardless their value and quantity, because the consequences of the equipment breakdown will override other operational costs.

Once these critical items are identified, the non-critical items can be treated according to the usual cost benefit analysis, perhaps still methoding the demand as a Poisson, compound Poisson, gamma, or some other adequate distribution, in case the demand exhibits an intermittent generation process .

4.5.1 Operational Availability and Probability of no Stock Disruption

Usually, there is no simple direct relation linking the operational availability Ao with the probability of no stock disruption Ps, because as seen in Section 4.2, the time off, TOFF, depends on many other factors, especially on the supplier's lead time. However, ceteris paribus, the greater the Ps is, the greater the Ao will be. According to the experience with critical and complex electronic systems, Ps plays a dominant role in defining A0 due to the strict standards that usually focus on the other intervening factors mentioned above.

Table 2, below, shows a suggestion, based on experience, to associate Ps with Ao.

Ao	Ps
0.95	0.95
0.96	0.97
0.97	0.98
0.98	0.99
0.99	0.995

Table 2 – Ps suggested as a function of Ao

Ideally a very large Ao would be desirable; but very high values may be economically infeasible. Because the cost of failures and time-off are hard to assess in a formal way, the more realistic approach is to leave availability as a discretionary value left to managerial judgment.

4.5.2 Consumption Forecast

The methodology for forecasting the consumption in a given time period is described below:

- a) Estimate time to decommissionement of the equipment, T_d ;
- b) Identify the operational availability or level of service contracted for the equipment under analysis. Estimate the no stock disruption rate that provides the level of service, Ao , contracted (such as the Table 2);
- c) Examine the failures of this item and the respective dates;
- d) Calculate the failure rate estimate λ_{eq} during the observation time, T (symbols as defined earlier):

$$\lambda_{eq} = \frac{f}{KNT},$$

- e) Calculate the number of parts s to be acquired for the desired probability of no stock disruption, where $F = \lambda_{eq} T_d$:

$$Ps(s) = \sum_{x=0}^s \frac{F^x}{x!} \times e^{-F}$$

It is important to stress that this approach calculates the probability of consuming no more than s units of an item during a given period of time. Therefore, it also useful for calculating the amount to be acquired to restore the stock, in case the initial stock is depleted.

5 Comparison between the SAGA method and the forecast method proposed

The forecast error measurement helps to assess the adequacy of method adopted by estimating the dispersion and the bias of the forecasted demand.

The forecast error, E_t , is given by:

$$Et = \text{Forecast} - \text{Actual Consumption}$$

The Mean Absolute Deviation (MAD) is used to evaluate the forecast error, given by:

$$MAD = \frac{1}{n} \sum_{t=1}^n |E_t|$$

The MAD was used to compare the forecasts between SAGA method and the method proposed.

Another important aspect to measure in the performance of the method is the number of "insufficient forecasts." More so for maintenance jobs that will be stopped (in case of lack of spare part to replace the item that has failed) until the part required becomes available. The lack of replacement parts is one of the most important factors compounding the time off. It is special importance for parts with long acquisition lead times. Cases, where the

acquisition time is more than 3 months are very frequent. Therefore, low occurrence of insufficient forecasts that lead to stockouts is an important performance measure.

The performances of the two methods were evaluated for three representative items of a flight protection system using historical data of four consecutive semesters. The values presented in Table 3 below show that the forecast method using the Poisson distribution presents smaller MAD when compared with the SAGA method. Using the Poisson model, no insufficient forecast was observed, while a total of nine occasions in two items were produced by the SAGA method in the period considered. In the case of the SAGA system, the stock manager would have to increase the safety stock to improve the availability of spare parts, meaning a substantial cost increase because, due to obsolescence, residual value of the leftover parts is negligible.

Item	Períod	Actual Consuption	Forecast SAGA	Forecast POISSON	Insufficient Forecast SAGA	Insufficient Forecast POISSON	MAD SAGA	MAD POISSON
Power Supply	2005-1	13	7	13	6	0	7,25	4,5
	2005-2	5	12	13	0	0		
	2006-1	8	24	15	0	0		
	2006-2	11	11	14	0	0		
	TOTAL	37	54	55	6	0		
Amp. Module	2005-1	8	7	10	1	0	4,25	3,5
	2005-2	6	4	9	2	0		
	2006-1	6	13	10	0	0		
	2006-2	5	12	10	0	0		
	TOTAL	25	36	39	3	0		
Local Oscill	2005-1	2	3	3	0	0	1,5	1,25
	2005-2	2	4	3	0	0		
	2006-1	3	3	3	0	0		
	2006-2	1	4	4	0	0		
	TOTAL	8	14	13	0	0		

Table 3 – Performance comparison: SAGA vs. POISSON

Conclusions

The objective of this study was to show an alternative method for use in sizing the initial stock of spare parts for expensive electronic control equipment. Consumption forecasts of critical electronic spare parts were based on the Poisson distribution as alternative to traditional methods that use classical exponential smoothing time series forecasting.

Through simulations of both methods applied in real cases, it was found that the simple and easy-to-implement methodology proposed may be considered more appropriate for forecasting the consumption of electronic parts, since it presented the smallest forecast errors when compared with the exponential smoothing method.

In addition, the alternative method provided considerable reduction of insufficient forecasts. The cost of insufficient forecasts is related to penalties for not meeting the operational availability contracted – due to the lack of spare parts for the ready operational recovery of the equipment – and, above all, to the company's image, which is priceless in complex systems maintenance business.

It is important to stress the importance of calculating the failure rate. When the manufacturer provides the failure rate of the equipment and its components, the value provided is obtained under standard conditions, and often does not match that found in the actual life of the item. In the case studied, the initial stock of replacement parts was dimensioned according to the failure rate informed by the manufacturer. However, after one year of operation, it was possible to reevaluate it to estimate future consumptions. The recalculated failure rate is an important forecast parameter and a description of the item behavior under the prevailing conditions. It should be noted, however, that the methodology proposed here assumes that the failure rate is stable. This means that the infant mortality phase and the wear out phase, of the "bath tube" shaped failure rate curve, are ruled out by an adequate burn in and by decommissioning the equipment before failure rate accelerates.

Future studies should consider the improvement of the failure rate calculation, especially in relation to the optimum calculation period and perhaps the application of a damping coefficient; so that failures occurred in a distant past have less weight than recent failures. Additionally, the operational availability as a function of the probability of no stock disruption also merits further investigation. The optimization of these factors could lead to results better than those already found with the application of the simple method proposed in this paper.

5 References

- Cohen, M. A.; Zheng, Y.-S. and Wang, Y. (1999) Identifying opportunities for improving teradyne's service-parts logistics system. *Interfaces*, Vol.29, No.4, pp.1-18, , July - August 1999.
- CISCEA, Comissão de Implantação do Sistema de Controle do Espaço Aéreo (1994). Documento de Especificações Logísticas, Rio de Janeiro, Brasil.
- C.C.T - COMITÉ DE COORDINATION DES TELECOMMUNICATIONS (1983) Maintenabilite des Équipements Électroniques, Paris, França.
- Croston, J. D. (1972) Forecasting and Stock Control for Intermittent Demands. *Operational Research Quarterly*, Vol.23, No 3, pp. 289-303, September.
- FIAR, Fábrica Italiana Apparecchiature Radioelettriche. (1988) Reliability, Maintainability and Failure Mode, Effects and Criticality Analysis for Precision Approach Radar (PAR), Roma, Itália.

- Fleischmann, M.; Van Nunen, J. A. E. E. and Gräve, B. (2003) Integrating closed-loop supply chains and spare-parts management at ibm. *INTERFACES*, Vol. 33, No.6, pp.44-56, November - December.
- PAME, Parque De Material De Eletrônica De Aeronáutica Do Rio De Janeiro. (1986) Instrução de Serviço, ISDS 013, Método Matemático do SAGA. Rio de Janeiro 1986.
- Shapiro, S.H. (1967) *Statistical Methods in Engineering*, New York, USA. John Wiley.
- S.I.T.T.I. (1999) System Design M600, SD 1099A, Milano, Itália.
- Strijbosch, L. W. G.; Heuts, R. M. J. and Schoot, E. H. M. V. D. (2000) A combined forecast-inventory control procedure for spare parts. *The Journal of the Operational Research Society*, Vol. 51, No. 10, pp. 1184-1192, October
- Villemeur A. (1992) *Reliability Availability Maintainability and Safety Assessment*, Chippenham, John Wiley & Sons.
- Wanke, P. (2005) Metodologia para gestão de estoques de peças de reposição: um estudo de caso em empresa brasileira. *Revista Tecnologística*. São Paulo, Dezembro.
- Wanke, P. (2006) Gestão de Estoques de Peças de Reposição de Baixo Giro. Disponível em: <http://www.cel-coppead.ufrj.br/fs-pesquisa.htm>, Rio de Janeiro, Brasil..
- Wanke, P. (2005) Método de Nível de Serviço e Otimização de Estoques na Cadeia de Suprimentos: Probabilidade de não Faltar Produtos e Vendas Perdidas. Disponível em: <http://www.centrodeelogistica.com.br/news/fs.public.html>, Rio de Janeiro 2005.
- Yeh, Q.J. (1997) A Practical Implementation of Gamma Distribution to the Reordering Decision of an Inventory Control Problem. *Production and Inventory Management Journal*. Vol. 38, No. 1. USA.

Validación del intervenciones ‘SOFT OR’

Carlos Raúl Navarro González †
 Sara Ojeda Benítez ‡

† Universidad Autónoma de Baja California
 Mexico
 carlos_raul_n@hotmail.com

‡ Universidad Autónoma de Baja California
 Mexico
 sara@iing.mxl.uabc.mx

Abstract

In order to subsist and be successful in this environment with greater complexity and competition, any organization is urged to start and keep a inquiry-learning-improvement cycle; Establish and increase this cycle is the objective of the soft systems approach –also know as ‘soft operations research’ ‘Soft OR’ or ‘soft system methodology’ ‘SSM’-. The big benefits of “Soft OR” and SSM is recognized by many authors, unfortunately any soft approach has an inherent difficulty to be validate in an objectivity way; thus, in many cases, it’s application not produces the expected cycle of inquiry-learning-improvement or worse only in isolated interventions that not always achieve success. In this way, this work gives a proposal methodology to evaluate in an objective manner the ‘Soft or’ approach trough the measurement of the organizational achieved; also to identify and evaluate elements (from organization, team members, or methodologist) than may limit the obtained success. An application case was provided and strong and weak elements from it’s applications was discussed.

Resumo

Para subsistir e ser bem sucedido neste ambiente de crescente complexidade e de competição, qualquer organização é convidada a iniciar e manter um do ciclo de inquérito-melhoria-aprendizagem; Estabelecer e aumentar este ciclo é o objetivo do soft-sistemas abordagem também conhecida como 'Soft OR' ou 'soft system methodology' (SSM). As grandes vantagens do "Soft Or" é reconhecido por muitos autores, infelizmente, tem uma abordagem suave qualquer dificuldade inerente ao ser validar uma objetividade em curso; assim, em muitos casos, é pedido que não produz o esperado ciclo de inquérito - -melhoria da aprendizagem ou pior só em intervenções isoladas que não sempre alcançar o sucesso. Desta forma, este trabalho apresenta uma metodologia proposta para avaliar de forma objectiva "Soft OR" abordagem vale a medição da organização alcançados; também para identificar e avaliar elementos (da organização, os membros da equipa, ou metodologia) que podem limitar ou condição o nível de sucesso. Um caso é apresentado de uma forma objectiva de avaliar o impacto alcançado organizacional e identifica os pontos fortes e fracos da metodologia que a interferência deve ser melhorado.

Resumen

Para que una organización pueda subsistir y tener éxito en este entorno de creciente complejidad y competencia requiere iniciar cuanto antes y mantenerse en una espiral de indagación, aprendizaje y mejoría; cuya instauración y acrecentamiento es el objetivo del pensamiento de sistemas suaves, representado por la investigación operativa suave ('Soft

OR') y la metodología de sistemas suaves ('SSM'). La gran utilidad de la 'Soft OR' y la 'SSM' ha sido documentada por múltiples autores, desgraciadamente cualquier intervención suave mantiene una dificultad inherente para valorarse objetivamente; por lo que en ocasiones su aplicación puede no redundar en un ciclo, sino más bien en intervenciones aisladas y que no siempre conducen al éxito. Este trabajo, presenta una propuesta para la evaluación objetiva de una intervención metodológica suave mediante la medición del impacto organizacional logrado; y con la identificación y valoración de aspectos propios de la organización, del equipo de trabajo y de la propia metodología que podrían impulsar o condicionar el nivel de éxito obtenido. Se presenta un caso que evalúa de manera objetiva el impacto organizacional logrado e identifica aspectos fuertes y débiles de la injerencia metodológica que conviene mejorar.

Keywords: System approach, Soft system, Soft OR, validation, evaluation

Title: Hard validation of soft methodologies

1 Introducción

La gran constante de los tiempos modernos es el acelerado cambio hacia un entorno de creciente complejidad y competencia en los sistemas organizacionales, lo que conlleva a la gran necesidad de mejoría dentro del ambiente industrial (Torras, 1997). En este sentido existe una mayor conciencia en las organizaciones de la importancia –para su subsistencia y éxito– de lo que implica el compartir sus conocimientos y desarrollar una cultura de aprendizaje (Dixon, 2001; Bartlett y Ghoshal, 2002), que cuestione los mecanismos usados para la resolución de problemas (Forrester, 1989; Richmond, 1994; Senge, 1998; Sterman, 2001), impulse una “intuición colectiva” (Yeung, 2000) y maneje adecuadamente la resistencia organizacional (Pardo y Martínez, 2003 y 2005).

Varios autores (Forrester, 1989; Checkland y Scholes, 1990; Dash, 1994; Flood, 1998; Senge, 1998; Argyris, 2001; Checkland, 2001; Roth, 2001; Sterman, 1991 y 2001; Grinstein, 2005) afirman que las soluciones hacia la complejidad dinámica creciente de nuestro entorno se encuentran en el “pensamiento de sistemas”. Su exitosa utilización en infinidad de casos ha sido documentada en los últimos treinta años; sin embargo su aplicación más bien resulta en intervenciones aisladas y que no siempre conducen al éxito (Salner, 2000), por lo que no se genera e impulsa una espiral de indagación, aprendizaje y mejoría (Checkland y Scholes, 1990).

2 Marco teórico

2.1 Metodologías sistémicas suaves

El pensamiento de sistemas suaves encarna un paradigma de aprendizaje (Wilson, 1993; Checkland, 2001), considera el mayor o menor grado de incertidumbre –que frecuentemente se tiene al abordar un problema que englobe seres humanos– en cuanto a los fines que se persiguen o el no tener una idea clara de cuál es la razón de las deficiencias observadas y de las posibilidades reales de que se modifiquen. Por lo que su interés es crear una metodología que guíe la acción para tratar de “administrar” situaciones problemáticas del mundo real, intentando cambiar situaciones reales constructivamente, conocer la situación real y definir (o redifinir) objetivos, recursos y restricciones para sobre esta base recomendar los cambios y acciones más convenientes (Checkland y Scholes, 1990; Fuentes, 1993; Checkland, 2001). Formando de un ciclo de indagación y aprendizaje que tiene como objetivo introducir mejoras en las organizaciones (Checkland y Scholes, 1990)

Aunque el pensamiento de sistemas “suaves” tiene su origen en la “metodología de sistemas suaves” (SSM) (Checkland y Scholes, 1990; Checkland, 2001), muchos investigadores combinan la SSM con otros modelos (Bergvall y Grahn, 1996; López y Delgado, 2003; Pidd, Brown y Koper, 2006); incluso, la propia SSM requiere –en una de sus etapas– el análisis e inclusión de otros modelos de pensamientos sistémicos que se consideren pertinentes y con ello adaptar la metodología a la situación problemática. Esto ha conducido al desarrollo de metodologías derivadas o complementarias al enfoque de sistemas suaves, tales como: Metodología Multi-Modal (Mirijamdotter, 1998); Dinámica de sistemas (Forrester, 1989; Sterman, 1991), Planeación Iteractiva (Ackoff 1979), teoría de los sistemas vivientes (Miller, 1982 y 1995), Intervención de Sistemas Totales (Flood, 1995), Heurísticas sistémicas críticas (Ulrich 1998 y 2000), etc. Por lo que la “Aproximación suave de sistemas” (Soft systems approach) o más genéricamente la Investigación operativa ‘suave’ (Soft OR) (Daellenbach, 2001) engloba una variedad de metodologías y complementos sistémicos.

Aunque la evolución del pensamiento de sistemas “suaves” ha sido vertiginosa, documentándose infinidad de aplicaciones exitosas, surgiendo nuevos seguidores y formándose asociaciones para su estudio; las intervenciones tienden a resultar aisladas dentro de una misma organización (Salner, 2000) y no se logra encarnar e impulsar la espiral de indagación, aprendizaje y mejoría.

2.2 Dificultades inherentes en las metodologías suaves

Los enfoques constituidos en la ‘Soft OR’ presentan puntos débiles y han recibido distintas críticas en su aplicación (Leonard y Beer, 1994; Bergvall y Grahn, 1996; Checkland, 2000; Salner, 2000; Valqui, 2005b; Pidd, Brown y Koper, 2006); una crítica generalizada al uso de los enfoques suaves radica en que su aplicación no es concluyente; este principio de ‘aplicabilidad inconcluyente’ caracteriza a toda metodología que dependa en su uso del sentir y forma de pensar de las personas, e implica la dificultad de probar que la metodología tal sea útil.

El principio de ‘aplicabilidad inconcluyente’ se puede ilustrar al conjeturar sobre la aplicación de alguna metodología que consecuentemente logre la resolución del problema, en este caso se puede opinar que la metodología ‘funciona’, y en caso contrario decir que esta ‘no funciona’. Pero bajo otro punto de vista, se puede apuntar a que los bajos resultados se deben a la falta de habilidades y competencia del usuario, y viceversa que los buenos resultados obtenidos se deben a las habilidades y tenacidad del aplicador y no por la metodología en sí (Checkland, 2000; Joldersma y Roelofs, 2003). Bajo esta perspectiva no existe una metodología que por sí sola pueda conducir a alguien al éxito, sino más bien, que ayuda a llegar al “éxito” mejor y más rápidamente de lo que se podría lograr sin sus principios y lineamientos (Checkland, 2000). Este principio también rige a cualquier metodología industrial, tal y como ‘Lean Manufacturing’ y ‘Six Sigma’.

Otra deficiencia implícita en las metodologías suaves se encuentra en la subjetividad en su aplicación, la cual puede verse reflejada en la formulación de un problema equivocado o subóptimo. Pues el problema expresado por el grupo de ‘usuarios’ puede ser el menos controversial y no ser el “mejor problema” del cual se puedan generar soluciones creativas y efectivas, con lo que se afecta la calidad e impacto las decisiones tomadas (Heslin y Moldoveanu, 2002). Así mismo, un segundo equipo de estudio hará un diferente análisis, lograra disparejos modelos y esquemas sistémicos que los orientara hacia otras decisiones con variados niveles de impacto, con lo que se muestra el problema de subjetividad que se tiene también en la validación de los lineamientos metodológicos suaves (Pidd, Brown y Cooper; 2006; Valqui 2005b; Daellenbach, 2001).

Es fácil observar que la evaluación de una intervención suave también presenta problemas de objetividad, pues esta se realiza bajo distintas perspectivas de “usuarios”

Connell (2001), esta subjetividad es mas latente al entender los diferentes tipos “usuarios” que algunos autores han explorado más detenidamente (Ormerod, 2001; Taylor, Robinson y Ladbrook; 2003 y 2005): ‘el hacedor’, ‘los receptores’, ‘los colaboradores’, ‘entrevistados’ y ‘no involucrados’.

Existen tres elementos interactivos que condicionan a una intervención metodológica suave (Checkland, 2000) y que cualquier evaluación o medición debe considerarlos (Connell, 2001): el nivel de capacitación y habilidades del usuario, la estructura y claridad en los procedimientos metodológicos y la sensibilidad y entendimiento hacia la situación percibida. Con lo que se deduce que es conveniente evaluar y retroalimentar la intervención metodológica con la identificación y el reconocimiento de aspectos propios de la organización, del equipo de trabajo y de la metodología que podrían condicionar los resultados (Tolvanen, 1998; Barnden y Darke, 2000; Connell, 2001; Valqui, 2005b; Blackman, 2006; Sorensen y Valqui, 2006); pues de lo contrario una intervención suave no redundara en un ciclo, sino en una aplicación aislada que pueda o no tener éxito.

2.3 Evaluación objetiva de una intervención suave

Joldersma y Roelofs (2004) indican que el impacto de una metodología ‘Soft OR’ en la estructuración de problemas puede hacerse desde cuatro distintas perspectivas (o una combinación de ellas): Experiencias con el método, la atracción por el uso del método, percepciones de la efectividad del método por los participantes y percepción de la efectividad por los observadores externos sobre la cantidad y calidad de las ideas generadas. Pidd, Brown y Cooper (2006) manifiestan que la validación pudiera ser hecha bajo el aspecto de si los modelos y esquemas conceptuales son coherentemente defendibles, consistentemente lógicos y factibles. En cambio Barnden y Darke (2000) (congruentemente con la meta de las metodologías ‘Soft OR’ de inducir una espiral de indagación, aprendizaje y mejoría) manifiestan que la evaluación deberá hacerse sobre el ‘producto de aprendizaje’ que es la realización del “cambio deseado y factible”, lo cual se puede manifestar superficialmente (por cambios en los diagramas, esquemas o procedimientos de la compañía) y más profundamente (por el cambio de la “teoría en uso” –creencias, asunciones y formas de trabajo–). Se puede fácilmente observar que las propuestas anteriores para validar y evaluar una intervención suave presentan problemas de objetividad, pues están sometidas a una evaluación subjetiva que debe realizarse bajo distintas perspectivas de “usuarios” Connell (2001).

La necesidad de disponer un mecanismo de validación y evaluación mas objetivo es manifestada por Blackman (2006), quien explica la falacia de medir los niveles de conocimiento y aprendizaje; para mejor calcular el nivel de impacto que debe presentarse en la organización; para en todo caso complementarlo con la información subjetiva de cómo se fue dando el aprendizaje (que se puede recolectar en forma de entrevistas y de revisión de perspectivas e implicaciones sobre los ‘usuarios’). Ello evidencia la necesidad de vincular una metodología ‘Soft OR’ con una metodología Dura –llamada simplemente ‘Hard OR’–; en este sentido diversos autores (Pauley y Ormerod, 1998; Salner, 2000; Mingers, 2000 y 2001; Daellenbach, 2001; Holst y Nidhall, 2001; Valqui, 2005a; Mingers, Liu y Meng, 2006) han manifestado las ventajas obtenidas de la vinculación entre una metodología ‘Soft OR’ (que utiliza información cualitativa) y otra ‘OR’ (que provea información cuantitativa) para con ello no sólo proporcionar validez al uso de la metodología; sino apoyar al proceso orientador y evaluador de su utilización.

Kushner (2002) indica que una efectiva medición del impacto organizacional no solo hace referencia a ‘hacerlo bien’, sino debe incluir responder a la pregunta “¿Qué tan bien lo hago?”. Y en esto, precisamente radica la diferencia entre validar la intervención metodológica (¿Lo hago bien?) y evaluarla (¿Qué tan bien lo hago?). Validar una intervención metodológica implica verificar que el problema “aminoro o se resolvió”, mientras que la evaluación busca proporcionar información más detallada de la

intervención metodológica al enfatizar debilidades y fortalezas en los elementos interactivos que condicionan el nivel de éxito o fracaso de toda intervención 'Soft OR' (Checkland, 2000). Por lo que más que validar se debe evaluar la intervención metodológica 'Soft OR' y con esto proporcionar una guía para el cambio (Kushner, 2002) e impulsar la espiral de indagación, aprendizaje y mejoría que toda metodología suave pretende crear.

2.4 Medición del impacto organizacional de una intervención suave

Para que una intervención suave redunde en una espiral de indagación, aprendizaje y mejoría, y no en intervenciones asiladas, es conveniente evaluarla y retroalimentarla con los aspectos propios de la organización, del equipo de trabajo o de la metodología que puedan condicionar los resultados (Tolvanen, 1998; Barnden y Darke, 2000; Connell, 2001; Valqui, 2005b; Blackman, 2006; Sorensen y Valqui, 2006); esta validación y evaluación ha de ser lo más objetiva posible (Blackman, 2006) y orientarse hacia calcular el nivel de impacto que se presente en la organización; pues esto puede representar una guía para el cambio (Kushner, 2002) e impulsar la espiral de indagación, aprendizaje y mejoría que toda metodología suave pretende crear.

Existen muchos modelos para la medición de la efectividad organizacional; el modelo más tradicional es el 'modelo de metas' (Etzioni, 1960 y 1987), el cual asume que la organización en su conjunto se orienta racionalmente hacia la obtención de metas, consecuentemente la efectividad será medida por el cumplimiento de sus salidas (objetivos y metas); presenta la desventaja de que solo puede ser aplicado en un entorno en el las metas son claras, consensuadas, que se conocen con certeza los resultados y fechas esperados; y no en un entorno en el que éstas no estén claramente definidas, sean complejas, cambiantes o contradictorias (Cameron, 1980; Murdaugh, 1998).

Otro modelo de la efectividad organizacional es desarrollado por Cameron (1980) al considerar que la organización está inmersa en un contexto de alta complejidad, con metas confusas, complejas, cambiantes y contradictorias; donde los medios para lograr los resultados no son claros y la capacidad de respuesta se vuelve crítica; todo ello contenido en lo que él llama "Anarquía Organizada". Para la cual presenta el 'modelo de inefectividades' para la efectividad organizacional (Cameron, 1980 y 1986), en el cual concibe a la organización como un conjunto de problemas y fallas, en la que la efectividad se puede evaluar mediante la ausencia de factores de inefectividad que inhiben el rendimiento exitoso de la organización (Henri, 2004), vinculando el rendimiento de la organización con la capacidad de establecer conexiones de alta calidad entre los empleados, que sean significativas y valiosas (Davel y Tremblay, 2003). Estas conexiones significativas y valiosas entre los empleados surgen en los elementos de la 'cultura organizacional' y pueden presentarse en forma de comportamientos, procedimientos, valores, asunciones o creencias; y buscan incrementar la capacidad de la gente de cooperar dentro y a través de sus unidades, facilitan una efectiva coordinación entre las distintas partes de la organización, afianzan el sentido de pertenencia de los empleados a su organización, facilitan la transmisión del propósito o meta de la empresa, promueven el dialogo y la deliberación para facilitar el aprendizaje organizacional, y mejorar la capacidad de la compañía de ser flexible y adaptarse al entorno. Por lo que algunos autores apoyan la relación negativa entre los índices de inefectividad con la 'cultura organizacional' (Baer y Freese, 2003; Bhattacharjya y Venable, 2006) o alguno de sus elementos (Stetzer, Morgeson y Anderson, 1997; Davel y Tremblay, 2003; Guerra, et al, 2004).

Existen otros modelos para entender y medir la efectividad organizacional en las organizaciones (Cameron, 1980 y 1986; Murdaught, 1998; Henri, 2004), pero los anteriormente presentados sirven para ilustrar el conflicto existente en la investigación de la efectividad organizacional, según el cual una misma organización puede juzgarse como

efectiva o inefectiva según diferentes modelos (Cameron, 1980). Otro aspecto importante a considerar es el marco de temporal: no es lo mismo medir la efectividad a corto plazo que a largo plazo, pues ello puede presentar ambigüedades e incluso contradicciones (Cameron, 1980); el mantener un enfoque en el rendimiento a corto plazo es cada vez más inadecuado para la nueva realidad de las organizaciones inmersas en un entorno de creciente necesidad de innovación y flexibilidad (Henri, 2004). Adicionalmente, existe un movimiento en la medición de la efectividad organizacional hacia consideraciones no financieras de dirección estratégica tales como participación del mercado, aprendizaje organizacional, satisfacción del cliente, satisfacción del empleado, calidad del producto; esto es debido a que se considera que estos indicadores no financieros pueden anticipar precozmente EN el rendimiento financiero a largo plazo (Banker, Potter y Srinivasan, 2000; Henri, 2004); esta postura es apoyada por algunos autores (Ittner y Larcker, 2001; Amir, Lev y Sougiannis, 2003; Lev y Radnakrishnan, 2005) al coincidir en que es crucial introducir indicadores no financieros y que nos orienten en el largo plazo al medir el movimiento hacia una dirección estratégica, en vez de una distancia a una meta (Henri, 2004).

La elección del adecuado modelo de efectividad organizacional es crítica y subordina la medición del impacto organizacional de cualquier intervención suave que se haga. En este aspecto, una postura (Murdaugh, 1998) es considerar que deben revisarse a fondo las fortalezas y debilidades conceptuales de cada modelo y compararlas con las circunstancias específicas, para en base a ello seleccionar el más adecuado; una mejor postura es la complementaria (Henri 2004), que indica que ninguno de estos modelos es apropiado en todas las circunstancias y organizaciones, por lo que deben usarse como guías y complementos para evaluar la efectividad de una organización. En este aspecto el “modelo de metas” está más relacionado al desempeño a corto plazo, y se complementa con el “modelo de inefectividades” que se relaciona más con el desempeño y supervivencia a largo plazo (Henri, 2004; Banker, Potter y Srinivasan, 2005).

En la intersección del “modelo de metas” y del “modelo de inefectividades” para la eficiencia organizacional se puede lograr una evaluación objetiva del impacto de una metodología suave; y con ello tener una guía para el cambio (Kushner, 2002) que impulse la espiral de indagación, aprendizaje y mejoría que se pretende crear en la organización. Bajo el paradigma del “modelo de metas” se implica la utilización de índices cuantitativos para medir la efectividad en el corto plazo y observar su distancia a una meta; mientras con la utilización del “modelo de inefectividades” es necesario incluir indicadores cualitativos que al evaluar el cambio en la ‘cultura organizacional’ nos orienten en el largo plazo hacia el movimiento en una dirección estratégica. Este doble proceso de validación proporciona información detallada de la intervención metodológica y ayuda a enfatizar debilidades y fortalezas en los elementos que condicionan el nivel de éxito o fracaso de una intervención ‘Soft OR’ y podrá apoyar la espiral de indagación, aprendizaje y mejoría que toda metodología ‘suave’ pretende crear.

Por lo tanto, para medir objetivamente el impacto de una metodología suave, es necesario introducir índices cuantitativos y cualitativos que valoren el cambio organizacional en el corto plazo (distancia a una meta) y en el largo plazo (en la capacidad de la organización de establecer conexiones de alta calidad entre sus miembros, que sean significativas y valiosas); y con ello valorar objetivamente su utilidad, además de deducir fortalezas y debilidades de la intervención metodológica.

3 Metodología

En este artículo se pretende evaluar el nivel de impacto organizacional logrado por una metodológica suave y no enfocarse en analizar el proceso de la intervención metodología suave en sí mismo; con lo que los resultados y conclusiones obtenidos serán más

generales y aplicables a cualquier metodología ‘Soft OR’. Por lo anterior, se describirá de manera general el procedimiento de la intervención suave, y posteriormente de manera más detallada la metodología de evaluación de impacto organizacional que se haya logrado.

3.1 Metodología para la intervención suave

Se presenta un caso de aplicación realizado en una empresas de la ciudad de Mexicali, Baja California, México; del tipo de maquiladora del ramo electrónico, donde se trabajo sobre el área de producción.

Se integró un equipo de trabajo de cuatro personas, el cual se reunió una vez por semana durante un periodo de cinco meses. Durante los cuales por requerimientos de la metodología, recibió entrenamiento en los principios subyacentes de las corrientes de “sistemas suaves”, “aprendizaje organizacional”, y de “dinámica de sistemas”; con lo que se buscó sensibilizar y modificar sus “modelos mentales” y “rutinas internalizadas”, para promover que el análisis y las mejorías realizadas no sean intervenciones aisladas sino que produzcan la espiral de mejoría esperada en la organización.

Esta propuesta metodológica ‘Soft OR’ es derivada de la SSM y se desarrolla en siete etapas que se describen a continuación:

Etapa 1 - Reconocer la complejidad del entorno

- Conformar un equipo de trabajo
- Entender la necesidad de una “intuición colectiva” para la resolución de problemas, y el entendimiento de las barreras para lograrlo.

Etapa 2 - Expresar el problema raíz

1. Deducir el problema principal que genera y abarca la problemática del área.
2. Generar un esquema o diagrama para sustentar y justificar el problema raíz expresado; que ilustra y mejora el entendimiento por otros miembros de la organización, los cuales pueden complementarlo.

Etapa 3 - Percibir el flujo del actuar organizacional

- Lograr una mayor comprensión de la estructura organizacional; y sensibilizar hacia los patrones y ciclos de comportamiento que pueden estar causando problemas en la organización.
- Expresar de manera la genérica la orientación de los esfuerzos colectivos hacia la eliminación de las causas estructurales del mal comportamiento organizacional.

Etapa 4 - Propuestas de cambio al entorno decisorio

- Expresar de manera específica tres alternativas concretas orientadas a contrarrestar de manera estructural el “problema raíz” del área.

Etapa 5 - Refinar y sustentar los cambios propuestos

- Sustentar y refinar los cambios específicos propuestos mediante la generación de un esquema de actividades, que muestre y clarifique las actividades y supuestos tácitos no expresados.

Etapa 6 - Implementación de cambios

- Compartir y mejorar el esquema de actividades por otros miembros de la organización, al revisar cada actividad propuesta para reconocerla o ajustarla para que sea deseable y factible de implementar
- Detallar y definir roles relacionados con las propuestas y sus esquemas de actividades para introducir los cambios en la organización.

- Realizar cambios en el entorno decisorio del departamento basados en las “propuestas específicas” expresadas y sustentadas en las dos etapas anteriores.

Etapa 7 - Monitoreo y evaluación del impacto

- Procurar y verificar que los resultados perseguidos en cada paso se alcancen de la mejor manera posible.
- Establecer pautas para la evaluación del impacto logrado en la mejoría organizacional.

3.2 Metodología para la evaluación del impacto organizacional logrado

Para la evaluación objetiva del impacto organizacional se utilizó de manera complementaria el ‘modelo de metas’ que implica la utilización de índices cuantitativos para medir la efectividad en el corto plazo al medir la distancia a una meta; en complemento con el ‘modelo de ineffectividades’ con la utilización de indicadores cualitativos para evaluar el cambio en la ‘cultura organizacional’ que nos oriente en el largo plazo hacia el movimiento en una dirección estratégica.

Para la evaluación de la efectividad en el corto plazo se proponen y revisan algunos indicadores cuantitativos de productividad específicos del área (tales como nivel de desperdicio, tiempos muertos, demoras de suministro, demoras de envíos, cantidad de quejas, utilización del equipo, etc.), y se verifica su cambio; en este caso se utilizó el indicador de ‘porcentaje de rechazos de calidad’.

Para medir el impacto metodológico en el largo plazo se evaluó el cambio organizacional en su habilidad de “establecer conexiones de alta calidad entre los empleados, que sean significativas y valiosas”. Para valorar la organización en su ausencia de factores de ineffectividad y en su movimiento en una dirección estratégica se utilizó el ‘modelo Denison’ para la medición de la cultura y efectividad organizacional (Denison Organizational Culture Survey questionnaire – DOCS); este evalúa a la organización en 12 dimensiones esenciales: Empowerment, Orientación de trabajo en equipo, Capacidad de desarrollo, Valores núcleo, Consensos y acuerdos, Integración y coordinación, Capacidad de crear cambios, Enfoque en el cliente, Aprendizaje organizacional, Visión, Claridad en objetivos y metas; y Manejo de dirección estrategia (Denison, 1990; Denison y Mishra, 1995; Juechter, Fisher y Alford, 1998; Fisher, 2000). El instrumento de medición es un cuestionario que consta de 60 preguntas y que ya ha sido validado en múltiples estudios (Fey y Denison, 2003; Davidson, 2003; Denison, et al, 2005). Una vez aplicado el instrumento, se comparan los resultados con la “base de datos normativa” que contiene las respuestas de más de 1,500 organizaciones de todo el mundo; y se genera una calificación en forma de percentil (porcentaje) que indica la proporción de empresas que se encuentran por debajo de una determinada puntuación.

El cuestionario DOCS se aplicó de manera inicial en los miembros del equipo de trabajo y adicionalmente sobre otros cinco miembros de la organización inmersos en la problemática del área; para aplicarlo nuevamente al final de los cinco meses de trabajo y medir su diferencia.

4 Resultados

El objeto de este estudio es evaluar de manera objetiva el impacto organizacional logrado luego de la intervención metodológica suave, y con ello poder identificar tanto aspectos fuertes en la metodología, como puntos débiles que conviene mejorar; con lo que se podrá disponer de una guía para el cambio e impulso de la espiral de indagación, aprendizaje y mejoría que toda metodología suave pretende crear. Por lo que primeramente se proporcionaran de manera general los resultados propios de la intervención metodológica

suave, para posteriormente detallar los resultados de la evaluación del impacto organizacional.

4.1 Resultados de la intervención suave

Como resultados específicos de la intervención suave se puede indicar que se recolectaron los problemas expresados del área de producción, para posteriormente interrelacionarlos y expresar el problema raíz: "Falta de atención por parte de Ingenieros y supervisores al fomentar la buena actitud y compromiso de los operadores hacia la calidad y mejora del proceso, a pesar de las urgencias de producción, falta de sentido de pertenencia y descuidos de los operadores".

Posteriormente, este "problema raíz" se estudio y discutió a buscando los patrones y ciclos de comportamiento que podrían estar causando problemas en la organización; para con ello obtener una propuesta general, que resulto en: "Fomentar políticas y lineamientos organizacionales que promuevan el compromiso hacia la calidad del producto y del proceso por parte de los operadores, del supervisor y de los ingenieros de soporte; a pesar de las urgencias del área, la falta de pertenencia y los descuidos de los operadores".

A partir de la "propuesta general" se definieron tres "propuestas específicas" de actuación orientadas a atacar y eliminar el "problema raíz enriquecido". Todo esto fue hecho por el equipo de trabajo a partir del entendimiento e indagación tanto de los procedimientos, procesos y formas actuales de trabajo; así como de las políticas y lineamientos existentes en la organización. Las propuestas específicas resultantes se proporcionan a continuación; y en la (FIGURA 1) se muestra el sistema de actividades que soporta y explica las propuestas específicas de actuación.

- 1) El soporte al área de producción viene dado por tres áreas: Calidad del producto (QA), Ingeniería y Calidad de Materiales (SQE), quienes respectivamente se encargan de la calidad del producto, la calidad del proceso y calidad de los materiales. Se tiene un cierto descontrol y mala actitud de los operadores debido al área de soporte, junto por problemas y errores por descuidos de los operadores. Por lo que se plantea nombrar como líder de soporte a QA y ponerlo más al alcance del proceso y de los operadores. Y así vincular más eficientemente las actividades de soporte con la producción.
- 2) Las dos líneas de producción existentes en el área tienen aprox. 50 operadores cada una, lo cual dificulta la detección rápida de los errores en los operadores y el mejorar la actitud y compromiso de estos. Para lo cual se plantea el segmentar la línea y su celda en tres secciones, y asignarlas a las dos guías y al supervisor.
- 3) Realización de juntas semanales de no más de $\frac{1}{2}$ hora entre los supervisores de producción del área, y los tres responsables de soporte (QA, Ingeniería y SQE). En la cual se revise el desempeño y la productividad semanal del área y se realicen propuestas y acciones orientadas a la eliminación de problemas. Con presentaciones bimestrales de resultados ante los gerentes.

Todo lo anterior se presentó ante gerencia, para iniciar una etapa de implementación y monitoreo, la cual duro dos meses, donde luego se inicia la etapa de medición del impacto organizacional logrado.

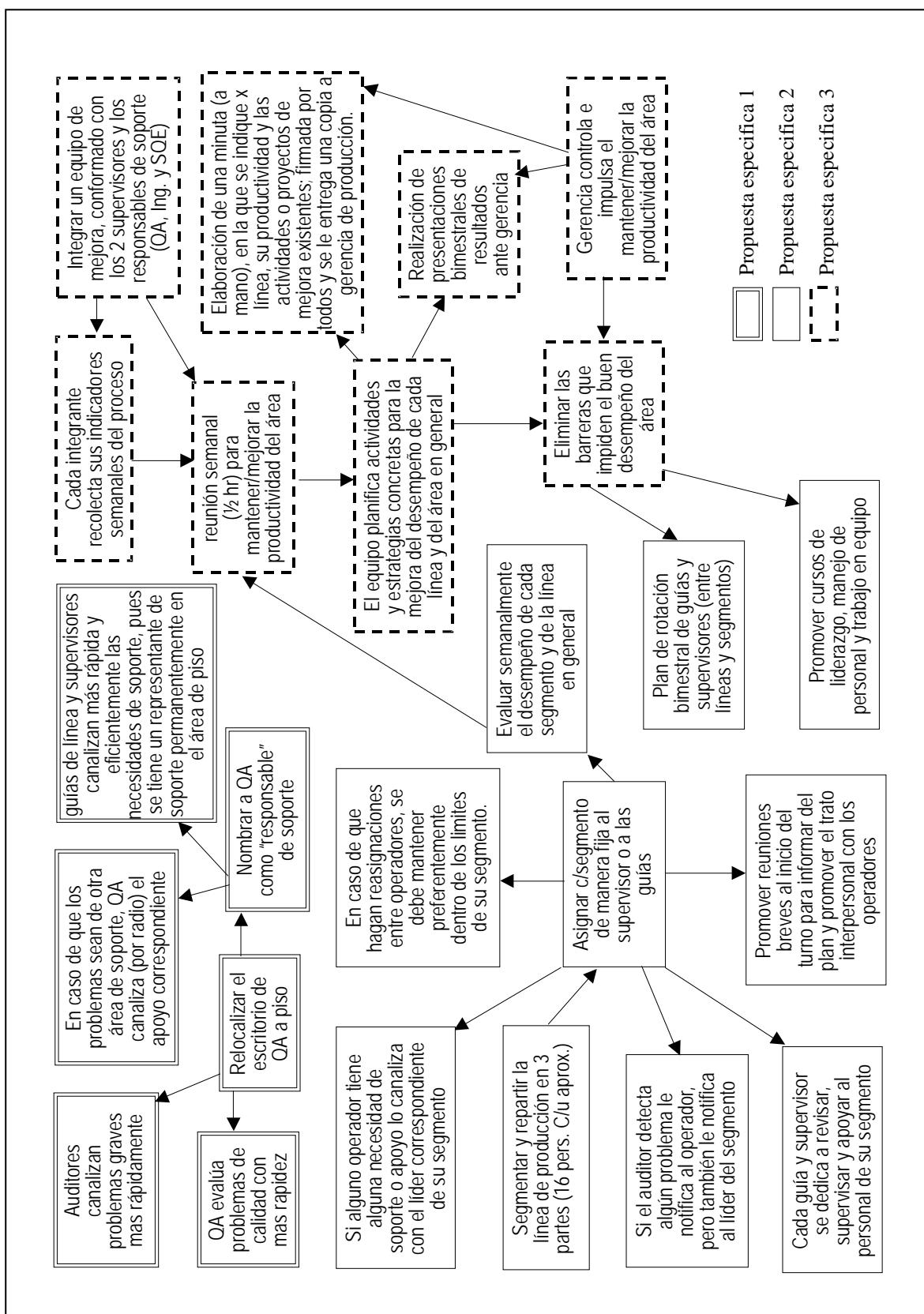


Figura 1 Caso A, Sistema de actividades que soporta las propuestas específicas

4.2 Resultados de la evaluación del impacto organizacional

Después de la implementación de las propuestas, para la evaluación de la efectividad en el corto plazo se monitoreó durante dos meses el indicador cuantitativo específico de 'porcentaje de rechazos de calidad', el cual evolucionó favorablemente, mostrando una reducción en su manejo, tal y como se ilustra en la (FIGURA 2).

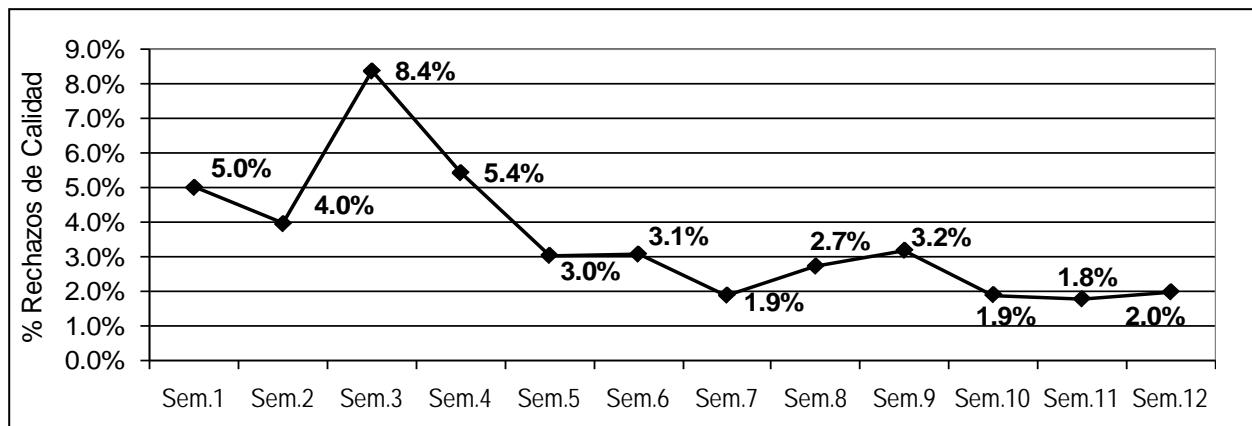


Figura 2 Grafica de la tasa de rechazos

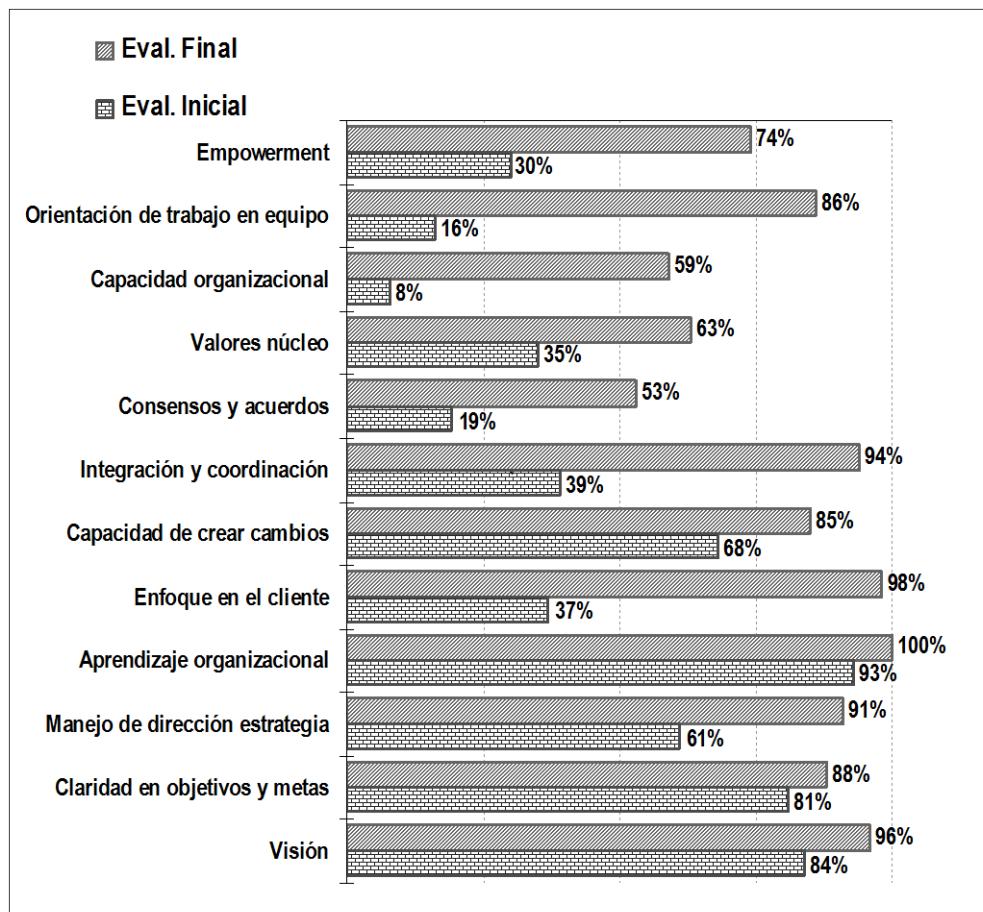


Figura 3 Comparativo de indicadores cualitativos estratégicos según el modelo Denison

Para la evaluación de la efectividad en el largo plazo se aplicó de manera inicial el cuestionario Denison a los cuatro miembros del equipo de trabajo y a otras cinco miembros inmersos en la problemática-, para nuevamente usar este instrumento después de la intervención metodológica, que duró cinco meses en esas mismas personas, su comparativo nos proporciona los indicadores cualitativos que se muestran en la (FIGURA 3), para cada una de las doce dimensiones críticas que indica el modelo DOCS. En la (FIGURA 4), se muestran los totales por rasgo estratégico, según el modelo 'Denison'

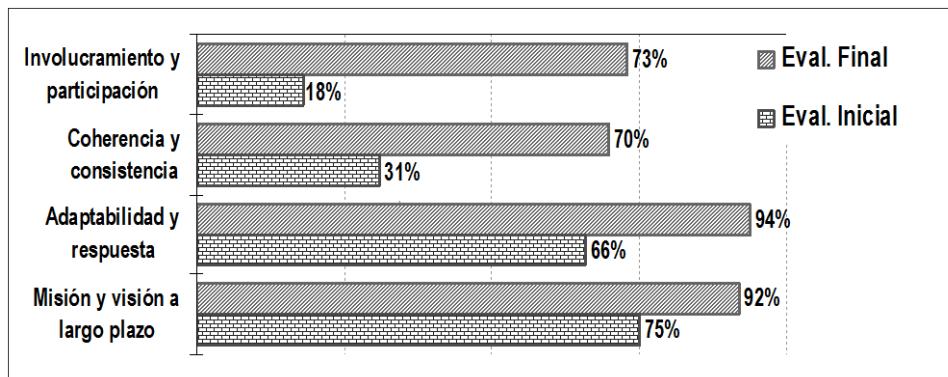


Figura 4 Comparativo de rasgos estratégicos cualitativos según el modelo Denison

5 Discusión

La metodología 'Soft OR' planteada es derivada de la metodología de sistemas suaves de Checkland a la que se le han agregado elementos para que la 'definición raíz' de sistemas relevantes tenga una vinculación más directa hacia la problemática existente; Ya que el enfatizar la participación de grupos de trabajo para resolver problemas e impulsar el mejoramiento es una tipología que orienta a las organizaciones al aprendizaje (Yeung; 2000). Así mismo, apoya en la eliminación de barreras del aprendizaje pues el esquema generado para la obtención del 'problema raíz' orienta esfuerzos y crea sinergias hacia el entendimiento y eliminación de la problemática existente en el área; pues "si las personas empiezan a compartir ideas acerca de los asuntos que perciben como realmente importantes, esto crea por sí mismo una cultura de aprendizaje" (Dixon, 2001, p.16)

Para la evaluación objetiva del impacto organizacional se utilizará de manera complementaria el 'modelo de metas' que implica la utilización de índices cuantitativos para medir la efectividad en el corto plazo al medir la distancia a una meta; en complemento con el 'modelo de inefectividades' con la utilización de indicadores cualitativos para evaluar el cambio en la 'cultura organizacional' que nos oriente en el largo plazo hacia el movimiento en una dirección estratégica.

Para la evaluación de la efectividad en el corto plazo se utilizó el 'porcentaje de rechazos de calidad' como indicador representativo del área, este se monitoreo durante dos meses indicando una evolución favorable, lo cual se ilustra en la figura 2. Desgraciadamente con la información de la figura 2, es difícil concluir algo adicional a que "el problema aminoró o se resolvió", mientras que la utilización de parámetros cualitativos a través del uso del 'modelo de inefectividades' proporciona más información, tal y como se observa en las figuras 3 y 4. Esto evidencia las limitantes del 'modelo de metas' al simplemente utilizar indicadores cualitativos para medir la distancia a una meta como indicador de la efectividad organizacional; por lo que la valoración no solo ha de ser cuantitativa, sino incluir aspectos cualitativos para poder responder a la pregunta "¿Qué tan bien lo hago?" Kushner (2002) y con ello apoyar al proceso orientador y de análisis de la intervención metodológica suave. Al medir el impacto metodológico en el largo plazo en el cambio organizacional en su habilidad de "establecer conexiones de alta calidad entre

los empleados, que sean significativas y valiosas" (Davel y Tremblay, 2003); con lo que se pretende orientar el largo plazo al medir el movimiento hacia una dirección estratégica (Henri, 2004), y no solo indicar la distancia a una meta.

Este doble proceso de validación (cualitativa y cuantitativa) mostrado en las figuras 2,3 y 4, elimina el componente subjetivo en la valoración de la metodología; al vincularla más directamente al nivel de impacto que debe presentarse en la organización y ser los más objetiva posible (Blackman; 2006), al emplear el 'modelo de metas' que esta mas relacionado al desempeño a corto plazo, en complemento con el 'modelo de ineffectividades' que se relaciona mas con el desempeño y supervivencia a largo plazo (Henri, 2004; Banker, Potter y Srinivasan, 2005). Esto independientemente de apoyar al proceso orientador y evaluador de la metodología, proporciona validez al uso de la metodología suave dentro de la organización.

En base a la información que se muestra en la figura 3, se puede afirmar que la metodología fortaleció la percepción de que es mejor trabajar en equipo que de manera individual, la sensibilidad hacia las necesidades del cliente de su proceso, y su capacidad y destrezas para resolver los problemas de la organización; pues todas estas habilidades crecieron más de un 50%. Lo que indica la utilidad de la intervención metodológica hacia la sensibilización de las necesidades del cliente, y sobre sus creencias sobre la conveniencia y gusto por el trabajo en equipo.

En la figura 4, se muestra que las áreas de 'Adaptabilidad y respuesta' y de 'Misión y visión a largo plazo' ya eran altas antes de la intervención metodológica, lo que significa que su capacidad de crear cambios en el proceso y de aprendizaje organizacional, así como el sentido de dirección estratégico y la claridad en objetivos y metas de la empresa era anteriormente a la intervención bastante alto, por lo que este aspecto de la organización pudo apoyar al nivel de éxito de la metodología; por lo que sería interesante evaluar el nivel de éxito en una organización que en la que no ocurriera esta característica; lo que podría resultar en la necesidad de reforzar estos rasgos en la capacitación a los 'usuarios' de la metodología.

6 Conclusiones

El lograr inducir una espiral de indagación, aprendizaje y mejoría dentro de la organización representa la mejor y más sostenible ventaja en un entorno de creciente complejidad y competencia; para lo cual es necesario enfatizar la participación de grupos de trabajo para resolver problemas y focalizar mejorías, fomentando el aprendizaje organizacional como fuente de ventaja competitiva.

Se concluye que si una intervención suave resulta aislada y no forma una espiral de indagación, aprendizaje y mejoría, e incluso no conduce al éxito; es porque su eficacia estuvo condicionada por aspectos propios de la organización y del equipo de trabajo e incluso por problemas inherentes a la propia metodología. Aspectos que es conveniente evaluar y retroalimentar en la intervención metodológica para con ello proporcionar una guía para el cambio que podría orientar a la obtención de mayores mejorías dentro de la empresa. Al enfatizar deficiencias en el entrenamiento o guía del equipo involucrado, o necesidad de estandarización y monitoreo de los procedimientos aplicados; para modificar mas efectivamente los patrones de comportamiento de la organización.

El evaluar el impacto organizacional logrado por una intervención metodológica permite demostrar objetivamente la fortaleza de una metodología suave, lo que promueve su utilización rutinaria dentro de la organización, con lo que en lugar de efectuarse intervenciones aisladas se podría formar una espiral de indagación, aprendizaje y mejoría dentro de la empresa. En la intersección del "modelo de metas" y del "modelo de ineffectividades" para la medición de la eficiencia organizacional se puede lograr una

evaluación mucho más objetiva del impacto de una metodología suave. Pues cualquier proyecto de mejora estará soportado tanto en el corto plazo al indicar una mejoría en los indicadores cuantitativos, como en el largo plazo por la utilización de indicadores cualitativos para evaluar a la organización en su orientación hacia una dirección estratégica.

Una importante área de investigación es sobre los mecanismos y modelos adecuados para una adecuada evaluación de una metodología ‘Soft OR’; la cual debe enfatizar sus fortalezas y debilidades en los tres elementos interactivos (Checkland, 2000) que condicionan una intervención ‘Soft OR’: ‘Usuario’, ‘procedimientos metodológicos’ y ‘situación percibida’. Lo que podría resultar la mejor selección o capacitación a los ‘usuarios’, clarificar los principios metodológicos o hacer que el usuario se apropie mas del método al considerarlo adecuado para su situación y organización particular.

7 Referencias

- Amir E., Lev V., Sougiannis T., (2003), “Do Financial Analysts Get Intangibles?”, *European Accounting Review*, 12(4):635-659
- Ackoff R.L., (1979), “Resurrectiong the Future of Operational Research”, *The Journal of the Operationa Research Society*, 30(3):189-199
- Argyris C. (2001), “Sobre el aprendizaje organizacional”, Oxford university press, México
- Baer M., Frese M., (2003), “Innovation is not enough: climates for initiative and psychological safety, process innovations, and firm performance”, *Journal of Organizational Behavior*, 24:45-68
- Banker R.D., Potter G., Srinivasan D., (2000), “An Empirical Investigation of an Incentive Plan that Includes Nonfinancial Performance Measures”, *The Accounting Review*, 75(1):65-92
- Banker R.D., Potter G., Srinivasan D., (2005), “Association of Nonfinancial Performance Measures with the Financial Performance of a Lodging Chain”, *Cornell Hotel and Restaurant Administration Quarterly*, 46(6):394-412
- Barnden A.W., Darke P., (2000), “A Comparison of SSM with an Organisational Learning Model”. in Proceedings of the International Conference on Systems Thinking in Management (ICSTM 2000), Australia
- Bartlett C., Ghoshal S., (2002), “Building competitive advantage through people”, *MIT Sloan Management Review* 43(2):34-41
- Bergvall K.B., Grahn A. (1996), “Expanding the framework for monitor and control in Soft Systems Methodology”, *Systemic Practice and Action Research*, 9(5):469-495
- Bhattacharjya J., Venable J., (2006), “The Mutual Influence of Organizational Culture and SSM Applied to SISP – An Action Research Study in a Non-Profit Organization”, The tenth Pacific Asia Conference of Information Systems (PACIS 2006), pp.516-530
- Blackman D., (2006), “How measuring learning may limit new knowledge creation”, *Journal of Knowledge Management Practice*, 7(3)
- Cameron K. (1980), “Critical Questions in Assessing Organizational Effectiveness”, *Organizational Dynamics*, 15:66-80
- Cameron K. (1986), “Effectiveness as paradox: Consensus and conflict in conceptions of organizational effectiveness”, *Management Science*, 32(5):539-553
- Checkland P., Scholes J., (1990), “Soft systems methodology in action”, Edit. Wiley, USA
- Checkland P., (2000), “Soft Systems Methodology: A thirty year retrospective”, *Systems Reseach and Behavioral Science*, 17(S1):11-58
- Checkland P., (2001), “Pensamiento de Sistemas, Practica de Sistemas”, Limusa, Mexico
- Connell N., (2001), “Evaluating soft OR: some reflections on an apparently ‘unsuccessful’

- implementation using a Soft Systems Methodology (SSM) based approach”, Journal of the Operational Research Society, 52(2):150-160
- Daellenbach H., (2001), “Hand OR, Soft OR, Problem Structuring Methods, Critical Systems Thinking: A Primer”, Conference Twenty Naught One of the Operational Research Society of New Zealand
- Dash D.P., (1994), “System Dinamics: Changing Perspectives”, *Systems Practice* 7(1):87-98
- Davel E., Tremblay D.G., (2003), “Organizational culture and social performance : insights from the experience of family organizations”, presented in Iberoamerican Academy of Management, 3rd International Conference, Sao Paulo
- Davidson G.M., (2003), “The relationship between organisational culture and financial performance in a South African investment Bank”, *Industrial and organisational psychology*, University of South Africa
- Denison D.R., Janovics J., Young J., Cho H.J., (2005), “Diagnosing Organizational Cultures: Validating a Model and Method”, International Institute for Management Development, Working paper 2005-11
- Denison D.R., Mishra A.K., (1995), “Toward a Theory of Organizational Culture and Effectiveness”, *Organization Science*, 6(2):204-223
- Denison D.R., (1990), “Corporate Culture and Organizational Effectiveness”, John Wiley & Sons
- Dixon N., (2001), “El conocimiento común: Cómo prosperan las compañías que comparten lo que saben”, Oxford university press, Mexico
- Etzioni A., (1987), “Normative-Affective factors: Toward a new decision-making model”, *Journal of Economic Psychology*, 9:125-150
- Etzioni A., (1960), “Two approaches to Organizational Analysis: A Critique and a Suggestion”, *Administrative Science Quarterly*, 5(2):257-278
- Fey C.F., Denison D.R., (2003), “Organizational Culture and Effectiveness: Can American Theory Be Applied in Russia?”, *Organization Science*, 14(6):686-706
- Fisher C.J., (2000), “Like It or Not ... Culture Matters”, *Employment Relations Today*, 27(2):43-52
- Flood R., (1998), “Fifth Discipline: Review and discussion”, *Systemic Practice and action research* 11(3):259-273
- Flood R., (1995), “Solving problem solving”, John wiley & Sons
- Forrester J.W., (1989), “The Beginning of Systems Dynamics”, MIT, Banquet Talk at the international meeting of the System Dynamics Society Stuttgart, Germany, D-4165-1, USA, p.16
- Fuentes A., (1993), “El pensamiento sistémico, Serie: Cuadernos de Planeación y Sistemas, Vol.3.”, DEPFI, UNAM, México
- Gomez P.J., Lorente J.C., Cabrera R.V., (2005), “Organizational learning capability: a proposal of measurement”, *Journal of Business Research*, 58:715-725
- Grinstein C., s/f, “El pensamiento sistémico: Parte 1”, <http://www.cepao.org/contenidoscepao/ArticuloMes.htm> (Con acceso en Octubre de 2005)
- Guerra J., Martinez I., Munduate L., Medina F., “A contingency perspective on the study of the consequences of conflict types: The role of organizational culture”, 17th Conference of the Internationale Association for Conflict Management (IACM 2004)
- Henri J.F., (2004), “Performance Measurement and Organizational Effectiveness: Bringing the gap”, *Managerial Finance*, 30(6):93-123
- Heslin P.A., Moldoveanu M., (2002), “What’s the ‘real’ problem here? A model of problem formulation”, presentado en “Managing the Complex IV: Conference on Complex Systems and the Management of Organizations”
- Holst M., Nidhall L., (2001), “Soft Systems Methodology and Organizational Informatics”,

- Institutionen for Industriell ekonomi och systemvetenskap
- Ittner C.D., Larcker D.F., (2001), "Determinants of performance measure choices in worker incentive plans", *Journal of Labor Economics*, 20(2):S58-S90
- Joldersma C., Roelofs E., (2004), "The impact of soft OR-methods on problem structuring", *European Journal of Operational Research*, 152:696-708
- Juechter W.M., Fisher C.J., Alford R.J., (1998), "Five Conditions for High-Performance Cultures", *Training & Development*, 52(5):63-67
- Kushner R.J., (2002), "Action Research Validation of an Inventory of Effectiveness Measures", Conference "Nonprofit Organizational Effectiveness and Performance"
- Leonard A., Beer S., (1994), "The Systems Perspective: Methods and Models for the Future", AC/UNU Project,
- Lev B., Radhakrishnan S., (2005), "The Valuation of Organization Capital," in Corrado C., Haltiwanger J., and Sichel D., eds., "Measuring Capital in a New Economy", National Bureau of Economic Research and University of Chicago Press, pp.73-99.
- Lopez M.P., Delgado A., (2003), "Application of a systems methodology in the diagnosis of the organizational culture in a telecommunications company", *International Journal of Computers, Systems and Signals*, 4(1):22-32
- Marmenout K., (2006), "Aligning your Culture with your Growth: Where to go and how to get there", McGill University - Faculty of Management, Working Paper Series, Available at SSRN: <http://ssrn.com/abstract=941991>
- Miller J.G., (1995), "Applications of Living Systems Theory", *Systemic Practice and Action Research*, 8(1):19-45
- Miller J.G., (1982), "The Earth as a System", *Behavioral Science*, 27(4):303-322
- Mirijamdotter A. (1998), "A multi-modal systems extension to soft systems methodology", Institutionen for Industriell ekonomi och samhällsvetenskap
- Mingers J., Liu W. y Meng W., (2006), "Combining SSM and DEA: Evaluating the Basic Research Performance of the Chinese Academy of Sciences", Kent Business School, Working Paper Series, No.128
- Mingers J., (2001), "Combining IS Research Methods: Towards a Pluralist Methodology", *Information Systems Research*, 12(3):240-259
- Mingers J., (2000), "Variety is the spice of life: combining soft and hard OR/MS methods", *International Transactions in Operational Research*, 7(6):673-691
- Murdaugh J., (1998), "Organizational Effectiveness and Executive Succession: Conclusions About and Implications for Florida's Municipal Police Chiefs", Senior Leadership Program Publications, SLP-5
- Ormerod R.J., (2001), "The success and failure of methodologies –a comment on Connell(2001): Evaluating Soft OR", *Journal of the Operational Research Society*, 52(10):1176-1179
- Pardo M, Martinez C., (2005), "Resistencias al cambio organizativo: un análisis empírico en cambios reactivos y anticipativos", *M@n@gement* 8(3):47-67
- Pardo M., Martinez C., (2003), "Resistance to change: A literature Review and empirical study", *Management Decision* 41(2):148-155
- Pauley G.S., Ormerod R.J. (1998), "The Evolution of a Performance Measurement Projecct at RTZ", *INTERFACES*, 28(4):94-118
- Pidd M., Brown J., Cooper C., (2006), "A taxing problem: The complementary use of hard and soft OR in public policy", *European Journal of Operational Research*, 172(2):666-679
- Richmond B., (1994), "System Dynamics/Systems thinking: Let's just get on with it", Delivered at the International Systems Dynamics Conference in Sterling, Scotland, USA, 25 pags.
- Roth G., (2001), "El lado humano del cambio: La innovación y el aprendizaje en la organización", Oxford university press, Mexico

- Salner M., (2000), "Beyond Checkland & Scholes: Improving SSM", Occasional Papers on Systemic Development, University of Western Sydney, 11:23-44
- Senge P., (1998), "La quinta disciplina", Granica, Mexico
- Sorensen L., Valqui R.V., (2006), "Evaluating six soft approaches", Informatics and Mathematical Modelling, Technical University of Denmark, DTU.
- Sterman J.D., (1991), "A Skeptic's guide to computer models", MIT, Wetview Press, D-4101-1, USA, 25 pages.
- Sterman J.D., (2001), "System Dynamics Modeling: Tools for learning in a complex World", California management review 43(4):8-25
- Stetzer A., Morgeson F.P., Anderson E.L., (1997), "Organizational Climate and Ineffectiveness: Evidence from 25 Outdoor Work Crew Divisions", Journal of Quality Management, 2(2):251-265
- Taylor S., Robinson S., Ladbrook J., (2005), "An investigation into the use of net-conferencing groupware in simulation modelling", Journal of Computing and Information Technology, 13(2):95-106
- Taylor S., Robinson S., Ladbrook J., (2003), "Towards collaborative simulation modelling: Improving human-to-human interaction through groupware", Proceedings of 17th European Simulation Multiconference, pp.474-482
- Tolvanen J.P., (1998), "Incremental Method Engineering with Modeling Tools", University of Jyväskylä, Finland
- Torras L., (1997), "Aprender: La ventaja competitiva mas sostenible en el tiempo", Alta Dirección, 191:13-19, España
- Ulrich W., (2000), "Reflective Practice in the Civil Society: The contribution of Critically Systemic Thinking", Reflective Practice 1(2):247-268
- Ulrich W., (1998), "Systems Thinking as if People Mattered: Critical Systems Thinking for Citizens and Managers", Lincoln School of Management, Woring Paper No. 23, <http://www.geocities.com/csh_home/downloads.html> (accessed in July 2007)
- Valqui R.V., (2005a), "Operational Reseach: A multidisciplinary discipline", Informatics and Mathematical Modelling, Technical University of Denmark, DTU.
- Valquil R.V., (2005b), "Soft OR approaches", Engevista, 7(1):4-20
- Wilson, B., (1993), "Sistemas: Conceptos, metodologías y aplicaciones", Megabyte - Grupo Noriega Editores, Mexico
- Young S., Tu Y.M., Tseng Y.T., (1999), "Organizational Learning as a feedback system: a Conceptual Framework", presented in The 17th International Conference of The System Dynamics Society and the 5th Australian & New Zealand Systems Conference
- Yeung A., Ulrich D., Nason S., VonGlinow M., (2000), "Las capacidades de aprendizaje en la organización", Oxford university press, México

Some rankings for the Athens Olympic Games using DEA models with a constant input

João Carlos Correia Baptista Soares de Mello †

Lidia Angulo Meza †

Brenda Branco da Silva †

† Federal Fluminense University
Brazil

jcsmello@producao.uff.br

lidia_a_meza@pq.cnpq.br

brendabranco@yahoo.com

Abstract

There is no official method to establish a final ranking for the Olympic games. The usual ranking is based on the Lexicographic Multicriteria Method, the main drawback of which is to overvalue gold medals. Furthermore it does not take in account that the various sports may be of different importance. This work proposes a ranking model to eliminate those drawbacks. First we use a modified cross evaluation DEA (Data Envelopment Analysis) model with weighted restrictions for each sport. The outputs are the number of gold, silver and bronze medals and the input is a unitary constant for all countries. After obtaining a rank for each and every sport we build a general ranking using a weighted sum. The weights are calculated taking in account the number of countries that participated in each sport. We use our model with the results of the Athens Olympic Games.

Keywords: DEA, Olympic, Ranking, weight restrictions, unitary input

1 Introduction

The first recorded Olympic Games occurred in Olympia, Greece in 776 ac (Fischer, 2003). The modern Games, were born from an initiative of Baron de Coubertin in 1892 and occurred in 1896 in Athens. The Games tried to maintain the initial spirit of individual competition. As noted by Lins et al (2003) the purpose clearly failed and the Games have become a national competition.

Despite their national character, the Olympic Committee has never issued an official ranking to pick an overall Olympic winner country. However the IOC presents the medals data in a table that suggests a ranking. As a matter of fact, the mass media do use this table as a ranking. This quasi-official ranking is based on the Lexicographic Multicriteria method, as explained in Lins et al (2003). This ranking does not deal properly with the

possible existence of countries that have won a large number of silver and bronze medals but no gold medal.

In the literature there are some studies for alternative rankings for the Olympic Games. In almost all of them the rankings were obtained by comparing the number of medals earned with explicative variables. In all of them there is no difference in the value of medals earned in different competitions.

We have two main goals in this study. The first one is to provide a rank (different from the quasi-official ranking) that takes into account only the number of medals won and not the explicative variables. This ranking must not over evaluate the gold medal. We must mention that there are already at least two rankings with this characteristic. We believe that our ranking has some theoretical advances in the DEA field.

The second and most important goal is to take into account that medals won in different competitions do not have the same value. As a matter of fact, the existing rankings do not take into account that in some disciplines there are more events than in others, and so there are more possibilities of winning a medal. For instance, in gymnastics there are a lot of gold medals to be earned and in football there are only two possibilities for a country to win a gold medal (one for men, the other for women). As far as we know this the first study on that matter at least concerning Olympic rankings.

To take into account the difference in winning values for different sports, we aggregate competitions into clusters, as done by the IOC (www.olympic.org): each discipline is a cluster.

Our method is illustrated using data from the 2004 Athens Olympic Games.

In the next section we review the literature about Olympic rankings and the literature of Data Envelopment Analysis rankings in sports. In section 3, we establish the fundamentals of our ranking, the rankings for each sport and the two general rankings. At last, section 4 presents some conclusions concerning the results we have obtained.

2 A Review on the Analysis of the Olympic Games results

The Lexicographic Method is not the sole method used to rank countries in the Olympic Games. Some newspapers produce a ranking determining the total number of medals earned by each country. They simply add up bronze, silver and gold medals. The obvious drawback of this method is to under-evaluate gold medals.

An alternative approach is to make an arbitrary evaluation of each medal, for instance, 1 point for bronze, 2 for silver and 3 for gold. This is a much unsophisticated approach, as it assumes all medals to be equally desired, albeit in proportion to their value.

The previous approaches follow contradictory assumptions. It is important to study alternative ways to rank competitors in the Olympic Games. Morton (2002) used statistical methods to determine "who wins the Olympics". Other statistical approach considering socio economical variables and the number of medal earned has been performed by Bernard and Busse (2004).

There are already some approaches using DEA to establish Olympic rankings. The very first one was proposed by Lozano et al (2002). They used population and GNP as inputs and the medals as outputs. In a similar approach, Lins et al (2003) built a new model taking in account one more constraint: the total amount of medals is a constant. This resulted in the development of a new model, the so-called Zero Sum Gains DEA model (ZSG-DEA). Churilov and Flitman (2006) used DEA to establish a ranking, the inputs of which were some social economics variables. Instead of using as outputs the number of gold, silver and bronze medal, they used four linear combinations of these figures. This approach eliminates the problem of nil valued weights. For each country, they determined which output has the greatest weight, in order to divide the countries into clusters. They also made a classical cluster analysis using socio economical

variables, and compared the two classifications. The authors emphasized the importance of Olympic rankings and asked for new studies on this subject.

All the works mentioned hereabove take into account the results in the Olympics and the socio economical conditions of each country.

Balmer et al (2001); Balmer et al (2003), Soares de Mello et al (2004) and Soares de Mello et al (2008) use only the results themselves. The former study the existence of "home advantage" and the latter want to establish a ranking for both Summer and Winter Games. Another work comparing summer and winter Games is the one of Johnson and Ali (2004). The uncertainty in Olympic Games was studied by Baimbridge (1998).

In all the above mentioned papers there is no difference in the value of a medal in different sports, i.e., the one hundred meters gold medal has the same value of a baseball gold medal. In this paper we will propose a method the takes into account that the medals are obtained in different sports.

We shall also mention that there are DEA based rankings for some other sports. Among them we can mention Espitia-Escuer and Garci-Cebrian (2006), Barros and Leach (2006), Haas (2003), Soares de Mello and Gomes Junior (2006) and Calôba and Lins (2006).

3 Building the new ranking

Our method is performed in two steps. The first one is to make a ranking for each sport independently. This is done to avoid the possibility of a country that has a good performance in a sport that has a great number of different competitions (athletics, gymnastics, swimming and so on) to be placed in a higher position than a country with a similar performance in a sport with a few number of different competitions (baseball, football, volleyball, and so on). In the second step, we must aggregate the different rankings obtained in step one. This is achieved using a weighted sum of the partial performances. We mention different forms to determine these weights and we carry out the calculation for two of them.

For the first step, we will use a DEA model for each sport. The DMUs (Decision Making Units that are the units under evaluation in DEA) are all the countries that won a medal in this sport. The three outputs are the number of gold, silver and bronze medal each country earned. We do not use inputs. According to Lovell and Pastor (1999) this leads to mathematical inconsistencies and so we adopt a unitary input for each DMU.

Owing to the existence of a single constant input, we use the Constant Returns to Scale DEA model (DEA CCR) Charnes et al (1978). In the particular case of a constant input the CCR model becomes (1) in which η_0 is the DMU 0 efficiency; y_{j0} is the j-th output ($j=1,\dots,s$) of the k-th DMU ($k=1,\dots,n$); x_{ij} is the i-th input ($i=1,\dots,r$) of the k-th DMU; w_{ij} and v_{jk} are the output and the input weights, respectively. As we have a unitary input it does not appear in the formulation.

$$\begin{aligned}
 & \text{Maximise } \eta_0 = \sum_{j=1}^s \mu_j y_{j0} \\
 & \text{subject to} \\
 & \sum_{j=1}^s \mu_j y_{jk} \leq 1, \quad k = 1, \dots, n \\
 & \mu_j \geq 0, \quad j = 1, \dots, s,
 \end{aligned} \tag{1}$$

Although this is an input oriented DEA model, this model allows other interpretation owing to the absence of the equality constraint. If this model was to be used with an input oriented interpretation, this model will become meaningless owing to the presence of a unitary input. The dual for this model is presented in (2)

$$\begin{aligned}
 & \text{Minimize} \sum_{k=1}^n \lambda_k \\
 & \text{subject to} \\
 & \sum_{k=1}^n y_{jk} \lambda_k \geq y_{j_0}, \quad j=1, \dots, s \\
 & \lambda_k \geq 0, \quad \forall k
 \end{aligned} \tag{2}$$

In this model there is no input reduction. The minimization of the share sum interpretation makes the model a meaningful one even in the presence of a constant input. This model has already been derived by Caporaletti et al (1999). The authors interpreted this model as a multi-attribute one, in the spirit of DEA only with outputs. This is the same as considering a unitary and constant input. Foroughia and Tamiz (2005) use an analogous model but they missed the theoretical considerations. A model with the same objective function and different constraints is used by Kao and Hung (2007).

For the Olympic ranking, model (1) is transformed into model (3), where g , s and b refer to the gold, silver and bronze medals in the Athens 2004 Olympic Games.

$$\begin{aligned}
 & \text{Maximise } h_0 = \mu_g y_{g0} + \mu_s y_{s0} + \mu_b y_{b0} \\
 & \text{subject to} \\
 & \mu_g y_{gk} + \mu_s y_{sk} + \mu_b y_{bk} \leq 1, \quad k = 1, \dots, n \\
 & \mu_j \geq 0, \quad j = g, s, b
 \end{aligned} \tag{3}$$

Obviously the medals are not equally important. To take that fact into account we will use weight restrictions in our DEA model. For sure, a gold medal is more important than a silver one and this one is more important than a bronze one. However, the difference in their relative importance is not the same. In opposition to Baron de Coubertin ideals, victory is the main goal of the competitors. So the difference in importance between gold and silver medals should be greater than the difference between silver and bronze medals. Having these assumptions in mind, the unitary input DEA model based on Soares de Mello and Gomes Junior (2006) model is shown in (4).

$$\begin{aligned}
 & \text{Maximise } h_0 = \mu_g y_{g0} + \mu_s y_{s0} + \mu_b y_{b0} \\
 & \text{subject to} \\
 & \mu_g y_{gk} + \mu_s y_{sk} + \mu_b y_{bk} \leq 1, \quad k = 1, \dots, n \\
 & \mu_g - \mu_s \geq 0.001 \\
 & \mu_s - \mu_b \geq 0.001 \\
 & \mu_g - 2\mu_s + \mu_b \geq 0.001 \\
 & \mu_j \geq 0, \quad j = g, s, b
 \end{aligned} \tag{4}$$

This is a weight restrictions model with non-homogeneous restrictions; such models were studied by Podinovsky (2004)

The non Archimedean constant 0.001 is required to avoid a critical distortion, i.e. in special conditions the three medals may be equally valued. Such a situation leads to non-suitable rankings. For instance, in table 1, medals for Baseball are shown.

Table 1. Medals for Baseball

Country	Gold	Silver	Bronze
Cuba	1	0	0
Australia	0	1	0
Japan	0	0	1

Source: International Olympic Committee

As shown in that table each one of the three countries earned only one medal. If we applied the model with homogenous weights restrictions, the three countries will be equally efficient. This is not a desirable result. Common sense will attribute the first position for Cuba, the second one for Australia and the third place to Japan.

Even with the constraints imposed in this model, there is a high degree of freedom for the weights. To avoid this freedom, we use a secondary model inspired on the Sexton et al (1986) Cross Evaluation model. As we have only one input, the cross evaluation model becomes a fixed weight model Anderson et al (2002). So, we use average weights, which are easier to calculate than using the aggressive and benevolent model of Doyle and Green (1994). Average weights are used in a model similar to cross evaluation used by Lins et al (2003). The use of both weight restrictions and cross evaluation combines two approaches for improving discrimination in DEA: the first one with decision maker value judgements and the second one is a fully objective one (Angulo-Meza and Lins, 2002)

On the other hand, it is common knowledge in DEA that for some DMUs the weights are not uniquely determined (see for instance, Rosen et al, (1998); Soares de Mello et al, (2002); Cooper et al, (2007)). We use an auxiliary linear programming model to determine a unique set of weights for each DMU. The aim of this model is to maximize the difference of the weights for gold and silver medals assuming that the efficiency previously determined in model (4) remains the same. This model is shown in (5).

$$\begin{aligned}
 & \text{Maximise } \mu_g - \mu_s \\
 & \text{subject to} \\
 & \mu_g y_{g0} + \mu_s y_{s0} + \mu_b y_{b0} = h_0, \\
 & \mu_g y_{gk} + \mu_s y_{sk} + \mu_b y_{bk} \leq 1, \\
 & \mu_g - \mu_s \geq 0.001 \\
 & \mu_g - \mu_s \geq 0.001 \\
 & \mu_g - 2\mu_s + \mu_b \geq 0.001 \\
 & \mu_j \geq 0, \quad j = g, s, b
 \end{aligned} \tag{5}$$

The objective function guarantees the maximization of the difference between the importance of the gold and silver medals. The invariance of the efficiency value is assured by the first constraint. The remaining constraints are the same as used in model (4).

In each sport, to obtain the gold medal weight we calculate the average of all the weights attributed to the gold medal by the complete set of DMUs. In a similar way we obtain the weights for silver and bronze medals. The performance of a particular DMU in a given sport is calculated using equation (6).

$$P_0 = \overline{\mu_g} Y_g + \overline{\mu_s} Y_s + \overline{\mu_b} Y_b \tag{6}$$

3.1 Some results for the first step

Using models (4) and (5) and equation (6), we obtain the results for every sport. In table 2 we show results for Sailing.

Table 2. Sailing results

DMU	y_g	y_s	y_b	μ_g	μ_s	μ_b	P_0
United Kingdom	2	1	2	0,49995	0,0001	0	1
Brazil	2			0,49995	0,0001	0	0,710103
Spain	1	2		0,40002	0,19996	0	0,655042
Austria	1	1		0,40002	0,19996	0	0,505047
Greece	1	1		0,40002	0,19996	0	0,505047
USA	1	1		0,40002	0,19996	0	0,505047
France	1		1	0,49995	0,0001	0	0,425003
Israel	1			0,49995	0,0001	0	0,355052
Norway	1			0,49995	0,0001	0	0,355052
Ukraine		2		0,40002	0,19996	0	0,29999
Canada		1		0,40002	0,19996	0	0,149995
China		1		0,40002	0,19996	0	0,149995
Czech Republic		1		0,40002	0,19996	0	0,149995
Denmark			2	0,20016	0,19996	0,19986	0,139902
Argentina			1	0,20016	0,19996	0,19986	0,069951
Italy			1	0,20016	0,19996	0,19986	0,069951
Japan			1	0,20016	0,19996	0,19986	0,069951
Poland			1	0,20016	0,19996	0,19986	0,069951
Slovenia			1	0,20016	0,19996	0,19986	0,069951
Sweden			1	0,20016	0,19996	0,19986	0,069951
<i>Average weights</i>				0,355052	0,149995	0,069951	

Table 3. The results for Archery

DMU	y_g	y_s	y_b	μ_g	μ_s	μ_b	P_0
Korea	3	1		0,3333	0,0001	0	1
Italy	1			0,3333	0,0001	0	0,270863
Chinese Taipei		1	1	0,25005	0,24985	0,24975	0,374725
China		1		0,25005	0,24985	0,24975	0,187413
Japan		1		0,25005	0,24985	0,24975	0,187413
Australia			1	0,25005	0,24985	0,24975	0,187313
United Kingdom			1	0,25005	0,24985	0,24975	0,187313
Ukraine			1	0,25005	0,24985	0,24975	0,187313
<i>Average weights</i>				0,270863	0,187413	0,187313	

As shown in table 2, in the case of sailing, the ranking obtained by our model and the lexicographic method is the same. A case where the two methods lead to different rankings is Archery, whose results are shown in table 3.

Table 4 shows the average weights for all disciplines as well as the differences between average weights.

Table 4. Medals Average Weights for every discipline.

Discipline	$\bar{\mu}_g$	$\bar{\mu}_s$	$\bar{\mu}_b$	$\bar{\mu}_g - \bar{\mu}_s$	$\bar{\mu}_s - \bar{\mu}_b$
Archery	0.270863	0.187413	0.187313	0.083450	0.000100
Athletics	0.078268	0.022021	0.021921	0.056248	0.000100
Badminton	0.222300	0.166600	0.166500	0.055700	0.000100
Baseball	1.000000	0.666567	0.666467	0.333433	0.000100
Basketball	0.900030	0.300010	0.099970	0.600020	0.200040
Beach volleyball	0.750000	0.250000	0.249900	0.500000	0.000100
Boxing	0.128831	0.118649	0.118549	0.010182	0.000100
Canoe/kayak flatwater	0.161822	0.117571	0.117471	0.044251	0.000100
Canoe/kayak slalom racing	0.388917	0.222167	0.222067	0.166750	0.000100
Cycling mountain bike	1.000000	0.666567	0.664667	0.333433	0.001900
Cycling road	0.703722	0.185211	0.111067	0.518511	0.074144
Cycling track	0.132667	0.086989	0.081343	0.045678	0.005646
Diving	0.127034	0.079306	0.079206	0.047728	0.000100
Equestrian	0.437398	0.125204	0.101746	0.312195	0.023458
Fencing	0.227012	0.082541	0.071343	0.144471	0.011198
Football	1.000000	0.666567	0.666467	0.333433	0.000100
Gymnastics Artistic	0.180873	0.049377	0.042793	0.131496	0.006584
Gymnastics Rythmic	0.375038	0.249925	0.249825	0.125113	0.000100
Trampoline	0.900030	0.300010	0.099970	0.600020	0.200040
Handball	1.000000	0.666567	0.666467	0.333433	0.000100
Hockey	0.875038	0.250025	0.124963	0.625013	0.125063
Judo	0.103157	0.087373	0.087273	0.015784	0.000100
Modern Pentathlon	1.000000	0.666567	0.666467	0.333433	0.000100
Rowing	0.282669	0.173878	0.166541	0.108791	0.007338
Sailing	0.355052	0.149995	0.069951	0.205057	0.080044
Shooting	0.181849	0.063653	0.036339	0.118196	0.027314
Softball	1.000000	0.666567	0.666467	0.333433	0.000100
Swimming	0.047813	0.028005	0.024888	0.019808	0.003117
Sync swimming	0.500000	0.333233	0.333133	0.166767	0.000100
Table tennis	0.200100	0.133300	0.133200	0.066800	0.000100
Taekwondo	0.365427	0.173088	0.134565	0.192338	0.038523
Tennis	0.383403	0.233293	0.233193	0.150110	0.000100
Triathlon	0.700020	0.299980	0.299880	0.400040	0.000100
Volleyball	1.000000	0.333387	0.266653	0.666613	0.066733
Water polo	1.000000	0.666567	0.666467	0.333433	0.000100
Weightlifting	0.140048	0.099920	0.099820	0.040128	0.000100
Wrestling	0.133407	0.071191	0.063528	0.062216	0.007663

The difference between the average weights for gold and silver is larger than the difference between silver and bronze, i.e., $\bar{\mu}_g - \bar{\mu}_s > \bar{\mu}_s - \bar{\mu}_b$. Although this is an obvious consequence of this constraint, it can be seen that for almost all the disciplines, $\bar{\mu}_g - \bar{\mu}_s \gg \bar{\mu}_s - \bar{\mu}_b$. This means that for the majority of countries the gold medal is much more important than the other medals.

On other hand, for a large number of disciplines, $\bar{\mu}_s - \bar{\mu}_b$ is very small. For 21 disciplines, this difference is 0.0001. This is the value chosen for the non-Archimedean

constant ε . So we conclude that this constant is very important to distinguish the weight values for silver and bronze medals. As a matter of fact, we may conclude that for a large number of countries winning a silver medal or one of bronze has almost the same value.

3.2 Aggregation of the partial results

As mentioned in the introduction of this paper, the final ranking is obtained using a weighted sum of the performance of each country in each sport as shown in equation (7).

$$I_o = \sum_{i=1}^{37} P_{oi} n_i \quad (7)$$

In which P_{oi} is the performance of the country O in sport i and n_i are the weights for sport i , $i = 1, \dots, 37$.

Different methods can be used to estimate the weights for each sport. In a first model, we can suppose that all sports are equally important and so the weights are all the same.

In a second model, we can measure the importance of each sport for their potential of attracting spectators, mainly when TV broadcastings are concerned. A direct approach would need the figures for TV audiences. This is a rather difficult task, so we may use as a proxy the number of countries participating in each sport. Despite the several drawbacks of this approach, we are justified to believe that the greater the number of countries participating in a sport, the greater the number of potential spectators.

Another method would weigh each sport according to its competitiveness Mitchell and Stewart (2007).

In this paper, we will use both the first and the second model taking into account the number of participating countries rather than TV audiences.

The number of participant countries in each sport is shown in table 5.

After using equation (7) to obtain the indexes, these are normalized using equation (8).

$$\text{Index}_o = \frac{I_o}{I_{max}} * 100 \quad (8)$$

The results for both models are shown in table 6 and 7.

Table 5. Participant countries for each sport

Sport	n_i	Sport	n_i
Archery	43	Handball	16
Athletics	196	Hockey	14
Badminton	32	Judo	94
Baseball	32	Modern Pentathlon	26
Basketball	18	Rowing	55
Beach volleyball	24	Sailing	61
Boxing	72	Shooting	106
Canoe/kayak flat-water	45	Softball	8
Canoe/kayak slalom racing	22	Swimming	154
Cycling mountain bike	34	Sync swimming	24
Cycling road	49	Table tennis	50
Cycling track	39	Tae-kwon-do	60
Diving	30	Tennis	53
Equestrian	68	Triathlon	33
Fencing	42	Volleyball	19
Football	22	Water polo	13
Gymnastics Artistic	42	Weightlifting	79
Gymnastics Rhythmic	21	Wrestling	99
Trampoline	19		

Table 6. Final Ranking using identical weights for all sports

Ranking	Country	Index	Ranking	Country	Index
1	USA	100,00	39	Belgium	4,321921
2	Russia	90,95	40	Uzbekistan	4,272584
3	Germany	71,21	41	North Korea	3,533214
4	Australia	54,34	42	Azerbaijan	3,292518
5	Italia	46,80	43	Israel	3,286145
6	France	42,13	44	Ireland	3,24955
7	Korea	37,37	45	Mexico	3,095507
8	United Kingdom	37,01	46	Georgia	2,984847
9	Japan	34,29	47	Slovenia	2,622701
10	Ukraine	28,83	48	South Africa	2,312637
11	Netherlands	27,68	49	Estonia	2,103016
12	Hungary	27,64	50	Cuba	1,99198
13	Brazil	27,48	51	China	1,982228
14	Greece	18,43	52	Ethiopia	1,97946
15	Argentina	17,48	53	Venezuela	1,741313
16	Canada	16,57	54	Spain	1,733198
17	Romania	15,64	55	Portugal	1,702438
18	Norway	14,82	56	Jamaica	1,652263
19	New Zealand	11,39	57	Kenya	1,561582
20	Czech Republic	10,49	58	U Arab Emirates	1,351006
21	Bulgaria	10,42	59	Morocco	1,326551
22	Belarus	10,41	60	Finland	1,001795
23	Austria	10,29	61	Egypt	0,991114
24	China Taipei	9,50	62	Hong Kong	0,990322

25	Slovakia	9,22	63	Denmark	0,989579
26	Turkey	7,67	64	Syria	0,880729
27	Chile	7,43	65	Zimbabwe	0,748165
28	Thailand	7,28	66	Bahrain	0,744332
29	Switzerland	7,24	67	Colombia	0,74159
30	Iran	6,28	68	Croatia	0,74159
31	Poland	6,28	69	Mongolia	0,648371
32	Latvia	6,22	70	Cameroon	0,581476
33	Lithuania	5,70	71	Dominican Republic	0,581476
34	SCG	5,43	72	India	0,472897
35	Kazakhstan	5,15	73	Nigeria	0,325711
36	Paraguay	4,95	74	Trinidad and Tobago	0,184898
37	Indonesia	4,87	75	Eritrea	0,162856
38	Sweden	4,75113			

Table 7. Final Ranking using different weights for all sports

Ranking	Country	Index	Ranking	Country	Index
1	USA	100,00	39	Egypt	5,77
2	Russia	88,08	40	Jamaica	5,37
3	China	65,79	41	Kenya	5,07
4	Germany	52,19	42	Slovakia	4,85
5	Australia	41,67	43	Azerbaijan	4,68
6	Japan	35,75	44	Lithuania	4,55
7	France	35,12	45	Georgia	4,44
8	United Kingdom	34,67	46	North Korea	4,33
9	Italy	34,12	47	Morocco	4,31
10	South Korea	32,08	48	Switzerland	4,22
11	Cuba	30,53	49	Belgium	4,18
12	Ukraine	25,13	50	South Africa	4,10
13	Netherlands	24,84	51	Latvia	3,89
14	Greece	18,26	52	Israel	3,68
15	Spain	18,08	53	Ireland	3,66
16	Hungary	17,98	54	Slovenia	3,24
17	Brazil	16,03	55	Mexico	3,22
18	Romania	14,93	56	Indonesia	3,16
19	Belarus	13,76	57	Estonia	2,72
20	Norway	12,20	58	Bahamas	2,42
21	Bulgaria	11,39	59	U Arab Emirates	2,37
22	Canada	11,02	60	Portugal	2,18
23	Turkey	10,33	61	Venezuela	1,97
24	Austria	9,18	62	Zimbabwe	1,91
25	Thailand	8,91	63	Serbia/Montenegro	1,90
26	Poland	8,82	64	Cameroon	1,89
27	Sweden	8,81	65	Dominican Republic	1,89
28	Czech Republic	8,76	66	Paraguay	1,81
29	Chinese Taipei	8,66	67	Finland	1,70
30	Iran	7,57	68	Nigeria	1,06
31	Denmark	7,47	69	Syria	1,05
32	Argentina	7,44	70	Mongolia	1,01

33	New Zealand	7,27	71	Colombia	0,97
34	Kazakhstan	7,26	72	India	0,83
35	Chile	6,53	73	Hong Kong	0,82
36	Ethiopia	6,43	74	Eritrea	0,53
37	Uzbekistan	6,22	75	Trinidad and Tobago	0,47
38	Croatia	6,17			

There are very few ties in both rankings, mostly due to numerical precision. In the quasi-official ranking all the countries are equally ranked that won only, for instance, one bronze medal. The largest difference between the quasi-official ranking and the second model of our proposed ranking concerned the Czech Republic that went up 14 rungs. The results of those two models are very similar mainly for the first positions.

A comparison between the quasi-official (lexicographic) ranking and the ranking obtained with model 1 shows some differences. The greatest one is Cuba, which was the eleventh country in the lexicographic ranking and drops to the fiftieth position in model 1 ranking. Another country that has a considerable change of rank is China that drops from the second position to the twenty seventh. On the other hand, Argentina that was ranked thirty fourth goes up to fifteenth.

As the quasi-official ranking and the ranking obtained using model 2 are very similar, we expect to obtain almost the same differences between model 1 and model 2. As a matter of fact, when using model 2 China is ranked third and Cuba eleventh and they drop to the previously mentioned positions when using model 1.

4 Final Comments

Two rankings were proposed: one of them takes into account that all sports are equally important. The other one takes into account the “impact” of each sport measured by the number of countries participating in each sport. This measure takes into account as well that the chances to win a medal are not the same in the different sports.

The two indexes obtained have two common characteristics. The first is that they do not overrate gold medals. The second is that they put on an equal footing sports that give away a large number of medals - athletics or gymnastics, for instance - and team sports (such as basket ball or volley ball) in which a single medal rewards a large number of athletes.

The first positions on the ranking using model 2 and according to the quasi-official ranking are almost the same. The major differences appear in the middle and bottom of the table due to other factors including the absence of ties in our method. The similarities of these two rankings can be explained by the approaches being used. In model 2 we used a weighing scheme that takes into account the number of countries disputing medals. The lexicographic method uses the number of medals to rank the countries. It is a fact that in the Olympic Games, the greater the number of medals, the greater the number of participant countries. So, in a way model 2 has a very similar approach than the lexicographic method without over evaluating the gold medal, which might explain the small differences that were found.

The major differences appear when comparing models 1 and 2, i.e., models considering equal and different weights for each sport. As mentioned earlier, China and Cuba go up several rungs as we move from model 1 to model 2. This happens because both countries concentrate their efforts in sports with a large number of medals. This shows an investment in higher “impact” sport. On the other hand, Argentina drops from fifteenth to thirtieth between model 1 and model 2. This shows an opposite strategy, because Argentina concentrates efforts in sports with few medals and few participants such as team sports (football, basketball, and others).

Further developments should include as well the tradition of a given sport in the Olympic Games.

5 References

- Anderson, T. R., Hollingworth, K., e Inman, L. (2002) The fixed weighting nature of a cross-evaluation model, *Journal of Productivity Analysis*, Vol 17, pp. 249-255.
- Baimbridge, M. (1998) Outcome uncertainty in sporting competition: The olympic games 1896-1996, *Applied Economics Letters*, Vol 5, No 3, pp. 161-164.
- Balmer, N. J., Nevill, A. M., e Williams, A. M. (2001) Home advantage in the winter olympics (1908-1998), *Journal of Sports Sciences*, Vol 19, No 2, pp. 129-139.
- Balmer, N. J., Nevill, A. M., e Williams, A. M. (2003) Modelling home advantage in the summer olympic games, *Journal of Sports Sciences*, Vol 21, No 6, pp. 469-478.
- Barros, C. P., e Leach, S. (2006) Performance evaluation of the english premier football league with data envelopment analysis, *Applied Economics*, Vol 38, No 12, pp. 1449-1458.
- Bernard, A. B., e Busse, M. R. (2004) Who wins the olympic games: Economic resources and medal totals, *Review of Economics and Statistics*, Vol 86, No 1, pp. 413-417.
- Caloba, G. M., e Lins, M. P. E. (2006) Performance assessment of the soccer teams in brazil using dea, *Pesquisa Operacional*, Vol 26, No 3, pp. 521-536.
- Caporaletti, L. E., Dulá, J. H., e Womer, N. K. (1999) Performance evaluation based on multiple attributes with nonparametric frontiers, *Omega*, Vol 27, No, pp. 637-645.
- Churilov, L., e Flitman, A. (2006) Towards fair ranking of olympics achievements: The case of sydney 2000, *Computers and Operations Research*, Vol 33, No 7, pp. 2057-2082.
- Cooper, W. W., Ruiz, J. L., e Sirvent, I. (2007) Choosing weights from alternative optimal solutions of dual multiplier models in dea, *European Journal of Operational Research*, Vol 180, No, pp. 443-458.
- Doyle, J., e Green, R. H. (1994) Efficiency and cross-efficiency in dea derivations, meanings and uses, *Journal of the Operational Research Society*, Vol 45, No, pp. 567-578.
- Fischer, D. 2003. *The encyclopedia of the summer olympics*: Franklin Watts.
- Foroughia, A. A., e Tamiz, M. (2005) An effective total ranking model for a ranked voting system, *Omega*, Vol 33, No, pp. 491 - 496.
- Haas, D. J. (2003) Productive efficiency of english football teams - a data envelopment analysis approach, *Managerial and Decision Economics*, Vol 24, No 5, pp. 403-410.
- Johnson, D. K. N., e Ali, A. (2004) A tale of two seasons: Participation and medal counts at the summer and winter olympic games, *Social Science Quarterly*, Vol 85, No 4, pp. 974-993.
- Kao, C., e Hung, H.-T. (2007) Management performance:An empirical study of the manufacturing companies intaiwan, *Omega*, Vol 35, No, pp. 152-160.
- Lins, M. P. E., Gomes, E. G., Soares de Mello, J. C. C. B., e Soares de Mello, A. J. R. (2003) Olympic ranking based on a zero sum gains dea model, *European Journal of Operational Research*, Vol 148, No 2, pp. 312-322.

- Lozano, S., Villa, G., Guerrero, F., e Cortés, P. (2002) Measuring the performance of nations at the summer olympics using data envelopment analysis, *Journal of the Operational Research Society*, Vol 53, No 5, pp. 501-511.
- Mitchell, H., e Stewart, M. F. (2007) A competitive index for international sport, *Applied Economics*, Vol 39, No 5, pp. 587-603.
- Morton, R. H. (2002) Who won the sydney 2000 olympics?: An allometric approach, *Journal of the Royal Statistical Society Series D: The Statistician*, Vol 51, No 2, pp. 147-155.
- Podinovsky, V. V. (2004) Suitability and redundancy of non-homogeneous weight restrictions for measuring the relative efficiency in dea, *European Journal of Operational Research*, Vol 154, No. 380-395.
- Rosen, Schaffnit, C., e Paradi, J. C. (1998) Marginal rates and two dimensional level curves in dea, *Journal of Productivity Analysis*, Vol 9, No 3, pp. 205-232.
- Sexton, T. R., Silkman, R. H., e Logan, A. J. 1986. Data envelopment analysis: Critique and extensions. In *Measuring efficiency: An assessment of data envelopment analysis*, edited by Silkman, H. San Francisco: Jossey-Bass Editor.
- Soares de Mello, J. C. C. B., Gomes, E. G., Angulo-Meza, L., e Biondi Neto, L. (2008) Cross evaluation using weight restrictions in unitary input dea models: Theoretical aspects and application to olympic games ranking, *WSEAS Transactions on Systems*, Vol 7, No 1, pp. 31-39.
- Soares de Mello, J. C. C. B., Lins, M. P. E., e Gomes, E. G. (2002) Construction of a smoothed dea frontier, *Pesquisa Operacional*, Vol 28, No 2, pp. 183-201.

Optimal Pricing and Ordering Policies For deteriorating items under progressive trade credit scheme

Nita H. Shah ¹
Hardik Soni ²

¹ Department of Mathematics,
Gujarat University, Ahmedabad – 380009, Gujarat.
India.
nita_shah@rediffmail.com

² Chimanbhai Patel Post Graduate Institute of Computer Applications,
Ahmedabad – 380051, Gujarat.
India.
Hardik_soni13@yahoo.com

Abstract

In this paper, a mathematical model is developed to formulate optimal pricing and ordering policies when the units in inventory are subject to constant rate of deteriorating and the supplier offers progressive credit periods to settle the account. The concept of progressive credit periods is as follows:

If the retailer settles the outstanding amount by M, the supplier does not charge any interest. If the retailer pays after M but before N ($M < N$), then the supplier charges the retailer on the un-paid balance at the rate I_{C_1} . If the retailer settles the account after N, then he will have to pay an interest rate of I_{C_2} ($I_{C_2} > I_{C_1}$) on the un-paid balance.

The objective is to maximize the net profit. The decision variables are selling price and ordering quantity. An algorithm is given to find the flow of optimal selling price and ordering policy. A numerical illustration is given to study the effect of offered two credit periods and deterioration on decision variables and the net profit of the retailer.

Resumo

Neste trabalho, um modelo matemático desenvolvido está optimizado para formular políticas de preços e encomendas, quando as unidades do inventário estão sujeitos à taxa constante de deterioração progressiva eo fornecedor oferece crédito períodos de liquidar a conta. O conceito de progressividade de crédito períodos é a seguinte: Se o varejista apurado o montante pendente por M, o fornecedor não cobra qualquer interesse. Se o revendedor paga após M, mas antes de N ($M < N$), em seguida, o fornecedor cobra o varejista sobre as un-pago à taxa equilíbrio I_{C_1} . Se o varejista liquidar a conta depois de N, então ele terá que pagar uma taxa de juro de I_{C_2} ($I_{C_2} > I_{C_1}$) sobre o saldo un-pagos. O objetivo é maximizar o lucro líquido. A decisão são variáveis preço de venda e ordenando quantidade. Um algoritmo é determinado a encontrar o fluxo otimizado de preço de venda e ordenação política. A ilustração é dado numérico para estudar o efeito do crédito oferecido dois períodos ea deterioração variáveis e decisão sobre o lucro líquido da varejista.

Keywords: EOQ, Progressive Credit Periods, deterioration, Selling Price, Ordering Policy.

Title: Optimal Preços e Encomenda Políticas Para deterioração progressiva itens sob regime de comércio de crédito

1 Introduction

The Wilson's lot – size model is derived with the assumption that the retailer pays for the goods as soon as it is received by the system. However, in practice, the supplier offers credit period to the retailer to settle his account within the fixed allowable credit period; which encourages retailer to buy more and also attracts more customers. Davis and Gaither (1985) derived a lot – size model when the supplier offers one time opportunity to delay the payments of order, in case, orders for additional units are placed. Shah et al (1988) extended Goyal's (1985) model by allowing shortages. Mandal and Phaujdar (1989) derived a mathematical model by including interest earned from the sales revenue on the stock remaining beyond the settlement period. Shah and Shah (1992) studied inventory model when supplier offers credit period to settle the retailer's account by considering stochastic demand. Jamal et al. (1997) developed an inventory model to allow for shortages under the permissible delay in payments. Shah (1997) derived a probabilistic order-level system with lead-time when delay in payments is permissible. Jamal et al. (2000) formulated a mathematical model when retailer can settle the payment either at the end of the credit period or later incurring interest charges on the un-paid balance for the over-due period. Hwang and Shinn (1997) developed the model for determining the retailer's lot-size and optimal selling price when the supplier permits delay in payments for an order of a product whose demand rate is a function of constant price elasticity.

Arcelus et al. (2001) compared retailer's response to special sales in two strategies viz. price discount and trade credit. Arcelus et al. (2003) derived mathematical model for retailer's maximum profit when supplier offers credit period and/or price discount on the purchase of regular order when units in inventory are subject to constant deterioration. Related articles are by Hadley and Higgins (1973), Kingsman (1983), Chapman et al. (1985), Daellenbach (1986, 1988), Ward and Chapman (1987), Chapman and Ward (1988), Raafat (1991), Wee (1995), Shinn et al. (1996), Chung (1998), Chu et al. (1998), Shah and Shah (2000), Goyal and Giri (2001), Teng (2002) etc.

In this article, an attempt is made to develop mathematical model when units in inventory are subject to constant rate of deterioration and supplier offers two progressive credit periods to the retailer to settle the account. The net profit is maximized with respect to optimal selling and ordering quantity. The effect of deterioration rate of units in inventory system and credit periods on objective function and decision variables are studied using hypothetical numerical example. An algorithm is given to explore the computational flow.

The paper is organized as follows: In section 2, assumptions and notations are given. Section 3 deals with development of mathematical model. In section 4, flowchart is given to search for optimal solution. Analytical results are stated in section 5. The numerical example and observations are given in section 6. The paper concludes with conclusion and bibliography at the end.

2 Assumptions and Notations

The following assumptions are used to develop aforesaid model:

1. The inventory system deals with the single item.
2. The demand is $R(p) = a - bp$, ($a, b > 0$, $a \gg b$). p denotes selling price of the item during the cycle time and a decision variable.
3. Shortages are not allowed and lead-time is zero.
4. Replenishment is instantaneous.
5. Replenishment rate is finite.
6. If the retailer pays by M , then supplier does not charge to the retailer. If the retailer pays after M and before N ($N > M$), he can keep the difference in the unit sale price and unit cost in an interest bearing account at the rate of

I_e /unit/year. During $[M, N]$, the supplier charges the retailer an interest rate of I_{C_1} /unit/year. If the retailer pays after N , then supplier charges the retailer an interest rate of I_{C_2} /unit/year ($I_{C_2} > I_{C_1}$) on un-paid balance.

7. The units on - hand deteriorate at a constant rate θ ($0 \leq \theta \leq 1$). The deteriorated units can neither be repaired nor replaced during the cycle time.

The notations are as follows:

- h = The inventory holding cost/unit/year excluding interest charges.
- p = The selling price/unit. (a decision variable).
- C = The unit purchase cost, with $C < p$.
- A = The ordering cost/order.
- M = The first offered credit period in settling the account without any extra charges.
- N = The second permissible delay period in settling the account $N > M$.
- I_{C_1} = The interest charged per \$ in stock per year by the supplier when retailer pays during $[M, N]$.
- I_{C_2} = The interest charged per \$ in stock per year by the supplier when retailer pays during $[N, T]$. ($I_{C_2} > I_{C_1}$)
- I_e = The interest earned/\$/year. ($I_{C_1} > I_e$)
- T = The replenishment cycle time (a decision variable).
- IHC = Inventory holding cost/cycle.
- PC = Purchase cost / cycle.
- OC = Ordering cost / cycle.
- IE = Interest earned / cycle.
- IC = Interest charged / cycle.
- $Q(t)$ = The on-hand inventory level at time t ($0 \leq t \leq T$).
- GR = Gross revenue.
- $NP(p, T)$ = Net profit / cycle.
- θ = Deterioration rate

3 Mathematical Formulation

The on-hand inventory depletes due to constant demand $R(p)$ and deterioration of units. The instantaneous state of inventory at any instant of time t is governed by the differential equation

$$\frac{dQ(t)}{dt} + \theta Q(t) = -R(p), \quad 0 \leq t \leq T \quad (3.1)$$

with initial condition $Q(0) = Q$ and boundary condition $Q(T) = 0$. Consequently, the solution of (3.1) is given by

$$Q(t) = \frac{R(p)(e^{\theta(T-t)} - 1)}{\theta}; \quad 0 \leq t \leq T \quad (3.2)$$

and the order quantity is

$$Q = \frac{R(p)(e^{\theta T} - 1)}{\theta} \quad (3.3)$$

The cost components per unit time are as follows:

- Ordering cost;

$$OC = \frac{A}{T} \quad (3.4)$$

- Inventory holding cost;

$$IHC = \frac{h}{T} \int_0^T Q(t)dt = \frac{h(a-bp)(e^{\theta T} - 1 - \theta T)}{\theta^2 T} \quad (3.5)$$

- Cost due to deterioration;

$$DC = \frac{C(a-bp)(e^{\theta T} - 1 - \theta T)}{\theta T} \quad (3.6)$$

- Gross revenue;

$$GR = (p - C)R(p) \quad (3.7)$$

Regarding interest charged and earned, based on the length of the cycle time T , three cases arise:

Case 1: $T \leq M$

Case 2: $M < T < N$

Case 3: $T \geq N$

We discuss each case in detail.

Case 1: $T \leq M$

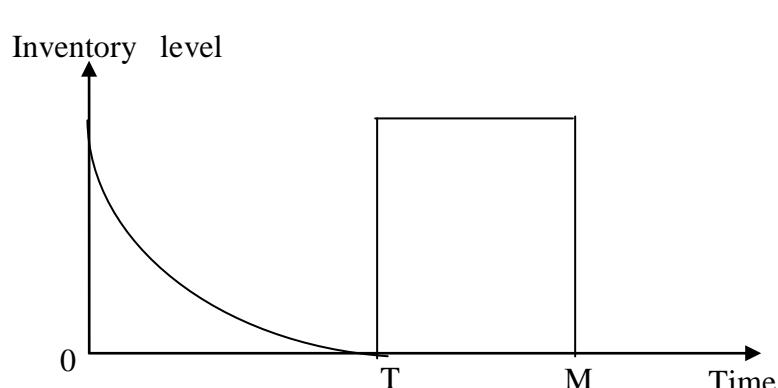


Figure 1: $T \leq M$

Here, the retailer sells Q -units during $[0, T]$ and is paying for CQ - units during $[0, T]$ and is paying for CQ in full to the supplier at time $M \geq T$. So interest charges are zero. i.e.

$$IC_1 = 0 \quad (3.8)$$

The retailer sells products during $[0, T]$ and deposits the revenue in an interest bearing account at the rate of $I_e/\$/year$. In the period, $[T, M]$ the retailer deposits revenue into the account that earns $I_e/\$/year$. Therefore interest earned per year is

$$IE_1 = \frac{pIe}{T} \left[\int_0^T R(p).t dt + R(p)T(M-T) \right] = \frac{pIe(a-bp)(2M-T)}{2} \quad (3.9)$$

the net profit; NP_1 is given by

$$NP_1(p, T) = GR - OC - IHC - DC - IC_1 + IE_1 \quad (3.10)$$

p and T are continuous variables. Hence, the optimal values of p and T can be obtained by setting

$$\frac{\partial NP_1(p, T)}{\partial p} = a - 2pb + bC + \frac{(e^{\theta T} - 1 - \theta T).b}{\theta T} \left(\frac{h}{\theta} + C \right) + \frac{Ie(2M-T)}{2}(a - 2pb) = 0 \quad (3.11)$$

and

$$\frac{\partial NP_1(p, T)}{\partial T} = \frac{A}{T^2} - \left(\frac{h}{\theta} + C \right) (a - bp) \left\{ \frac{(e^{\theta T} - 1)}{T} - \frac{(e^{\theta T} - 1 - \theta T)}{\theta T^2} \right\} - \frac{pIe(a-bp)}{2} = 0 \quad (3.12)$$

The obtained $T = T_1$ and $p = p_1$, maximizes the net profit; provided

$$XY - Z^2 < 0 \quad (3.13)$$

where $X = \frac{\partial^2 NP_1(p,T)}{\partial p^2} = -2b - Ie b(2M - T)$;

$$Y = \frac{\partial^2 NP_1(p,T)}{\partial T^2} = \frac{-2A}{T^3} \left(\frac{h}{\theta} + C \right) (a - bp) \left\{ \frac{\theta e^{\theta T}}{T} - \frac{2(e^{\theta T} - 1)}{T^2} + \frac{2(e^{\theta T} - 1 - \theta T)}{\theta T^3} \right\}$$

$$Z = \frac{\partial^2 NP_1(p,T)}{\partial T \partial p} = \frac{(e^{\theta T} - 1)b(h + C)}{T} - \frac{b(e^{\theta T} - 1 - \theta T)}{\theta T^2} \left(\frac{h}{\theta} + C \right) - \frac{Ie(a - 2bp)}{2}$$

Case 2: $M < T < N$

Inventory level

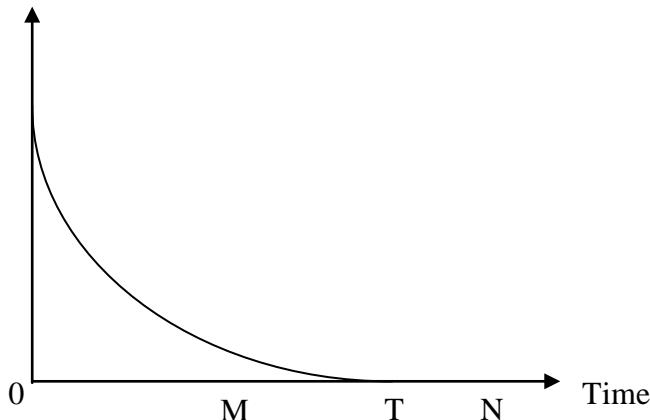


Figure 2: $M < T < N$

The retailer's sells units and deposits the revenue into an interest bearing account at an interest rate Ie /unit/year during $[0, M]$. Therefore, interest earned during $[0, M]$ is given by

$$IE_2 = pIe \int_0^M R(p)t dt = \frac{1}{2}pIe(a-bp)M^2 \quad (3.14)$$

Buyer has to pay for $Q = R(p)T$ units purchased in the beginning of the cycle at the rate of C \$/unit to the supplier during $[0, M]$. The retailer sells $R(p)M$ -units at sale price p \$/unit. So he has generated revenue of pM $R(p)$ plus the interest earned, $IE_{2.1}$, during $[0, M]$. Two sub-cases may arise:

Sub-case 2.1: Let $pR(p)M + IE_2 \geq CQ$, i.e. the retailer has enough money to pay for all Q -units procured. Then, interest charges;

$$IC_{2.1} = 0 \quad (3.15)$$

and interest earned; $IE_{2.1}$ per time unit is

$$IE_{2.1} = \frac{IE_2}{T} \quad (3.16)$$

Using equations (3.4) to (3.7), (3.15) and (3.16), the net profit; $NP_{2.1}(p, T)$ is given by

$$NP_{2.1}(p, T) = GR - OC - IHC - DC - IC_{2.1} + IE_{2.1} \quad (3.17)$$

The optimum values of $p = p_{2.1}$ and $T = T_{2.1}$ are solutions of

$$\frac{\partial NP_{2.1}(p,T)}{\partial p} = a - 2pb + bC + \frac{(e^{\theta T} - 1 - \theta T)b}{\theta T} \left(\frac{h}{\theta} + C \right) + \frac{Ie M^2(a - 2pb)}{2T} = 0 \quad (3.18)$$

and

$$\frac{\partial NP_{2.1}(p,T)}{\partial T} = \frac{A}{T^2} - \left(\frac{h}{\theta} + C \right) (a - bp) \left\{ \frac{(e^{\theta T} - 1)}{T} - \frac{(e^{\theta T} - 1 - \theta T)}{\theta T^2} \right\} - \frac{pIe(a - bp)M^2}{2T^2} = 0 \quad (3.19)$$

The obtained $p = p_{2.1}$ and $T = T_{2.1}$ maximizes the net profit provided

$$XY - Z^2 < 0 \quad (3.20)$$

Where, $X = \frac{\partial^2 NP_{2.1}(p,T)}{\partial p^2} = -2b - \frac{Ie b M^2}{T}$;

$$Y = \frac{\partial^2 NP_{2.1}(p,T)}{\partial T^2} = \frac{-2A}{T^3} - \left(\frac{h}{\theta} + C \right) (a - bp) \left\{ \frac{\theta e^{\theta T}}{T} - \frac{2(e^{\theta T} - 1)}{T^2} + \frac{2(e^{\theta T} - 1 - \theta T)}{\theta T^3} \right\} + \frac{pIe(a - bp)M^2}{2T^3}$$

$$Z = \frac{\partial^2 NP_{2.1}(p,T)}{\partial T \partial p} = \frac{(e^{\theta T} - 1)b(h + C)}{T} - \frac{b(e^{\theta T} - 1 - \theta T)}{\theta T^2} \left(\frac{h}{\theta} + C \right) - \frac{Ie(a - 2bp)M^2}{2T^2}$$

Sub case 2.2: Let $p R(p) M + IE_2 < CQ$

Here, retailer will have to pay interest on un-paid balance $U_1 = C R(p) - [p R(p)M + IE_2]$ at rate of Ic_1 at time M to supplier. The interest to be paid; $IC_{2.2}$ per time unit is:

$$IC_{2.2} = \frac{U_1^2}{pR(p)T} Ic_1 \int_M^T Q(t) dt = \frac{Ic_1 U_1^2 (e^{\theta(T-M)} + M\theta - 1 - \theta T)}{2\theta^2 p T} \quad (3.21)$$

and interest earned;

$$IE_{2.2} = \frac{IE_2}{T} \quad (3.22)$$

Using equations (3.4) to (3.7), (3.21) and (3.22), the net profit; $NP_{2.2}(p, T)$ is given by

$$NP_{2.2}(p, T) = GR - OC - IHC - DC - IC_{2.2} + IE_{2.2} \quad (3.23)$$

The optimum values of $p = p_{2.2}$ and $T = T_{2.2}$ are solutions of

$$\begin{aligned} \frac{\partial NP_{2.2}(p,T)}{\partial p} &= a - 2pb + bC + \frac{(e^{\theta T} - 1 - \theta T)b}{\theta T} \left(\frac{h}{\theta} + C \right) \\ &\quad - \frac{Ic_1 U_1 (e^{\theta(T-M)} + M\theta - 1 - \theta T) (-CbT - (a - 2bp)M - \frac{Ie M^2(a - 2bp)}{2})}{\theta^2 p T} \\ &\quad + \frac{Ic_1 U_1^2 (e^{\theta(T-M)} + M\theta - 1 - \theta T)}{2\theta^2 p^2 T} + \frac{Ie M^2(a - 2pb)}{2T} = 0 \end{aligned} \quad (3.24)$$

and

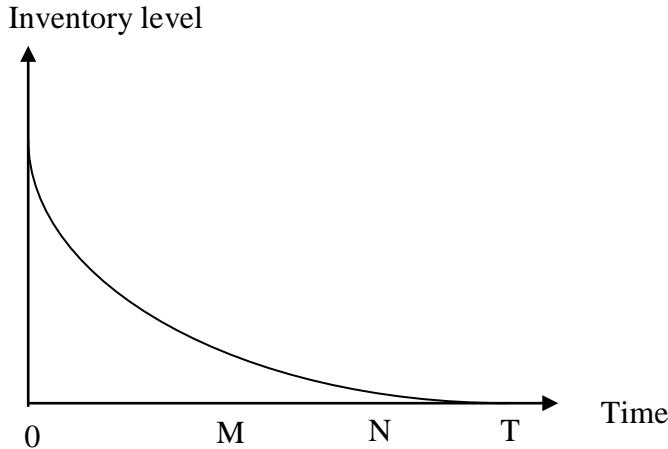
$$\begin{aligned} \frac{\partial NP_{2.2}(p,T)}{\partial T} &= \frac{A}{T^2} - \left(\frac{h}{\theta} + C \right) (a - bp) \left\{ \frac{(e^{\theta T} - 1)}{T} - \frac{(e^{\theta T} - 1 - \theta T)}{\theta T^2} \right\} - \frac{Ic_1 U_1 (e^{\theta(T-M)} + M\theta - 1 - \theta T) C(a - bp)}{\theta^2 p T} \\ &\quad - \frac{Ic_1 U_1^2 (e^{\theta(T-M)} - 1)}{2\theta p T} + \frac{Ic_1 U_1^2 (e^{\theta(T-M)} + M\theta - 1 - \theta T)}{\theta^2 p T^2} - \frac{pIe(a - bp)M^2}{2T^2} = 0 \end{aligned} \quad (3.25)$$

The obtained $p = p_{2.2}$ and $T = T_{2.2}$ maximizes the net profit provided

$$EF - G^2 < 0 \quad (3.26)$$

Where

$$\begin{aligned}
E &= \frac{\partial^2 NP_{2,2}(p,T)}{\partial p^2} = -2b - \frac{Ic_1(e^{\theta(T-M)} + M\theta - 1 - \theta T)(-CbT - (a-2bp)M - \frac{Ie M^2(a-2bp)}{2})^2}{\theta^2 p T} \\
&\quad + \frac{2Ic_1 U_1(e^{\theta(T-M)} + M\theta - 1 - \theta T)(-CbT - (a-2bp)M - \frac{Ie M^2(a-2bp)}{2})}{\theta^2 p^2 T} \\
&\quad - \frac{Ic_1 U_1(e^{\theta(T-M)} + M\theta - 1 - \theta T)(2bM + Ie b M^2)}{\theta^2 p T} - \frac{Ic_1 U_1^2(e^{\theta(T-M)} + M\theta - 1 - \theta T)}{\theta^2 p^3 T} - \frac{Ie b M^2}{T}; \\
F &= \frac{\partial^2 NP_{2,2}(p,T)}{\partial T^2} = \frac{-2A}{T^3} - \left(\frac{h}{\theta} + C \right) (a - bp) \left\{ \frac{\theta e^{\theta T}}{T} - \frac{2(e^{\theta T} - 1)}{T^2} + \frac{2(e^{\theta T} - 1 - \theta T)}{\theta T^3} \right\} \\
&\quad - \frac{Ic_1 C^2(a-bp)^2(e^{\theta(T-M)} + M\theta - 1 - \theta T)}{\theta^2 p T} - \frac{2Ic_1 U_1(e^{\theta(T-M)} - 1)C(a-bp)}{\theta p T} \\
&\quad + \frac{2Ic_1 U_1(e^{\theta(T-M)} + M\theta - 1 - \theta T)C(a-bp)}{\theta^2 p T^2} - \frac{Ic_1 U_1^2 e^{\theta(T-M)}}{2 p T} + \frac{Ic_1 U_1^2(e^{\theta(T-M)} - 1)}{\theta p T^2} \\
&\quad - \frac{Ic_1 U_1^2(e^{\theta(T-M)} + M\theta - 1 - \theta T)}{\theta^2 p T^3} + \frac{p Ie(a-bp)M^2}{2T^3} \\
G &= \frac{\partial^2 NP_{2,2}(p,T)}{\partial T \partial p} = \frac{(e^{\theta T} - 1)b(h+C)}{T} - \frac{b(e^{\theta T} - 1 - \theta T)}{\theta T^2} \left(\frac{h}{\theta} + C \right) \\
&\quad - \frac{Ic_1(e^{\theta(T-M)} + M\theta - 1 - \theta T)C(a-bp)(-CbT - (a-2bp)M - \frac{Ie M^2(a-2bp)}{2})}{\theta^2 p T} \\
&\quad + \frac{Ic_1 U_1(e^{\theta(T-M)} + M\theta - 1 - \theta T)C(a-bp)}{\theta^2 p^2 T} + \frac{Ic_1 U_1(e^{\theta(T-M)} + M\theta - 1 - \theta T)Cb}{\theta^2 p T} \\
&\quad - \frac{Ic_1 U_1(e^{\theta(T-M)} - 1)(-CbT - (a-2bp)M - \frac{Ie M^2(a-2bp)}{2})}{\theta p T} + \frac{Ic_1 U_1^2(e^{\theta(T-M)} - 1)}{2\theta p^2 T} \\
&\quad + \frac{Ic_1 U_1(e^{\theta(T-M)} + M\theta - 1 - \theta T)C(a-bp)(-CbT - (a-2bp)M - \frac{Ie M^2(a-2bp)}{2})}{\theta^2 p T^2} \\
&\quad - \frac{Ic_1 U_1^2(e^{\theta(T-M)} + M\theta - 1 - \theta T)}{2\theta^2 p^2 T^2} - \frac{Ie(a-2bp)M^2}{2T^2}
\end{aligned}$$

Case 3: $T \geq N$
Figure 3: $T \geq N$

Based on the total purchase cost, CQ , the total money in account at M is $p R(p) M + IE_2$ and total money in account at N is $p R(p) N + p Ie R(p) N^2/2$, three sub-cases may arise:

Sub-case 3.1: Let $p R(p) M + IE_2 \geq CQ$

This sub-case is same as sub-case 2.1. (Note: Decision variables and objective function are designated by 3.1)

Sub-case 3.2: Let $p R(p) M + IE_2 < CQ$ and

$$p R(p) (N - M) + \frac{p Ie R(p)(N-M)^2}{2} \geq CQ - (p R(p) M + IE_2)$$

This sub-case coincides with sub-case 2.2. (Note: Decision variables and objective function are designated by 3.2)

Sub-case 3.3: Let $p R(p)N + \frac{p R(p)Ie N^2}{2} < CQ$ and

$$p R(p) (N - M) + \frac{p Ie R(p)(N-M)^2}{2} < CQ - (p R(p) M + IE_2)$$

Here, retailer does not have money in his account to pay off total purchase cost at time N . He will do payment of $p R(p) M + IE_2$ at M and $p R(p) (N-M) + \frac{p Ie R(p)(N-M)^2}{2}$ at N .

So, he has to pay interest charges on the un-paid balance $U_1 = CQ - (p R(p) M + IE_2)$ with interest rate Ic_1 during $[M, N]$ and un-paid balance, $U_2 = U_1 - \left(p R(p) (N - M) + \frac{p Ie R(p)(N-M)^2}{2} \right)$ with interest rate Ic_2 during $[N, T]$. Therefore total interest charges; $IC_{3.3}$; per time unit is given by

$$\begin{aligned} IC_{3.3} &= \frac{U_1 Ic_1 (N - M)}{T} + \frac{U_2^2}{p R(p) T} Ic_2 \int_N^T Q(t) dt \\ &= \frac{U_1 Ic_1 (N - M)}{T} + \frac{Ic_2 U_2^2 (e^{\theta(T-N)} + N\theta - 1 - \theta T)}{\theta^2 p T} \end{aligned} \quad (3.27)$$

and interest earned;

$$IE_{3.3} = \frac{IE_2}{T} \quad (3.28)$$

Using equations (3.4) to (3.7), (3.27) and (3.28), the net profit; $NP_{3.3}(p, T)$ is given by

$$NP_{3.3}(p, T) = GR - OC - IHC - DC - IC_{3.3} + IE_{3.3} \quad (3.29)$$

The optimum values of $p = p_{3.3}$ and $T = T_{3.3}$ are solutions of

$$\begin{aligned} \frac{\partial NP_{3.3}(p, T)}{\partial p} &= a - 2pb + bC + \frac{(e^{\theta T} - 1 - \theta T)b}{\theta T} \left(\frac{h}{\theta} + C \right) - \frac{Ic_1(N-M)(-CbT - (a-2bp)M - \frac{Ie M^2(a-2bp)}{2})}{T} \\ &\quad - \frac{2Ic_2U_2(e^{\theta(T-N)} + N\theta - 1 - \theta T)}{\theta^2 p T} + \frac{Ic_2U_2^2(e^{\theta(T-N)} + N\theta - 1 - \theta T)}{\theta^2 p^2 T} + \frac{Ie M^2(a-2pb)}{2T} = 0 \quad (3.30) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial NP_{3.3}(p, T)}{\partial T} &= \frac{A}{T^2} - \left(\frac{h}{\theta} + C \right) (a - bp) \left\{ \frac{(e^{\theta T} - 1)}{T} - \frac{(e^{\theta T} - 1 - \theta T)}{\theta T^2} \right\} - \frac{Ic_1(a-bp)C(N-M)}{T} + \frac{U_1Ic_1(N-M)}{T^2} \\ &\quad - \frac{2Ic_2U_2(e^{\theta(T-N)} + N\theta - 1 - \theta T)C(a-bp)}{\theta^2 p T} - \frac{Ic_2U_2^2(e^{\theta(T-N)} - 1)}{\theta p T} + \frac{Ic_2U_2^2(e^{\theta(T-N)} + N\theta - 1 - \theta T)}{\theta^2 p T^2} \\ &\quad - \frac{pIe(a-bp)M^2}{2T^2} = 0 \quad (3.31) \end{aligned}$$

The obtained $p = p_{3.3}$ and $T = T_{3.3}$ maximizes the net profit provided

$$BK - J^2 < 0 \quad (3.32)$$

where

$$\begin{aligned} B &= \frac{\partial^2 NP_{3.3}(p, T)}{\partial p^2} = -2b - \frac{(2bM + IebM^2)Ic_1(N-M)}{T} - \frac{2Ic_2(e^{\theta(T-N)} + N\theta - 1 - \theta T) \% 1^2}{\theta^2 p T} \\ &\quad + \frac{4Ic_2U_2(e^{\theta(T-N)} + N\theta - 1 - \theta T)\% 1}{\theta^2 p^2 T} - \frac{2Ic_2U_2(e^{\theta(T-N)} + N\theta - 1 - \theta T)(2bN + Ie b M^2 + Ie b (N-M)^2)}{\theta^2 p T} \\ &\quad - \frac{2Ic_2U_2^2(e^{\theta(T-N)} + N\theta - 1 - \theta T)}{\theta^2 p^3 T} - \frac{Ie b M^2}{T}; \end{aligned}$$

$$\begin{aligned} K &= \frac{\partial^2 NP_{3.3}(p, T)}{\partial T^2} = \frac{-2A}{T^3} - \left(\frac{h}{\theta} + C \right) (a - bp) \left\{ \frac{\theta e^{\theta T}}{T} - \frac{2(e^{\theta T} - 1)}{T^2} + \frac{2(e^{\theta T} - 1 - \theta T)}{\theta T^3} \right\} \\ &\quad + \frac{2C(a-bp)Ic_1(N-M)}{T^2} - \frac{2U_1Ic_1(N-M)}{T^3} \\ &\quad - \frac{2Ic_2C^2(a-bp)^2(e^{\theta(T-N)} + N\theta - 1 - \theta T)}{\theta^2 p T} - \frac{4Ic_2U_2(e^{\theta(T-N)} - 1)C(a-bp)}{\theta p T} \\ &\quad + \frac{4Ic_2U_2(e^{\theta(T-N)} + N\theta - 1 - \theta T)C(a-bp)}{\theta^2 p T^2} - \frac{Ic_2U_2^2 e^{\theta(T-N)}}{p T} \\ &\quad + \frac{2Ic_2U_2^2(e^{\theta(T-N)} - 1)}{\theta p T^2} - \frac{Ic_2U_2^2(e^{\theta(T-N)} + N\theta - 1 - \theta T)}{\theta^2 p T^3} + \frac{pIe(a-bp)M^2}{2T^3} \end{aligned}$$

$$\begin{aligned}
J = \frac{\partial^2 NP_{3,3}(p,T)}{\partial T \partial p} &= \frac{(e^{\theta T} - 1)b(h+C)}{T} - \frac{b(e^{\theta T} - 1 - \theta T)}{\theta T^2} \left(\frac{h}{\theta} + C \right) + \frac{CbIc_1(N-M)}{T} \\
&+ \frac{(-CbT - (a-2bp)M - \frac{Ie M^2 (a-2bp)}{2})Ic_1(N-M)}{T^2} \\
&- \frac{2Ic_2 \% 2C(a-bp)\% 1 + 2Ic_2 U_2 \% 2C(a-bp)}{\theta^2 p T} \\
&+ \frac{2Ic_2 U_2 \% 2Cb}{\theta^2 p T} - \frac{2Ic_2 U_2 (e^{\theta(T-N)} - 1)\% 1}{\theta p T} + \frac{Ic_2 U_2^2 (e^{\theta(T-N)} - 1)}{\theta p^2 T} \\
&+ \frac{Ic_2 U_2 \% 2\% 1}{\theta^2 p T^2} - \frac{Ic_2 U_2^2 \% 2}{\theta^2 p^2 T^2} - \frac{Ie(a-2bp)M^2}{2T^2}
\end{aligned}$$

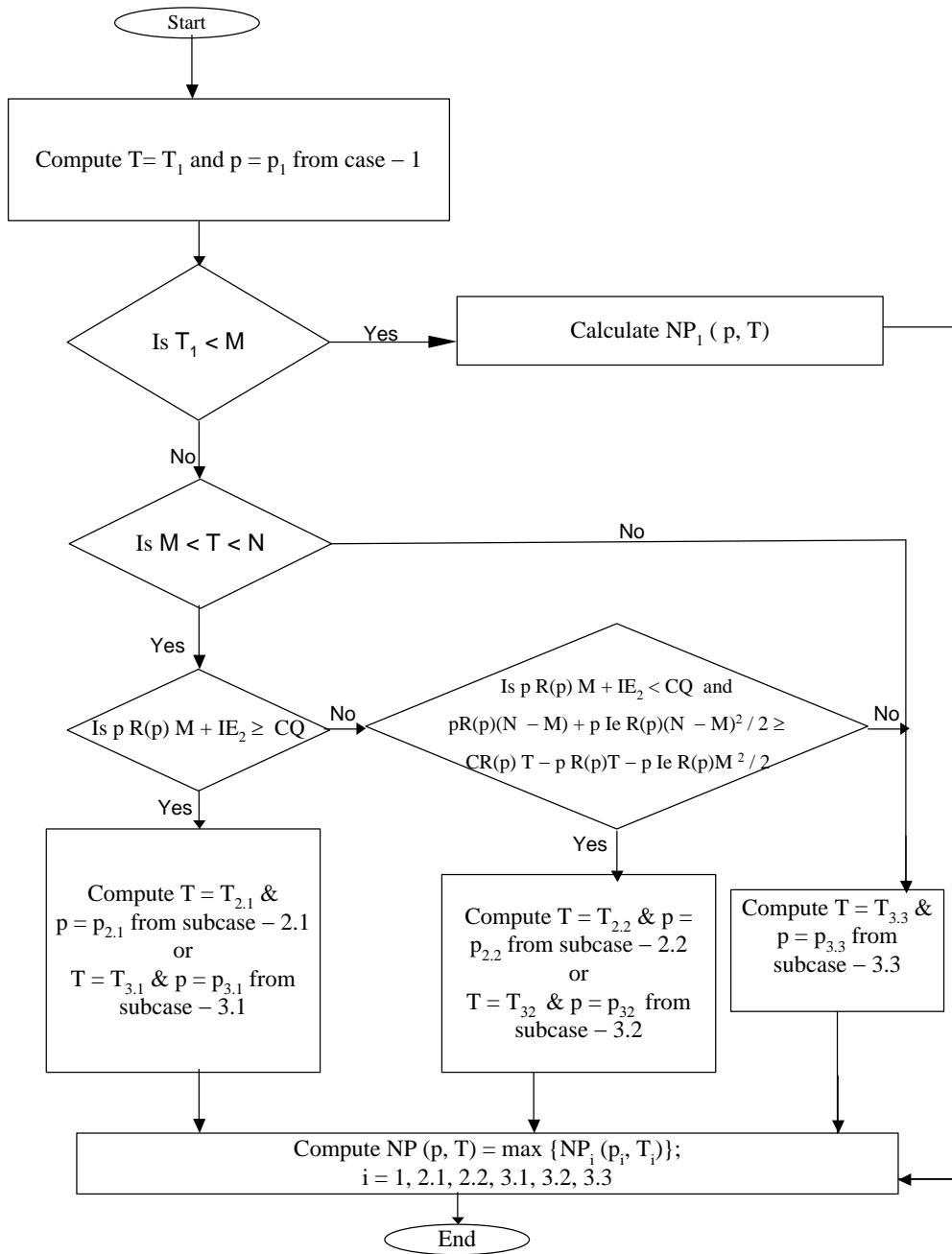
where

$$\% 1 = \{-CbT - (a-2bp)M - \frac{Ie M^2 (a-2bp)}{2} - (a-2bp)(N-M) - \frac{Ie (N-M)^2 (a-2bp)}{2}\}$$

$$\% 2 = (e^{\theta(T-N)} + N\theta - 1 - \theta T)$$

In the next section, computational flowchart is given to search for optimal solution.

4 Flowchart



5 Theoretical Results

Proposition 5.1: $NP_i(p_i, T_i)$ is maximum for $i = 1, 2.1, 2.2, 3.1, 3.2, 3.3$.

Proof: It follows from equations (3.13), (3.20), (3.26), (3.32).

Proposition 5.2: For $T > N$, $NP_{3,3}(p, T)$ is increasing function of M and N .

Proof:

$$\begin{aligned} \frac{\partial NP_{3,3}(p,T)}{\partial M} &= \frac{p(a-bp)(1+M)Ic_1(N-M)}{T} + \frac{U_1Ic_1}{T} \\ &\quad - \frac{2Ic_2U_2\%2}{\theta^2 p T} [pIe(a-bp)(N-2M)] + \frac{pIe(a-bp)M}{T} \\ &> 0 \\ \frac{\partial NP_{3,3}(p,T)}{\partial N} &= -\frac{U_1Ic_1}{T} + \frac{2Ic_2U_2\%2}{\theta^2 p T} [p(a-bp)(1+Ie(N-M))] + \frac{Ic_2U_2^2(e^{\theta(T-N)}-1)}{\theta p T} \\ &> 0 \\ \%2 &= (e^{\theta(T-N)} + N\theta - 1 - \theta T) \end{aligned}$$

Proposition 5.3: $NP_i(p_i, T_i)$ is a decreasing function of θ .

Proof:

$$\begin{aligned} \frac{\partial NP_1(p,T)}{\partial \theta} &= -\frac{(a-bp)(e^{\theta T}-1)}{\theta} \left(\frac{h}{\theta} + C \right) + \frac{(a-bp)(e^{\theta T}-1-\theta T)}{\theta^2 T} \left(\frac{2h}{\theta} + C \right) \\ &< 0 \\ \frac{\partial NP_{2,1}(p,T)}{\partial \theta} &= -\frac{(a-bp)(e^{\theta T}-1)}{\theta} \left(\frac{h}{\theta} + C \right) + \frac{(a-bp)(e^{\theta T}-1-\theta T)}{\theta^2 T} \left(\frac{2h}{\theta} + C \right) \\ &< 0 \\ \frac{\partial NP_{2,2}(p,T)}{\partial \theta} &= -\frac{(a-bp)(e^{\theta T}-1)}{\theta} \left(\frac{h}{\theta} + C \right) + \frac{(a-bp)(e^{\theta T}-1-\theta T)}{\theta^2 T} \left(\frac{2h}{\theta} + C \right) \\ &\quad - \frac{Ic_1 U_1^2 ((T-M)e^{\theta(T-M)} + M - T)}{2\theta^2 p T} + \frac{Ic_1 U_1^2 (e^{\theta(T-M)} + M\theta - 1 - T)}{\theta^3 p T} \\ &< 0 \\ \frac{\partial NP_{3,3}(p,T)}{\partial \theta} &= -\frac{(a-bp)(e^{\theta T}-1)}{\theta} \left(\frac{h}{\theta} + C \right) + \frac{(a-bp)(e^{\theta T}-1-\theta T)}{\theta^2 T} \left(\frac{2h}{\theta} + C \right) \\ &\quad - \frac{Ic_2 U_2^2 ((T-N)e^{\theta(T-N)} + N - T)}{\theta^2 p T} + \frac{2Ic_2 U_2^2 (e^{\theta(T-N)} + N\theta - 1 - T)}{\theta^3 p T} \\ &< 0 \end{aligned}$$

6 Numerical Illustration

Using data $[A, C, h, a, b] = [200, 20, 0.2, 1000, 10]$ in appropriate units, we study effect of progressive trade credits and deterioration in following tables:

Table 1: Optimal decision variables when $\theta = 0.01$

		15/365	20/365	25/365
		Ic ₁ = 15%	Ic ₁ = 16%	Ic ₁ = 17%
30/365	Ic ₂ = 18%	T = 0.3058 P = 60.59 Q = 120.70 R = 394.09 NP = 15250.85	T = 0.3055 P = 60.57 Q = 120.65 R = 394.34 NP = 15269.16	T = 0.3054 P = 60.54 Q = 120.69 R = 394.56 NP = 15287.32

35/365	$Ic_2 = 20\%$	$T = 0.3332$ $P = 60.56$ $Q = 131.65$ $R = 394.44$ $NP = 15306.35$	$T = 0.3327$ $P = 60.53$ $Q = 131.54$ $R = 394.69$ $NP = 15327.64$	$T = 0.3325$ $P = 60.51$ $Q = 131.52$ $R = 394.93$ $NP = 15346.31$
40/365	$Ic_2 = 22\%$	$T = 0.3637$ $P = 60.53$ $Q = 143.80$ $R = 394.71$ $NP = 15353.73$	$T = 0.3630$ $P = 60.50$ $Q = 143.65$ $R = 394.96$ $NP = 15374.53$	$T = 0.3626$ $P = 60.48$ $Q = 143.59$ $R = 395.22$ $NP = 15394.32$

Table 2: Optimal decision variables when $\theta = 0.02$

		$M \rightarrow$	15/365	20/365	25/365
		$Ic_1 = 15\%$	$Ic_1 = 16\%$	$Ic_1 = 17\%$	
30/365	$Ic_2 = 18\%$	$T = 0.3052$ $P = 60.60$ $Q = 120.62$ $R = 394.04$ $NP = 15238.73$	$T = 0.3049$ $P = 60.57$ $Q = 120.57$ $R = 394.28$ $NP = 15257.04$	$T = 0.3048$ $P = 60.55$ $Q = 120.55$ $R = 394.50$ $NP = 15275.19$	
35/365	$Ic_2 = 20\%$	$T = 0.3326$ $P = 60.56$ $Q = 131.59$ $R = 394.38$ $NP = 15293.16$	$T = 0.3321$ $P = 60.54$ $Q = 131.48$ $R = 394.63$ $NP = 15313.44$	$T = 0.3318$ $P = 60.51$ $Q = 131.46$ $R = 394.87$ $NP = 15333.11$	
40/365	$Ic_2 = 22\%$	$T = 0.3630$ $P = 60.54$ $Q = 143.78$ $R = 394.64$ $NP = 15338.30$	$T = 0.3624$ $P = 60.51$ $Q = 143.63$ $R = 394.90$ $NP = 15359.17$	$T = 0.3620$ $P = 60.48$ $Q = 143.57$ $R = 395.16$ $NP = 15380.91$	

It is observed that increase in deterioration of units in inventory decreases the net profit and cycle time and increases selling price. The results exhibited in tables coincide with analytical propositions.

7 Conclusion

In this article, an attempt is made to develop an EOQ model in which demand is assumed to be decreasing function of selling price (a decision variable) and units in inventory deteriorate at a constant rate when supplier offers two progressive credit periods, if retailer could not settle his account. An easy-to-use computational algorithm is given to search for optimal policy. The observed managerial issues are as follows:

1. Increase in first allowable credit period decreasing the order quantity and increases net profit whereas selling price is insensitive.
2. Increase in extended permissible credit period lowers cycle time and selling price. Net profit decreases significantly.
3. Increase in deterioration rate reduces cycle time and net profit whereas selling price leisurely goes up.

The proposed model can be extended by taking demand as a function of time, product quality, stock etc. It can also be generalized to allow for shortages, partial lost-sales.

8 References

- Arcelus, F. J., Shah, Nita H. and Srinivasan, G. (2001) Retailer's response to special sales: Price discount vs. trade credit, OMEGA, Vol 29, No 5, pp. 417 – 428.
- Arcelus, F. J., Shah, Nita H. and Srinivasan, G. (2003) Retailer's pricing, credit and inventory policies for deteriorating items in response to temporary price / credit incentives, International Journal of Production Economics, Vol 81-82, pp. 153 – 162.
- Chapman, C. B. and Ward, S. C. (1988) Inventory control and trade credit – A Further Reply, Journal of Operational Research Society, Vol 39, pp. 219 – 220.
- Chu, P., Chung, K. J., and Lan, S. P. (1998) Economic order quantity of deteriorating items under permissible delay in payments, Computers and Operations Research, Vol 25, pp. 817 – 824.
- Chung, K. J.; (1998): A theorem on the determination of economic order quantity under conditions of permissible delay in payments, Computers and Operations Research, Vol 25, pp. 49 – 52.
- Daellenbach, H. G.; (1986): Inventory control and trade credit, Journal of Operational Research Society, Vol 37, pp. 525 – 528.
- Daellenbach, H. G.; (1988): Inventory control and trade credit – a rejoinder, Journal of Operational Research Society, Vol 39, pp. 218 – 219.
- Davis, R. A. and Gaither, N.; (1985): Optimal ordering policies under conditions of extended payment privileges, Management Science, Vol 31, pp. 499 – 509.
- Goyal, S. K. and Giri, B. C.; (2001): Recent trends in modeling of deteriorating inventory. European Journal of Operational Research, Vol 134, pp. 1 – 16.
- Goyal, S. K.; (1985): Economic order quantity under conditions of permissible delay in payments, Journal of the Operational Research Society, Vol 36, pp. 335 – 338.
- Haley, C. W. and Higgins, R. C.; (1973): Inventory policy and trade credit financing, Management Science, Vol 20, pp. 464 – 471.
- Hwang, H. and Shinn, S. W.; (1997): Retailer's pricing and lot sizing policy for exponentially deteriorating products under the condition of permissible delay in payments, Computers and Operations Research, Vol 24, pp. 539 – 547.
- Jamal, A. M., Sarker, B. R. and Wang, S.; (1997): An ordering policy for deteriorating items with allowable shortages and permissible delay in payment, Journal of the Operational Research Society, Vol 48, pp. 826 – 833.
- Jamal, A. M. M., Sarker, B. R. and Wang, S.; (2000): Optimal payment time for a retailer under permitted delay of payment by the wholesaler, International Journal of Production Economics, Vol 66, No 1, pp. 59 – 66.
- Kingsman, B. G.; (1983): The effect of payment rules on ordering and stockholding in purchasing, Journal of the Operational Research Society, Vol 34, pp. 1085 – 1098.
- Mandal, B. N. and Phaujdar, S.; (1989-a): Some EOQ models under permissible delay in payments, International Journal of Management and Systems, Vol 5, No 2, pp. 99 – 108.
- Raafat F.; (1991): Survey of literature on continuously deteriorating inventory model, Journal of the Operational Research Society, Vol 42, pp. 27 – 37.

Shah, Nita H. and Shah, Y. K.; (1992): A probabilistic order level system when delay in payments is permissible, presented at Operational Research Society of India, Convention - 25, India.

Shah, Nita H. and Shah, Y.K.; (2000): Literature survey on inventory model for deteriorating items. *Economic Annals (Yugoslavia)*, XLIV, pp. 221 – 237.

Shah, Nita H.; (1997): Probabilistic order level system with lead time when delay in payments is permissible, *TOP (Spain)*, Vol 5, No 2, pp. 297 – 305.

Shah, V. R., Patel, H. C. and Shah, Y. K.; (1988): Economic ordering quantity when delay in payments of order and shortages are permitted, *Gujarat Statistical Review*, Vol 15, No 2, pp. 51 – 56.

Shinn, S. W., Hwang, H. P. and Sung, S., (1996): Joint price and lot-size determination under conditions of permissible delay in payments and quantity discounts for freight cost, *Euporean Journal of Operation Research*, Vol 91, pp. 528-542.

Teng, J. T.; (2002): On economic order quantity under conditions of permissible delay in payments, *Journal of Operational Research Society*, Vol 53, pp. 915 – 918.

Ward, S. C. and Chapman, C. B.; (1987): Inventory control and trade credit – a reply to Daellenbach, *Journal of Operational Research Society*, Vol 32, pp.1081 – 1084.

Wee, H. M.; (1995): Joint pricing and replenishment policy for deteriorating inventory with declining market, *International Journal of Production Economics*, Vol 40, pp. 163 – 171.

REVISTA INVESTIGAÇÃO OPERACIONAL

Política Editorial

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

Processo de Aceitação

Todos os manuscritos submetidos para publicação são revistos e aceites apenas com base na avaliação da sua qualidade, importância e adequação à política editorial. Será responsabilidade do Editor interpretar a avaliação dos revisores. A contribuição de cada artigo deve estar claramente evidenciada na Introdução. Critérios como a relação com literatura existente, comprimento e estilo do artigo são tidos em consideração. Uma indicação clara da viabilidade de aceitação do artigo é habitualmente dada na primeira fase de revisão do artigo.

Será requerido aos autores de um artigo aceite que transfiram os direitos de autoria para a APDIO, que assegurará a mais ampla disseminação possível de informação. Os volumes da Revista são publicados em papel, e distribuídos a todos os associados da APDIO, e em formato electrónico na rede SciELO - Scientific Electronic Library Online.

Resumos dos Artigos indexados em

IAOR - International Abstracts in Operations Research

Instruções aos Autores

1. *Submeter artigos* para publicação ao editor principal, de preferência por e-mail em Microsoft Word ou “Portable Document Format” (PDF) para jfo@fe.up.pt, ou por correio normal (quatro cópias) para o seguinte endereço: Prof. José Fernando Oliveira, Departamento de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 PORTO, Portugal.
2. *Língua*. Os artigos devem ser escritos em Português, Inglês ou Espanhol.
3. *Os Manuscritos* devem ser impressos. Numerar as páginas consecutivamente.
4. A *primeira página* do manuscrito escrito em português ou em espanhol deve ter a seguinte informação: (a) Título; (b) nome, e-mail e afiliação institucional dos autores; (c) um resumo; (d) palavras-chave; (e) título em inglês (f) um resumo em inglês; (g) palavras-chave em inglês; (h) identificação do autor correspondente. Se o manuscrito for escrito em inglês, a primeira página deve ter a seguinte informação: (a) Título em inglês; (b) nome, e-mail e afiliação institucional dos autores; (c) um resumo em inglês; (d) palavras-chave em inglês; (e) identificação do autor correspondente.

5. *Agradecimentos*, incluindo informação sobre apoios, dever ser colocados imediatamente antes da secção de referências.
6. *Notas de rodapé* devem ser evitadas.
7. *Formulas* que são referenciadas devem ser numeradas consecutivamente ao longo do manuscrito como (1), (2), etc. do lado direito.
8. *Figuras*, incluindo grafos e diagramas, devem ser numerados consecutivamente em numeração árabe.
9. *Tabelas* devem ser numeradas consecutivamente em numeração árabe.
10. *Referências*. Citar apenas as mais relevantes e listar só as que são citadas no texto. Indicar as citações no texto através de parênteses rectos, e.g., [4]. No final do artigo listar as referências alfabeticamente por apelido do primeiro autor e numerá-las consecutivamente, de acordo com o seguinte formato: Artigos: autore(s), título, nome e volume da revista (ou livro, mas neste caso incluir o nome dos editores), ano e páginas. Livros: Autor(es), título, editor, ano.
11. *Artigos aceites* devem ser enviados pelo autor ao editor, de preferência na forma de um ficheiro fonte em LaTeX com ficheiros EPS para as figuras, juntamente com um ficheiro PDF ou Postscript. Em alternativa, ficheiros fonte em Word são também aceites. Para garantir uma boa qualidade gráfica, as figuras devem ser em formato vectorial; formatos raster como JPG, BMP, GIF, etc. devem ser evitados.
12. *Provas dos artigos* serão enviadas por e-mail como ficheiros PDF para o autor correspondente. Corrigir as provas cuidadosamente, e restringir as correções apenas aos pontos em que as provas diferem do manuscrito. Desvios à versão aceite pelo editor são apenas possíveis com a autorização prévia e explícita do editor. Trinta separatas de cada artigo são enviados gratuitamente ao autor correspondente.

Informação sobre a Publicação

Investigação Operacional (ISSN 0874-5161) está registada na Secretaria de Estado da Comunicação Social sob o número 108335. Os volumes da Revista são publicados em papel, e distribuídos a todos os associados da APDIO, e em formato electrónico na rede SciELO - Scientific Electronic Library Online. O preço da assinatura anual é de 25 euros. Os volumes são enviados por correio normal. Informação adicional sobre assinaturas pode ser solicitada ao Secretariado da APDIO- CESUR, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 LISBOA, Portugal. Tel. +351 218 407 455 - www.apdio.pt - apdio@cesur.civil.ist.utl.pt

JOURNAL INVESTIGAÇÃO OPERACIONAL

Editorial Policy

Investigação Operacional (IO) is the scientific journal of APDIO - Associação Portuguesa de Investigação Operacional (the Portuguese Operational Research Association). The editorial policy of IO is to publish high quality and original articles that contribute to theory, methodology, techniques and software of Operational Research (OR) and its application to different fields. It also publishes articles with relevant reviews of OR subjects. Cases of successful application of OR to practical problems are specially welcome.

Acceptance Process

All manuscripts submitted for publication are refereed and accepted only on the basis of its quality, importance and adequacy to the editorial policy. It will be the responsibility of the Editor to interpret the referee's assessment. The contribution of each paper should be clearly stated in the introduction. Criteria such as relationship with existing literature, length and style are taken into account. A clear indication on the suitability of a manuscript is usually provided after the first round of refereeing. The authors of an accepted paper will be asked to transfer its copyright to the publisher, which will ensure the widest possible dissemination of information. The volumes of the journal are published in hardcopies, which are distributed to all APDIO associates, and in electronic format in SciELO - Scientific Electronic Library Online.

Articles are abstracted/indexed in

IAOR - International Abstracts in Operations Research

Instructions to Authors

1. *Submit papers* for publication to the main editor, preferably by e-mail in Microsoft Word or "Portable Document Format"(PDF) to jfo@fe.up.pt, or by ordinary mail (four copies) to the following address: Prof. José Fernando Oliveira, Departamento de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 PORTO, Portugal.
2. *Language.* Papers must be in written in Portuguese, English or Spanish.
3. *Manuscripts* should be typewritten or typeset. Number the pages consecutively.
4. *The first page* of the manuscript written in English should contain the following information: (a) Title; (b) names, e-mails and institutional affiliations of the authors; (c) an abstract; (d) keywords (f) identification of the corresponding author.
5. *Acknowledgements*, including support information, should be placed prior to the references section.
6. *Footnotes* should be avoided.

7. *Formulas* that are referred to should be numbered consecutively throughout the manuscript as (1), (2), etc. on the right.
8. *Figures*, including graphs and diagrams, should be numbered consecutively in Arabic numbers.
9. *Tables* should be numbered consecutively in Arabic numbers.
10. *References*. Cite only the most relevant references and list only those cited in the text. Indicate citations in the text by bracketed numbers, e.g., [4]. At the end of the paper list the references alphabetically by the surname of the first author and number them consecutively, according to the following formats: Articles: author(s), title, name and number of the journal (or book, but in this case include the editors names), year, pages. Books: Author(s), title, publisher, year.
11. *Accepted papers* are to be sent by the author to the editor, preferably in the form of a source file in LaTeX and EPS files for the figures together with a PDF or postscript file. Alternatively, source files in Word are also accepted. To ensure good publishing quality the figures should be in vector formats; raster formats like JPG, BMP, GIF, etc. should be avoided.
12. *Page proofs* will be e-mailed as a PDF file to the corresponding author. Correct proofs carefully, and restrict corrections to points at which the proof is at variance with the manuscript. Deviations from the version accepted by the editor are only possible with the prior and explicit approval of the editor. Thirty offprints of each paper are supplied free of charge to the corresponding author.

Publication information

Investigação Operacional (ISSN 0874-5161) is registered in the *Secretaria de Estado da Comunicação Social* under number 108335. The volumes of the journal are published in hardcopies, which are distributed free of charge to all APDIO associates, and in electronic format in SciELO - Scientific Electronic Library Online. Subscription price is 25 euros. Issues are sent by standard mail. Additional subscription information is available upon request from APDIO Secretariat - CESUR, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 LISBOA, Portugal. Tel. +351 218 407 455 - www.apdio.pt - apdio@cesur.civil.ist.utl.pt

ÍNDICE

R. Fragoso, M. Carvalho e P. Henriques <i>A Programação Matemática Positiva como Instrumento de Calibração e Prescrição dos Modelos de Oferta Agrícola</i>	1
A. Correia, J. Matias e C. Serôdio <i>Métodos de penalidade exacta para resolução de problemas de optimização não linear</i>	17
V. Silva, A. Coelho e S. Mayerle <i>Proposta de um Modelo Heurístico Utilizando Algoritmos Genéticos para a Resolução de um Problema Operacional Portuário</i>	31
G. Neves, M. Diallo e L. Lustosa <i>Electronic Spare Parts Stock Forecasting.....</i>	45
C. González e S. Benítez <i>Validación Del Intervenciones ‘Soft OR’</i>	59
J. Mello, L. Angulo-Meza e B. Silva <i>Some rankings for the Athens Olympic Games using DEA models with a constant input</i>	77
H. Soni e N. Shah <i>Optimal Pricing and Ordering Policies For deteriorating items under progressive trade credit scheme</i>	91