IO 2011
15º Congresso da Associação Portuguesa
de Investigação Operacional

Livro de Actas

18 a 20 de Abril de 2011
Faculdade de Economia da
Universidade de Coimbra

Este volume contém os artigos submetidos e apresentados no 15º Congresso da Associação Portuguesa de Investigação Operacional, realizado em Coimbra, Portugal, de 18 a 20 de Abril de 2011.

Título: **Livro de Actas do 15º Congresso da Associação Portuguesa de Investigação Operacional, IO2011**

Coordenadores:
Domingos Cardoso e João Clímaco

Edição:
INESC-Coimbra

Produção e Design:
Luís Alçada, Abril de 2011

ISBN: **978-989-95055-4-4**
Depósito Legal: **325755/11**

Impressão: Ediliber-Editora de Publicações e Artes Gráficas, Lda. – Coimbra

Comissão de Organização

Presidente:

João Clímaco (Presidente), Universidade de Coimbra (UC)

Carlos Gomes da Silva, Instituto Politécnico de Leiria (IPL)

Carlos Henggeler Antunes, Universidade de Coimbra (UC)

Joana Dias, Universidade de Coimbra (UC)

João Paulo Costa, Universidade de Coimbra (UC)

José Craveirinha, Universidade de Coimbra (UC)

Lúcia Martins, Universidade de Coimbra (UC)

Luís Alçada-Almeida, Universidade de Coimbra (UC)

Luís Dias, Universidade de Coimbra (UC)

Maria João Alves, Universidade de Coimbra (UC)

Marta Pascoal, Universidade de Coimbra (UC)

Paulo Melo, Universidade de Coimbra (UC)

Rita Girão Silva, Universidade de Coimbra (UC)

Rui Lourenço, Universidade de Coimbra (UC)

Teresa Martinez Gomes, Universidade de Coimbra (UC)

Comissão de Programa

Presidente:

Domingos Moreira Cardoso (Presidente), Universidade de Aveiro (UA)

José Valério de Carvalho, Universidade do Minho (UM)

Jorge Orestes Cerdeira, Instituto Superior de Agronomia (ISA, UTL)

João Clímaco, Universidade de Coimbra (UC)

Ruy Costa, Universidade Nova de Lisboa (UNL)

João Paulo Costa, Universidade de Coimbra (UC)

João Coutinho Rodrigues, Universidade de Coimbra (UC)

Edite M. G. P. Fernandes, Universidade do Minho (UM)

Carlos Ferreira, Universidade de Aveiro (UA)

Carlos Luz, Instituto Politécnico de Setúbal (IPS)

João Matias, Universidade de Trás-os-Montes e Alto Douro (UTAD)

José Fernando Oliveira, Universidade do Porto (UP)

Rui Oliveira, Instituto Superior Técnico (IST, UTL)

Margarida Vaz Pato, Instituto Superior de Economia e Gestão (ISEG, UTL)

Marília Pires, Universidade do Algarve (UAlg)

José Pinto Paixão, Universidade de Lisboa (UL)

António José Rodrigues, Universidade de Lisboa (UL)

Jorge Freire de Sousa, Universidade do Porto (UP)

Luís Nunes Vicente, Universidade de Coimbra (UC)

Índice

Comissão de Organização..... iii

Comissão de Programa..... v

Índice..... vii

Nota Editorial..... ix

Ana Paula Teixeira, Maria Manuel Nascimento, Helena Monteiro
Programação Linear: uma experiência na Escola de Verão de Matemática..... 1

André Lopes Fradinho, Maria Isabel Gomes, Ana Paula Ferreira Dias Barbosa Póvoa
Optimization of Batteries Sustainable Distribution Network with Reverse Flows..... 13

Andreia Filipa Antunes, Margarida Moz
Optimização do Escalonamento de Enfermeiros numa Unidade Hospitalar..... 25

Andreia Zanella, Ana S. Camanho, Teresa G. Dias
Countries environmental performance assessment..... 37

Carlos Alberto Rodrigues, Maria Helena Correia
Problema de definição de rotas na prestação de serviços de exploração de sistemas de saneamento..... 53

Clara Vaz, Ana S. Camanho
Evolução do desempenho de lojas de retalho ao longo do tempo utilizando o índice de Malmquist..... 65

Cláudia Duarte, J. Valério de Carvalho, Ana Paula Ferreira Dias Barbosa Póvoa
Car sequencing approaches: a literature survey..... 79

Elsa Silva, Filipe Pereira e Alvelos, José Manuel Valério de Carvalho
Problema integrado de corte bidimensional e dimensionamento de lotes..... 91

Frederico Gonçalves, Sónia R. Cardoso, Susana Relvas, Ana Paula Ferreira Dias Barbosa Póvoa
Optimization of a Distribution Network using Electric Vehicles: a VRP Problem..... 103

Isabel M. Horta, Ana S. Camanho, Jorge Moreira da Costa
Assessment of performance and innovation of Portuguese construction companies..... 115

Leão José Fernandes, Susana Relvas, Ana Paula Ferreira Dias Barbosa Póvoa
Optimal Network Design of Petroleum Supply Chains..... 129

Manuel Rodrigues, Mónica Oliveira, Teresa Peña
Developing Methods to Evaluate Benefits and Costs of Arterial Spin Labeling in Alzheimer's Disease..... 141

Maria Cláudia Sávio Masiero, Edméa Cássia Baptista, Aguinaldo Aparecido Vieira, Vanusa Alves de Sousa
Uma variação da função Lagrangiana Aumentada Barreira Modificada para resolução de problemas de Programação Não Linear..... 153

Marisa Oliveira, Eduarda Pinto Ferreira, A. Miguel Gomes
Rect-TOPOS: Uma heurística construtiva para a minimização da área da envolvente rectangular no Posicionamento de formas com componentes rectangulares..... 165

Mónica D. Oliveira, Teresa C. Rodrigues, Carlos A. Bana e Costa, Armando Brito de Sá
Prioritizing Health Care Interventions: A Multicriteria Resource Allocation Model to Inform the Choice of Community Care Programmes *177*

Níssia Carvalho Rosa Bergiante, Mariana Vieira Rangel Nunes, João Carlos Correia Baptista Soares de Mello
Índice Técnico para Pentatlo Moderno baseado em DEA *187*

Nuno Miguel Santana, Maria Isabel Gomes, Ana Paula Ferreira Dias Barbosa Póvoa
Supply Chain Optimization: Application to a real case *199*

Reinaldo Castro Souza, André Luís Marques Marcato, Bruno Henriques Dias, Fernando Luíz Cyrino Oliveira
Operação Ótima de Sistemas Hidrotérmicos com Geração de Cenários Hidrológicos Através de Modelos Auto-Regressivos Periódicos identificados via Bootstrap *213*

Sandra Silva, Luís Alçada-Almeida, Luís Dias
Avaliação da Gestão de Efluentes em Explorações Leiteiras: uma abordagem com integração de Análise Multicritério num Sistema de Informação Geográfica *225*

Sónia R. Cardoso, Ana Paula Ferreira Dias Barbosa Póvoa, Susana Relvas
Optimal Network Design of Supply Chains with Reverse Flows *237*

Tânia Pinto-Varela, Ana Paula Ferreira Dias Barbosa Póvoa, Augusto Queiroz Novais
A Goal Programming Approach for the Retrofit of Supply Chain Networks *249*

Tânia Rodrigues Pereira Ramos, Maria Isabel Gomes, Ana Paula Ferreira Dias Barbosa Póvoa
Solving a multi-product, multi-depot vehicle routing problem by a hybrid method *261*

Tiago Costa Gomes, Filipe Pereira e Alvelos, Maria Sameiro Carvalho
Modelação e optimização de cadeias de abastecimento de biomassa florestal *275*

Vítor Mendonça, António J. L. Rodrigues
Adaptive forecasting methods for time series with complex trend patterns *287*

Índice de Autores *299*

Nota Editorial

A Comissão de Programa e a Comissão Organizadora do Congresso IO 2011, dando cumprimento a uma decisão da Direcção da APDIO, prepararam o volume de Actas que agora apresentamos, constituído por trabalhos submetidos a um processo de revisão de "arbitragem pelos pares". Os artigos incluídos neste volume correspondem a vinte e quatro, das cento e vinte e seis comunicações apresentadas no congresso, podendo de forma muito breve dizer-se que se distribuem por várias áreas de aplicação da IO, onde se incluem as que estão associadas a problemas de logística, saúde, desporto, energia, avaliação de impacto ambiental, avaliação de desempenho, ensino da investigação operacional, e corte e empacotamento. Neste conjunto de artigos as abordagens seguidas incluem a programação linear, programação inteira, programação não linear, optimização combinatória, DEA, análise multicritério, goal programming, metaheurísticas, métodos de previsão, sistemas de informação geográfica, etc.. A diversidade, quer das aplicações, quer dos métodos utilizados, é fruto do dinamismo da comunidade portuguesa de investigação operacional, aqui reforçada pelas contribuições dos nossos colegas e amigos brasileiros.

Este congresso tem por lema "A Investigação Operacional na Economia e Gestão da maior zona económica exclusiva da Europa" e, embora os artigos apresentados não se tenham focado neste tema, o que aliás era previsível dada a escassez de estudos no âmbito do tópico, do programa científico consta a realização de uma mesa redonda sobre a Economia do Mar no desenvolvimento de Portugal e os desafios que se criam para a comunidade de Investigação Operacional. Esta mesa redonda conta com a participação de Carlos Borrego (Universidade de Aveiro), João Carlos Marques (F.C.T.–Universidade de Coimbra), João Ferreira do Amaral (I.S.E.G.–Universidade Técnica de Lisboa) e Peter Roebling (Universidade de Aveiro) e será moderada por João Clímaco (F.E. – Universidade de Coimbra).

Estamos convictos de que este volume poderá vir a contribuir para uma maior divulgação das potencialidades da IO, enquanto área científica na procura de soluções adequadas em problemas de decisão, num período em que os recursos são particularmente escassos.

Finalmente, não queremos deixar de expressar o nosso reconhecimento a todos os autores, pelos trabalhos apresentados e também a todos os membros da Comissão de Programa e outros colegas que colaboraram no processo de revisão, com leituras muito cuidadas e um conjunto de correcções e sugestões que contribuíram de forma decisiva para a elevada qualidade do produto final.

Domingos Cardoso e João Clímaco

Programação Linear: uma experiência na Escola de Verão de Matemática

Teixeira A.[a*], *Nascimento M. M.*[b], *Monteiro H.*[c]

[a]*Universidade de Trás-os-Montes e Alto Douro / CIO, Portugal*
ateixeir@utad.pt

[b]*Universidade de Trás-os-Montes e Alto Douro / CM-UTAl, Portugal*
mmsn@utad.pt

[c]*Escola Secundária de S. Pedro, Vila Real, Portugal*
hagarez@gmail.com

[*]*Corresponding Author*

Resumo

Neste trabalho descreve-se um estudo em que se propôs aos participantes de um mini-curso, PMC, de uma Escola de Verão de Matemática (alunos do 10º ao 12º ano do Ensino Secundário), a formulação de alguns problemas de Programação Linear, PL, e a respectiva resolução utilizando um *software* livre. Este estudo teve por objectivo, não só sensibilizar os participantes para a possibilidade de utilizar a PL na resolução de problemas do quotidiano, mas também averiguar se os alunos conseguem utilizar um *software* na resolução de problemas de PL. Para além disso, pretende-se conhecer a opinião dos PMC relativamente à utilização do *software* e identificar as principais dificuldades por eles sentidas no decorrer do processo.

Palavras chave: Programação Linear, formulação de problemas, *software*

1- Introdução

Segundo [Hillier e Lieberman, 2005],

A Programação Linear, PL, usa um modelo matemático para descrever os problemas a resolver – problemas de atribuição de recursos limitados entre diferentes actividades da melhor forma possível, isto é óptima. O adjectivo linear significa que todas as funções deste modelo são lineares e a palavra programação não se refere a programação de computadores, mas é sinónimo de Planeamento.

Os problemas de PL, PPL, modelam situações reais de diversas áreas e envolvem frequentemente dezenas a milhões de variáveis e/ou restrições, tornando-se indispensável o recurso a *software* de PL para a sua resolução.

A Programação Linear foi introduzida como conteúdo programático obrigatório da disciplina de Matemática do 11º ano do Ensino Secundário no ano lectivo 2004/05. Embora quase todos os manuais escolares, por exemplo [Costa et al, 2004], [Jorge et al, 2004], [Neves et al, 2005] e [Soveral e Silva, 2004], façam referência à necessidade da utilização de *software* de PL para resolver problemas do quotidiano (em virtude da sua elevada dimensão), nenhum explora ou incentiva a utilização deste recurso. Para além disso, os métodos de resolução de PPL, contemplados nos conteúdos programáticos do Ensino Secundário [Ministério da Educação, 2004], apenas podem ser aplicados na resolução de problemas com duas ou três variáveis, ficando, assim, os alunos com uma perspectiva bastante limitada da aplicabilidade da PL.

Estudos em que alunos do 10º e do 12º anos do Ensino Secundário resolveram alguns PPL com recurso a *software* com resultados bastante positivos são descritos em [Monteiro e Teixeira, 2009a], [Nascimento e Nascimento, 2003] e [Nascimento, 2004]. Fica, assim, a ideia de que estes alunos estão aptos a utilizar *software* de PL, o que lhes permite explorar a resolução de problemas mais interessantes e mais diversificados, tomando deste modo consciência da abrangência do tema. Outros estudos envolvendo a resolução de PPL no Ensino Secundário são descritos em [Filipe, 1998] e [Monteiro e Teixeira, 2009b]. não só sensibilizar os participantes para a possibilidade de utilizar a PL na resolução de problemas do quotidiano

O objectivo deste mini-curso foi, não só sensibilizar os participantes para a utilidade da PL, mas também perceber se os participantes conseguiam formular PPL, utilizar o *software* livre *WinQSB* [WinQSB, 2009] na sua resolução e identificar as principais dificuldades por eles sentidas.

Com este trabalho pretende-se dar a conhecer a experiência, levada a cabo no mini-curso, realizado no âmbito da Escola de Verão de Matemática 2009 da Universidade de Trás-os-Montes e Alto Douro, UTAD, em que foi proposta aos participantes a formulação de alguns PPL e a sua resolução utilizando o *software* proposto.

2- Descrição do estudo

Esta experiência decorreu de 21 a 23 de Julho de 2009 na UTAD, ao longo das três sessões (de 1 hora cada) que constituíram o mini-curso intitulado “Compras e produções

automatizadas ou não? Nem se coloca a questão!!!". Este MC foi dinamizado no âmbito da Escola de Verão de Matemática 2009 da UTAD e nele participaram catorze alunos do Ensino Secundário das zonas norte e centro de Portugal, três dos quais do 12º ano com a disciplina de Matemática Aplicada às Ciências Sociais concluída há um ano, oito do 10º ano e penas três do 11º ou 12º ano. Ou seja, a maior parte dos alunos nunca tinha estudado este tema.

Os trabalhos realizaram-se na sala de computadores do Departamento de Matemática da UTAD, que continha catorze computadores, onde foi previamente instalado o *software WinQSB*. A escolha deste *software* deveu-se ao facto de ser livre e, na opinião das responsáveis pelo MC, simples de utilizar. As tarefas foram executadas individualmente, tendo sido entregue a cada participante:

- uma ficha de trabalho com os enunciados de três PPL e fazendo referência às etapas a percorrer para formular um PPL,
- um manual de utilização do *software WinQSB*, que explica sucintamente o seu funcionamento usando como exemplo o primeiro PPL proposto na ficha de trabalho mencionada no ponto anterior,
- um questionário de opinião.

Tanto a ficha de trabalho como o questionário de opinião encontram-se em anexo, na Secção 6. A ficha de trabalho foi elaborada com base no texto de [Nascimento, 2004], que explicita não só os conteúdos e conceitos de exploração de cada tarefa, mas também os objectivos e competências de cada uma delas, em particular os referentes à formulação e resolução de PPL. Assim, os conceitos e conteúdos envolvidos nesta tarefa são: *variáveis de decisão*, *função objectivo* e *restrições*. Os objectivos e competências a alcançar são: interpretar o enunciado de um problema do dia a dia, tão popular como o das "Dietas", e estruturar a sua formulação matemática; apreender a ideia da PL e da sua aplicação a problemas reais; usar a linguagem própria da PL. O segundo problema foi escolhido por ser adaptado de um manual escolar. O último problema foi seleccionado devido à sua complexidade e às potencialidades para justificar a utilização de um *software.*

Uma das docentes (que será designada ao longo do texto por DDA) dirigiu as actividades em todas as sessões enquanto as outras duas auxiliaram os PMC na resolução das tarefas propostas, tendo estas decorrido do seguinte modo:

1ª Sessão: A DDA começou por transmitir aos PMC o objectivo do MC e introduziu as noções de PL e de PPL. Com o objectivo explicar as etapas necessárias à formulação de um PPL, prosseguiu com a formulação, passo a passo, de um problema académico simples.

De seguida, entregou a cada participante a ficha com os enunciados dos PPL a formular/resolver ao longo das três sessões do MC. Foi-lhes pedido que se debruçassem sobre o primeiro exercício desta ficha e que tentassem formular um PPL que traduzisse o enunciado desse problema. Após algum tempo, em que os PMC tentaram formular um PPL e foram colocando questões às docentes responsáveis pelo MC, a DDA, juntamente com os alunos, construiu uma formulação, fazendo referência específica a cada etapa. O segundo problema foi formulado pelos PMC de forma mais autónoma, embora sob a supervisão de todas as docentes, tendo a formulação do terceiro sido proposta como trabalho de casa, a discutir na sessão seguinte.

2ª Sessão: Começou-se por discutir a formulação do terceiro PPL, proposto na sessão anterior e que seis alunos tinham apresentado. Os alunos conseguiram sem dificuldade formular a função objectivo, bem como a produção mínima anual requerida por espécie. Contudo, alguns trocaram a desigualdade nesta restrição. Além disso, a maior parte dos alunos que tentaram formular sozinhos esqueceram-se das restrições da área disponível por região, o que foi prontamente corrigido depois de realçado pelas professoras. Destaca-se o caso de um aluno que não conseguiu definir as variáveis de decisão.

De seguida passou-se à fase relativa à resolução dos PPL previamente formulados. Após a distribuição do manual de utilização do *WinQSB*, a DDA explicou o funcionamento do *software* e utilizou-o para resolver, passo a passo, o primeiro problema proposto – procedimento que os PMC acompanharam através da projecção do ecrã do computador numa tela e reproduziram nos seus computadores. Após ter sido obtida a solução óptima e o respectivo valor óptimo, procedeu-se a uma breve interpretação dos resultados obtidos no contexto do problema.

3ª Sessão: Os PMC utilizaram o *WinQSB* para resolver os outros dois PPL propostos na ficha de trabalho, recorrendo ao auxílio do manual fornecido e às docentes quando necessário. Para além da interpretação dos resultados obtidos no contexto de cada problema, foram exploradas diversas potencialidades deste *software*, nomeadamente inserir/eliminar variáveis/restrições e a resolução gráfica dos problemas envolvendo duas variáveis.

No final dos trabalhos, foi entregue o questionário de opinião, que cada aluno preencheu e que se encontra em anexo, na Subsecção 6.2.

É ainda de referir que os PMC foram muito colaborantes e se mostraram muito interessados tanto em aprender a formular PPL como a resolvê-los utilizando este *software*. Ficou patente o interesse e a rapidez com que a generalidade dos alunos apreendeu os procedimentos a efectuar e a facilidade com que os utilizaram na resolução dos restantes problemas.

3- Apresentação e análise dos resultados

A análise das opiniões dos alunos relativamente ao decurso dos trabalhos, baseia-se nas respostas às questões formuladas no questionário de opinião.

Os Gráficos 1 e 2 dizem respeito às respostas dadas pelos PMC às diferentes questões colocadas no Ponto 1 do questionário de opinião. O Gráfico 1 permite verificar que 9 alunos consideram que por vezes a formulação de PPL é difícil. Tal deve-se, como se pode observar no Gráfico 2, às dificuldades que sentem em interpretar o enunciado e a extrair os dados correspondentes (Questão I). No entanto, 11 são de opinião que nem a função objectivo, nem as expressões matemáticas que traduzem as restrições do problema são difíceis de construir (Questões II e III).

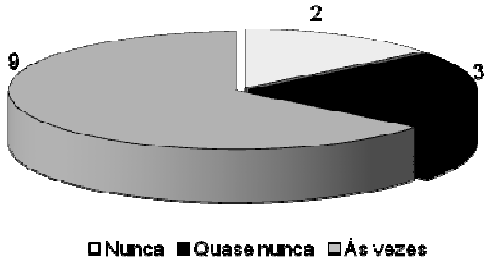
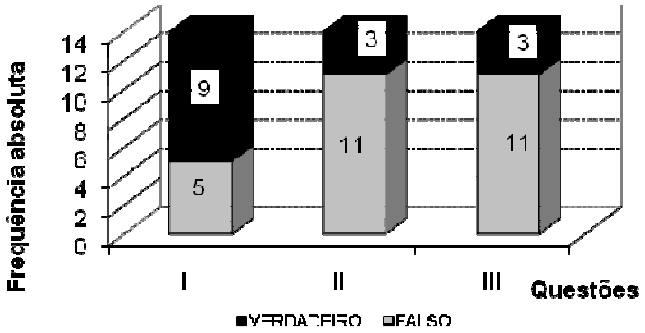


Gráfico 1 – "Consideras que a formulação de PPL é difícil?"



Questões:

A formulação de PPL é difícil porque:

I- a interpretação do texto é difícil, sendo complicado retirar os dados do problema.

II- é difícil construir a função matemática que traduz a Função Objectivo.

III- é difícil construir as expressões matemáticas que traduzem as restrições do problema.

Gráfico 2 – Opiniões dos alunos acerca das dificuldades sentidas ao formular PPL.

Convém recordar que apenas três dos catorze PMC já tinham frequentado o 11º ou o 12º ano da área de Ciências, tendo dois dos outros onze alunos escrito: *"Na resolução dos problemas surgiram conceitos novos, para nós desconhecidos, o que levou ao aparecimento de algumas dificuldades."* e, ainda, *"Senti dificuldades porque, às vezes, o português das perguntas era um bocado complicado"*.

Foi a opinião das docentes que os alunos revelaram muitas dificuldades na interpretação dos enunciados dos problemas. As suas leituras eram de início muito superficiais, quando era pedido para relerem os textos com mais cuidado ou até em voz alta, a reflexão surgia e a compreensão também. Contudo, algumas destas dificuldades também tiveram que ver com o uso da linguagem própria da PL, como é o caso dos conceitos: *função objectivo, restrições, variáveis de decisão,* entre outros.

Pode verificar-se, através do Gráfico 3, que todos os PMC consideram útil a utilização de *software* na resolução de PPL, sendo mesmo a maioria da opinião que é sempre ou quase sempre útil.

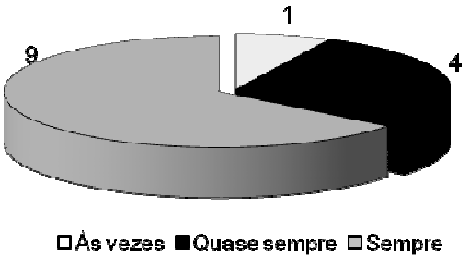
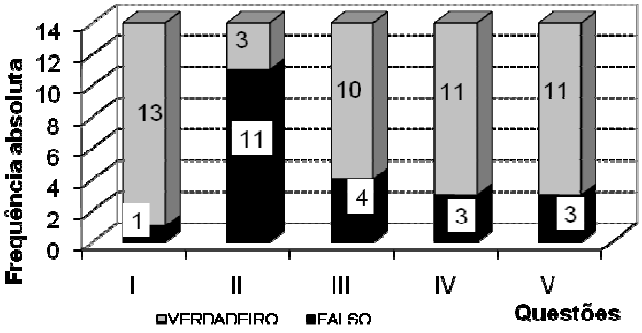


Gráfico 3 – "Consideras que a utilização de *software* para resolver PPL é útil?"

Como se pode observar no Gráfico 4, quase todos os alunos concordam que a utilização de *software* é um processo rápido, pouco trabalhoso e geral de resolução de PPL, que permite efectuar alterações nos dados com facilidade. No entanto,

consideraram que a sua utilização implica a necessidade do utilizador do *software* ter muitos conhecimentos sobre o tema da PL. Os alunos comentaram que o *WinQSB* permitia resolver problemas com um grande número de variáveis e sem se enganarem tanto



Questões:

A utilização de software para resolver problemas de PL é útil porque:

I- é um processo de resolução rápido.

II- não são necessários muitos conhecimentos sobre o tema da PL por parte do utilizador.

III- é um processo de resolução pouco trabalhoso.

IV- é um processo de resolução geral, visto permitir a resolução de problemas de várias variáveis.

V- permite efectuar alterações nos dados com facilidade.

Gráfico 4 – Opiniões dos PMC acerca da utilidade de *software* na resolução de PPL.

Relativamente ao Ponto 3 do questionário, os PMC foram unânimes em considerar que, em geral, o *WinQSB* é simples utilizar. Na opinião das professoras, os alunos revelaram-se muito à vontade em lidar com o *software* e as questões que surgiram foram mais relacionadas com a PL e a interpretação dos resultados do que com os procedimentos do programa.

Finalmente, pode verificar-se, através dos Gráficos 5 e 6, que 11 participantes são da opinião que o que aprenderam no decorrer deste MC pode ser útil na resolução de problemas do quotidiano (Ponto 4 do questionário), estando 9 deles interessados em aprofundar o tema da PL (Ponto 5 do questionário). Face ao entusiasmo dos alunos com a utilização do *software*, as docentes lamentaram não poder realizar, pelo menos, mais uma sessão.

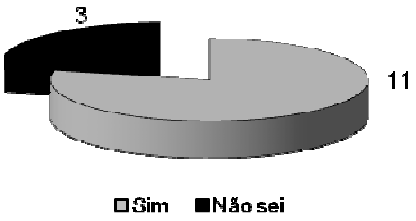


Gráfico 5 – "Achas que o que aprendeste pode ser útil na resolução de problemas do dia a dia?"

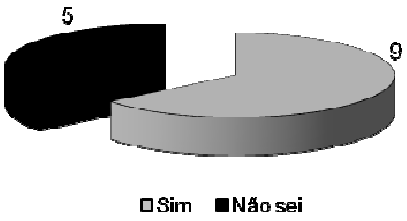


Gráfico 6 – "Gostarias de explorar mais o tema da PL?"

4- Conclusões

Na opinião da maioria dos PMC, a formulação de PPL é por vezes difícil, em virtude de considerarem complicado interpretar o enunciado do problema e extrair os dados correspondentes. No entanto, referiram não ter sentido dificuldades significativas tanto na construção da função objectivo, como na das expressões matemáticas que traduzem as restrições do problema, apesar destes conceitos serem desconhecidos para a maioria deles.

Todos os PMC consideraram útil a utilização de *software* na resolução de PPL e, na sua maioria, referiram mesmo ser sempre ou quase sempre útil utilizá-lo, visto terem concluído ser um processo rápido, pouco trabalhoso e geral de resolução de problemas. No entanto, foram de opinião que para utilizar um *software* é necessário possuir conhecimentos consideráveis sobre o tema da PL. Para além disso, foram unânimes em considerar que o *software* proposto, o *WinQSB*, é simples de utilizar.

A maioria dos alunos também é de opinião que os conhecimentos adquiridos ao longo deste MC podem ser úteis na resolução de problemas do quotidiano, estando a maioria interessada em aprofundar o tema da PL.

Em resumo, os alunos que participaram nesta experiência, embora tenham sentido algumas dificuldades na formulação de PPL, adaptaram-se com facilidade à utilização do *WinQSB* para os resolver.

Tal como já tinha sido constatado noutros trabalhos, por exemplo em [Monteiro e Teixeira, 2009a], [Monteiro e Teixeira, 2009b], [Nascimento e Nascimento, 2003] e [Nascimento, 2004], mantém-se a convicção de que o uso de *software* no Ensino Secundário, como complemento no processo de resolução de PPL, seria bem aceite pelos alunos, para além de estimular o seu interesse e de lhes dar uma visão mais abrangente do tema da PL.

5- Referências

Costa, B., Resende, L. C. e Rodrigues, E. (2004) *Espaço 11*, Asa Editora, Porto.

Filipe, J. (1998). Programação Linear: relato de uma experiência, *Educação e Matemática*, 49, 25-27, 32.

Hillier, F. S., Lieberman, G. J.. (2005) *Introduction to Operations Research*, (Eigth edition) McGraw-Hill, Singapura.

Jorge, A. M. B., Alves, C., Fonseca, G. e Barbedo, J. (2004) *Infinito 11, Parte 1*, Areal Editores, Porto.

Ministério da Educação (2004) http://sitio.dgidc.min-edu.pt/secundario/Paginas/Programas_ES_M.aspx. Junho 2009

Monteiro, H., Teixeira, A. (2009). Estudo experimental relativo à utilização de software de Programação Linear no Ensino Secundário. *Livro de Actas do 14º Congresso da Associação Portuguesa de Investigação Operacional, APDIO*, 211-218.

Monteiro, M H., Teixeira, A. (2009). An experimental study on the resolution of Linear Programming problems on High School classrooms. *Actas do IX Congreso Galego de Estatística e Investigación de Operacións*, 53-58.

Murty, Katta G. (1983). *Linear Programming.* John Wiley & Sons., EUA, 2. Ed.

Nascimento. M. J. (2004) Programação Linear: Uma proposta de intervenção didáctica no Ensino Secundário. *Tese de Mestrado.* UTAD, Portugal.

Nascimento, M. J. T. do e Nascimento M. M da S., (2003). Programação Linear: Tomada de Decisões, Uma Nova Abordagem no Ensino Secundários. *Actas do VI Congreso Galego de Estatística e Investigación de Operacións*, 369-374.

Neves, M. A. F., Guerreiro, L. e Moura, A. (2005) *Geometria 11º ano*, Porto Editora, Porto.

Ramalhete, M., Guerreiro, J., Magalhães, A. (1984). *Programação Linear* (Vol. I). McGraw-Hill de Portugal, Lda.

Shamblin, J. E. e G. T. Stevens, (1987). *Pesquisa Operacional: uma abordagem básica*. Atlas S.A., São Paulo.

Soveral, A. A. e Silva, C. V. (2004) *Matemática – 11º ano*, Vol. 1, Texto Editora, Lisboa.

WinQSB (2009) http://winqsb.10001downloads.com, Novembro 2009.

6- Anexos

Nesta secção disponibilizam-se a ficha com os enunciados dos três PPL propostos e o questionário de opinião, entregues aos PMC.

6.1- Ficha de trabalho

As etapas necessárias para se ***formular matematicamente um Problema de Programação Linear*** (às vezes também abreviado como ***PPL***) são:

1ª ETAPA: Definir as variáveis de decisão.
2ª ETAPA: Definir a função objectivo, usando as variáveis de decisão definidas.

3ª ETAPA: Definir as restrições (equações ou inequações), usando as variáveis de decisão definidas.

Problema 1 (adaptado de [Ramalhete et al, 1984])

Um nutricionista pretende determinar as quantidades de barras nutritivas, Superbom e Dáme+, que devem ser dadas diariamente a cada pessoa nesta dieta de forma a conseguir uma certa qualidade nutritiva a um custo mínimo. Os dados relativos ao custo de cada tipo de barra (em unidades monetárias por quilograma, u.m./kg), às quantidades mínimas diárias de ingredientes nutritivos básicos a fornecer a cada pessoa nesta dieta, bem como às quantidades destes existentes em cada barra (em gramas por quilograma, g/kg) constam do quadro seguinte:

	Barra		
Ingredientes nutritivos	Superbom	Dáme+	Quantidade mínima requerida
*****Hidratos de Carbono	20	50	200
*****Vitaminas	50	10	150
*****Proteínas	30	30	210
Custos (u.m./kg)	10	5	

Formule o modelo de programação linear que permita ao nutricionista estabelecer a dieta para estas pessoas a um custo mínimo.

1ª ETAPA: Definir as variáveis de decisão.

x = quilogramas de barras Superbom a comprar

y = _____

2ª ETAPA: Definir a função objectivo, usando as variáveis de decisão definidas.

Optimização pretendida → custo mínimo → minimização dos custos (gastar o menos possível):

* Para os quilogramas de barras Superbom a comprar tem-se:

custo da compra de cada quilograma de barras Superbom = 10 u.m. (unidades monetárias)

custo da compra de ***x*** quilogramas de barras Superbom = 10 ***x*** u.m.

* Para os quilogramas de barras Dáme+ a comprar tem-se:

custo da compra de cada quilograma de barras Dáme+ = _____ u.m.

custo da compra de **y** quilogramas de barras Dáme+ = _____ u.m.

* Logo a **função objectivo** (abreviada por FO), aqui chamada **z**, poderá escrever-se

_____ **z** = _____ x + _____ y

3ª ETAPA: Definir as restrições (equações ou inequações), usando as variáveis de decisão definidas.

***Hidratos de Carbono:** *gramas de hidratos de carbono fornecidos nesta dieta: gramas por quilograma que as barras Superbom fornecem x quilogramas de barras Superbom compradas **+** gramas por quilograma que as as barras Dáme+ fornecem x quilogramas de barras Dáme+ comprados **≥** gramas de hidratos de carbono requeridos nesta dieta

* ou seja, resumindo **matematicamente**: _____ x + _____ y **≥** _____

***Vitaminas**: _____

***Proteínas**: _____

***Não – negatividade das variáveis de decisão**: _____

Problema 2 (adaptado de [Jorge et al, 2004])

Um fabricante produz quatro tipos de bicicletas: B1, B2, B3 e B4, com lucro unitário de 100 € (Euros). Na produção semanal usam-se três oficinas, cuja disponibilidade horária semanal é 200, 240 e 600, respectivamente para a Oficina A, B e C. O número de horas necessárias, em cada oficina, para produzir uma bicicleta de cada tipo é o seguinte:

	B1	**B2**	**B3**	**B4**
Oficina A	**0**	**2**	**1**	**0**
Oficina B	**4**	**0**	**3**	**2**
Oficina C	**20**	**0**	**0**	**19**

Q_1 – Devido a um problema com as máquinas, a fábrica, nessa semana, passou a produzir apenas bicicletas dos tipos B2 e B3. Formule o PPL de modo a que o dono da fábrica saiba quantas bicicletas de cada tipo deve a fábrica produzir para obter o lucro máximo?

1ª ETAPA: Definir as variáveis de decisão.

*O que é que se pretende produzir na fábrica?

x = _____

y = _____

2ª ETAPA: Definir a função objectivo, usando as variáveis de decisão definidas.

Optimização pretendida → _____ → _____:

*Logo a **função objectivo** (abreviada por FO), aqui chamada **z**, poderá escrever-se

_____ **z** = _____ x + _____ y

3ª ETAPA: Definir as restrições (equações ou inequações), usando as variáveis de decisão definidas.

***Oficina A: _____**

***Oficina B: _____**

***Oficina C: _____**

***Não – negatividade das variáveis de decisão**: _____

Q_2 – Na semana seguinte, após a resolução parcial do problema, a fábrica conseguiu passar a produzir bicicletas dos tipos B1, B2 e B3. Agora, quantas bicicletas de cada tipo deve a fábrica produzir para obter o lucro máximo?

1ª ETAPA: Definir as variáveis de decisão.

2ª ETAPA: Definir a função objectivo, usando as variáveis de decisão definidas.

3ª ETAPA: Definir as restrições (equações ou inequações), usando as variáveis de decisão

Q_3 – Quando a fábrica estiver apta a produzir os quatro tipos de bicicletas, quantas bicicletas de cada tipo deve a fábrica produzir para obter o lucro máximo?

Q_4 – Após a resolução total do problema (estando a fábrica estava apta a produzir os quatro tipos de bicicletas) uma multinacional contactou a fábrica e fez uma encomenda de bicicletas, impondo como condição a entrega de pelo menos 10 bicicletas do tipo B1. Quantas bicicletas de cada tipo deve agora a fábrica produzir para obter o lucro máximo?

TPC (trazer feito para a próxima sessão)

(adaptado de [Murty, 1983]) Uma determinada área florestal está dividida em quatro regiões disponíveis para a plantação de quatro espécies florestais: pinheiro (P), abeto (A), nogueira (N) e carvalho (C). No

quadro seguinte, compilaram-se os dados referentes a essa área florestal – regiões e espécies florestais –. O valor esperado de produção por região é dado em metros cúbicos por hectare (m^3 /ha) e o valor esperado do rendimento por espécie é dado em unidade monetárias por hectare (um/ha). Além disso, a área disponível por região é dada em hectares, há, e a produção mínima requerida em metros cúbicos, m^3.

		Valor esperado da produção por região				Valor esperado do rendimento por espécie			
Região	Área disponível por região	P	A	N	C	P	A	N	C
1	1500	17	14	10	09	16	12	20	18
2	1700	15	16	12	11	14	13	24	20
3	900	13	12	14	08	17	10	28	20
4	600	10	11	08	06	12	11	18	17
Produção mínima anual requerida por espécie		22.5	09	4.8	3.5				

Formule o modelo de programação linear que permita maximizar o rendimento anual de cada espécie em cada região disponível, atendendo aos valores das produções esperadas e às produções mínimas requeridas.

6.2- Questionário de opinião

Resolveste alguns Problemas de Programação Linear, PPL. Responde agora às questões:

1. Consideras que a formulação de PPL é difícil?

□ Nunca □ Quase nunca □ Às vezes □ Quase sempre □ Sempre

Em caso afirmativo (se não assinalaste a primeira hipótese) tal acontece porque **(preenche cada quadrícula com V, no caso da afirmação ser verdadeira, ou F, no caso de ser falsa)**:

I) a interpretação do texto é difícil, sendo complicado retirar os dados do problema. □
II) é difícil construir a função matemática que traduz a Função Objectivo. □
III) é difícil construir as expressões matemáticas que traduzem as restrições do problema. □

2. Consideras que a utilização de *software* para resolver PPL é útil?

□ Nunca □ Quase nunca □ Às vezes □ Quase sempre □ Sempre

Em caso afirmativo (se não assinalaste a primeira hipótese) tal acontece porque **(preenche cada quadrícula com V, no caso da afirmação ser verdadeira, ou F, no caso de ser falsa)**:

I) é um processo de resolução rápido. □
II) não são necessários muitos conhecimentos sobre o tema da PL por parte do utilizador. □
III) é um processo de resolução pouco trabalhoso. □
IV) é um processo de resolução geral, visto permitir a resolução de problemas de várias variáveis □
V) permite efectuar alterações nos dados com facilidade. □

3. Consideras que, em geral, foi fácil utilizar o *WinQSB* para resolver PPL? □ Sim □ Não

Em caso negativo tal deveu-se a teres dificuldade em **(preenche cada uma das quadrículas seguintes com V, no caso da afirmação ser verdadeira, ou F, no caso de ser falsa):**:

I) trabalhar com o computador. □
II) introduzir os dados do problema. □
III) executar os procedimentos necessários nas diferentes janelas. □
IV) compreender os conceitos de PL necessários à utilização do *software*. □

4. O que aprendeste pode ser útil na resolução de problemas do dia a dia? □ Sim □ Não □ Não sei

5. Gostarias de explorar mais o tema da PL? □ Sim □ Não □ Não sei

Optimization of Batteries Sustainable Distribution Network with Reverse Flows

André Lopes Fradinho[a], *Maria Isabel Gomes*[b*], *Ana Paula Barbosa Póvoa*[c]

[a]*IST, UTL, Lisboa, Portugal*
alfradinho@gmail.com

[b]*CMA, FCT, Universidade Nova de Lisboa, Monte da Caparica, Portugal*
mirg@fct.unl.pt

[c]*Centro de Estudos de Gestão do IST, UTL, Lisboa, Portugal*
apovoa@ist.utl.pt

* Corresponding author

Abstract

Optimized supply chains, both economically as well as environmentally, are significant instruments to be adopted by companies if they want to succeed in the current competitive world. However, without the support of decision tools to help, at least approximately, the integration of these two aspects in the company's decision making such aim is difficult to reach.

The purpose of this study is to develop a mixed integer linear programming model (MILP), which allows the optimization of closed loop supply chains, considering not only economical aspects, but also environmental ones. The model developed was applied to a Portuguese company where their logistic network structure was redesigned.

Keywords: Close-loop supply chain, economical optimization, environmental optimization, warehouses location.

1- Introduction

Consumers and modern organizations are changing their goals. These are no longer strictly connected to economic factors, quality and level of service. They have also started to look into environmental aspects. The business community increasingly recognizes the significance of products and processes environmental performance. As an immediate consequence, the supply chain management is under transformation process.

Numbers are quite impressive and they represent well this urgency for a change. The increasing emission of GHG (greenhouse gases) into the atmosphere reached menacing and frightening values over the last years. According to the *Stern Report*, in 2006, transportations accounted for 14% of greenhouse gases emissions, from which three quarters are related to road transportation (Stern, 2006). It should be noted there is still a shortness of scientific literature related to the quantification of these emissions for further improvement of supply chains performance (Peacyk and McKinnon, 2009). However, efforts have been made in order to find new ways to optimize supply chains taking this kind of concerns into consideration.

The decrease of products life-cycle, the growing environmental concerns, and a tendency towards the diminishing of profit margins has lead companies to increasingly consider the return of used products. Companies are becoming more aware, or legally obligated, to take responsibility for the entire product life-cycle, including a proper recovery and disposal (Kleindorfer *et al*., 2005).

Literature, concerning the connection of environmental issues with the activities of a supply chain is rather limited. However, this subject has been explored in a some research studies (Bojarski *et al*., 2009). Among the published works, it must be mentioned the work of Duke et al. (2007), which developed a MILP model for the definition of the best supply chain structure transportation routes, taking into account the design of the product process, environmental impacts and economical metrics. Later on, this work was improved, by adding an environmental indicator, the Eco-indicator 99 (EI99), and therefore being able to quantify the supply chain environmental impact. Guillén-Gosálbez and Grossmann (2008) have also proposed a MILP model that considers profits and environmental impacts of a chemical plant supply chain. They access the environmental performance using the Eco-Indicator 99. Neto *et al.* (2007) developed a framework and applied it to the design of a paper industry logistic chain. Profits and environmental impacts of logistic networks are balanced by means of a multi-objective analysis. These authors provide an algorithm for a multi-objective linear problem with three objectives: reduce impacts and costs, and achieve target sets by an EU directive. The CED (Cumulative Energy Demand) was chosen as the environmental impacts indicator.

With the present work, we aim to contribute to the development of a decision tool that allows for the design of supply chains with return flows (*Closed Loop Supply Chains*), taking into account not only the economical aspects, but also the environmental ones. The model developed is then applied to a real case study of a batteries distribution supply chain where reverse flows are also considered.

2- Case Study

The model developed is applied to a Portuguese company, which manufactures and distributes batteries. The company wishes to optimize its logistics network structure considering not only economical performance metrics but also environmental ones.

This logistics network is composed of one factory and twelve warehouses, spread along Portugal. The latter have different features in terms of functions and dimensions. All of these facilities are responsible for the capillary distribution (distribution to costumers from the warehouse) and act as a direct point of sale to the public.

In terms of selling, this company acts as direct points of sale to the public, performs capillary distribution, or sells trough two licensed agents.

Moreover, the company also manages its reverse flow, which means that it collects and sells the end-of-life batteries released by the customers. These reverse flows represent raw materials to produce new batteries.

The transportation of the batteries can be done in two different ways. The primary distribution is outsourced. Subcontracted companies take products between the factory and the warehouses. To deliver batteries to customers, the company uses in-house vehicles (figure 1). By doing this direct supply (and collect when EOL – End Of Life – batteries are available) the company assures that orders are fulfilled within 24 hours after having been placed. This service level is of major importance for the company.

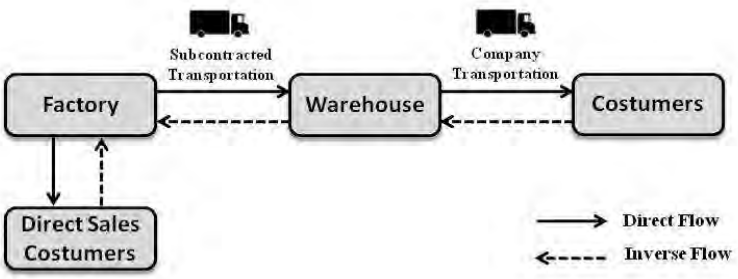


Figure 1: Diagram of company's supply chain.

The current network structure was developed along the years. It now allows the company to achieve their service level targets. However, since this operating structure was not object of a global analysis, the company aims at accessing its economical as well as environmental efficiency. This study will focus on warehouse location while considering the associated activity planning. This will be achieved by means of a MILP model that will be summarized in the next section.

3- Mathematic Model

The developed model was based on the model proposed by Salema et al. (2010). This is a multi-period location model that provides information regarding the supply chain planning. By using a graph representation and since the model integrates both forward (flows from factories to customer) and reverse flows (flows from customers back to factories), it is easily adapted to the context in study. Within this graph representation, nodes act as any entity of a supply chain (factories, warehouses, customers, among others) and arcs establish the link between two nodes, thus defining a flow. The developed model considers three types of entities (factory, warehouses and costumers), each one defined by its geographical location.

The costs and environmental impacts accounted in the model are the following: fixed costs of warehouses and environmental impact associated with its opening, batteries supplying costs and environmental impacts related to their production, costs and environmental impacts associated with the activity of the two fleets (primary and secondary), the amount paid to customers for the EOL products, and transportation costs and environmental benefits to this collecting activity.

The final purpose consists in redesigning the logistic network with the smallest cost and minimum environmental impacts. Therefore two objective functions are defined. On the cost minimization, the objective function considers fixed costs of opening/renting warehouses, the costs of raw materials needed for production, not only acquired from suppliers, but also the raw materials produced by recycling the EOL products purchased from customers. In addition, the transportation costs are estimated differently, depending on the specific flow. For the primary distribution in the forward flow, since it is subcontract transportation, the cost accounts for the distance from the factory to the warehoused and the number of trips performed by the subcontracted company. For the transportation of EOL batteries from warehouses back to the factory, the transportation cost is given by a fraction of its direct cost. In terms of the capillary distribution (between warehouses to the costumers) since the company uses in-house fleet, the cost is proportional to the distance traveled and the number of necessary trips to deliver products. Note that the cost of the vehicle return also adds to this transportation cost. All data used to estimate the above mentioned costs were provided by the company. Due to confidentiality reasons these values can not be published. Concerning the minimization of the environmental impacts, the Eco Indicator 99 (Goedkoop and Spriensma, 1999) was the choice for accessing the values. This indicator is based on a Life Cycle Assessment and the final value is a weighted sum of damages in three categories: human health, ecosystem quality and resources.

Each one of the damage categories is analyzed in a similar way to costs. Namely, the environmental objective function considers the effect of entities installation (land-use), the production of batteries, and several terms related with the transportation activity. Since the company does not consider changing the location of the factory, the impact of land-use by the factory is not considered. Regarding the definition of the impacts related to the production of batteries, the impacts considered result directly from the production of batteries having in mind that the recovery of old batteries leads to a reduction in the volume of purchased raw materials. In terms of transportation, the environmental burden comes mostly from gases emissions. All transportation are included, even the primary distributions that is outsourced.

The problem characteristics are modeled through a set of constraints that accounted for:

- Balance of products across all the facilities that represent the supply chain throughout the time;
- Satisfaction of costumers demand;
- Collection of used batteries (limited to a fraction);
- Storage profiles accounting for storage limits;
- Amounts of direct and reverse flows along the chain;
- Installation of new entities when required.

A Mixed Integer Linear Model Program (MILP) resulted. This was implemented in the GAMS software and solved through the CPLEX V.23.5 solver in a Intel (R) Core ™ i7 CPU, 6.00 GB, 2.8 GHz. All the values related to the environmental impacts are obtained in the data library of *SimaPro 7.1*.

4- Results

The model previously explained is applied to the case study where both the economical as well as the environmental objectives are considered. Since this the first step of a more deep study, each objective is optimized separately. A bi-objective analysis is the next step to take. Therefore, two scenarios were studied where a distribution network from grassroots was considered. These are compared with the actual network. The first scenario (scenario A1) considered the design of the distribution network for the minimization of an economical function. The second case (scenario A2) designed also the distribution network but in this case the minimization of the environmental impacts was considered. For each scenario, different solutions are obtained. The difference will be below analyzed.

Due to confidentiality reasons, all values presented in this section were modified but they reflect a context similar with the real case.

4.1 Logistics Network

The logistic network achieved in the scenario A1 consists on a logistic structure with 13 warehouses (the same, in number, as the actual network) contrasting with the logistic network in the scenario A2, which has only 8 warehouses.

The location of warehouses for the three networks (current, cost effective and environmental effective) is presented in Figure 2.

The structure of the network A2, compared with the existing network, is rather inferior in the number of warehouses, given the strong environmental impact caused by the opening of warehouses. The opening impact of a warehouse with an average capacity is substantially larger than the impact of other activities, which led the environmental objective function to be extremely sensitive to the opening of such facilities.

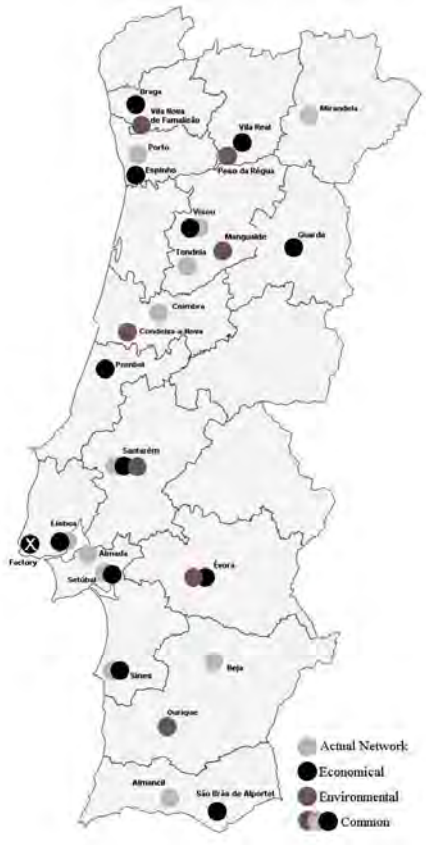


Figure 2: Logistics network obtained

4.2 Costs

As expected, the total cost in the scenario A1 is inferior to the cost of scenario A2. However, the analysis of different types of costs (Table 1) shows some singularities.

Table 1: Cost values for scenarios A1 and A2.

Costs (u.m.)	A1	A2
Fixed	81404,64	57366
Transp.f1(factory → warehouse)	80348,35	9398,638
Transp.f2+r1 (warehouse ↔ costumers)	177720	791620
Transp.r2 (warehouse → factory)	2630,977	-

Suppliers	1296800	1288800
Bat. Recover	13989,95	17305,95
Total	1652900	2164500

The most sticking aspects are related with transportation costs. The first one is the massive reduction observe in the costs values of Transp. f1, which refers to the forward connection between the factory and warehouses. The value decreases almost 90%. This is explained by the smaller number warehouses open in scenario A2, thus fewer subcontracted vehicles are needed. One second reason not so obvious is related with the supply plans. The scenarios show optimal supplying plans completely different. In scenario A2 (environmental minimization), the warehouse located in the factory processes much more batteries than all other seven warehouses. The reason is the non existence of generation of environmental impacts associated with the transportation between the factory and this warehouse, which for longer distances are associated with large impact values due to the type of transportation used in the primary distribution. This will be later on detailed. Also this connection does not have any cost associated. Since this is the largest warehouse of the company it has enough capacity to serve an enormous amount of demand.

Another consequence is the higher value of the secondary transportation costs (transp. f2 + r1) in the scenario A2 (almost 4 times larger), which is explained again by the smaller number of warehouses opened in scenario A2. Therefore there is a greater need for secondary transportation.

As for the transportation costs from warehouses back to the factory (Transp. r2), we have a curious situation. Apparently, the scenario A2 chooses not to transport the collected batteries from the warehouses to the factory, this essentially due to the environmental impacts as it will be detailed later on. However, costs associated to the collection itself are higher in this scenario, in other words, the company collects more batteries. After analyzing the volume of products sent from the warehouse to the factory in the two scenarios, it is possible to verify that the scenario A2 collects 20% more batteries than the scenario A1. The non-existence of transportation cost related with this flow is explained by the fact that all products are collected by the warehouse located in the factory, which connection has no associated cost.

Finally, the scenario A2 achieves a 1% reduction in raw material costs when compared with the scenario A1, which despite seeming as a residual value, it represents a decrease of 3300 u.m. per year in the company accounts.

4.3 Environmental Impacts

The environmental impact, measured by the eco-indicator 99, in the scenario A2 is naturally smaller than in the scenario A1. However, in order to achieve a 1% reduction in these impacts when compared with the A1 scenario, there was a 30% increase in costs (see Table 2). Cleary this is an unfavorable relation.

Table 2: Environmental impact values for scenarios A1 and A2.

EI99		**A1**	**A2**
Warehouses		930	655
Batteries		7890000	7840000
Bat. Recover		- 61110	- 73155
Transp (total)		30180	22500
	Transp.f1(factory → warehouse)	23500	3150
	Transp.f2+r1 (warehouse ↔ costumers)	4350	19400
	Transp.r2 (warehouse → factory)	2330	0
Total		7860000	7790000

The main difference between costs minimization and the minimization of the environmental impacts is related with the extreme importance of the batteries production activity, since the environmental impacts presented by the model in both scenarios are dominated by the impact of production. Although the impact of a single battery is very small when compared, for example, with the impact of opening a warehouse, the total production represents 95% and 99% of the total environmental impact of this supply chain, in the scenarios A1 and A2, respectively.

It is now clear why the scenario A2 collects all the possible EOL batteries. The main difference between the minimization of the two functions lies in this fact. Considering its strong impact in the chain, minimizing the environmental function A2 favors more the collection of batteries than the economical function (A1), therefore the minimization is strongly focused on this reduction. That is why the warehouse located within the factory is essential in scenario A2. The transportation cost for batteries collected at this location is zero since factory and warehouse share the same location. Also the environmental impact between these two points is zero. As a result, the flow of this

warehouse is massive and subsequently the secondary distribution increases exponentially, thus creating a huge gap in costs between the two scenarios. Another fact that helps increasing the secondary distribution, rather than the primary, is the small amount of effects that the secondary fleet has (21 times lower than the primary fleet) has on the environmental impacts when compared to the primary distribution.

Regarding the optimal cost scenario, it is possible to achieve a 34% decrease of the transportation impacts in scenario A2. Despite the higher impact created by the secondary fleet in this scenario, the primary offsets it, because it is much less used (flow r2).

In short, scenario A1 privileges the primary transportation with regard to the secondary, since economically that relationship is much more favorable. In scenario A2, one observes the opposite phenomenon due to the associated environmental impacts. Therefore, one may say that it is of advantage to transport less with the sub-contracted fleet in favor the secondary transportation considering its small environmental impact.

Overall it can be said the supply chain when the total cost is at stake has a decentralized structure, while when environmental is the driver, the structure changes to a centralized one.

For a more sustained analysis the values of some parameter should careful analyzed in future work. The above results show that they have a considerable influence on the supply chain structure.

5- Conclusions

The present paper studied the design and planning of a logistics network where both economical as well as environmental aspects are considered. The model is applied to a real logistics network where production, distribution as well as recovery of batteries are considered.

As final results it can be seen that the current operating network can be further optimized not only in terms of costs but mostly in environmental terms. Nonetheless a more deep analysis needs to be performed since both optimal supply chains present very different structures.

The high environmental impact of the primary fleet, combined with the extreme importance of the batteries collected in the model, given the environmental benefits of this activity, entails as the best option the intensive use of the fleet responsible for the secondary distribution. However, these optimal results lead to a supply chain structure that does not favor the service quality policy followed by the company. The agility and flexibility in the delivery of the batteries, key features for the company, could be at risk. Another aspect concerns the costs involved if the optimal environmental structure is chosen. However, the tendency towards centralization without damaging the service level should be considered by the company.

As future work, a bi-objective model is developed to account simultaneously with the minimization of both economical and environmental objectives. Moreover, a careful sensitivity analysis should be performed on several parameters that come out critical in final solutions. Furthermore, given the different structure in both solutions, a bi-objective analysis is the next step to take. This analysis will provide further insight into what is the best compromise solution that the company should follow.

Acknowledgment

The authors gratefully acknowledge the support of the Portuguese National Science Foundation through the project PTDC/SEN-ENR/102869/2008.

References

Bojarski, A.D., Laínez, J.M., Espuna, A. Puigjaner, L. (2009). Incorporating environmental impacts and regulations in a holistic supply chains modeling: An LCA approach. Computers and Chemical Engineering, 33(10), 1747-1759.

Duque, J., Barbosa-Póvoa, A.P.F.D, Novais, A.Q. (2007). Synthesis and optimization of the recovery route for residual products under uncertain product demand. Computers & Operations Research, 34, 1463-1490.

Goedkoop, M.J., Spriensma, R.S. (1999). The Eco-indicator 99, Methodology report. A damage oriented LCIA Method; VROM, The Hague, The Netherlands.

Guillén-Gosálbez, G., Grossman, I.E. (2008). Optimal Design and Planning of Sustainable Chemical Supply Chains Under Uncertainty, AIChe Journal, 55(1), 99-121.

Kleindorfer, P.R, Singhal, K, Van Wassenhove, L.N. (2005). Sustainable Operations Management, Production and Operations Management, 14(4), 482-492.

Neto, Q.F, Bloemhof, J., Van Nunen, J.A.A.E, Van Heck, E. (2007). Designing and evaluating sustainable logisticis networks. International Journal of Production Economics, 111, 195-208.

Piecyk, M.I., McKinnon, A.C. (2009). Forecasting the carbon footprint of road freight transport in 2020. International Journal of Production Economics, 128, 1-12.

Salema, M.I.G., Barbosa-Póvoa, A., Novais, A.Q. (2010). Simultaneous design and planning of supply chains with reverse flows: A generic modeling framework. European Journal of Operational Research, 203, 336-349.

Stern, N. (2006). Stern Review: The Economics of Climate Change. HM Treasury, London.

OPTIMIZAÇÃO DO ESCALONAMENTO DE ENFERMEIROS NUMA UNIDADE HOSPITALAR

Andreia Filipa Antunes, Margarida Moz
Centro de Investigação Operacional (Universidade de Lisboa),Portugal
andreiafilipantunes@gmail.com
Centro de Investigação Operacional (Universidade de Lisboa) e ISEG (Universidade Técnica de Lisboa),Portugal
mmoz@iseg.utl.pt

1. Introdução

O presente trabalho teve como objecto de estudo o planeamento de escalas de enfermeiros de um Serviço de Cirurgia de um Hospital Português.

O escalonamento de enfermeiros numa unidade hospitalar consiste na atribuição de diferentes tarefas a cada enfermeiro, com diferentes aptidões e restrições laborais, satisfazendo na medida do possível as preferências de cada um, a legislação em vigor e condições impostas pelo próprio Hospital, obtendo um plano de escalas, para um certo período de tempo, que cobre as necessidades de pessoal assegurando um eficiente funcionamento do Serviço 24 horas por dia.

O Serviço de Cirurgia do Hospital Português estudado também funciona 24 horas, 365 dias por ano, e recai sobre a Enfermeira Chefe o árduo, e muitas vezes habilidoso, trabalho de elaborar as escalas de Serviço de 28 dias, para cada um dos enfermeiros, que irão assegurar o bom funcionamento do Serviço. Nestas escalas, para além de um conjunto de restrições obrigatórias, a Enfermeira Chefe tem ainda que ter em conta as preferências e competências de cada enfermeiro.

O escalonamento de pessoal de enfermagem tem atraído o interesse de muitos investigadores. Desde a década de sessenta do século passado (Howell, 1966) até hoje, têm sido publicados artigos que dão conta de abordagens para resolver problemas específicos de unidades hospitalares (Burke et al., 2004). Estando o escalonamento muito dependente das normas estipuladas pela legislação de cada país não existe uma metodologia única que permita resolver a diversidade de problemas que surgem na prática. Continua por isso a ser hoje um desafio para a comunidade de investigação operacional a procura quer de metodologias de solução de problemas específicos, quer o estudo teórico da estrutura dos

problemas com vista a definir um enquadramento formal aplicável a vários casos (Bilgin et al., 2011).

Seguidamente enuncia-se o problema a tratar. Na secção 3 descreve-se a forma como actualmente é tratado o escalonamento. Na secção 4 descreve-se a abordagem e a formulação matemática propostas e, por fim, apresenta-se a solução obtida e algumas conclusões.

2. Descrição do Problema

O Serviço de Cirurgia Geral do Hospital Português que foi objecto de estudo neste trabalho funciona todo o ano 24 horas por dia e recebe pacientes com doenças de vários tipos.

Existem dois tipos de cirurgia, convencional e adicional que, respectivamente, ocorrem durante o período do dia (entre as 8 e as 16 horas) ou resultam do sistema de listas de espera e podem dar-se fora do horário convencional. Actualmente, este Serviço tem capacidade para receber 42 doentes, sendo que de segunda a sexta-feira se operam em média 8 a 10 pessoas por dia. O espaço é composto por 12 enfermarias e uma sala para doentes em pós-operatório (sala de recuperação).

Presentemente, o Serviço possui um efectivo de 50 enfermeiros divididos em quatro equipas, cada uma das quais (A, B, C e D) é constituída por 10 ou 11 elementos, perfazendo um total de 42 enfermeiros. Os restantes 8 enfermeiros estão incluídos numa quinta equipa (F) devido às suas características que os distinguem dos outros. Desta equipa fazem parte o Enfermeiro Chefe, enfermeiros especialistas, enfermeiros da sala de recuperação e outros que vão assistir a consultas. Estes enfermeiros não possuem o mesmo tipo de carga laboral e funções dos restantes e, uma vez que têm um horário de tipo fixo em que trabalham Manhãs todos os dias úteis e folgam ao fim-de-semana, o seu escalonamento está sempre determinado e por isso não vai ser incluído neste estudo.

A carga horária semanal não é a mesma para os 50 enfermeiros, sendo mais frequentes as 35 horas, 40 horas e 42 horas semanais, existindo no entanto uma pequena percentagem com horários de 30 horas, 31 horas ou 32 horas semanais. Esta variação advém do facto do Hospital possuir um estatuto especial para os enfermeiros dos Serviços de internamento que prevê a redução do horário de trabalho em 1 hora, até um máximo de 30 horas, por períodos de 3 anos. Este estatuto especial prevê ainda que ao fim de um ano de trabalho efectivo neste Serviço, os enfermeiros tenham direito a um período adicional de cinco dias úteis de férias a gozar no ano seguinte.

Na Tabela 1, apresenta-se um resumo da situação quanto ao número de enfermeiros por equipa (A a D) e por carga horária semanal.

Tabela 1: Número de enfermeiros por carga horária semanal e por equipa

Horas semanais	Nº enfermeiros na equipa A	B	C	D	Nº horas período
42	1	1	2	1	168
40	8	8	8	7	160
33	1	0	1	0	132
32	1	1	0	0	128
31	0	0	0	2	124
	11	**10**	**11**	**10**	

Com o intuito de organizar mais eficientemente o trabalho, o dia é dividido em 3 turnos, Manhã (M), Tarde (T) e Noite (N), pelos quais a Enfermeira Chefe distribui os enfermeiros. Esta divisão visa cobrir da melhor forma as diferentes necessidades de pessoal ao longo do dia. Consideram-se ainda dois outros tipos de turnos (turnos de não trabalho), o turno Folga (F) e o turno Descanso (D) em que os enfermeiros se encontram ausentes do Serviço. Cada um dos turnos de trabalho tem duração fixa, iniciando-se e terminando sempre à mesma hora, a saber: o turno da Manhã (M) decorre das 8h às 16h; o turno da Tarde (T) das 15:30h às 23h e o da Noite (N) estende-se das 22:30h às 8:30h. Esta sobreposição de horas serve para que os enfermeiros do turno cessante passem toda a informação relevante aos enfermeiros que vão trabalhar no turno seguinte. Nos períodos entre as 20h e as 7h de segunda a sábado, e das 13h de sábado às 7h de segunda-feira as remunerações por hora são consideravelmente mais elevadas, pelo que tem de haver uma justa distribuição destes períodos tão almejados pelo conjunto de todos os enfermeiros.

Cada unidade do Serviço (enfermarias e sala de recuperação) tem diferentes necessidades de pessoal consoante as horas do dia. Para além das necessidades de Serviço, quando se elabora a escala do período tem de se contemplar ainda as necessidades dos enfermeiros se deslocarem do Serviço e irem assistir nas consultas. Estes enfermeiros integram a denominada equipa F.

Na tabela 2 apresentam-se as necessidades mínimas e máximas por turno e por dia da semana para os enfermeiros das equipas A a D.

Os enfermeiros destas quatro equipas, quando se encontram a trabalhar sem restrições, têm um horário rotativo, isto é, a sequência de turnos de trabalho não é composta por um só tipo de turno; os enfermeiros da equipa F distinguem-se precisamente pelo seu horário fixo.

Tabela 2: Número de enfermeiros por turno e por dia da semana necessários ao Serviço

Turnos/dias	2ª	3ª	4ª	5ª	6ª	sábado e domingo
Manhã	12/13	12	13	12/13	13	8/9
Tarde	7/8	8	8	8	8/9	7/8
Noite	6/7	6/7	6/7	6/7	6/7	6/7

3. Planeamento das escalas na unidade

Designa-se por escala uma sequência de turnos (de trabalho ou não trabalho), um por dia, atribuída a cada enfermeiro no período de escalonamento. Cada plano de escalas é o conjunto das escalas de todos os enfermeiros do Serviço e tem de obedecer a um conjunto de regras impostas pela legislação laboral bem como pelo próprio Hospital e, simultaneamente, ir de encontro às preferências que os enfermeiros mostram por certas sequências de turnos.
Assim, os planos de escalas são efectuados para períodos de 28 dias, isto é, 4 semanas, a iniciar sempre numa segunda-feira e terminar num domingo, e as restrições a que têm de obedecer são:

- a cada enfermeiro é atribuída uma e uma só tarefa (turno) por dia;
- a semana de trabalho é em regra de 35 horas e de cinco dias, podendo sofrer alterações consoante as necessidades do Serviço ou a situação laboral do enfermeiro, como atrás referido;
- os enfermeiros têm de estar um número mínimo de horas sem trabalhar entre turnos de trabalho consecutivos (neste caso serão 10 horas), o que determina a existência de certas sequências de turno de trabalho inadmissíveis;
- não se pode atribuir um turno a enfermeiros que não possam efectuar esse turno nesse dia;
- os enfermeiros têm direito a um dia de descanso semanal (Folga) acrescido de um dia de descanso complementar (Descanso), devendo, em cada período de 28 dias, pelo menos um dos dias de descanso coincidir com um sábado ou domingo;
- o planeamento deve ser feito para um período de quatro semanas;
- na elaboração das escalas, os feriados nacionais e municipais que recaiam em dias úteis têm de ser considerados pois, caso se encontrem a trabalhar, os enfermeiros devem ter direito posteriormente a gozar aquelas horas;
- os enfermeiros não podem trabalhar mais de seis dias consecutivos;

- o número de horas de trabalho de cada enfermeiro, por período, deve-se aproximar o mais possível da sua carga horária contratual;
- existe um requisito mínimo e um máximo de enfermeiros por turno de trabalho e por dia da semana.

Actualmente, no processo de escalonamento, a Enfermeira Chefe começa por construir um plano inicial - o Plano Ideal. Este plano baseia-se num esquema, isto é, uma sequência de turnos que os enfermeiros em horário rotativo preferem.

Figura 1: Esquema inicial

M	M	T	T	D	N	N	F

A posição relativa de turnos no esquema é apresentada na Figura 1 e considera-se que a seguir ao oitavo turno está o primeiro do esquema. Se se iniciar este esquema em posições diferentes, para cada enfermeiro por equipa (ver Figura 2), e repetir as sequências de oito dias parando no dia vinte e oito, obtém-se um Plano Ideal. Este plano pode ainda sofrer alterações pois a Enfermeira Chefe tem de assinalar as ausências conhecidas previamente, por motivo de férias, doença ou outro.

Figura 2: Sequência de turnos de um esquema por equipa para os primeiros 8 dias

A	D N N F M M T T
B	N F M M T T D N
C	T T D N N F M M
D	M M T T D N N F

O Plano Ideal assim obtido satisfaz algumas restrições mas não satisfaz, em geral, todas as restrições obrigatórias e a Enfermeira Chefe tem de iniciar um procedimento de trocas de turnos para satisfazer todas as restrições. Este procedimento manual é complexo e moroso até se atingir o plano que irá vigorar para o mês seguinte, o que motivou a abordagem de optimização proposta.

4. Abordagem proposta e formulação do problema

Com o objectivo de reproduzir o mais fielmente possível o procedimento actual de planeamento de escalas, propôs-se então uma formulação em Programação Linear Inteira Mista (Taha, 2006). Definiu-se a variável booleana y_{td}^{ve} com o seguinte significado:

$$y_{td}^{ve} = \begin{cases} 1, \text{se o enfermeiro } v \text{ da equipa } e \text{ for afecto ao turno } t \text{ no dia } d \\ 0, \text{se o enfermeiro } v \text{ da equipa } e \text{ não for afecto ao turno } t \text{ no dia } d \end{cases}$$

$e = 1, 2, 3, 4$ respectivamente A, B, C, D;
$v = 1, 2, \dots, 11$; para $e = 1, 3$ e $v = 1, 2, \dots, 10$; para $e = 2, 4$;
$t = 1, 2, 3, 4, 5$ respectivamente M, T, N, F, D; $d = 1, 2, \dots, 28$.

Os turnos de trabalho são: Manhã (M) representado por 1; Tarde (T) representado por 2 e Noite (N) representado por 3. Os turnos de não trabalho ou descanso, são: Folga (F) representado por 4 e Descanso (D) representado por 5. O modelo não contempla a equipa F (de 8 enfermeiros) pois esta possui um horário fixo (MMMMMFF) que não é possível alterar ou ser sujeito a melhoramentos.

Definiram-se ainda as variáveis desvio que assumirão o desvio das horas trabalhadas para as que, contratualmente, deveriam ter sido trabalhadas. Δ_{-}^{ve} é a variável de desvio que representa o número de horas que o enfermeiro v da equipa e trabalhará a menos no período de escalonamento ($0 \leq \Delta_{-}^{ve} \leq 8$ horas). Δ_{+}^{ve} é a variável de desvio que representa o número de horas que o enfermeiro v da equipa e trabalhará a mais no período de escalonamento ($0 \leq \Delta_{+}^{ve} \leq 8$ horas).

Qualquer solução admissível para o problema descrito corresponde a um plano de escalas que satisfaz todas as condições enunciadas. No entanto pretende-se obter o plano que mais se aproxima do Plano Ideal. Deste modo, vai-se definir uma função objectivo a minimizar, que representa o valor da dissemelhança entre as soluções admissíveis e o Plano Ideal.

Considere-se c_{td}^{ve} o custo de afectar o enfermeiro v da equipa e ao turno t do dia d com os seguintes valores e respectivos significados.

$$c_{td}^{ve} = \begin{cases} 0, \text{se o turno afecto ao enfermeiro } v \text{ da equipa } e \text{ for igual ao do Plano Ideal} \\ 1, \text{se o turno afecto ao enfermeiro } v \text{ da equipa } e \text{ for diferente do Plano Ideal} \\ +\infty, \text{se o enfermeiro } v \text{ da equipa } e \text{ não puder fazer o turno } t \text{ no dia } d \end{cases}$$

A função objectivo

$$z = \sum_{e \in \{1,3\}} \sum_{v=1}^{11} \sum_{d=1}^{28} \sum_{t=1}^{5} c_{td}^{ve} y_{td}^{ve} + \sum_{e \in \{2,4\}} \sum_{v=1}^{10} \sum_{d=1}^{28} \sum_{t=1}^{5} c_{td}^{ve} y_{td}^{ve}$$

com os custos acima definidos representa a dissemelhança entre o plano de escalas correspondente a cada solução admissível e o Plano Ideal medida em número de turnos atribuídos diferentes daqueles que constam no Plano Ideal.

No modelo de Programação Linear Inteira Mista (PLIM), adiante apresentado, as restrições 4.1.1 a 4.1.3 impõem a exigência do número mínimo de horas de descanso entre turnos de trabalho consecutivos, isto é, evitam sequências de turnos de trabalho não admissíveis. Um enfermeiro que tenha trabalhado o turno Tarde num dia não pode ser afecto a um turno

Noite no dia seguinte (4.1.1). Esta situação repete-se para o caso de se trabalhar Manhã depois de Tarde (4.1.2) ou Noite depois de Manhã (4.1.3).

$$\min z = \sum_{e \in \{1,3\}} \sum_{v=1}^{11} \sum_{d=1}^{28} \sum_{t=1}^{5} c_{td}^{ve} y_{td}^{ve} + \sum_{e \in \{2,4\}} \sum_{v=1}^{10} \sum_{d=1}^{28} \sum_{t=1}^{5} c_{td}^{ve} y_{td}^{ve}$$

$$s.a \begin{cases}
y_{2d}^{ve} + y_{3d+1}^{ve} \leq 1 \qquad \forall v, \forall e, d = 1, \dots, 27 & (4.1.1)\\
y_{2d}^{ve} + y_{1d+1}^{ve} \leq 1 \qquad \forall v, \forall e, d = 1, \dots, 27 & (4.1.2)\\
y_{1d}^{ve} + y_{3d+1}^{ve} \leq 1 \qquad \forall v, \forall e, d = 1, \dots, 27 & (4.1.3)\\
\sum_{v=1}^{11} y_{1d}^{v1} + \sum_{v=1}^{10} y_{1d}^{v2} + \sum_{v=1}^{11} y_{1d}^{v3} + \sum_{v=1}^{10} y_{1d}^{v4} \leq a_{1k} \quad \forall d, k = d \bmod 7 & (4.2.1)\\
\sum_{v=1}^{11} y_{1d}^{v1} + \sum_{v=1}^{10} y_{1d}^{v2} + \sum_{v=1}^{11} y_{1d}^{v3} + \sum_{v=1}^{10} y_{1d}^{v4} \geq b_{1k} \quad \forall d, k = d \bmod 7 & (4.2.2)\\
\sum_{v=1}^{11} y_{2d}^{v1} + \sum_{v=1}^{10} y_{2d}^{v2} + \sum_{v=1}^{11} y_{2d}^{v3} + \sum_{v=1}^{10} y_{2d}^{v4} \leq a_{2k} \quad \forall d, k = d \bmod 7 & (4.2.3)\\
\sum_{v=1}^{11} y_{2d}^{v1} + \sum_{v=1}^{10} y_{2d}^{v2} + \sum_{v=1}^{11} y_{2d}^{v3} + \sum_{v=1}^{10} y_{2d}^{v4} \geq b_{2k} \quad \forall d, k = d \bmod 7 & (4.2.4)\\
\sum_{v=1}^{11} y_{3d}^{v1} + \sum_{v=1}^{10} y_{3d}^{v2} + \sum_{v=1}^{11} y_{3d}^{v3} + \sum_{v=1}^{10} y_{3d}^{v4} \leq a_{3k} \quad \forall d, k = d \bmod 7 & (4.2.5)\\
\sum_{v=1}^{11} y_{3d}^{v1} + \sum_{v=1}^{10} y_{3d}^{v2} + \sum_{v=1}^{11} y_{3d}^{v3} + \sum_{v=1}^{10} y_{3d}^{v4} \geq b_{3k} \quad \forall d, k = d \bmod 7 & (4.2.6)\\
8 \times \sum_{d=1}^{28} y_{1d}^{ve} + 7{,}5 \times \sum_{d=1}^{28} y_{2d}^{ve} + 10 \times \sum_{d=1}^{28} y_{3d}^{ve} + \Delta_{-}^{ve} - \Delta_{+}^{ve} = H^{ve} - S^{ve} \quad \forall v, \forall e & (4.3)\\
\sum_{t=1}^{5} y_{td}^{ve} = 1 \qquad \forall v, \forall e, \forall d & (4.4)\\
\sum_{d=1+l}^{7+l} (y_{1d}^{ve} + y_{2d}^{ve} + y_{3d}^{ve}) \leq 5 \qquad \forall v, \forall e, \quad l = 0{,}7{,}14{,}21 & (4.5)\\
\sum_{t=4}^{5} (y_{t6}^{ve} + y_{t7}^{ve} + y_{t13}^{ve} + y_{t14}^{ve} + y_{t20}^{ve} + y_{t21}^{ve} + y_{t27}^{ve} + y_{t28}^{ve}) \geq 1 \qquad \forall v, \forall e & (4.6)\\
\sum_{t=1}^{3} \sum_{l=0}^{6} y_{t,d+l}^{ve} \leq 6 \qquad \forall v, \forall e, d = 1, \dots, 22 & (4.7.1)\\
\sum_{l=d-6}^{0} \sum_{t=1}^{3} p_{tl}^{ve} + \sum_{t=1}^{3} \sum_{l=1}^{d} y_{t,l}^{ve} \leq 6 \qquad \forall v, \forall e, d = 1, \dots, 6 & (4.7.2)\\
\Delta_{-}^{ve} \leq 8 \qquad \forall v, \forall e & (4.8.1)\\
\Delta_{-}^{ve} \geq 0 \qquad \forall v, \forall e & (4.8.2)\\
\Delta_{+}^{ve} \leq 8 \qquad \forall v, \forall e & (4.8.3)\\
\Delta_{+}^{ve} \geq 0 \qquad \forall v, \forall e & (4.8.4)\\
y_{td}^{ve} \in \{0,1\} \qquad \forall v, \forall e, \forall t, \forall d & (4.9)\\
y_{t1}^{ve} = 0 \quad \text{consoante o valor de } t, \forall v, \forall e \text{ e a informação do plano anterior} & (4.10)
\end{cases}$$

Para as restrições 4.2.1, 4.2.3 e 4.2.5 foi definido o vector a_{tk} de dimensão 1×7 que contém o número máximo de enfermeiros a atribuir ao turno t ($t = 1,2,3$ consoante se trata do turno da Manhã, Tarde ou Noite) do dia da semana k, com $k = 0, \dots, 6$; com $k = 0$ a corresponder a domingo. A dimensão 1×7 justifica-se porque as necessidades de enfermeiros por turno são as mesmas por dia de semana. O primeiro membro destas restrições representa o número de enfermeiros das quatro equipas afectos ao turno t do dia d do período. Só se consideram os turnos M, T e N, uma vez que se está a determinar as necessidades de pessoal para os turnos de trabalho.

De forma semelhante, para 4.2.2, 4.2.4 e 4.2.6; referentes à satisfação dos requisitos mínimos por turno, definiu-se um novo vector b_{tk} de dimensão 1×7 que representa as necessidades mínimas de enfermeiros no Serviço por turno e por dia da semana.

As restrições 4.3 reflectem o desvio entre as horas trabalhadas no período e a carga horária contratual do período, para cada enfermeiro. Sendo H^{ve} o número de horas que o enfermeiro v da equipa e deve trabalhar por período. Esta assume os valores 168h, 160h, 132h, 128h e 124 horas, consoante a carga horária semanal seja de 42h, 40h, 33h, 32h ou 31 horas, respectivamente; e S^{ve} o saldo de horas do período anterior. Note-se que as variáveis Δ_{-}^{ve} e Δ_{+}^{ve} em qualquer solução básica admissível ou são ambas nulas, caso em que não há desvio, ou se uma é positiva a outra é nula, indicando que há desvio por defeito ($\Delta_{-}^{ve} > 0$) ou há desvio por excesso ($\Delta_{+}^{ve} > 0$).

As restrições 4.4 indicam que a cada enfermeiro se atribui um e um só turno por dia, quer seja de trabalho ou descanso. As restrições 4.5 impõem que os enfermeiros devem gozar no mínimo duas folgas por semana (do tipo Descanso ou Folga), o que implica que no máximo cada enfermeiro deverá trabalhar 5 dias. As restrições 4.6 exigem que, para cada enfermeiro, um dia de não trabalho (Folga ou Descanso) em cada período de 4 semanas, coincida com o sábado ou domingo. Os sábados e domingos são os dias $6, 7, 13, 14, 20, 21, 27$ e 28 do período.

As restrições 4.7.1 determinam que no máximo cada enfermeiro só poderá trabalhar 6 dias seguidos, isto é, não pode haver uma sequência de turnos de trabalho (M, T ou N) com mais de 6 elementos e as restrições 4.7.2 referem-se igualmente ao número máximo de dias de trabalho seguidos, mas considerando a informação dos últimos dias do plano anterior. Para estas definiu-se, o parâmetro p_{td}^{ve} que indica o tipo de turno afecto ao enfermeiro v da equipa

e no dia d do período anterior. Para o efeito, interessa apenas saber os turnos atribuídos entre o 23º e 28º dias do plano anterior.

$p_{td}^{ve} = \begin{cases} 1, \text{se o enfermeiro } v \text{ da equipa } e \text{ trabalhou } (t = 1,2,3) \text{ no dia } d \text{ do plano anterior} \\ 0, \text{caso contrário} \end{cases}$

($l = 0, -1, -2, -3, -4, -5$ representam os últimos dias do plano anterior).

As restrições de 4.8.1 a 4.8.4 e 4.9 indicam o domínio das variáveis. Por sua vez as restrições 4.10 são restrições de fronteira que, considerando a informação do último dia do plano anterior, fixam as variáveis do dia 1 no valor 0 nas situações em que a afectação do enfermeiro ao turno constituiria uma inadmissibilidade.

5. Plano de Escalas Óptimo e Análise

A programação necessária para a resolução do problema em Programação Linear Inteira Mista, pelo CPLEX (IBM ILOG, 2010), foi feita em C++, tendo sido utilizado o compilador wx Dev-C++ obtido de forma livre através da internet (GNU, 2007).

Para testar o modelo utilizaram-se os dados do Serviço de Cirurgia em estudo referentes a Setembro/Outubro de 2009. A experiência foi realizada para o mês de Outubro para poder comparar com o plano de escalas fornecido para esse mesmo período de tempo pela Enfermeira Chefe.

Do Plano Ideal para Outubro deduzem-se as férias e as situações de atestado que ocorrerão no período a que se refere. Só assim se poderá obter uma representação real das disponibilidades de enfermeiros para aquele período e determinar uma afectação mais próxima possível à ideal. Para isso são fixadas a zero as variáveis que correspondem a afectação de trabalho aos enfermeiros que estão de férias.

A instância do problema de Programação Linear Inteira Mista resultante tem 5964 variáveis e 6258 restrições, o CPLEX obteve a solução óptima em 0,38 segundos. Esta solução corresponde ao Plano de Escalas Óptimo e é apresentada em Anexo 1. O valor óptimo da função objectivo é 143. Isto significa que no Plano de Escalas Óptimo existem 143 turnos que não coincidem com aqueles que seriam considerados ideais.

Estes resultados são considerados muito satisfatórios, quer quanto à rapidez de obtenção de um Plano de Escalas Óptimo, quer quanto ao valor de dissemelhança, isto é, 143 diferenças num máximo possível de 1176 é uma solução de muito boa qualidade. Note-se que, no universo de 42 enfermeiros a escalonar num período de 28 dias, poderia obter-se uma

dissemelhança de 1176, se o Plano de Escalas não coincidisse em nenhum dia e em nenhum turno com o Plano Ideal.

Quando comparado o Plano Realizado em Outubro com o Plano Ideal a dissemelhança que se obtém é de 184 turnos, um pior resultado do que o Plano de Escalas Óptimo. Quanto ao tempo de elaboração de um Plano de Escalas manualmente, segundo informação da Enfermeira Chefe, nunca é menor do que 8 horas.

Esta primeira análise permite inferir que o Plano de Escalas Óptimo é melhor do que o Plano Realizado pela Enfermeira Chefe do Serviço, aproximando-se mais do Plano Ideal.

No entanto, pode comparar-se o Plano de Escalas Óptimo e o Plano Realizado em Outubro (Anexo 2) de outros pontos de vista. Com efeito, ao contabilizar o número de enfermeiros das quatro equipas afectos a cada turno de trabalho por dia num e noutro plano, expresso nas tabelas 3 e 4, para cada um dos planos, verifica-se que o Plano de Escalas Óptimo satisfaz as necessidades de enfermeiros por turno que foram apresentadas na Tabela 2, enquanto no Plano Realizado em Outubro essas restrições são violadas.

Tabela 3: Nº de enfermeiros por turno e por dia do período no Plano de Escalas Óptimo

M	12	12	13	12	13	8	9	12	12	13	12	13	9	8	12	12	13	12	13	8	8	12	12	13	12	13	8	8
T	7	8	8	8	8	7	7	7	8	8	8	8	7	7	8	8	8	8	8	8	7	8	8	8	8	9	8	7
N	6	7	7	7	6	7	6	6	6	7	7	6	6	7	6	6	6	7	7	6	6	6	6	6	6	7	6	7

Tabela 4: Número de enfermeiros por turno e por dia do período no Plano de Escalas Realizado em Outubro

M	12	10	11	11	13	8	8	12	13	16	11	9	8	8	8	8	8	9	10	8	9	11	5	8	7	9	7	8
T	9	9	9	9	11	7	8	8	9	9	10	9	7	7	9	9	8	8	9	8	7	7	9	10	7	9	8	8
N	7	8	8	7	7	7	7	6	5	3	7	8	7	7	7	7	7	8	8	6	6	7	9	7	8	8	8	7

Por outro lado, o valor das variáveis de desvio $\Delta_{-}^{ve}, \Delta_{+}^{ve}$ que representam respectivamente o desvio por defeito e por excesso entre o número de horas efectivamente trabalhadas no período e o que corresponde à carga horária por período do enfermeiro *v* da equipa *e,* pode-se concluir que, no caso do Plano de Escalas Óptimo, há um maior equilíbrio na distribuição entre enfermeiros destes desvios quando comparados com os registados no Plano de Escalas Realizado em Outubro. No Plano de Escalas Óptimo o valor das variáveis de desvio nunca excede as 8 horas, o que é preferível aos desvios de 10, 12, 14, 17 e mesmo 19 horas observáveis no Plano Realizado em Outubro. Desvios tão grandes tornam-se gradualmente mais difíceis de gerir nos planos de escalas seguintes, para além de difíceis de compensar pelas enfermeiras. Pode afirmar-se que o Plano de Escalas Óptimo é preferível ao Plano de

Escalas Realizado em Outubro e o modelo de PLIM proposto constitui uma boa formulação do problema apresentado.

De notar que o presente modelo não visa resolver situações inesperadas de faltas que obriguem ao reescalonamento do presente plano. No entanto, ele é facilmente adaptável a tais situações.

6. Conclusão

O objectivo deste trabalho foi a aplicação de técnicas de Investigação Operacional ao escalonamento de enfermeiros num Serviço de Cirurgia num Hospital Português.

Após análise aprofundada da situação no referido Serviço, da situação laboral dos enfermeiros e dos requisitos em cuidados de saúde por dia e por turno, propôs-se uma formulação em Programação Linear Inteira Mista para produzir planos de escalas que satisfizessem a legislação em vigor, as condições impostas internamente pelo Hospital e, na medida do possível, as preferências dos enfermeiros por certas sequências de turnos.

A análise e comparação do Plano de Escalas Óptimo obtido através do CPLEX e do Plano de Escalas Realizado permite concluir que o modelo de PLIM representa bem o problema de elaboração de planos de escalas em estudo. Nesta experiência em poucos segundos obteve-se um plano. É importante notar que para validar o modelo serão ainda necessários mais testes computacionais com dados reais de outras instâncias, mas a vantagem desta metodologia a ser validada contribuirá para a facilidade com que a Enfermeira Chefe pode testar cenários e gerir os saldos individuais de horas e satisfazer tanto quanto possível as preferências dos enfermeiros.

7. Referências

Burke, E.K., P.D. Causmaecker, G.V. Berghe and H.V. Landeghem (2004). The state of the art of nurse rostering. *Journal of Scheduling*, 7, 441-499.

Bilgin, B., P. De Causmaecker, G. Vanden Berghe (2011).Modelling Issues in Nurse Rostering. *Annals of Operations Research*. (accepted for publication) https://lirias.kuleuven.be/ bitstream/ 123456789/285229/1/ModellingIssues.pdf.

GNU General Public License Copyright (C) (2007) http://wxdsgn.sourceforge.net/.

Howell, J.P. (1966). Cyclical scheduling of nursing personnel. *Hospitals*, 40: 77-85.

IBM ILOG CPLEX (2010) http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/.

Taha, H.A. (2006) *Operations Research: An Introduction*, Prentice Hall; 8th. Ed,Upper Saddle River.

Agradecimentos

As autoras agradecem aos revisores anónimos, as correcções e valiosas sugestões.

Anexo 1 – Plano de Escalas Óptimo:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Férias	Horário	Desvio	Horas	Delta
A1	D	N	N	M	M	D	D	F	F	F	F	F	F	F	M	M	M	M	M	F	F	M	M	M	D	N	N	F	43,0	42	-3	120	8
A2	F	F	F	F	F	F	F	F	F	F	F	F	F	F	T	T	D	N	N	F	D	M	T	T	D	N	N	F	87,0	40	-9,2	78	4,2
A3	D	M	M	F	M	M	T	T	D	F	M	F	M	M	F	M	D	F	N	F	M	M	T	T	D	M	D	F	7,5	32	-3,8	128	-3,7
A4	D	N	M	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	M	F	M	M	T	T	D	M	D	F	-	40	7,1	157	-4,1
A5	M	M	M	F	M	D	T	T	D	N	N	F	M	M	F	M	D	N	N	F	M	M	T	T	D	N	N	F	-	40	4,1	162	-6,1
A6	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	D	F	F	F	F	F	F	F	43,0	40	-2,7	122	-2,3
A7	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	M	D	F	-	40	3,9	161	-4,9
A8	D	N	N	F	M	M	D	M	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	-	40	-1,9	166	-4,1
A9	D	F	N	F	M	M	T	T	D	M	M	F	M	M	T	T	D	F	M	F	M	M	T	T	D	N	D	F	-	33	-0,4	137	-4,6
A10	D	N	N	F	M	M	D	M	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	-	40	-8,1	166	2,1
A11	D	N	N	F	M	M	T	T	D	N	N	F	M	D	T	T	D	M	M	F	M	M	M	M	M	D	N	N	F	-	40	3,1	162	-5,1
B1	N	F	M	M	T	T	D	N	F	F	M	M	T	T	D	N	N	F	F	F	F	F	F	F	F	F	F	F	67,0	42	-1	102	0
B2	F	F	M	M	T	F	D	M	F	F	M	M	T	T	D	F	N	F	M	M	T	T	D	M	F	F	M	M	-	32	5	127,5	-4,5
B3	N	F	M	M	T	T	D	N	N	F	M	F	F	F	F	N	N	F	M	M	T	T	D	N	N	F	M	F	23,0	40	-3,2	148	-7,8
B4	M	F	M	M	T	T	D	N	N	F	M	M	T	F	D	N	M	F	M	M	T	T	D	N	N	F	M	M	-	40	0,4	167,5	-7,9
B5	N	F	M	M	T	T	D	F	N	F	M	M	T	T	D	N	N	F	M	M	F	T	D	N	N	F	M	M	-	40	3,3	161,5	-4,8
B6	N	F	M	M	T	T	D	N	N	F	M	M	F	T	D	N	N	F	M	M	T	T	D	M	M	F	M	M	-	40	0,2	167,5	-7,7
B7	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	160,0	40	5,7	0	-5,7
B8	M	F	M	M	T	T	D	N	N	F	M	M	F	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	-	40	-1,5	169,5	-8
B9	N	F	M	M	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	M	M	T	T	D	N	N	F	M	M	66,0	40	5,3	93	-4,3
B10	N	F	M	M	T	T	D	N	N	F	M	M	F	T	D	M	M	F	M	M	T	T	D	N	N	F	M	M	-	40	-0,3	167,5	-7,2
C1	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	F	T	T	D	N	N	F	M	M	T	T	D	N	-	42	-2,3	170	0,3
C2	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	D	F	M	M	T	T	D	N	-	42	-0,5	168	0,5
C3	F	F	F	F	F	D	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	132,0	33	-0,5	0	0,5
C4	F	M	D	M	M	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	D	F	M	M	M	M	T	D	N	7,5	40	-1,5	157,5	-3,5
C5	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	D	N	F	M	M	T	T	D	N	-	40	-3,8	168	-4,2
C6	T	T	D	M	M	F	M	M	T	T	D	M	D	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	92,0	40	2,8	70	-4,8
C7	T	T	D	N	N	F	M	M	T	T	D	M	T	F	M	M	T	T	D	N	D	F	M	M	T	T	D	M	-	40	5,2	161,5	-6,7
C8	T	T	D	N	M	F	M	M	M	M	D	N	N	F	M	M	F	T	D	N	N	F	M	M	T	T	D	N	-	40	-3,1	169,5	-6,4
C9	F	F	F	F	F	F	F	F	T	T	D	M	T	F	M	M	T	T	D	D	N	F	M	M	T	T	D	N	51,0	40	-3,6	112,5	0,1
C10	T	T	D	N	N	F	M	M	T	T	D	M	T	F	M	M	T	F	D	N	N	F	M	M	T	T	D	N	-	40	-2	166	-4
C11	M	T	D	N	N	F	M	M	M	M	D	N	N	F	M	M	T	T	D	D	N	F	M	M	T	T	D	D	-	40	4,1	159,5	-3,6
D1	M	M	T	T	D	F	N	F	M	M	T	T	D	N	M	F	M	M	T	T	D	N	F	F	M	M	T	T	-	42	-0,2	162	6,2
D2	M	M	F	T	D	N	F	F	M	M	T	T	D	F	M	F	M	M	F	F	D	M	F	F	M	M	T	T	-	31	1,4	127,5	-4,9
D3	M	M	T	F	D	N	N	F	M	M	M	M	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	F	-	40	-2,6	170	-7,4
D4	M	M	T	F	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	F	T	-	40	-3,1	169	-5,9
D5	M	M	T	T	D	F	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	F	F	M	M	T	T	-	40	2,3	164	-6,3
D6	M	M	T	T	D	N	F	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	F	-	40	-2,8	166,5	-3,7
D7	M	M	T	T	D	F	N	F	M	M	M	M	D	N	N	F	M	M	T	T	D	N	N	F	M	M	F	T	-	40	0,3	167,5	-7,8
D8	F	F	M	T	D	N	N	F	M	M	T	T	D	T	F	F	M	M	T	T	D	M	M	F	M	M	T	F	16,0	40	-0,4	144,5	-0,1
D9	F	M	T	T	D	N	F	F	M	M	T	T	D	F	F	F	M	M	T	F	D	F	N	F	M	M	T	T	-	31	1,5	128,5	-6
D10	M	M	T	T	D	N	F	F	M	M	T	T	D	N	N	F	M	M	M	T	D	F	N	F	M	M	T	T	-	40	2	164,5	-6,5

Anexo 2 – Plano de Escalas Realizado em Outubro:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	Férias	Horário	Desvio	Horas	Delta
A1	D	N	N	F	M	F	F	F	F	F	F	F	F	T	T	D	N	N	F	M	M	T	T	D	N	N	F		43,0	42	-3	114	14
A2	F	F	F	F	F	F	F	F	F	F	F	F	F	T	T	D	N	N	F	M	M	T	T	D	N	N	F		87,0	40	-9,2	86	-3,8
A3	M	F	F	F	M	M	T	T	F	M	M	F	F	M	F	T	F	F	M	F	M	M	T	T	F	T	F	F	7,5	32	-3,8	117	7,3
A4	D	N	N	F	M	M	T	T	D	M	N	F	M	M	T	T	F	F	F	F	M	M	T	T	D	N	N	F	0,0	40	7,1	151	1,9
A5	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	F	F	T	T	D	N	N	F	0,0	40	4,1	157	-1,1
A6	D	N	N	F	M	M	T	T	D	M	N	F	M	M	T	T	D	N	N	F	F	F	F	F	F	F	F	F	43,0	40	-2,7	120	-0,3
A7	D	N	N	F	M	M	T	T	D	M	N	F	M	M	T	F	D	N	N	F	M	M	T	T	D	N	N	F	0,0	40	3,9	163,5	-7,4
A8	M	F	F	M	T	M	T	M	F	D	N	M	M	M	F	F	D	N	N	F	M	M	F	T	D	N	N	F	0,0	40	-1,9	144,5	17,4
A9	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	F	T	F	F	M	M	T	T	F	F	F	F	0,0	33	-0,4	140,5	-8,1
A10	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	0,0	40	-8,1	173	-4,9
A11	D	N	N	F	M	F	F	M	M	M	M	M	F	F	T	T	D	N	N	F	M	M	T	T	D	N	N	F	0,0	40	3,1	154	2,9
B1	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	F	F	F	F	F	F	F	F	F	F	67,0	42	-1	112	-10
B2	T	F	M	T	T	T	F	F	F	F	M	M	T	T	F	F	F	F	M	M	F	M	F	M	F	F	M	M	-	32	5	117	6
B3	N	F	M	M	T	T	D	N	N	F	M	F	F	F	F	D	N	F	M	M	T	T	D	N	N	F	M	M	23,0	40	-3,2	146	-5,8
B4	N	F	M	M	T	T	D	N	N	F	M	F	T	T	D	N	N	F	F	M	T	T	D	N	N	F	M	M	-	40	0,4	163	-3,4
B5	M	M	M	M	T	F	F	M	M	M	M	M	F	F	D	N	N	F	M	M	T	T	D	N	N	F	M	M	-	40	3,3	166,5	-9,8
B6	N	F	F	M	T	T	D	N	M	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	-	40	0,2	169	-9,2
B7	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	160,0	40	5,7	0	-5,7
B8	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	F	M	-	40	-1,5	171	-9,5
B9	N	F	M	M	F	F	F	F	F	F	F	F	F	F	F	F	F	F	M	M	T	T	D	N	N	F	M	M	66,0	40	5,3	93	-4,3
B10	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	F	T	T	D	N	N	F	M	M	-	40	-0,3	171	-10,7
C1	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	F	T	T	D	N	N	F	M	M	T	T	D	N	-	42	-2,3	170	0,3
C2	T	T	M	M	M	F	M	M	M	F	M	M	M	F	M	M	F	M	M	M	F	M	M	M	M	N	M	F	F	-	42	-0,5	169	-0,5
C3	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	132,0	33	-0,5	0	0,5
C4	F	T	D	N	N	F	M	T	T	T	D	N	N	F	M	M	T	T	D	N	F	F	N	M	T	T	D	N	7,5	40	-1,5	162	-8
C5	T	T	D	N	M	F	M	M	T	T	D	N	N	F	M	M	T	T	M	F	F	F	N	M	T	T	D	N	-	40	-3,8	166	-2,2
C6	T	T	M	M	M	F	F	M	T	T	T	D	N	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	92,0	40	2,8	79,5	-14,3
C7	T	T	D	N	N	F	M	M	T	T	D	N	F	F	M	M	T	T	D	N	N	F	M	F	F	T	F	F	-	40	5,2	142,5	12,3
C8	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	T	F	D	N	N	F	M	M	T	T	D	N		-	40	-3,1	162,5	0,6
C9	F	F	F	F	F	F	F	F	T	T	D	N	N	F	M	M	T	T	D	N	N	F	F	M	T	T	D	N	51,0	40	-3,6	119	-6,4
C10	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	F	F	T	T	D	N	-	40	-2	162	0
C11	T	T	D	N	N	F	M	M	T	T	D	N	F	F	M	M	T	T	F	D	N	F	M	M	T	T	D	N	-	40	4,1	158	-2,1
D1	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	F	F	M	M	T	T	-	42	-0,2	174	-5,8
D2	M	M	T	F	M	F	F	M	M	F	F	T	T	F	F	M	M	T	T	F	D	N	F	M	M	M	F	F	-	31	1,4	119,5	3,1
D3	M	M	T	T	T	D	N	M	M	M	T	T	D	N	N	F	M	M	T	F	D	N	N	F	M	M	T	T	-	40	-2,6	182	-19,4
D4	M	M	T	T	T	F	F	F	M	M	T	T	D	N	N	F	F	M	T	T	D	N	N	F	F	M	T	T	-	40	-3,1	155,5	7,6
D5	M	M	T	T	D	N	N	F	M	M	T	T	F	D	N	F	M	M	T	T	D	N	N	F	F	M	F	T	-	40	2,3	158,5	-0,8
D6	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	F	F	T	T	-	40	-2,8	168	-5,2
D7	M	M	M	D	N	N	F	F	M	M	T	T	D	N	F	F	M	M	T	T	D	N	F	F	M	M	T	T	-	40	0,3	157	2,7
D8	F	F	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	M	F	F	D	N	N	F	M	M	T	T	16,0	40	-0,4	153	-8,6
D9	M	M	T	T	D	N	N	F	F	M	T	T	F	F	D	N	F	F	T	T	F	F	F	F	M	M	T	F	-	31	1,5	122,5	0
D10	M	M	T	T	D	N	N	F	M	M	T	T	D	N	N	F	M	F	T	T	F	D	N	F	M	F	T	T	-	40	2	158	0

Countries environmental performance assessment

A. Zanella [a*], *A. S. Camanho.* [b], *T. G. Dias* [c]

[a] *Faculdade de Engenharia da Universidade do Porto, Porto, Portugal*
andreia.zanella@fe.up.pt

[b] *Faculdade de Engenharia da Universidade do Porto, Porto, Portugal*
acamanho@fe.up.pt

[c] *Faculdade de Engenharia da Universidade do Porto, Porto, Portugal*
tgalvao@fe.up.pt

[*] *Corresponding Author*

Abstract

Environmental performance assessments are often conducted using environmental indicators. Although these indicators provide a starting point for performance assessments, they do not provide guidelines that countries should follow to improve performance. This paper develops an enhanced DEA model that provides a single summary measure of countries environmental performance, based on the aggregation of the indicators that underlie the estimation of the Environmental Performance Index (EPI). The main contribution of the methodology used in this paper is to enable benchmarking, such that it becomes possible to identify the strengths and weaknesses of each country, as well as the peers that inefficient countries should follow to learn the best environmental practices.

Keywords: environmental performance, Data Envelopment Analysis, benchmarking

1- Introduction

The environmental concerns have increased dramatically in the past few years and are now among the most serious challenges affecting people's well-being. Besides the climate change that has been widely discussed, other environmental problems such as local air and water pollution, soil erosion, water scarcity, deforestation, and loss of biodiversity are also becoming more serious [World Bank, 2008].

Countries are facing new challenges to control their waste production and to reduce the consumption of natural resources, in order to achieve the environmental targets

imposed by international agreements such as the Kyoto Protocol [United Nations, 1998] or the European Union climate and energy package [European Union, 2008]. In this context, it is imperative that countries become able to monitor their environmental performance in order to understand how they are doing compared to others and to identify the potential for improvements.

Environmental performance assessments are often conducted using environmental indicators that are able to measure the pressures on the environment, to appraise the state of the ecosystem and to evaluate the impacts on human activity resulting from changes in environmental quality. These indicators usually measure particular features of the environment, and provide a starting point for performance assessments. It is also common to use composite indicators to aggregate several individual indicators in a summary measure of performance. Although the indicators, individual or aggregated can be extremely useful to guide discussions on environmental issues and attract public interest, they do not provide guidelines that countries should follow to improve performance. The main contribution of this paper is to develop an enhanced CI that, besides assigning a summary measure of performance for each country, can be used to benchmark countries environmental performance. This procedure enables the identification of peers that countries should look in order to search for the best practices. The information obtained from the identification of best practices can guide decision makers in the definition of better environmental policies to improve the performance of countries.

2- The construction of composite indicators

In the past few years, efforts to assess environmental performance of organizations, cities and countries have generated a large number of indicators related to gas emissions, water quality, green space area, waste production, among others.

Due to the large amount of individual indicators, the process of analyzing and understanding the information is difficult for the decision makers. So, many of these indicators have been aggregated into composite indicators (CI).

A CI is formed when individual indicators are compiled into a single index on the basis of an underlying model. It should measure multidimensional concepts which

cannot be captured by a single indicator. CI have benefits such as the capacity to summarize information, the facility to interpret results compared with a battery of many separate indicators, and the capacity to reduce the visible size of a set of indicators without dropping the underlying base information [Nardo et al., 2008].

According to OCDE report by Nardo et al. [2008] the construction of composite indicators involves several stages: the selection of sub-indicators, the treatment of missing values, the understanding of the data, and the specification the weights for the sub-indicators to be used in the aggregation model.

Concerning the selection of sub-indicators, they should be selected according to their analytical soundness, measurability, coverage, and relevance to the phenomenon being assessed. Regarding to the treatment of missing values, there are two possibilities: to impute values to replace the missing fields or to remove the whole observation with missing data from the analysis. Concerning the understanding of the data, an exploratory analysis should be performed to study the overall structure of the dataset. It is necessary to understand the relationship among indicators and observations, and to identify outliers and extreme values that can introduce bias in the results.

Concerning the specification of weights for the sub-indicators, they can be specified from quantitative methods or from expert judgment. The weights can reflect policy priorities, or reward the factors deemed to be more important to the performance assessment. The choice of weights inevitably impacts the results of the CI, so it should be robust to avoid undermining the credibility of the CI results. Since the specification of the weights is often subject of criticism and disagreement, most composite indicators rely on equal weighting to minimize the subjectivity. One such example is the calculation of the Human Development Index [UNDP, 2000], which is based on the use of equal weights for its component indices, which reflect longevity (measured by life expectancy at birth), education attainment (measured by adult literacy and enrolment rate) and standard of living (measured by GDP per capita).

In the context of environment performance assessments, the Environmental Performance Index (EPI), and the Climate Change Performance Index (CCPI) are

examples of well established composite indicators that aggregate individual indicators in a summary measure.

The EPI provides a global index based in 25 indicators, grouped in 10 categories, covering Environmental Health and Ecosystem Vitality. The EPI requires a predefinition of weights and targets for each indicator. After defining the indicators' weights (mostly resorting to expert opinion) and the proximity-to-target of each country in each indicator, a weighted average is calculated to obtain the global performance index for each country [Emerson et al., 2010].

The CCPI is measured via 12 different indicators, which can be classified in the following categories: emissions trend, emissions level and climate policy. This index compares countries that together are responsible for more than 90% of global energy-related CO_2 emissions. Similarly to the EPI, the CCPI needs the pre-definition of weights for the indicators. The countries ranking is calculated from the weighted average of the scores achieved by the countries evaluated in the indicators considered [Burck et al., 2009].

Both the EPI and CCPI use a weighted average to provide an overall measure of performance, and rely on expert opinion to specify the weights. An alternative to overcome the difficulties in the selection of weights is to use Data Envelopment Analysis (DEA) to determine the weights. Using DEA, the individual indicators weights result from an optimizing process, based on linear programming, so they are less subjective and thus less prone to controversy.

Although the DEA technique has been used to construct CI in different fields, such as the evaluation of urban quality of life [Morais and Camanho, 2011] and human development [Despotis, 2005], its application for environmental performance assessment is scarce. This paper contributes to the literature in this field by proposing a methodology to evaluate countries environmental performance based on the construction of a composite indicator whose weights are specified using DEA.

3- Data used

In order to assess countries environmental performance we will use the indicators that underlie the estimation of the EPI. According to Emerson et al. [2010], the indicators used in the EPI reflect state-of-the-art data and the best current thinking in environmental health and ecological science. Some of the indicators represent direct measures and others are proxy measures that give an indication of policy progress by tracking a correlated variable.

The EPI provides data of 163 countries in 25 environmental indicators grouped in 10 policy categories, as showed in Table 1. The indicators cover two core objectives of environmental policy: Environmental Health and Ecosystem Vitality. The first core objective measures the environmental effects to human health, whereas the second measures the state of the ecosystem and the natural resources management.

Concerning the EPI indicators presented in Table 1, the first three policy categories are related with environmental public health objective, while the last seven with the ecosystem vitality. Each of these two major objectives contributes with a weight of 50% to the overall score of EPI.

The construction of EPI requires a predefinition of weights and targets for the indicators. For each country and each indicator, a proximity-to-target value was calculated based on the gap between the current results presented by each country and the target previously identified. The targets were defined based on four sources: treaties or other internationally agreed goals, standards defined by international organizations, national regulatory requirements, or expert judgment.

The weights assigned to the indicators were determined through expert judgment and also, for some indicators, using principal component analysis (PCA). Table 1 shows the weights set to the EPI indicators. After defining the indicators' weights and the proximity-to-target of each country, a weighted average is calculated to obtain the EPI score for each country [Emerson et al., 2010].

In this study, the environmental performance evaluation is based on the values of each indicator for the countries assessed. The methodology used does not require

the use of a proximity-to-target measure. In addition, the model used enables benchmarking, as for each country considered inefficient in the environmental assessment it can be found a set of peer countries with better environmental performance.

Table 1 – EPI indicators

Policy Categories	Indicators	Weight (p_r)	Mean	SD	Min	Max
Environmental burden of disease	(1) Disability Life Adjusted Years	25%	58.80	53.42	13.00	231.34
Water (effects on humans)	(2) Access to adequate sanitation	6.3%	70.61	29.34	5.00	100.00
	(3) Access to drinking water	6.3%	84.97	16.89	40.00	100.00
Air pollution (effects on humans)	(4) Indoor air pollution	6.3%	35.29	35.35	5.00	95.00
	(5) Outdoor air pollution - Urban Particulates	6.3%	46.25	31.97	6.48	145.91
Air Pollution (effects on ecosystem)	(6) Sulfur dioxide emissions	2.1%	47796956.2	146554836.3	0	861361889.9
	(7) Nitrogen oxides emissions	0.7%	62.05	19.95	23.93	100.00
	(8) Non-methane volatile organic compound emissions	0.7%	3.94	11.64	0	94.22
	(9) Ozone Exceedance	0.7%	3.62	9.21	0.02	76.86
Water (effects on ecosystem)	(10) Water quality index	2.1%	4.45	9.00	0.01	57.39
	(11) Water stress index	1.0%	0.19	0.79	0	4.87
	(12) Water scarcity index	1.0%	13.60	17.44	0	68.55
Biodiversity & Habitat	(13) Biome protection	2.1%	6.60	3.44	0	10.00
	(14) Critical habitat protection	1.0%	46.93	32.94	0	100.00
	(15) Marine protection	1.0%	1.51	3.21	0	16.72
Forestry	(16) Growing stock change	2.1%	1.01	0.09	0.65	1.48
	(17) Forest cover change	2.1%	-0.17	1.25	-4.22	3.92
Fisheries	(18) Marine trophic index	2.1%	0.00	0.01	-0.04	0.04
	(19) Trawling intensity	2.1%	39.94	32.23	0.04	100.00
Agriculture	(20) Agricultural water intensity	0.8%	25.74	70.82	0	398.91
	(21) Agricultural subsidies	1.3%	0.06	0.12	0	0.55
	(22) Pesticide regulation	2.1%	13.06	8.61	0	22.00
Climate Change	(23) Greenhouse gas emissions per capita	12.5%	10.05	9.58	0	45.89
	(24) CO2 emissions per electricity generation	6.3%	76.27	68.42	0	290.86
	(25) Industrial greenhouse gas emissions intensity	6.3%	487.82	273.66	0.01	1348.91

4- Methodology

This section describes the Data Envelopment Analysis (DEA) model used to assess environmental performance. DEA is a linear programming technique that measures the relative efficiency of a homogeneous set of Decision Making Units (DMUs), in their use of multiple inputs to produce multiple outputs. It derives a single summary measure of efficiency for each DMU, based on direct comparisons with other DMUs in the sample. This feature makes DEA particularly interesting for benchmarking purposes, as it evaluates performance by comparison with what was actually observed. In addition, DEA is able to handle data measured in different measurement units, without requiring prior normalisation of measurement scales. This is a particularly interesting feature in the context of environmental performance assessments, as the indicators cannot be converted to common units, such as monetary values.

The output oriented model introduced by Charnes et al. [1978], assuming constant returns to scale, can be formulated as follows:

$$\text{Max } h_{j_0} = \delta \qquad (1)$$

$$\text{s.t. } \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{ij_0} \qquad i = 1,...,m$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \ge \delta y_{rj_0} \qquad r = 1,...,s$$

$$\lambda_j \ge 0\,, \forall j$$

The formulation above is known as the envelopment formulation of the DEA model. x_{ij} ($i = 1,...,m$) are the inputs consumed by the DMU j ($j = 1,...,n$) to produce the outputs y_{rj} ($r = 1,...,s$). The variables λ_j ($j = 1,...,m$) identify the peers for the DMU under assessment. A value of $\lambda_j > 0$ means that DMU j is efficient and a benchmark for the DMU under assessment. For each DMU j_0 under assessment, model (1) seeks to identify a comparator, i.e., a composite DMU corresponding to a linear combination of efficient DMUs $\left(\sum_{j=1}^{n} \lambda_j^* x_{ij}, \sum_{j=1}^{n} \lambda_j^* y_{rj} \right)$, that dominates DMU j_0 in all input and output dimensions. This composite DMU corresponds to the targets that the DMU under assessment should try to achieve in order to operate efficiently. The objective function value at the optimal solution corresponds to the factor (δ^*) by which all outputs of DMU under assessment can be proportionally increased to reach the target output values. The efficiency score of DMU j_0 under assessment is the reciprocal of this value. Therefore, the DMUs considered efficient are those for which there is no evidence that it is possible to expand their outputs, such that the value of δ^* is equal to 1.

In this research, model (1) was adapted to the specificities imposed by an environmental performance assessment. One important assumption of the DEA model is that higher output values correspond to better performance. Since some of the indicators used in the EPI do not satisfy this assumption, we had to modify the measurement scale of these variables. The indicators negatively related to environmental performance (e.g., ozone emissions), were transformed into positive aspects by subtracting from a large number, as suggested by Dyson et al. [2001]. In

the empirical assessment of countries environmental performance reported in the next section, we subtracted these non-isotonic outputs from the largest value observed in the sample for the corresponding output.

Another particularity that had to be dealt with concerned the use of output variables measured as ratios or indexes. The assessment of environmental performance based on the EPI indicators implied a direct comparison of countries without explicitly taking into account any inputs. This is equivalent to assume that all DMUs face similar conditions, which can be translated in the mathematical formulation of the DEA model by considering identical input levels for all DMUs. As a result, the input restriction of the DEA model only needs to impose that the composite DMU against which a DMU is assessed must also have an input level identical to all other DMUs. This can be stated as follows: $\sum_{j=1}^{n} \lambda_j x_i = x_{ij_0} \Leftrightarrow \sum_{j=1}^{n} \lambda_j = \frac{x_{ij_0}}{x_i} \Leftrightarrow \sum_{j=1}^{n} \lambda_j = 1.$

Note that although this restriction could also be used to impose variable returns to scale [Banker et al., 1994], in this context it is required due to the use of ratios or percentages as output variables, as noted by Hollingsworth and Smith [2003, p.735].

The dual of model (1), with the input restriction replaced by $\sum_{j=1}^{n} \lambda_j = 1$ is as follows:

$$\text{Min } h_{j_0} = w \qquad (2)$$

$$\text{s.t. } \sum_{r=1}^{s} u_r y_{rj_0} = 1$$

$$\sum_{r=1}^{s} u_r y_{rj} - w \le 0, \qquad \text{j} = 1,...,\text{n}$$

$$u_r \ge 0, \qquad \text{r} = 1,...,\text{s}$$

In the formulation above, u_r are the weights attached to the outputs in the performance assessment, and these are the variables of the model. The variable w corresponds to the weight assigned to the input. In the case of a composite indicator, the input is assumed to be identical for all DMUs. The output weights at the optimal solution can be used to indicate the relative importance of the output variables in determining the efficiency level of the DMU j_0 under assessment. However, as these 'raw' weights depend on the scaling of each output, 'virtual'

outputs weights ($u_r^* y_{rj_0}$) can be used instead. These virtual weights are in fact normalized weights, adding up to one for all DMUs. Therefore, for each output, the value of $u_r^* y_{rj_0}$ represents the relative importance of the output in the performance evaluation of DMU j_0, expressed as a percentage.

The original DEA model was based on the assumption that each DMU should have freedom in selecting the weights that show its performance in the best possible light. The flexibility in the choice of weights, which is a strength of a DEA analysis, may also be a weakness, as it allows some indicators to be assigned a zero weight. This means that these factors are in fact ignored in the performance assessment, such that a DMU may appear to have good performance due to a judicious choice of weights. Therefore, in the context of this paper, we decided to adopt weight restrictions that mimic the value judgements implicit in the EPI, to enhance the discriminatory power of the performance assessment. Since the relative importance of the indicators used in the EPI is expressed as a percentage, we decided to impose virtual weight restrictions in the DEA assessment as follows:

$$\frac{u_r \bar{y}_r}{\sum_{r=1}^{s} u_r \bar{y}_r} = p_r \qquad r = 1, \ldots, s \qquad (3)$$

The value of p_r imposed for each output restriction is identical to the weight assigned to each indicator in the estimation of the EPI, as shown in the Table 1. These weight restrictions allow reflecting the relative importance of each indicator, and ensure that all environmental categories are reflected on the performance measure. Note that we did not impose DMU specific weight restrictions, but used the average value of the output indicator instead ($\bar{y}_r$). With this specification, the weights correspond to assurance regions type I (see Allen et al., [1997] for details on the use of weight restrictions in DEA).

5- Results and discussion

The first stage of the performance evaluation consisted on the exploratory analysis of the EPI indicators using descriptive statistics. In this stage, extreme values were

replaced by 3 standard deviations from the mean, as they could bias the performance evaluation.

As argued by Dyson et al. [2001], a common pitfall in the application of DEA arises from simply attempting to compare non-homogenous DMUs, which can be avoided by clustering the DMUs in homogenous sets prior to the assessment. This was the approach we followed, such that the DEA model was applied to evaluate performance within clusters that group countries with similar environmental features.

Therefore, the first stage of the analysis consisted on the use of Cluster Analysis (using the k-means algorithm) to group the countries in relative homogeneous groups, and the application of Decision Trees (using the C4.5 algorithm) to identify the specific features that led the countries to be in the same cluster.

As the k-means algorithm requires the prior definition of the number of clusters, we used the Davies-Bouldin index to identify the best number of clusters to split the countries, and concluded that seven clusters would be the most appropriate. Using this number of partitions we obtained four larger clusters and three clusters with a small number of countries.

The cluster analysis was followed by the construction of a decision tree to reveal the main similarities among the indicators of the countries belonging to the same group. Table 2 shows the main characteristics found in each cluster, as well as the percentage of countries that presented these characteristics.

In order to illustrate the results found by this procedure consider, for example, the second cluster, which mainly includes the Occidental Europe. This cluster presents a rule that cover 81.40% of the countries and is characterized by low levels of Indoor air pollution, Water scarcity, Ozone emissions and high values for Critical habitat protection and Agricultural Subsidies. From the environmental point of view, the last variable is a bad characteristic, because the magnitude of the subsidies are positive correlated with the pressure that they exert. The first four indicators presented good characteristics for the countries of this cluster. The same interpretation was done for the others clusters. The good and bad characteristics are distinctly signed in Table 2.

Table 2 – Characteristics of each cluster

Cluster	Characteristics		% of countries
C1: Majority of Africa, three countries of South-eastern Asia, Papua New Guinea, Nicaragua and Haiti	✘	High Indoor air pollution (> 49.59)	94.44
	✘	High Disability Life Adjusted Years (> 69.50)	
C2: Countries of Occidental Europe, six countries of Latin America, Russia, Japan, South Korea, Malaysia, New Zealand and Canada.	✓	Low Indoor air pollution (≤ 49.59)	81.40
	✓	Low Water scarcity (≤ 1.45)	
	✓	Low Ozone emissions (≤ 646068941.12)	
	✓	High Critical habitat protection (> 40.15).	
	✘	High Agricultural Subsidies (> 0.089)	
C3: Majority of Central Asia and Middle East, some countries of Africa and Latin America, Brunei, Solomon Islands and Australia	✓	Low Indoor air pollution (≤ 49.59)	89.13
	✓	Low Water scarcity index (≤ 1.45)	
	✘	Low Water quality index (≤ 63.23)	
	✘	High CO2 emissions (> 51.84)	
C4: The most part of Asia (Southern, South-eastern and Eastern), seven countries of Latin America, Fiji and Namibia	✘	High Indoor air pollution (> 49.59)	48.15
	✓	Low Disability Life Adjusted Years (≤ 69.50)	
C5: Kuwait, Qatar, Saudi Arabia, United Arab Emirates and Libya	✓	Low Indoor air pollution (≤ 49.59)	100
	✘	High Water Scarcity index (> 1.45)	
	✘	High Agricultural water intensity (> 336.77)	
C6: Bahrain and Singapore	✓	Low Indoor air pollution (≤ 49.59)	100
	✘	Low Water Scarcity index (≤ 1.45)	
	✘	High SO2 emissions (> 60.13)	
	✘	High NOx emissions (> 76.86)	
	✘	High NMVOC emissions (> 57.38)	
C7: Angola, Brazil, Democratic Republic of Congo and United States of America	✓	Low Indoor air pollution (≤ 49.59)	75
	✓	Low Water Scarcity index (≤ 1.45)	
	✘	High Ozone Exceedance (> 646068941.12)	

The last three clusters have few countries that presented different behaviour compared with others. For each of these clusters we identified the features that led then to be isolated in these specific clusters.

Figure 1 shows the co-plot of the countries, identified with different colours according to the clusters they belong. The co-plot enables a two-dimensional graphical representation of observations (countries) and evaluation criteria (indicators) in a simple way. The graph represents the countries as points and indicators as arrows, relative to the same axis and origin. The countries' position in the graph is determined by all indicators simultaneously. This enables obtaining preliminary insights about the similarities among countries, as countries that are close to each other in the graph have similar values of the indicators considered. In addition, countries that are isolated in the graph may correspond to outliers and extreme values, which may need further analysis. The graph presentation of data can also be used to evaluate the representativeness of each indicator in the analysis. Some indicators, corresponding to arrows close to each other in the graph, can provide similar information.

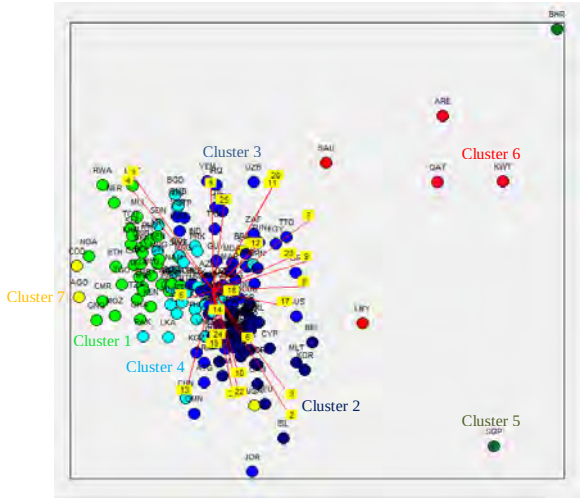


Figure 1 – Countries by cluster

From Figure 1 we can see that the clusters are quite separated in the plot. It is possible to confirm some of characteristics previously identified with the decision tree. For instance, the cluster C1 (that represent mainly the countries from Africa) have high values in the indicators related with the indoor air pollutions and premature death because they are in the direction of these two indicators (1 and 4), confirming the information provided by the Decision Tree. In addition, we can confirm that the countries with outlier behaviour were grouped together in small clusters (cluster C5, C6 e C7), due to the dissimilarities with the remaining countries studied.

The final step of the analysis consisted on the estimation of the performance score. In this section we will only present the results obtained for one of the clusters (cluster C2) to illustrate the potential of approach proposed for evaluating environmental performance. The performance scores obtained using model (1) with restrictions (3) are shown in Table 3.

The best environmental performance of this cluster, assessed taking into account the relative importance of the environmental indicators defined for the EPI, was observed in a country from Central America (Costa Rica). For the remaining countries, the score lower than 100% indicates that there is potential for improvement. For example, in the case of Cyprus, the environmental indicators can all be proportionally increased by a factor of 1.3 (=1/0.782). This number results from a comparison with Costa Rica, which can be considered the benchmark of this

cluster, whose practices can be followed in order to improve their environmental performance of inefficient countries.

Table 3 – Environmental performance scores for coutries of cluster 1

Country	Performance score	Country	Performance score	Country	Performance score
Costa Rica	100.0%	Norway	92.2%	Netherlands	87.6%
Ecuador	98.3%	Japan	92.0%	Slovakia	86.6%
Iceland	98.2%	Switzerland	91.1%	Hungary	86.6%
Colombia	97.5%	Latvia	90.9%	Bulgaria	86.6%
France	97.3%	Finland	90.6%	Poland	86.4%
Germany	96.9%	Denmark	90.2%	Malta	86.4%
Portugal	96.8%	Lithuania	90.0%	Estonia	86.2%
Dominican Republic	95.7%	Romania	89.8%	Greece	86.1%
Sweden	95.2%	Croatia	89.1%	Slovenia	86.0%
Italy	95.2%	Malaysia	88.9%	Czech Republic	85.5%
Spain	94.2%	Canada	88.8%	South Korea	84.2%
United Kingdom	94.2%	Ireland	88.8%	Belgium	84.1%
Cuba	93.9%	Austria	88.6%	Luxembourg	81.3%
New Zealand	92.6%	Russia	88.4%	Cyprus	78.2%
Panama	92.3%				

The ranking presents a satisfactory level of discrimination between the countries, which was possible due to the imposition of weight restrictions in the DEA model.

The methodology used to construct the CI is useful also to analyse the strengths and weaknesses of each country. For this purpose, we decided to relax the weight restrictions, imposing them only at the category level. Furthermore, for each category, we allowed a flexibility of plus and minus 5% around the EPI weight. Through the virtual weights chosen by each country, within the limits allowed, we were able to identify the areas in which countries are specialized and have better environmental performance.

The Figure 2 shows the virtual weights selected by Ireland, which was considered an efficient country allowing for flexibility in the relative importance assigned to each environmental category. The pie chart (a) has the fixed weights imposed to the EPI categories in the countries performance assessment, and the pie chart (b) has the weights chosen by Ireland, given the boundaries of flexibility allowed.

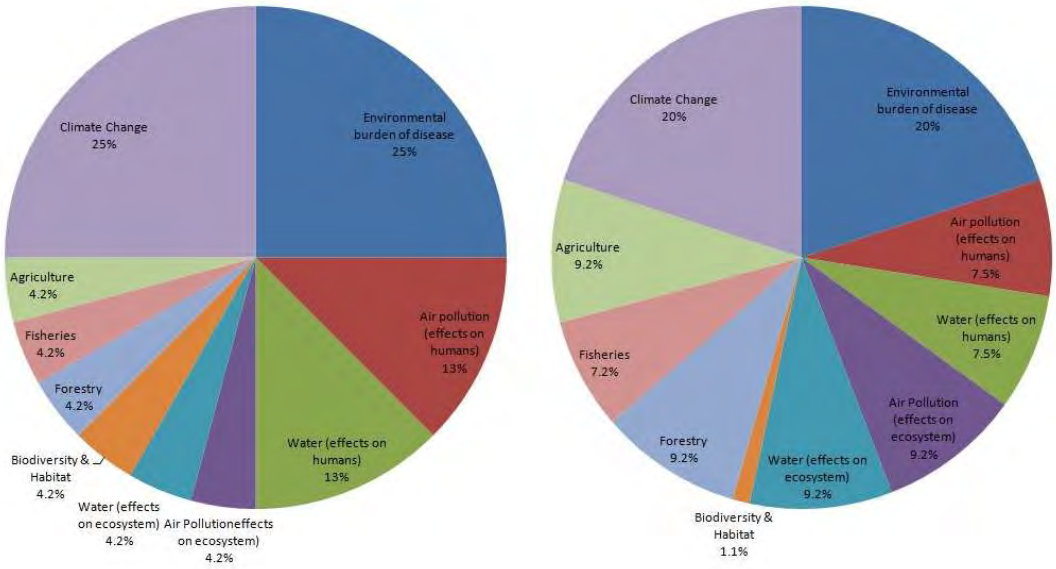


Figure 2 – Original weights for EPI categories (a) and Contributions of the categories to the CI for Ireland (b)

Ireland selected the higher virtual weights in the categories Forestry, Fisheries, Agriculture, Air Pollution (effects on ecosystem) and Water (effects on ecosystem). This is evidence that Ireland is specialised in these categories. On the other hand, Ireland selected lower weights to the others categories: Environmental burden of disease, Biodiversity & Habitat, Air Pollution (effects on humans), Water (effects on humans) and Climate Change.

6- Conclusions

This study provided an enhanced picture of countries environmental performance. It developed a methodology, based on DEA, for the construction of a composite indicator of environmental performance. Besides the possibility of assigning a summary measure of performance for each country, the model described in this paper can be used for benchmarking purposes, as it provides for each inefficient country a set of peers whose environmental practices are examples to be followed. In addition, it is possible to include in the environmental assessment information concerning the relative importance of the different environmental dimensions, allowing each country to select the set of weights that shows it in the best possible light. This enables the identification of the strengths and weaknesses of each

country, which can support decision makers in the definition of better environmental policies to improve the performance of countries.

7- References

Allen R, Athanassopoulos A, Dyson RG, Thanassoulis E. (1997). Weights restrictions and value judgments in data envelopment analysis: evolution, development and future directions. Annals of Operations Research, 73, 13–64.

Banker RD, Charnes A, Cooper WW. (1994). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30, 1078-1092.

Burck, J., Bals, C., and Rossow, V. (2009). Climate change performance index. Technical report, Germanwatch and Climate Action Network Europe.

Charnes A., Cooper W., Rhodes E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research. 2, 429–44.

Cherchye, L., Moesen, W., and Puyenbroeck, T.V. (2004). Legitimately diverse, yet comparable: on synthesizing social inclusion performance in the EU. Journal of common market studies, 42:919-955.

Cherchye, L., Moesen W., Rogge, N. and Puyenbroeck, T. (2007). An introduction to 'benefit of the doubt' composite indicators. Social Indicators Research, 82, 111-145.

Despotis D.K. (2005). Measuring human development via data envelopment analysis: the case of Asia and the Pacific. Omega - The International Journal of Management Science, 33, 385-390.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico C. S., Shale E. A. (2001). Pitfalls and protocols in DEA. European Journal of Operational Research, 132, 245-259.

Emerson, J., Esty, D. C., Levy, M. A., Kim, C. H., Mara, V., Sherbinin, A., and Srebotnjak, T. (2010). 2010 environmental performance index. Technical report, Yale Center for Environmental Law and Policy, New Haven.

European Union (2008). Energy and climate change elements of the final compromise. Technical report, European Union.

Hollingsworth B, Smith P. (2003). Use of ratios in data envelopment analysis. Applied Economics Letters, 10, 733–735.

Morais P. and Camanho A.S. (2011). Evaluation of performance of European cities with the aim to promote quality of life improvements. Omega – International Journal of management science, 39, 4, 398-409.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., and Giovannini, E. (2008). Handbook on Constructing Composite Indicators: Methodology and user guide. Organisation for Economic Co-operation and Development.

United Nations (1998). Kyoto protocol to the united nations framework convention on climate chance. Technical report, United Nations.

UNDP (2000). Human development report 2000, United Nations development program. Technical report, Oxford University Press.

World Bank (2008). Environmental sustainability: An evaluation of world bank group support. Technical report, World Bank, Washington.

World Wildlife Fund (2006). Living planet report 2006. Technical report, World Wildlife Fund.

Problema de definição de rotas na prestação de serviços de exploração de sistemas de saneamento

Carlos Alberto Rodrigues[a], *Maria Helena Correia*[b*]

[a]*Luságua – Serviços Ambientais S.A., Braga, Portugal*
carlosrodrigues@lusagua.pt

[b]*Faculdade de Economia e Gestão – Universidade Católica Portuguesa, Porto, Portugal*
mcorreia@porto.ucp.pt

[*]*Corresponding Author*

Resumo

Este trabalho aborda as questões fundamentais inerentes ao planeamento da prestação de serviços em regime descentralizado, no contexto da exploração de sistemas de saneamento básico. As fases de identificação de pontos de falha, o isolamento do problema, a procura da melhor solução técnica, a implementação em termos operacionais e os resultados em termos de mais-valias constituem as vertentes fundamentais do estudo realizado. O planeamento das actividades de operação e manutenção foi identificado como um dos pontos de falha no desenho do serviço, sendo o planeamento das rotas das equipas operacionais um factor crucial na gestão efectiva do processo de prestação de serviços. A aplicação das técnicas seleccionadas foi orientada por objectivos de aumento da qualidade e capacidade de resposta, associados à diminuição dos custos totais, sem negligenciar a qualidade do serviço prestado.

Os resultados apresentados enquadram-se na operação de uma empresa de serviços responsável pela gestão de contrato de exploração de sistemas de saneamento com tipologia descentralizada, no município de Braga.

O sistema de planeamento desenvolvido permitiu que a organização prestadora de serviços se tornasse capaz de cumprir as premissas assumidas de aumento da qualidade do serviço e capacidade de resposta acompanhada de uma diminuição dos custos totais.

Palavras-chave: Gestão de operações em serviços, logística, problema de rotas de veículos.

1- Introdução

Nos últimos anos tem-se verificado um aumento da subcontratação de determinados serviços por parte de organizações públicas e privadas em regime de o*utsourcing*. Este processo permite às organizações aumentar a eficácia e eficiência da gestão de determinadas operações [Clark, 2008]. Os prestadores de serviço criaram áreas particulares de saberes, e reposicionaram a sua estratégia, de modo a oferecer ao cliente um pacote de serviços especializados para satisfação das suas expectativas e necessidades.

O presente trabalho considera o caso de um cliente público que optou pela subcontratação da execução de um serviço específico - a exploração de estações de elevação e tratamento de águas residuais urbanas -, de entre a globalidade dos serviços desempenhados na sua área de actividade – a gestão municipal de águas e resíduos. Como exploração entende-se, neste contexto, a execução de serviços de operação e manutenção, realizadas nas instalações do cliente. A opção pela estratégia de subcontratação de serviços justifica-se, no presente caso, não só pelo atrás referido, mas também pelo facto de, na maior parte nos casos, os clientes municipais optarem por investir em sistemas de tratamento de águas residuais avançados, com equipamento de processo complexo e específico e, contrariamente ao que seria de esperar, não possuírem recursos humanos dotados de capacidade técnica (muitas vezes, por falta de verba para investimento em formação) para colocar em prática programas operacionais e de manutenção concordantes com o investimento realizado.

Os contratos de serviços de operação e manutenção neste sector, permitem, ao cliente, dispor dos serviços de técnicos dotados de conhecimentos especializados, ficando a cargo do prestador de serviços a responsabilidade de garantir o cumprimento dos requisitos legais do cliente e requisitos nacionais em vigor [Hartman, 2003].

O trabalho desenvolvido tem como enquadramento o desenho de contratos de prestação de serviços e pretende apoiar o planeamento e execução das actividades de operação (tarefas de limpeza, conservação e controlo) e manutenção (tarefas de cariz correctivo e preventivo) de sistemas de saneamento básico, em termos de optimização de percursos de entrega do serviço. Desta forma, pretendeu-se seleccionar os melhores meios que se traduzam, para o gestor de contrato, na redução de custos de exploração, ao nível do

consumo de combustível, desgaste de viaturas e percentagens de alocação de recursos humanos às várias instalações alvo de contrato. Assumindo-se o objectivo de melhoria da performance da prestação de serviços, foram seleccionados métodos quantitativos de apoio à decisão para melhorar/optimizar os percursos de execução dos serviços de exploração.

Nesta secção, pretendeu-se caracterizar genericamente a envolvente do trabalho realizado, descrevendo-se o tipo de contrato de prestação de serviços em causa e definindo-se as directrizes que deram início à análise e formulação do problema em estudo. Na secção seguinte apresenta-se a organização prestadora de serviços, descrevendo-se o seu enquadramento sectorial, e definem-se os contornos reais do problema que se pretendeu resolver. A secção 3 descreve o método de solução adoptado na resolução do problema em causa. Na secção 4, apresentam-se os resultados operacionais alcançados. Por fim, na secção 5, conclui-se com a apresentação das principais conclusões do trabalho e a formulação de propostas para o desenvolvimento de projectos futuros neste sector de prestação de serviços.

2- Descrição do caso empresarial

Quando a entrega de determinada tipologia de serviços a clientes tem como necessidades a deslocação a uma ou mais instalações, alvo de contrato, uma importante consideração a ter no desenho do serviço, é a aplicação/desenvolvimento de modelos de optimização/melhoria dos circuitos de execução dos serviços, de modo a minimizar a distância e/ou o tempo percorrido [Fitzsimmons, 1994]. O transporte é considerado um elemento de extrema importância na cadeia de fornecimento de produtos e serviços. A logística, sobretudo do lado empresarial, relaciona-se estreitamente com o transporte, uma vez que este confere ao produto/serviço/matéria–prima uma mudança posicional, aproximando-o do mercado, por forma a que cada trajecto, quando eficiente e devidamente concebido, gere um ganho efectivo de valor [Carvalho, 2004].

No presente estudo, o serviço a realizar localiza-se no concelho de Braga, sede do distrito homónimo, inserido na região do Minho. Tem uma área de aproximadamente 183,51 Km^2 (dados do INE), distribuída por 62 freguesias. O cliente do serviço é a empresa pública municipal AGERE - Empresa de Águas, Efluentes e Resíduos de

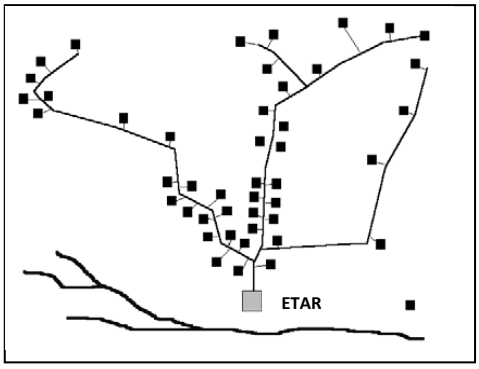
Braga, EM, que surgiu da transformação dos Serviços Municipalizados de Águas e Saneamento em empresa pública municipal, no dia 1 de Janeiro de 1999, com o objectivo de melhor servir os munícipes do concelho.

Em 2008 a EMASISA - Exploração e Manutenção de Sistemas Ambientais Lda, empresa prestadora de serviços da área ambiental, mais concretamente no sector de abastecimento de água e saneamento de águas residuais, arrancou com dois contratos de prestação de serviços no Município de Braga, cujo objectivo consistia, resumidamente, na exploração do sistema de saneamento municipal, constituído por 37 estações elevatórias (EE) e 10 estações de tratamento de águas residuais (ETAR). Cada uma das infra-estruturas de saneamento básico alvo de contrato incorpora um conjunto de equipamentos electromecânicos e processos de tratamento que têm de ser alvo de actividades específicas, operacionais e de manutenção.

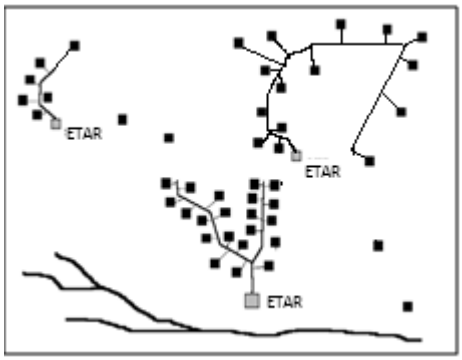
Os serviços de operação considerados no âmbito do contrato englobam a prestação de todos os serviços (actividades e tarefas) necessários e convenientes com vista a assegurar o melhor rendimento possível das instalações. Para realização destas actividades, o prestador de serviços possui um gestor de contrato (engenheiro especialista), um engenheiro coordenador e quatro equipas de operacionais formadas, cada uma, por dois elementos. Os serviços de manutenção abrangem a resolução de problemas inesperados que possam vir a ocorrer nas instalações do cliente (normalmente detectadas pelas equipas operacionais e/ou pelo cliente) e a realização de acções programadas. Deste modo, no caso dos serviços de manutenção, as actividades encontram-se divididas em duas áreas funcionais, com especificidades próprias e requerendo tratamento independente: a manutenção correctiva e a manutenção preventiva. A programação de funções de manutenção correctiva implica a previsão de horas de manutenção a realizar na correcção de avarias imprevistas. No caso da manutenção preventiva, o número de horas é previamente conhecida, com base no *Plano Anual de Manutenção Preventiva* dos equipamentos de todas as instalações, sendo as ordens de trabalho entregues aleatoriamente às equipas operacionais.

No momento da concepção de sistemas de saneamento básico, coloca-se a questão quanto à tipologia de sistema a adoptar para servir um município, ou parte dele. Normalmente, esta decisão implica um grande conhecimento do município alvo,

consistindo fundamentalmente na opção entre uma tipologia centralizada ou descentralizada (Figura 1).



Sistema de Saneamento Centralizado



Sistema de Saneamento Descentralizado

Figura 1 – Tipologias de serviço de recolha e tratamento de águas residuais em regime centralizado (uma ETAR) e descentralizado (três ETAR's).

Existem vários problemas associados à centralização de sistemas de saneamento básico sendo a tipologia descentralizada uma clara alternativa ao sistema convencional, em locais com densidade populacional reduzida.

As características económicas e demográficas do Munícipio de Braga determinaram a adopção da tipologia de sistemas de saneamento de águas residuais em regime descentralizado. Apesar de, em termos de custos de manutenção (reparações/substituições de equipamento), este tipo de sistema apresentar custos baixos, em termos de custos operacionais o mesmo não se verifica, assumindo aqui a dispersão geográfica das instalações a explorar um papel fundamental. A distribuição geográfica das instalações por diferentes zonas do município obriga à elaboração de um planeamento mais complexo da execução dos diferentes trabalhos, mais especificamente, em termos de desenho de circuitos e alocação de recursos para a execução dos serviços de operação e manutenção. Desta forma, o planeamento dos circuitos de execução dos serviços de operação e manutenção torna-se uma fase de extrema importância na gestão efectiva deste processo de prestação de serviços.

Inicialmente, os serviços de exploração eram realizados de acordo com um plano que agrupava as instalações do município em 8 rotas (R_1, R_2, R_3, R_4, R_5, R_6, R_7 e R_8), sendo

estas rotas divididas por 4 equipas operacionais (A, B, C e D). Este plano de exploração foi elaborado com base exclusiva na experiência do gestor operacional e em dados recolhidos no terreno, considerando como base 8 horas de trabalho diário. A sua elaboração considerou as distâncias entre instalações e o tempo médio de percurso entre as mesmas, pressupondo ainda o cumprimento de algumas premissas, parcialmente devidas a requisitos contratualizados com o cliente, nomeadamente, a imposição de visita diária a todas as instalações e a natureza das tarefas operacionais a realizar, com tempos de operação médio diferente. A definição deste plano de exploração terá sido feita de modo estritamente empírico, quer em termos de alocação das instalações e rotas às equipas operacionais, quer em termos de realização das tarefas programadas.

3- Definição do problema e método de solução

Com o trabalho realizado pretendeu-se desenvolver uma ferramenta de apoio à tomada de decisão no âmbito da gestão de contratos de prestação de serviços de operação e manutenção de sistemas de saneamento básico, considerando as características do caso em estudo, descritas na secção anterior. Assim, o problema consiste na determinação de um conjunto de rotas que permita visitar diariamente todas as estações consideradas no contrato de serviços de exploração e executar as tarefas contratadas. Cada rota deverá ser realizada por apenas um veículo, considerando-se que todas as rotas terminam e acabam num mesmo ponto – depósito de veículos. O objectivo assumido na definição destas rotas consiste na minimização da distância percorrida diariamente, garantindo-se o cumprimento dos requisitos referidos – todas as estações são visitadas e todas as tarfeas realizadas. Neste contexto, foram implicitamente assumidos objectivos de minimização dos recursos despendidos na execução do serviço relativos, fundamentalmente, à distância percorrida (Km/ano), consumo combustível (litros/ano), afectação dos colaboradores (% horas.homem) e tempo despendido entre instalações (h/ano) sem descurar a conservação, ou mesmo melhoria da qualidade do serviço prestado ao cliente e indirectamente à população.

Importa salientar que a abordagem apenas foi realizada aos serviços operacionais da organização prestadora de serviços, pois quando falamos nos serviços de manutenção, a complexidade de análise aumenta uma vez que o problema terá que ser decomposto em

dois sub-problemas com características distintas: gestão pontual de avarias (manutenção correctiva) e gestão programada da manutenção de equipamentos.

No processo de desenvolvimento do método de solução podem distinguir-se os 4 estágios sequenciais representados na Figura 2.

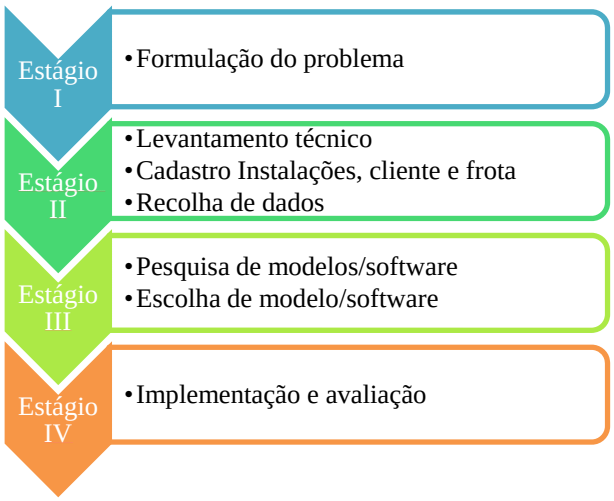


Figura 2 - Metodologia de resolução do problema.

Na execução do Estágio II foi utilizada uma ferramenta de *Global Positioning System* (GPS) que permitiu cadastrar todas as instalações do cliente (alvo do contrato) em termos de georenferenciação no modelo de ortofotomapas. O GPS *Amaryllo*[1] permitiu retirar no terreno as coordenadas das instalações, definir circuitos e planear percursos. A grande vantagem da utilização deste GPS reside na possibilidade de exportação de dados directamente para o software *Google Earth*, através da latitude e longitude de cada nó (coordenadas geográficas descritas em graus, minutos, segundos - *DMS*). Desta forma, foi possível catalogar as quarenta e sete instalações do cliente.

O problema em causa foi identificado com um problema de rotas de veículos (*Vehicle Routing Problem* – VRP) [Vigo, 2002]. O problema de rotas de veículos tem como objectivo determinar um conjunto de rotas que cobrem um conjunto pré-definido de nós, com o objectivo de minimização de custos relativamente aos recursos afectos,

[1] *Amaryllo's Trip Tracker* é um pequeno dispositivo que utiliza a tecnologia de navegação por satélite para monitorar e registar informação sobre uma viagem/deslocação. O equipamento tem um *display* no qual é continuamente monitorizada a velocidade actual, a velocidade máxima, a velocidade média, tempo de deslocação e distância percorrida relativamente ao percurso. http://www.amaryllo.com/almooj/sports-products/amaryllo-trip-tracker-gps-sports-device-6.html).

considerando a localização de um ou mais pontos de origem (centros de distribuição, armazéns, depósitos), de forma a visitar a globalidade dos nós, respeitando restrições duração do percurso e capacidade dos veículos. O problema de rotas de veículos clássico pode ser considerado como uma evolução que o modelo do caixeiro-viajante (*Traveling Salesman Problem* - TSP) sofreu ao longo dos anos, constituindo uma generalização deste último [Bellmore, 1968]. O problema real que se pretendeu resolver exigiu a consideração adicional do tempo de execução do serviço em cada nó, enquadrando-se, desta forma, no caso particular do problema de rotas de veículos com janelas de tempo (*Vehicle Routing Problem with Time Windows*– VRPTW, na terminologia anglo-saxónica).

Os problemas de rotas de veículos são considerados problemas *NP – díficil,* tendo sido desenvolvidas ao longo do tempo várias abordagens à sua resolução, desde técnicas exactas a técnicas heurísticas e metaheurísticas [Cordeau, 2007].

O estudo destes problemas tem vindo a ser alvo de atenção crescente, à medida que a definição das rotas de veículos de forma optimizada ganha cada vez mais protagonismo devido à abertura dos mercados e aumento de competitividade associada. Desta forma, custos relacionados com deslocalização de viaturas começam a ter um peso cada vez maior na globalidade dos custos das organizações prestadoras de serviços em regime descentralizado.

O principal obstáculo associado à resolução dos problemas de rotas de veículos reside na sua elevada complexidade computacional. Para a resolução do problema em em questão foi utilizado o software comercial *Viamente Route Planner*, da VIAMENTE s.r.l.[2] A escolha deste software orientou-se fundamentalmente por critérios de disponibilidade de licença e adequação ao problema a resolver, contemplando um conjunto de premissas de elevada importância para a sua resolução:

- ✓ planeamento e optimização de escalas de serviço;
- ✓ utilização eficiente do transporte e serviços de terreno com uma ideal alocação dos recursos;
- ✓ optimização da gestão de frotas e gestão de serviços de campo; e

[2] Fonte: http://vrp.viamente.com/

- ✓ minimização de custos e recursos, e maximização do seu nível de serviço.

O módulo de optimização do *VRP Planner* utiliza técnicas exactas (b*ranch-and-bound* e *branch-and-cut*) para construir o circuito inicial e, em seguida, utiliza técnicas heurísticas especiais de melhoria dos circuitos baseadas na troca entre os nós a visitar, tendo como objectivo a melhoria do circuito global.

O processo de obtenção da solução consistiu fundamentalmente nas seguintes tarefas:

- Carregamento de dados, efectuado em ficheiro *Excel* facultado no *site* da VIAMENTE, com dados relativos ao tempo de serviço, janela do tempo de trabalho, tempo de deslocação e distância percorrida, referentes à execução dos serviços operacionais;
- Cadastro de dados e sua importação para o *VRP Planner*, com marcação do ponto inicial e final dos percursos e todos os nós e veículos existentes;
- Resolução do problema de optimização dos percursos de acordo com as premissas assumidas;
- Exportação e interpretação de dados, com as diferentes rotas marcadas nos mapas finais a cor distinta e pormenorizadas em modelo de ortofotomapas, subdivididas por diferentes circuitos.

4- Resultados operacionais

O Quadro 1 apresenta os tempos e distâncias relativamente a três fases do processo de resolução do problema: situação inicial, solução computacional e implementação da solução. Esta última fase seria crucial para avaliar a solução em termos de mais-valias de planeamento e execução de operações, tendo-se decidido proceder à sua implementação prática com registo de tempos efectivos de deslocação e de serviço, e distância real percorrida entre cada instalação.

Variável em Estudo	*FASE I – Situação Inicial*	*FASE II – Resultado da Optimização*	*FASE III – Implementação da solução*
N.º de veículos utilizados:	4/4	4/4	4/4
N.º de nós cobertos:	47/47	47/47	47/47
Tempo total de trabalho:	33 hr 55 min	34 hr 23 min	15 hr 59 min
Tempo total de condução:	6 hr 23 min	6 hr 45 min	6 hr 23 min
Tempo total de serviço:	27 hr 32 min	27 hr 38 min	9 hr 36 min
Total de Distância Circuitos:	211 Km	191 Km	184 Km

Quadro 1 - Resumo dos resultados em termos de tempos e distância.

Conforme pode ser observado no quadro, o *VRP Planner* foi capaz de identificar circuitos que acabaram por se tornar mais curtos em termos de distância total percorrida, conseguindo-se uma diminuição em cerca de 10%. Em alternativa ao plano empírico de 8 rotas inicialmente considerado, a nova solução define o agrupamento das 47 estações em apenas 4 rotas. A Figura 3 representa graficamente estas 4 rotas.

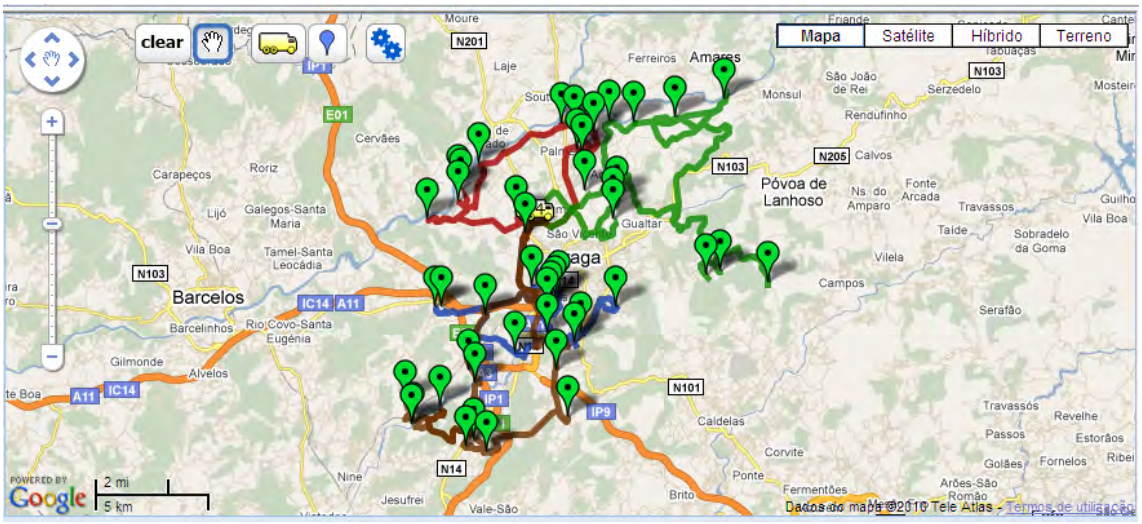


Figura 3 – Representação gráfica da solução.

Note-se, no entanto, que esta redução da distância é acompanhada de um acréscimo do tempo total de condução. Os desvios de sentido inesperado no tempo total de condução, entre a Fase II (resultado da optimização) e as Fase I e III justificam-se, em parte, pelo facto da base cartográfica do *Google Maps* não se encontrar actualizada, sendo as equipas operacionais capazes de identificar trajectos alternativos, mais curtos e/ou mais rápidos. Por outro lado, o valor final do tempo de condução irá depender necessariamente da velocidade média assumida. O *VRPPlanner*, à semelhança do que

acontece no *Google Maps,* atribui velocidades médias em função do tipo de itinerário/estradas a percorrer, sendo os valores ajustados após *feed-back* dos clientes. Para o problema em questão, e dada natureza rural dos trajectos, o *VRP Planner* assumiu uma velocidade média de 30 km/h na realizaçção dos percursos, sendo esta velocidade geralmente excedida pelas equipas operacionais.

A justificação da redução do tempo total de serviço entre a Fase II e a Fase III não se relaciona directamente com o objectivo da optimização, mas deve-se ao facto dos tempos de serviço inicialmente associados a cada tipo de instalação (dados recolhidos no campo), não estarem efectivamente a ser cumpridos. Este facto poderá ser resultado da pressão que o gestor de contrato colocou nas equipas operacionais quanto à necessidade de optimizar a prestação de serviços em termos de logística, tendo sido esta informação erradamente interpretada, o que resultou no incumprimento parcial das tarefas planeadas. Acresce a este facto a possibilidade de os mapas de operação estarem sobredimensionados em termos de tempos de execução de tarefas, devendo proceder-se à sua correcção.

5- Conclusões e desenvolvimentos futuros

O trabalho desenvolvido pretendeu demonstrar as potencialidades da logística na gestão de contratos de serviços no contexto do sector de saneamento básico, permitindo apoiar o planeamento das operações com vista à racionalização dos custos directos associados, fundamentalmente, à gestão da frota de veículos e alocação de pessoal. Os resultados obtidos reflectem-se em termos de custo na gestão de recursos, gestão de operações e gestão económico-financeira do contrato.

A técnica de planeamento desenvolvida requer apenas que o prestador de serviços conheça as instalações em termos de execução de tarefas operacionais e a sua georreferenciação, permitindo apoiar decisões relativas à gestão de contrato de prestação de serviços no planeamento de operações futuras.

O estudo desenvolvido constituí uma base para futuros trabalhos numa perspectiva de evolução para uma plataforma global que, utilizando como base um *software* de optimização de rotas, funcione como ferramenta integradora de optimização da globalidade dos serviços, considerando ambas as subáreas de actividade identificadas no

sector de exploração de sistemas de saneamento básico: serviços de operação e manutenção. A gestão informática destas actividades (por exemplo, via web/PDA), implicaria que cada viatura dispusesse de um sistema GPS com a finalidade de permitir o controlo em tempo real do dia-a-dia das equipas, através do conhecimento prévio do plano de instalações a visitar mensalmente, elaborado por georreferenciação em ortofotomapas.

A aplicação da metodologia desenvolvida não se confina ao contexto da exploração de sistemas de saneamento básico adoptado no âmbito deste trabalho, estendendo-se, de forma natural, a toda a actividade de prestação de serviços com características similares à abordada.

6- Referências

Carvalho, J. M. C. d. (2004). *Logística*, Edições Sílabo, Lda. 3.ª Edição, 189-221, ISBN: 972-618-279-4.

Clark, R. J. (2008). *Service Operations Management - Improving Service Delivery, Prentice Hall.* ISBN: 978-1-4058-4732-2.

Fitzsimmons, J. A., Fitzsimmons, M. J. (1994). *Service Management for Competitive Advantage*, MacGraw-Hill International Editions. ISBN: 0-07-021217-1.

Vigo, D., Toth, Paolo (2002). *The Vehicle Routing Problem*, SIAM Monographs on Discrete Mathematics and Applications. SIAM, Philadelphia.

Bellmore, M., Nemhauser, G. L. (1968). *The traveling salesman problem: A survey*. Operations Research. vol. 16, pp. 538-558.

Cordeau, J-F, Laporte, G., Savelsbergh, M.W.P., Vigo, D. (2007). Vehicle Routing, In Barnhart, C., Laporte, G. (Eds.), *Handbook in OR & MS*. vol.14, pp 367-428.

Hartman, R. (2003). Contracting Water and Wastewater Utility Operations. How-to Guide No. 8, 1-18.

Evolução do desempenho de lojas de retalho ao longo do tempo utilizando o índice de *Malmquist*

Clara Vaz[a*], *Ana Camanho*[b]

[a]*Escola Superior de Tecnologia e Gestão*
Instituto Politécnico de Bragança
Campus Santa Apolónia, 134, 5301-857 Bragança, Portugal
clvaz@ipb.pt

[b]*Faculdade de Engenharia da Universidade do Porto*
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
acamanho@fe.up.pt

[*]*Corresponding Author*

Abstract

Este artigo avalia a evolução do desempenho de lojas de base alimentar de um retalhista Europeu. A técnica de *Data Envelopment Analysis* e o índice de *Malmquist* são combinados com a técnica de *bootstrap* para medir as alterações de produtividade entre os anos de 2002 e 2004, e identificar as alterações consideradas estatisticamente significativas. Na avaliação de desempenho ao longo do tempo há que ter em conta dois efeitos: a variação de eficiência técnica de cada loja e a alteração da fronteira da tecnologia. O índice de *Malmquist* permite captar correctamente estes dois efeitos. A variação da eficiência técnica mede a evolução da capacidade de cada loja em aproximar-se dos melhores níveis observados nas unidades de referência. A alteração da fronteira da tecnologia traduz a evolução verificada nas práticas das melhores lojas, que indicam os melhores níveis de produtividade alcançáveis no período em análise, e o rumo o seguir numa perspectiva de melhoria contínua.

Keywords: *Data Envelopment Analysis, Efficiency, Malmquist index, Bootstrap, retail.*

1- Introdução

A técnica de *Data Envelopment Analysis* (DEA) e o índice de produtividade total de *Malmquist* têm sido largamente usados para a avaliar a evolução de produtividade de organizações, em vários sectores da economia. As suas vantagens prendem-se com o facto de permiterem acomodar tecnologias de produção com múltiplos inputs e outputs e de não necessitarem de informação relativa aos preços dos recursos utilizados e resultados produzidos. A decomposição do índice de produtividade total de *Malmquist*

permite, adicionalmente, obter informações sobre as causas de variação da evolução do desempenho.

No entanto, a técnica de DEA tem sido criticada na literatura pelo facto da sua natureza determinística não permitir acomodar o erro aleatório inerente aos dados, não fornecendo informação sobre o erro associado aos valores de eficiência obtidos pelos modelos. Para ultrapassar este problema, Simar e Wilson [1998, 1999] desenvolveram a técnica de *bootstrap* aplicada aos resultados dos modelos de DEA. A utilização combinada da técnica de DEA, índice de *Malmquist* e *bootstrap* torna possível a caracterização da evolução do desempenho das unidades de decisão (DMUs - *Decision Making Units*) considerando apenas os efeitos estatisticamente significativos. Esta abordagem tem sido efectivamente pouco utilizada na literatura como referiu Odeck (2009). Neste estudo, pretende-se utilizar esta abordagem para avaliar a evolução do desempenho das lojas de uma empresa de retalho (hipermercados e supermercados) entre os anos 2002 e 2004.

A secção seguinte fornece uma breve descrição da empresa utilizada como Caso de Estudo. A secção 3 descreve a metodologia usada na avaliação do desempenho das lojas de retalho ao longo do tempo. Os resultados obtidos e as suas implicações são discutidos na secção 4. Finalmente a secção 5 apresenta as principais conclusões.

2- Caso de Estudo

O sector de retalho em Portugal é dominado por quatro grupos comerciais, que operam num mercado oligopolista caracterizado por uma intensa concorrência. Esta tem sido avivada pelas ofensivas das lojas de desconto, que se diferenciam pelos preços baixos. Neste contexto de intensa concorrência do sector e consequente diminuição de margens, a avaliação e melhoria dos níveis de eficiência é um factor chave na gestão das cadeias de retalho.

A cadeia de retalho analisada neste artigo surge com dois formatos de lojas no mercado Português: hipermercados e supermercados. A operação da loja é gerida a dois níveis, pelo departamento central e pelos responsáveis da loja. O departamento central define a estrutura de funcionamento comum a todas as lojas e controla o seu desempenho. Cada

loja tem por objectivo maximizar as vendas e gerir da melhor forma os recursos disponíveis (área, stock, pessoal e encargos gerais).

A análise de desempenho das lojas é baseada numa política de gestão por objectivos. Os indicadores de desempenho utilizados pela empresa consistem em rácios que comparam os valores efectivamente obtidos por cada loja com o histórico e com os objectivos definidos pelo orçamento. Genericamente, o orçamento da loja define o objectivo relativo aos diversos indicadores de desempenho.

Verifica-se que a avaliação de desempenho realizada pela empresa implica a preparação e análise de múltiplos relatórios relativos aos vários indicadores de desempenho das lojas o que torna a análise do histórico e o processo de orçamentação complexo. Para facilitar e agilizar estes processos, pretende-se desenvolver uma metodologia que permita caracterizar a evolução do desempenho das lojas. Numa primeira fase, pretende-se utilizar a técnica de DEA para definir uma medida de eficiência para cada loja que tenha em conta as várias vertentes do seu desempenho. Numa segunda fase, pretende-se caracterizar o desempenho das lojas ao longo do tempo. Na avaliação de desempenho das lojas ao longo do tempo é necessário distinguir dois efeitos: a variação de eficiência técnica de cada loja e a alteração da fronteira da tecnologia. O índice de *Malmquist* permite captar correctamente estes dois efeitos. A variação da eficiência técnica mede a evolução da capacidade de cada loja em aproximar-se dos melhores níveis observados nas unidades de referência. A alteração da fronteira da tecnologia traduz a evolução verificada nas práticas das melhores lojas, que indicam os melhores níveis de produtividade alcançáveis no período em análise. Esta informação é importante porque permite caracterizar o histórico do desempenho das lojas e apoiar o processo de definição de objectivos para cada loja.

Este estudo avalia o desempenho de 70 lojas que incluem 14 hipermercados (hiper) e 56 supermercados (super).

3-Metodologia

3.1-Modelo de DEA

A técnica de DEA foi utilizada para avaliar a eficiência das lojas. Considere-se o modelo original de Charnes et al [1978], com economias de escala constantes (CRS:

Constant Returns to Scale) e orientação aos outputs, o que decorre do objectivo primordial do retalho, correspondente à maximização das vendas.

Considere-se que n DMUs designadas por $j\ (j = 1, \dots, n)$ utilizam m inputs $x_{ij}\ (x_{1j}, \dots, x_{mj}) \in \mathbb{R}_+^m$ na obtenção de s outputs $y_{rj}\ (y_{1j}, \dots, y_{sj}) \in \mathbb{R}_+^s$. A eficiência relativa da DMU_o pode ser avaliada usando o modelo de programação linear (1):

$$Max\{h_o | x_{io} \geq \sum_{j=1}^{n} \lambda_j x_{ij}, i = 1, \dots, m,\ h_o y_{ro} \leq \sum_{j=1}^{n} \lambda_j y_{rj}, r = 1, \dots, s,\ \lambda_j \geq 0\} \quad (1)$$

O modelo (1) avalia a eficiência técnica (E) da DMU_o que corresponde ao inverso de h_o^*.A solução óptima do modelo (1), h_o^*, corresponde ao factor máximo com que todos os outputs da DMU_o podem ser expandidos tendo em conta os recursos utilizados. A medida de eficiência atingirá o valor um quando a DMU avaliada for considerada eficiente, enquanto valores inferiores indicarão a existência de ineficiências. As unidades com um nível de eficiência igual a um estão situadas na fronteira do conjunto das possibilidades de produção (PPS) identificado pelo modelo de DEA. Uma unidade ineficiente é definida como aquela em que existe evidência de que, utilizando um nível igual ou inferior de todos os inputs, poderia produzir um nível de outputs superior ao que actualmente produz. Note-se que o valor de eficiência técnica pura da DMU_o é calculado incluindo a restrição adicional $\sum_{j=1}^{n} \lambda_j = 1$ no modelo (1). Neste caso, a fronteira do conjunto de possibilidades de produção é caracterizada por rendimentos variáveis à escala (VRS - *Variable Returns to Scale*).

Para corrigir os valores de eficiência tendo em conta o erro aleatório inerente aos dados, utilizou-se a abordagem proposta por Simar e Wilson [1998]. Estes autores propuseram o método de *bootstrap* aplicado à técnica de DEA (*smooth bootstrap*) para proceder à inferência estatística dos valores de eficiência calculados pelos modelos DEA. Assim, é possível estimar para cada DMU, o intervalo de confiança da eficiência, o factor de correcção de eficiência (*bias*) e a eficiência corrigida (*bias-corrected efficiency*). Os valores da eficiência corrigida foram utilizados para a avaliação do desempenho das lojas.

3.2-Índice de *Malmquist*

Para avaliar a evolução da produtividade das lojas entre os períodos t e $t + 1$, usou-se o índice de produtividade total de *Malmquist.* O índice de produtividade total de

Malmquist de Färe et al. [1994a] para uma DMU entre os períodos t e $t+1$ é calculado pela expressão (2).

$$I = \left[\frac{E^t(t+1)}{E^t(t)}\frac{E^{t+1}(t+1)}{E^{t+1}(t)}\right]^{\frac{1}{2}} \quad (2)$$

Em termos de notação, E representa a eficiência técnica de cada DMU avaliada. O valor entre parêntesis associado à eficiencia E representa o período em que a DMU é avaliada. O expoente de E representa a tecnologia de referência utilizada para calcular a eficiência da DMU. Por exemplo, $E^t(t)$ corresponde à eficiência técnica de uma DMU observada no período t calculada em referência à fronteira do período t. $E^t(t)$ é igual ao inverso da solução óptima obtida no modelo (1), isto é, $E^t(t) = 1/h_o^*$. Observe-se que os valores de eficiência das DMUs num dado período calculados com referência a outro período, $E^t(t+1)$ e $E^{t+1}(t)$, podem ter um valor superior a 1. Por exemplo, $E^{t+1}(t)$, será superior a 1 se o vector input-output da DMU do período t estiver fora do conjunto de possibilidades de produção relativo ao período $t+1$.

Em termos de interpretação, valores do índice de produtividade total (I) superiores a 1 correspondem a melhoria do desempenho. O cálculo do índice de *Malmquist* deve ser baseado na tecnologia de rendimentos constantes à escala, podendo estimar-se as variações de produtividade associadas a mudanças de escalas através da decomposição do índice em vários factores, tal como proposto em Färe e Grosskopf [1996].

Färe et al. [1994a,b] propuseram uma decomposição do índice de *Malmquist* (I) em três componentes: (i) IE: índice de mudança da eficiência técnica pura medida segundo rendimentos variáveis à escala, (ii) SE: índice de mudança de eficiência de escala e (iii) IF: índice de mudança de tecnologia, medido segundo rendimentos constantes à escala. A decomposição resultante é apresentada na expressão (3).

$$I = IE \times SE \times IF \quad (3)$$

As fórmulas de cálculo de cada componente, IE, SE e IF são apresentadas nas expressões (4), (5) e (6), respectivamente. As eficiências que estão assinaladas pela notação $|v$ são medidas utilizando como referência a fronteira de rendimentos variáveis à escala (fronteira VRS).

$$IE = \frac{E^{t+1}(t+1)|v}{E^{t}(t)|v} \qquad (4)$$

$$SE = \frac{S^{t+1}(t+1)}{S^{t}(t)} = \frac{\frac{E^{t+1}(t+1)}{E^{t+1}(t+1)|v}}{\frac{E^{t}(t)}{E^{t}(t)|v}} \quad (5)$$

$$IF = \left[\frac{E^{t}(t)}{E^{t+1}(t)}\frac{E^{t}(t+1)}{E^{t+1}(t+1)}\right]^{\frac{1}{2}} \qquad (6)$$

A mudança da eficiência técnica pura (4) avalia em que medida a produção da DMU se aproxima dos melhores valores observados da fronteira VRS em cada período, isto é, se a sua eficiência técnica pura aumenta entre t e $t+1$. Por esta razão, essa componente é designada por factor de *catching up*. Assim, valores maiores do que 1 indicam que a produção da DMU em análise em $t+1$ está mais próxima da fronteira VRS do que em t.

A mudança de eficiência de escala (5) de uma DMU é definida pelo rácio entre as eficiências de escala verificadas nos períodos $t+1$ e t, isto é, $S^{t+1}(t+1)$ e $S^{t}(t)$, respectivamente. Por exemplo, $S^{t}(t)$ é a eficiência de escala para o período t, sendo definida pela razão entre a eficiência técnica medida à fronteira CRS do período t, e a eficiência técnica pura medida à fronteira VRS do período t [Banker et al., 1984]. A eficiência de escala mede o impacto da escala na operação da DMU e avalia a capacidade de obtenção da produtividade máxima. Assim, SE avalia em que medida a DMU se aproxima dos valores da produtividade máxima (MPSS - *Most Productive Scale Size*) entre o período t e $t+1$. Caso o índice SE seja maior do que 1 significa que a posição da DMU em $t+1$ está mais próxima dos valores de MPSS do que no período t.

A mudança de tecnologia (6) mede radialmente a distância entre as fronteiras CRS relativas aos períodos t e $t+1$. Valores do índice IF maiores do que 1 significam que a produtividade da fronteira em $t+1$ é superior à produtividade da fronteira em t, isto é, ocorreu progresso técnico (ou tecnológico). O índice de mudança de tecnologia (IF) permite, para cada DMU, comparar o posicionamento das fronteiras da tecnologia em cada um dos períodos de tempo.

Para além de analisar individualmente a mudança da tecnologia (IF) para cada DMU, é possível avaliar globalmente se a fronteira do período $t+1$ relativamente à fronteira

do período t progride, regride ou se se verificam simultaneamente as duas situações (zonas com cruzamento de fronteiras). Para isso, é necessário avaliar a posição relativa das fronteiras para o conjunto das DMUs observadas. Observe-se que o mais relevante é saber se as DMUs no momento mais recente $(t+1)$ estão localizadas em zonas onde a fronteira de $t+1$ é mais produtiva do que a fronteira do período anterior. Isto significa que o cruzamento das fronteiras deve ser explorado apenas para a combinação de recursos e resultados das DMUs observadas no período mais recente, o que corresponde à análise do rácio seguinte: $E^t(t+1)/E^{t+1}(t+1)$. Na prática, podem surgir as seguintes situações entre t e $t+1$: (i) se o rácio é sempre superior a um para todas as DMUs avaliadas significa que ocorreu progresso tecnológico. (ii) se o rácio for sempre inferior a um para todas as DMUs avaliadas significa que ocorreu retrocesso tecnológico. (iii) Se existir pelo menos um rácio inferior a 1 e outro superior a 1 significa que há combinações produtivas onde se observou progresso tecnológico e outras em que se observou retrocesso. Isto implica que as fronteiras de t e $t+1$ se cruzam.

Para avaliar a robustez dos índices estimados para cada DMU, calcularam-se intervalos de confiança para I, IE, SE, IF e $E^t(t+1)/E^{t+1}(t+1)$, usando a técnica de *bootstrap* proposta por Simar e Wilson [1999]. Os intervalos de confiança permitem verificar se cada índice é significativamente diferente de 1. Isto significa que se o intervalo contém o valor um não há evidência estatística de que tenham ocorrido alterações significativas. Em oposição, se os limites do intervalo forem inferiores (ou superiores) ao valor 1 significa que há evidência estatística de que ocorreu um declínio (ou progresso). Esta abordagem tem sido relativamente pouco utilizada na literatura, embora se encontrem alguns estudos que apliquem a técnica de *bootstrap* neste contexto [Gilbert e Wilson, 1998; Odeck, 2009; Tortosaausina et al., 2008].

4- Resultados

4.1- Avaliação da eficiência das lojas

Esta secção pretende avaliar o desempenho da actividade operacional das lojas. Tendo em conta que cada loja é responsável pelos resultados que obtém e pela gestão dos recursos disponíveis, definiram-se como inputs a área, o stock, os encargos gerais, os

encargos de pessoal e a quebra, e como output o valor das vendas brutas, como mostra a Figura 1.

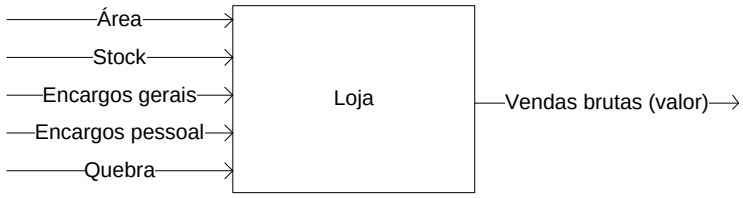


Figura 1- Modelo da actividade operacional da loja

A área representa a dimensão da loja e influencia criticamente as vendas. O stock reflecte o valor médio mensal investido em produtos que cada loja tem disponível para vender. Os encargos gerais referem-se aos custos que cada loja incorre com veículos e deslocações, publicidade, manutenção, energia, gastos administrativos, comunicação, segurança e limpeza. Os encargos de pessoal representam as despesas totais com os funcionários que trabalham na loja. A quebra é o montante anual perdido na loja relativo aos produtos que não se venderam por razões de furto, mau manuseamento e expiração do prazo de validade. Esta variável é um resultado da actividade que se deseja minimizar (i.e., output indesejável). Entre as várias alternativas para a inclusão deste tipo de dados nos modelos de DEA descritas na literatura, optou-se por definir a quebra como um input do modelo. O valor das vendas brutas da loja representa o valor total anual dos produtos vendidos.

Este modelo pretende analisar a capacidade de cada loja em maximizar as vendas com o nível de recursos existentes, o que corresponde a uma análise orientada pelos outputs. O modelo utilizado para avaliar a evolução do desempenho das lojas não considera o efeito da envolvente. Tal deve-se ao facto de a população se manter praticamente constante na envolvente das lojas e da concorrência se alterar apenas para um número reduzido de lojas em 2004.

Utilizou-se o teste proposto por Simar e Wilson [2002], baseado no *bootstrap*, para identificar os rendimentos à escala da actividade das lojas. Os resultados indicaram que há evidência estatística que a actividade das lojas nos anos de 2002 e 2004 exibiu rendimentos variáveis à escala. A técnica de *bootstrap* foi implementada no software R utilizando a *package* FEAR.

Inicialmente calculou-se a eficiência técnica pura para cada loja em cada período, utilizando o modelo (1), seguindo-se uma análise de *bootstrap* para obter os valores da

eficiência corrigida e o factor de correcção de eficiência. Os resultados para as lojas de cada ano resumem-se na Tabela 1.

Tabela 1: Resultados da eficiência técnica pura (Ef. técnica pura), eficiência corrigida (Ef. corrigida) e factor de correcção de eficiência (Factor correcção) das lojas nos períodos 2002 e 2004

		Ano de 2002			Ano de 2004		
		Ef. técnica pura	Ef. corrigida	Factor correcção	Ef. técnica pura	Ef. corrigida.	Factor correcção
Supermercados	Média	0.9027	0.8655	0.0333	0.8951	0.8563	0.0319
	Desvio Padrão	0.0890	0.0810	0.0193	0.0824	0.0713	0.0110
Hipermercados	Média	0.9117	0.8534	0.0738	0.9253	0.8595	0.0936
	Desvio Padrão	0.1019	0.0830	0.0362	0.0783	0.0608	0.0237

Os resultados indicam que é insuficiente fazer a avaliação de desempenho com base apenas nos valores de eficiência calculados com os modelos de DEA. Os valores de eficiência corrigida apresentam alguns desvios, com diferentes magnitudes, relativamente aos valores de eficiência originais, devendo ser preferencialmente utilizados. Nos dois anos analisados, as maiores diferenças entre os valores de eficiência originais e os valores corrigidos correspondem maioritariamente às lojas consideradas eficientes no modelo (1). Considerando os valores de eficiência corrigida, os dois formatos das lojas, nos dois anos, mantêm o valor médio de eficiência e apresentam uma dispersão de eficiência idêntica.

4.2- Avaliação da evolução do desempenho das lojas

A utilização integrada da técnica de DEA, índice de *Malmquist* e *bootstrap* permite avaliar a evolução do desempenho das lojas entre o ano 2002, e o ano 2004. Numa primeira análise, calculou-se o índice de *Malmquist*, I, e as suas componentes, IE, SE e IF para as várias lojas, utilizando as expressões (3), (4), (5) e (6) respectivamente. Na segunda análise, para cada loja, estimaram-se os intervalos de confiança relativos ao valor do índice de *Malmquist* e componentes, utilizando a técnica de *bootstrap* proposta por Simar e Wilson [1999]. Esta abordagem permite detectar se as diferenças de produtividade captadas pelo índice de *Malmquist* são significativas. Finalmente, para cada loja, estimaram-se os intervalos de confiança para o rácio $E^{02}(04)/E^{04}(04)$, que constitui o índice de mudança da tecnologia (IF) para as lojas de 2004 com base na técnica de *bootstrap* proposta por Simar e Wilson [1999]. Com este procedimento é possível tirar conclusões robustas sobre a posição relativa das fronteiras dos dois períodos de tempo considerados.

Os resultados obtidos no *bootstrap* para os índices de *Malmquist* e suas componentes (I, IE, SE e IF) resumem-se na Tabela 2 para os 14 hipermercados e os 56 supermercados. Os valores dos índices considerados significativamente diferentes de 1 a um nível de confiança de 90% foram designados por estimativas significativas. Nesta tabela, para cada formato das lojas, a primeira coluna indica o número de vezes que ocorreram valores superiores, inferiores ou iguais a 1 em cada um dos índices, e a segunda coluna indica o número de lojas para as quais as estimativas significativas do índice correspondiam a valores maiores ou menores do que 1.

Tabela 2 -Síntese dos resultados do *bootstrap* para *I*, *IE*, *SE* e *IF* para as lojas

		Hipermercados		Supermercados	
		Estimativas originais	**Estimativas significativas**	**Estimativas originais**	**Estimativas significativas**
Produtividade total (*I*)	Melhoria	12	10	24	17
	Declínio	2	1	32	23
	Estagnação	0	-	0	-
Eficiência técnica pura (*IE*)	Melhoria	7	2	21	7
	Declínio	2	1	29	12
	Estagnação	5	-	6	-
Eficiência de escala (*SE*)	Melhoria	4	1	14	7
	Declínio	6	1	41	16
	Estagnação	4	-	1	-
Tecnologia (*IF*)	Melhoria	14	5	45	15
	Declínio	0	0	11	0
	Estagnação	0	-	0	-

No caso dos hipermercados, dez melhoraram significativamente o desempenho (*I*), em que metade dos casos se deveu à melhoria da produtividade da fronteira (*IF*). Isto significa que se verificou uma evolução nas práticas dos melhores hipermercados em 2004. Verificou-se que um hipermercado piorou significativamente o desempenho devido à diminuição significativa da eficiência de escala (*SE*). Esta loja corresponde ao hipermercado mais pequeno, que por restrições da estrutura física não pode aumentar de dimensão. Neste caso a posição da loja em 2004 está mais afastada dos valores de produtividade máxima do que em 2002. Nos três restantes hipermercados não se observaram alterações significativas de desempenho entre os dois períodos.

No caso dos supermercados, 17 melhoraram significativamente o desempenho (*I*), na maioria dos casos devido à melhoria da produtividade associada à fronteira (*IF*). Isto significa que também se registou uma evolução nas práticas dos melhores

supermercados em 2004. Observou-se que 23 supermercados pioraram significativamente o desempenho (*I*), na maioria dos casos devido à diminuição significativa da eficiência de escala (*SE*) e/ou da eficiência técnica pura (*IE*). Isto significa, que alguns supermercados se afastaram dos melhores níveis observados nas lojas de referência em 2004. As lojas que pioraram a eficiência de escala incluem a unidade de menor dimensão, 2 unidades de dimensão média e 2 lojas de grande dimensão (uma é a de maior dimensão) e 11 unidades que perderam área disponível de venda para estruturas de apoio à área comercial. Nas restantes 16 lojas não se observaram alterações significativas de desempenho entre os dois períodos.

Globalmente os hipermercados tiveram uma evolução mais favorável do que os supermercados. Enquanto nos hipermercados mais de 70% melhoraram o desempenho, nos supermercados apenas 30% melhoraram, e mais de 40% pioraram significativamente. Esta análise e a desagregação em componentes permite à gestão da empresa caracterizar a evolução do desempenho de cada loja identificando as causas significativas que influenciaram o seu desempenho. Esta informação é útil para apoiar o processo de definição de objectivos para cada loja e o rumo a seguir numa perspectiva de melhoria contínua.

Para determinar o posicionamento relativo das fronteiras em 2004 e em 2002, analisou-se o rácio $E^{02}(04)/E^{04}(04)$ correspondente às lojas no ano de 2004, cujos resultados são apresentados na Tabela 3. Apesar de existirem rácios maiores e menores do que 1 para as lojas, verifica-se que apenas existem rácios significativamente diferentes de 1 em 30 lojas (5 hiper e 25 super). Nestas, os valores dos rácios são sempre superiores a 1, o que significa que não há evidência estatística do cruzamento de fronteiras. Conclui-se que a fronteira do período 2004 progrediu relativamente à fronteira do período 2002, não havendo evidência de que as fronteiras de 2002 e 2004 se cruzem.

Tabela 3 - Síntese dos resultados do *bootstrap* para o rácio $E^{02}(04)/E^{04}(04)$ para as lojas

		Estimativas originais	Estimativas significativas
Hipers	Melhoria	14	5
	Declínio	0	0
	Estagnação	0	-
Supers	Melhoria	50	25
	Declínio	6	0
	Estagnação	0	-

4- Conclusões

Este artigo avalia a evolução do desempenho de lojas entre os anos 2002 e 2004 utilizando o índice de *Malmquist* e a técnica de *bootstrap* [Simar e Wilson, 1999]. Para cada loja, o *bootstrap* permite calcular os intervalos de confiança relativos ao índice de *Malmquis*t e suas componentes, de forma a identificar as alterações que foram estatisticamente significativas na evolução (ou não) do desempenho de cada loja.

Esta abordagem permite também averiguar se houve evolução das práticas das lojas entre os dois períodos, dando resposta à expectativa da empresa de saber, efectivamente, se a produtividade das lojas melhorou em cada período. Neste caso, é possível determinar as lojas que melhoraram a produtividade, sendo possível identificar as boas práticas que devem ser implementadas pelas lojas com pior desempenho, e o rumo a seguir numa perspectiva de melhoria contínua.

Em relação ao posicionamento relativo das fronteiras nos dois períodos, este pode não ser identificado com rigor pela análise do valor médio do índice de mudança de tecnologia. Por isso, deve-se analisar o rácio que compõe o índice da mudança da tecnologia para a combinação de recursos e resultados das DMUs observadas no período mais recente de forma a verificar se houve progresso ou retrocesso da fronteira, ou se se verificam ambas as situações (fronteiras cruzadas). Para caracterisar a posição relativa das fronteiras de uma forma estatísticamente significativa, utilizou-se a técnica de *bootstrap* para estimar o intervalo de confiança para o rácio do índice da mudança da tecnologia, $E^t(t+1)/E^{t+1}(t+1)$.

Globalmente os hipermercados tiveram uma evolução mais favorável do que os supermercados. Enquanto nos hipermercados mais de 70% melhoraram o desempenho, nos supermercados apenas 30% melhoraram, e mais de 40% pioraram significativamente. Verificou-se que as lojas melhoraram significativamente o desempenho nomeadamente devido à melhoria da produtividade associada à fronteira, o que significa que houve evolução nas práticas das lojas. Observou-se que os supermercados que pioraram significativamente o desempenho, a maioria dos casos deveu-se à diminuição significativa da eficiência, o que significa que se afastaram dos melhores níveis observados nas lojas de referência e/ou dos valores de produtividade máxima em 2004. Esta análise permite caracterizar o histórico do desempenho das lojas

e fornece informação útil para apoiar o processo de definição de objectivos para cada loja no sentido da melhoria contínua.

5- Referências

Banker, R. D., Charnes, A. e Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science*. 30(9), 1078–1092.

Charnes, A., Cooper, W. W. and Rhodes, E. (1978) Measuring the efficiency of decision making units, *European Journal of Operational Research*. 2(6), 429-444.

Färe, R. e Grosskopf, S. (1996). *Intertemporal production frontier: with dynamic DEA*. Kluwer Academic Publishers, Boston.

Färe, R., Grosskopf, S., Lindgren, B. e Roos, P. (1994a). *Productivity developments in Swedish hospitals: a Malmquist output index approach*, pages 253–272. Kluwer Academic Publishers, Boston.

Färe, R., Grosskopf, S., Norris, M. e Zhang, Z. (1994b). Productivity growth, technical progress, and efficiency change in industrialized countries. *The American Economic Review*. 84(1), 66–83.

Gilbert, R. A. e Wilson, P. W. (1998). Effects of deregulation on the productivity of Korean banks. *Journal of Economics and Business*. 50(2), 133–155.

Odeck, J. (2009). Statistical precision of DEA and Malmquist indices: A bootstrap application to norwegian grain producers. *Omega*. 37(5), 1007–1017.

Simar, L. e Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*. 44(1), 49–61

Simar, L. e Wilson, P. W. (1999). Estimating and bootstrapping Malmquist indices. *European Journal of Operational Research*. 115, 459–471.

Simar, L. e Wilson, P. W. (2002). Non-parametric tests of returns to scale. *European Journal of Operational Research*. 139(1), 115–132.

Tortosaausina, E., Grifelltatje, E., Armero, C. e Conesa, D. (2008). Sensitivity analysis of efficiency and Malmquist productivity indices: An application to spanish savings banks. *European Journal of Operational Research*. 184(3), 1062–1084

Car sequencing approaches: a literature survey

Cláudia Duarte[a], *J. Valério de Carvalho*[b], *Ana Paula Póvoa*[c]

[a] *University of Minho, Braga, Portugal*
claudia.s.r.duarte@gmail.com

[b] *University of Minho, Braga, Portugal*
vc@dps.uminho.pt

[c] *IST, Lisbon, Portugal*
apovoa@ist.utl.pt

Abstract

In our days companies production systems have to be flexible and capable to change the production quickly so as to answer to the client's demand where product variety is as constant. This implies that setup times and costs associated to intermixed product sequences have to be reduced to a minimum in order to not disturb the production system. Finding efficient intermixed product sequences is a challenge. This document presents a review on the approaches presented in the literature to address the car sequencing problem. The focus is the car sequencing problem that arise from practical applications required by the automotive industry and intends to minimize sequence-dependent work overload taking into account just the sequence and the work overload.

Keywords: Car Sequencing, Sequencing Models, Exact techniques, Heuristics techniques, Meta-heuristics techniques.

1- Introduction

Henry Ford started producing Model-T in assembly lines in 1908. In those days assembly lines were designed for standardized products with a high level of demand (mass production). Nowadays, the conditions are different because customers are able to choose between a high number of car options. This fact increases the product variety and forces companies to find solutions that try to achieve benefits of an efficient flow production while enabling a highly diversified number of products. This kind of assembly lines are called mixed-model assembly lines.

In summary, this problem aims at finding an intermixed sequence of different types of products that satisfies the demand of all models and fulfils the company objectives in terms of costs and time [Scholl 1999].

Boysen et al. [2009] divided the sequencing models in three different approaches assuming that the operations are assigned to workstations and the system is only dependent on the models sequence. These are Mixed-Model Sequencing [Boysen et al. 2009; Scholl 1999], Car Sequencing [Boysen et al. 2009; Monden 1993] and Level Scheduling [Boysen et al. 2009; Monden 1993]. Mixed-Model Sequencing aims to avoid/minimize sequence-dependent work overload based on operational characteristics (operation times and station borders, for example). Car Sequencing aims to minimize sequence-dependent work overload taking into account the sequence. The objective of Level Scheduling is to minimize the differences between actual and ideal rates while keeping a constant rate of usage of the parts used in the line (JIT - Just in Time - concepts).

Nowadays, companies need solutions to apply in real situations where the objective is to minimize the work overload and the material requirements. So, Car Sequencing and Level Scheduling appear as very important for automotive companies. To meet this need, an extra type of models, called Hybrid Mixed-Models have appeared whose goal is to achieve simultaneously the objectives of the three models presented previously. So, the minimization of work overload and levelling part usage are the targets to be achieved by this approach. Consequently, other operational characteristics need to be added as setup operations, due dates and assembly line balancing [Boysen et al. 2009]. These models consider capacity and material aspects which are important for practical application and bring more complexity to the existing models. Drexl and Kimms [2001] developed an approach that considers Car Sequencing and Level Scheduling applying an exact technique called Column Generation, used in Branch and Price techniques.

The present article focus on the Car Sequencing problem which is explained in detail in section 2. Some authors as Brailsford et al. [1999] and Prandstetter and Raidl [2008] used Constraint Programming and integer programming software to help solving the problem. However this kind of software reaches its limit when one hundred or so vehicles with few options are considered [Estellon et al. 2008]. To prove its complexity,

Kis [2004], proved that this problem is NP-hard, non-deterministic polynomial-time hard. Hence, in order to solve it effectively, several heuristics have been proposed as Ant Colony Optimization, Greedy algorithms and Local Search [Gottlieb et al. 2003]. The techniques often used to solve the problem are described in sections 2.1 (exact techniques), 2.2 (heuristic techniques) and 2.3 (meta-heuristic techniques).

2- Car Sequencing problem

The Car Sequencing problem is usually defined as a linear programming problem as the one following described.

$$\min\left[\sum_{o}^{O}\sum_{t=1}^{T} \mathrm{pen}[o,s] * w_o\right] \quad (1)$$

Subject to

$$\sum_{v}^{V}\sum_{j=s}^{s+N_o-1} x_{v,j} * o_{o,v} - \mathrm{pen}[o,s] \leq H_o \quad o \in \quad s \in \{ \quad _o \quad \} \quad (2)$$

(...)

This problem intends to minimize the penalties, $pen_{o,s}$, associated with the violation of capacity constraints in one range of car slots. These capacity constraints are usually defined using sequencing rules of the type H_o:N_o. These rules mean that at most H_o out of N_o successively sequenced cars may require option $o \in O$, where O is the set of available options. In the objective function is associated a weight w_o to the penalty, representing the cost of violation capacity constraints.

In terms of objectives two kinds of problems can appear [Boysen et al. 2009]:

- *Without an objective function* - this is a feasibility problem. Is usually seen in the literature as a Constraint Satisfaction Programming problem, where the constraints represent a subset of the problem, declaring allowed or forbidden value combinations and each one providing a local view of the whole problem. The solution of these kind of problems satisfies all the constraints.
- *With an objective function* - the goal is to minimize the violations of sequencing rules and there are a lot of approaches as [Boysen et al. 2009]:
  - minimization of the number of violated N_o know as the "sliding window" approach [Gottlieb et al. 2003]
  - minimization of the number of all excessive options in violated N_o
  - etc.

The relevant operational characteristics can be classified in terms of *Number of options, Hard and soft sequencing rules, Kind of sequencing rules* and *Assignment restrictions*.

Hard rules are related with critical options and cannot be violated, soft rules can be violated. These two rules can be used in the same model or not depending on the problem. Other rules, can be used, in addition to conventional H_o:N_o rules as, for example, the maximum number of direct successions is restricted [Boysen et al. 2009]. Assignment restrictions are related with the production cycles available for assign model copies. For example, for a group of models it is specified the cycles in which these models can be assigned to [Boysen et al. 2009].

Car Sequencing Models arise from practical applications required by the automotive industry, nevertheless these models can be applied to other types of industries. These models are usually applied in automotive assembly lines, but can also be applied in the three consecutive areas of automotive industries: body, paint and assembly line. In the body area, metal stampings and pressings are welded together to build the car body. Setup times here are high, hence longer sequences with similar vehicles are desirable. In the paint area, the car body is treated, prepared and painted. Changes in the colors cause delays and losses of materials facts that represent a high cost that should be minimized [Gravel et al. 2005]. In the assembly line, the goal is to control the succession of work intensive model options as sunroof, to avoid work overload [Boysen et al. 2009] using sequencing rules of the type H_o:N_o. The use of these rules will allow to express the problem formulation as a constraint satisfaction problem that enforces the sequencing rules observance.

On average 70% of the car value is added in the final assembly line. Consequently, car sequence problems in the literature are more focused on finding a good sequence using the final assembly constraints that ensure load balancing and component supply [Gravel et al. 2005].

In 2005, Renault proposed a challenge, the ROADEF'2005, to researchers in this field. This challenge is an extension of the classical Car Sequencing problem. The goal is to schedule cars along an assembly line considering two types of capacity constraints, imposed by the assembly line, taking into account their priority but also the paint batching constraints [Solnon et al. 2008].

The efficiency of the approaches used in Car Sequencing problems (CSP) decreases as the problem size and difficulty increases. Another relevant aspect is that when there is

not a feasible solution, long computation times are needed to achieve this conclusion [Gravel et al. 2005]. To solve these models exact, heuristics and meta-heuristics approaches are used. A description of some of these techniques is done in the following chapters.

2.1- Exact techniques

Exact techniques provide an optimal solution meaning that there is not a better value for the objective function in this problem. In the Car Sequencing problem this means, for example, that the optimal solution minimises the number of capacity constraints violated and the number of color changes. If there is more than one optimal solution only one needs to be obtained [Prandstetter 2008].

Examples of exact techniques are the Branch and Bound [Fliedner and Boysen 2008], Branch and Cut [Barnhart et al. 1998], Branch and Price [Barnhart et al. 1998; Drexl and Kimms 2001]; and Constraint Programming [Solnon et al. 2008]. However, two major exact solution techniques are used in the solution of combinatorial problems. These are Branch and Bound and Constraint Programming. To address the complexity of this problem, [Boysen et al. 2009] suggest a mix between traditional combinatorial optimization and Constraint Programming, as well as the inclusion of more efficient heuristics and exact solution procedures. Gravel et al. [2005] and Prandstetter [2008] proposed also an integer linear programming approach that solves benchmark instances of combinatorial problems in acceptable time and proves the solutions optimality.

The Branch and Bound technique is probably the most used exact solution technique for mixed integer linear problems [Fliedner and Boysen 2008]. This algorithm enumerates all candidate solutions to an optimal integer solution doing successive partitions of the solution space and cutting the search tree by considering limits calculated along the enumeration. These cuts are calculated using upper and lower estimated bounds of the quantity being optimized. When problems are large or complex, strong cutting planes embedded into the Branch and Bound tree, resulting in Branch and Cut algorithms, typically improve the algorithm efficiency [Bradley et al. 1977].

The Branch and Cut algorithm is a mix between the Branch and Bound algorithm and the Cutting plane method. Is also defined as a generalization of the Branch and Bound

technique with LP relaxations that allows the separation and cut throughout the Branch and Bound tree [Barnhart et al. 1998]. In this technique classes of valid inequalities are not used in the LP relaxation because of the huge number of constraints to handle efficiently that will not lead to an optimal solution. During the LP relaxation if an optimal solution to the LP relaxation is infeasible, a separation problem (subproblem) is solved trying to find violated inequalities and add them to the LP to cut off the infeasible solution. This is done until no violated inequalities be found. After this the tree branching is performed [Barnhart et al. 1998].

The Branch and Price algorithm uses a similar strategy as the Branch and Cut algorithm. The difference is that in Branch and Price the procedure focuses on column generation and in the Branch and Cut the focuses is in row generation. Column generation is used to solve the LP relaxation within the Branch and Bound technique [Alvelos 2005]. However, Column generation and row generation are complementary procedures very useful to improve LP relaxation. There are some algorithms that use the two approaches [Barnhart et al. 1998]. In Branch and Price sets of columns are not used in the LP relaxation because of the huge number of columns to handle efficiently that will have their associated variable equal to zero in an optimal solution [Barnhart et al. 1998]. To verify the optimality of the LP solution a pricing problem (subproblem) that is a separation problem for the dual LP is solved to identify columns to enter the basis. If the columns are found the LP is reoptimized. After this is done the tree branching.

Apart from the above techniques also Constraint Programming is used. This is a generic technique that aims to optimize a function or to find a feasible solution subject to constraints over discrete and/or continuous variables: This is then a technique used to solve constraint satisfaction problems. Through Constraint Programming language it is possible to solve a constraint satisfaction problem specifying only the variables and constraints that the algorithm is going to solve using generic algorithms called constraint solvers. These solvers usually employ a systematic exploration of the search space that enumerates assignments of values to variables [Solnon et al. 2008]. Lookahead algorithms are effective reducing the size of the search space. After this a constraint propagation is applied to restrict the domains of other variables whose values are not yet fixed [Brailsford et al. 1999]. This is performed until a solution is found or until is proven that the problem has no solution [Solnon et al. 2008]. Constraint

Programming is an effective technique to solve easy or small instances as the others exact techniques but it has not yet been useful for harder or larger instances. Nevertheless, it has the advantage of its declarative nature allowing constraints to be expressed more easily.

2.2- Heuristic techniques

Heuristic techniques are designed to solve a given problem faster than the exact techniques. So heuristics are used to solve difficult problems in a reasonable amount of time. Different types of heuristics exist such as Greedy heuristics and Local Search heuristics.

As the name suggests, with Greedy heuristics the sequence is built in a greedy way. This means that at each stage a locally optimum choice is made which may or may not lead to a globally optimum solution to some problem. Once a car is placed into a position, it cannot be removed.

Gottlieb et al [2003] made an evaluation of some Greedy heuristics as Random Choice, Static Highest Utilization Rates, Dynamic Highest Utilization Rates, Static Sum of Utilization Rates, Dynamic Sum of Utilization Rates and Dynamic Even Distribution.

Local Search, Ant Colony Optimization (ACO) and many other techniques, use Greedy heuristics, similar to the ones presented before, to compute initial solutions [Bautista et al. 2008; Gottlieb et al. 2003].

In Local Search algorithms the search space is explored moving from solution to solution in the space of candidate solutions until an optimal sequence is found, a time bound is elapsed or a maximum number of moves is performed. The algorithm starts from a candidate solution and then iteratively moves to a neighbor solution. This is only possible if a neighborhood relation is defined on the search space. The choice of which one to move to is taken using only information about the solutions in the neighborhood of the current one, hence the name Local Search [Puchta and Gottlieb 2002].

The approaches depend on [Solnon et al. 2008]:

- *the way of construction of the initial sequence* - usually random permutation of the vehicles to produce

- *the neighbourhood considered at each move* - different moves can be considered [Puchta and Gottlieb 2002; Solnon et al. 2008]: Insert or Forward/backward Insert; Swap that includes two special cases Transposition or SwapT and SwapS; Lin2Opt; and Random or Shuffle.
- *the search strategy* - given a neighbourhood different heuristics or meta-heuristics can be chosen to decide which is the next move in the next iteration [Solnon et al. 2008].

The different moves considered affect only part of the sequence. For this reason, it is faster to evaluate the change in the objective function due to the change in positions instead of evaluating the whole sequence after each move. Transposition and SwapS, in particular, are evaluated very quickly because Transposition changes only two neighbouring positions and SwapS changes one or two options meaning that only constraints for these options must be evaluated [Puchta and Gottlieb 2002].

2.3- Meta-heuristic techniques

Meta-heuristics are a result of combining heuristics to optimize a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. These techniques have been most generally applied to NP-Hard or NP-Complete problems. Here are presented some of these techniques as Ant Colony Optimization algorithms [Gottlieb et al. 2003], Genetic algorithms [Solnon et al. 2008], Simulated Annealing [Reis 2007] and Tabu Search [Cordeau et al. 2008; Reis 2007; Warwick and Tsang 1995].

The Ant Colony Optimization algorithm is based on the behaviour of real ant colonies. The idea is to solve the problem as a search for a minimum cost path and use artificial ants to search for good paths [Gottlieb et al. 2003]. The behaviour of these artificial ants is similar to the behaviour of real ant colonies because in artificial ant colonies they lay pheromone trails on components of the minimum cost path and they choose the best path taking into account the probabilities that depend on pheromone trails that have been previously laid by the colony. These pheromone trails disappear progressively by evaporation in real life [Solnon et al. 2008]. Different authors applied this method to the Car Sequencing problem. In the algorithm of Gottlieb et al [2003] a Greedy heuristic is used to achieve pheromone trails. It is also shown that Local Search performance of [Puchta and Gottlieb 2002] is worse than the Ant colony algorithm, for small

computation time limits. For larger limits both approaches have similar results. Gravel et al. [2005] presented another Ant colony algorithm that integrates a Local Search procedure to improve the solutions constructed by ants.

The Genetic algorithms are inspired in the nature biological processes and evolution and has been also applied to the Car Sequencing problem [Joly and Frein 2008]. This search technique explores combinatorial search spaces simulating the evolution and recombining candidate solutions (called population) which are associated with fitness values related with a specific domain of the objective function [Warwick and Tsang 1995]. The goal is to combine solutions to obtain new ones considering the existence of mutations [Joly and Frein 2008]. In Warwick and Tsang [1995] the search space is explored through selection, cross-over and mutation, the main genetic operators, and at each generation, sequences are combined by cross-over operations. The new sequences may not satisfy the constraints, so they are greedily repaired by mutation operators and each offspring hill-climbed by a swap function similar to the one used Local Search approaches. This approach shows to work well in with low utilization percentages. However, with higher ones the number of successful runs decreased. The main steps of this technique are [Joly and Frein 2008]:

- *Generate an initial solution* - usually random permutation of the vehicles to produce but other techniques can be used
- *Parent selection* - considering for example that better the objective function, greater the probability to be selected
- *Determine the crossover operator* - some well known include [Joly and Frein 2008]: Order based crossover; Partially mapped crossover operator; Cycle crossover; and Uniform order crossover
- *Determine the mutation operator* - determine it using the following operators often used in literature [Zinflou et al. 2008]: Reflection; Random_swap; Group_exchange; and Block_reflection.
- *Population evaluation* - this evaluation is made using the value of the objective function for each individual
- *Selection of surviving population* - this selection is made considering that better the individuals, the greater the probability for them to survive in the next generation.

As Zinflou et al [2008] mentioned Genetic algorithms are efficient approaches if the different mechanisms of the algorithm, as crossover and/or mutation operators, were designed to deal with the specificities of the problem.

The Simulated Annealing technique takes inspiration from the annealing process in metallurgy. By analogy with this process the set of solutions of a certain problem is associated with the energy state, the objective function corresponds to the physical energy of the solid and the ground state corresponds to a global optimal solution [Reis 2007]. This neighbourhood search heuristic allows an escape from local optima by accepting solutions that may not be better than the last solution found. The probability of preserving a worse solution depends on the temperature parameter that initially is defined by the user to a high value that decreases during the run of the algorithm and consequently the probability to accept worsening moves decreases [Joly and Frein 2008]. The result is good approximation to the global optimum of a given function in a large search space. Simulated Annealing guarantees a convergence after running a sufficiently large number of iterations. However this is not very helpful, since the annealing time required to ensure a significant probability of success will usually exceed the time required for a complete search of the decision space [Reis 2007]. If an acceptable good solution in a fixed amount of time is rather than the best possible solution, Simulated Annealing may be more effective than exhaustive enumeration [Brailsford et al. 1999]. Comparing with other heuristics, Simulated Annealing is usually expensive in a computational sense. Nevertheless, the other heuristics tend to generate solutions far from the optimum [Reis 2007].

The Tabu Search technique uses a Local Search procedure to move from a solution to another in the neighbourhood until a defined criteria has been satisfied. This Local Search procedure uses a tabu list to restrict the moves between a solution and other [Warwick and Tsang 1995]. The tabu list is a short-term memory that contains solutions that have been visited recently and forbidding moves. The way that is managed plays a crucial part in the effectiveness and efficiency of the technique. For instance, if the number of forbidden moves is too high the algorithm may miss good solutions. When solving the Car Sequencing problem to avoid this situation the algorithm, usually, has a criteria that override the tabu status of certain moves [Reis 2007]. The efficiency can also be improved using methods that exploits the long-term memory of the search process either by recovering best solutions obtained so far or recovering the attributes of that solutions [Reis 2007].

3- Conclusion

In this paper a review was performed on the different methods developed to solve the Car Sequencing problem. In particular were identified exact methods, heuristics and meta-heuristics. It can be sais that exact techniques work reasonably fast for small problems. However, in larger problems, usually real problems, they are not useful due to the high computational times. One approach is to use exact techniques to prove new bounds integrating them with heuristic and meta-heuristic techniques. Heuristics and meta-heuristics do not guarantee the optimal solution of the problems even though they allows a gain in computational performance and/or conceptual simplicity. When optimal solutions are needed usually heuristics are coupled with exact methods. This provide good initial solutions and allow a reduction of the search space.

Based on the survey performed we can conclude that the Car Sequencing problem is still a problem that needs further research where more efficient methods should be explored. Also larger characteristics within this problem should also be explored in future research.

4- References

Alvelos, F.P.P.C. Branch-and-Price and Multicommodity Flows. In *Production and Systems Department.* Guimarães: University of Minho, 2005, vol. PhD, p. 196.

Barnhart, C., Johnson, E.L., Nemhauser, G.L., Savelsbergh, M.W.P. and Vance, P.H. Branch and Price: Column Generation for solving huge integer programs. Operations Research, 1998, vol. 46, no. 3, p. 316-329.

Bautista, J., Pereira, J. and Adenso-Díaz, B. A GRASP approach for the extended car sequencing problem. *Journal of Scheduling*, 2008, vol. 11, p. 3-16.

Boysen, N., Fliedner, M. and Scholl, A. Sequencing mixed-model assembly lines: Survey, classification and model critique. *European Journal of Operational Research*, 2009, vol. 192, p. 349-373.

Bradley, S.P., Hax, A.C. and Magnanti, T.L. *Applied mathematical programming*. Edtion ed.: Addison-Wesley Publishing Co. Inc., 1977. 716 p.

Brailsford, S.C., Potts, C.N. and Smith, B.M. Constraint satisfaction problems: Algorithms and applications. European Journal of Operational Research, 1999, vol. 119, p. 557-581.

Cordeau, J.F., Laporte, G. and Pasin, F. Iterated tabu search for the car sequencing problem. European Journal of Operational Research, 2008, vol. 191, p. 945-956.

Drexl, A. and Kimms, A. Sequencing JIT Mixed-Model Assembly Lines Under Station-Load and Part-Usage Constraints. *Management Science*, 2001, vol. 47, no. 3, p. 480–491.

Estellon, B., Gardi, F. and Nouioua, K. Two local search approaches for solving real-life car sequencing problems. *European Journal of Operational Research*, 2008, vol. 191, p. 928-944.

Fliedner, M. and Boysen, N. Solving the car sequencing problem via Branch & Bound. European Journal of Operational Research, 2008, vol. 191, p. 1023-1042.

Gottlieb, J., Pucjta, M. and Solnon, C. A study of greedy, local search and ant colony optimization approaches for car sequencing problems. In *Applications of Evolutionary Computing.* Springer Berlin / Heidelberg, 2003, vol. 2611/2003.

Gravel, M., Gagné, C. and Price, W.L. Review and comparision of three methods for the solution of the car sequencing problem. *Journal of the Operational Research Society*, 2005, vol. 56, p. 1287-1295.

Joly, A. and Frein, Y. Heuristics for an industrial car sequencing problem considering paint and assembly shop objectives. Computers & Industrial Engineering, 2008, vol. 55, p. 295-310.

Kis, T. On the complexity of the car sequencing problem. Operations Research Letters, 2004, vol. 32, p. 331-335.

Monden, Y. *Toyota production system*. Edtion ed. Norcross: Industrial Engineering Press, Institute of Industrial Engineers, 1993.

Prandstetter, M. Exact and heuristic methods for solving the Car Sequencing Problem. In *Institute of Computer Graphics an Algorithms.* Vienna: Vienna University, 2008, vol. Master, p. 66.

Prandstetter, M. and Raidl, G.R. An integer linear programming approach and a hybrid variable neighborhood search for the car sequencing problem. European Journal of Operational Research, 2008, vol. 191, p. 1004–1022.

Puchta, M. and Gottlieb, J. Solving Car Sequencing Problems by Local Optimization. Applications of Evolutionary Computing, 2002, vol. 2279, p. 181-188.

Reis, R.J.D.O Solving the Car Sequencing Problem from a Multiobjective Perspective. In. Lisbon: Instituto Superior Técnico, 2007, vol. Master, p. 90.

Scholl, A. *Balancing and Sequencing of Assembly Lines*. Edtion ed. New York: Physica-Verlag Heidelberg, 1999.

Solnon, C., Cung, V.D., Nguyen, A. and Artigues, C. The car sequencing problem: Overview of state-of-the-art methods and industrial case-study of the ROADEF'2005 challenge problem. *European Journal of Operational Research*, 2008, vol. 191, p. 912-927.

Warwick, T. and Tsang, E.P.K. Tackling Car Sequencing Problems Using a Generic Genetic Algorithm. Evolutionary Computation, 1995, vol. 3, no. 3, p. 267-298.

Zinflou, A., Gagné, C. and Gravel, M. Design of an Efficient Genetic Algorithm to Solve the Industrial Car Sequencing Problem. In W. K SIŃSKI. Vienna: I-Tech Education and Publishing, 2008, p. 377-400.

Problema integrado de corte bidimensional e dimensionamento de lotes

Elsa Silva[a*], *Filipe Alvelos*[b], *J.M. Valério de Carvalho*[b]

[a]*Centro de Investigação Algoritmi, Universidade do Minho*
elsa@dps.uminho.pt

[b]*Centro de Investigação Algoritmi e Departamento de Produção e Sistemas, Universidade do Minho*
falvelos@dps.uminho.pt
vc@dps.uminho.pt

[*]*Autor para correspondência*

Resumo

O problema de corte bidimensional aqui abordado consiste na minimização do número de placas utilizadas para o corte de um conjunto de itens rectangulares. Tipicamente, na indústria, este problema é abordado de modo repetitivo, sendo resolvido no início de cada período de planeamento. Neste artigo, o problema de corte bidimensional é estendido da resolução em apenas um período de planeamento para a resolução num conjunto de períodos de planeamento. Um modelo de programação inteira, que optimiza o problema combinado de corte a duas dimensões e o dimensionamento de lotes, minimizando os desperdícios e os custos de armazenamento, é proposto.

Palavras chave: problema de corte bidimensional, problema de dimensionamento de lotes, programação inteira.

1- Introdução

Os problemas de corte bidimensionais (*2CSP)* têm sido amplamente estudados nos últimos cinquenta anos, devido à sua vasta aplicabilidade prática. No entanto, na maior parte dos casos, têm sido resolvidos individualmente [Erjavec et al, 2009] não sendo estudados como um processo integrado nas indústrias. O problema de planeamento do

processo de corte é um processo repetitivo, acontecendo no início de cada período de planeamento, baseado nas procuras dos clientes. Procurar uma solução óptima para o problema de *2CSP* em cada período do horizonte de planeamento é diferente de uma avaliação completa de todo o horizonte de planeamento.

Considerar a antecipação da produção de alguns itens e o armazenamento em stock de objectos que sobram de cortes (*leftovers*), com tamanho suficiente para serem utilizados posteriormente, implica a mudança de uma política de melhoria local para uma estratégia óptima global. Deste modo, uma interacção entre períodos de planeamento dá origem a uma melhoria da eficiência da produção e a uma redução dos custos.

Neste trabalho será estudado o problema integrado de corte a duas dimensões e o problema de dimensionamento de lotes (2CS-LSP). Dada a procura de itens rectangulares durante um dado horizonte de planeamento, o problema é determinar a quantidade de itens que deve ser cortada de placas rectangulares em cada período do horizonte de planeamento, de modo a que os custos de produção, de armazenamento e os desperdícios das placas sejam minimizados.

Ambos os problemas *2CSP* e dimensionamento de lotes (*LSP*) são NP-difíceis, e do mesmo modo a sua integração é também um problema NP-difícil.

No problema de *2CSP* considerado, existe um conjunto de placas com a mesma largura e altura, de onde se pretende cortar um conjunto de itens rectangulares mais pequenos. As placas grandes estão disponíveis em número (virtualmente) infinito e o conjunto de itens a cortar está agrupado por tipo, de acordo com o tamanho (largura e altura). O objectivo é minimizar o número de placas grandes utilizadas ou o desperdício. No problema de *LSP* multi-item considerado, pretende-se determinar as quantidades a produzir de cada item num dado número de períodos de tempo, de modo a satisfazer as procuras de cada item em cada período, minimizando os custos de armazenamento e de produção.

Integrar os problemas *2CSP* e *LSP* pode ser visto como um balanceamento entre antecipar a produção de alguns itens, fornecendo uma melhor combinação dos itens nas placas, promovendo a minimização dos desperdícios, e não antecipar a produção dos itens, fomentando a minimização dos custos de armazenamento.

Resolver estes dois problemas separadamente pode levar a soluções não-óptimas para o problema integrado [Gramani e França, 2006].

Um método para obter soluções óptimas para o problema integrado *2CS-LSP*, que não tem por base a geração de padrões de corte é proposto. Um modelo de programação inteira é apresentado, como uma extensão do modelo de programação inteira para o problema *2CSP* proposto por Silva et al. [2010]. O modelo para o problema *2CSP* permite resolver problemas de dois e três-estágios e permite também a consideração de aspectos que podem ser muito relevantes em termos práticos, como a existência de placas com diferentes tamanhos, a atribuição de valores às placas residuais, o cálculo do comprimento dos cortes e a rotação dos itens.

O modelo que programação inteira proposto permite resolver de forma exacta o problema integrado de *2CS-LSP*, modelando a antecipação da produção dos itens e também o armazenamento dos objectos que sobraram dos cortes, no final de cada período do horizonte de planeamento.

2- Revisão da literatura

Não existe na literatura um número elevado de artigos abordando a integração dos problemas de *2CS-LSP* e a maioria dos trabalhos está muito relacionada com a aplicação industrial em que surge o problema.

Possivelmente, o primeiro trabalho abordando o problema combinado foi de Farley [1988]. O problema surge na indústria do vestuário, um modelo de programação inteira e um modelo de programação quadrática foram propostos. O principal objectivo é minimizar os custos globais de produção, tendo em conta os custos de corte, de costura, de armazenagem e também de superprodução e subprodução. Embora tenham sido considerados vários períodos de produção, o problema integrado de corte e dimensionamento de lotes não foi explicitamente abordado, em vez disso o problema foi formulado como um problema de programação da produção ao longo de um único período.

O processo de produção na indústria do cobre foi também analisado de forma integrada por Hendry et al. [1996]. Um procedimento em duas fases foi proposto; na primeira fase, uma solução agregada é determinada para o problema de corte a uma dimensão, e, na segunda fase, é decidido o escalonamento da produção para cada dia do período de tempo. Procedimentos heurísticos foram utilizados para resolver o primeiro estágio e um modelo de programação inteira foi considerado no segundo estágio.

Um problema de *2CSP* com figuras irregulares e um problema de dimensionamento de lotes com procura estocástica, com origem numa empresa de *dumpers*, foram resolvidos de forma integrada por Nonas e Thorstenson [2000]. A minimização dos desperdícios e dos custos de armazenagem e de preparação foi o principal objectivo. O problema foi formulado de modo não-linear, com uma função objectivo côncava com restrições lineares. Foram utilizadas heurísticas para resolver o problema e também um procedimento de geração de colunas. Posteriormente em Nonas e Thorstenson [2008] a abordagem proposta foi integrada com a heurística sequencial apresentada por Haessler [1971].

Em Arbib e Marinelli [2005] o problema de corte a uma dimensão foi integrado com o problema de planeamento de inventário. Os autores propõem um modelo de programação inteira e uma metodologia para integrar um sistema hierárquico de planeamento da produção. O modelo é uma extensão de um optimizador de dia-a-dia, permitindo a inclusão de custos de transporte e de armazenagem, que pode ser utilizado para gerir "encomendas prioritárias" (quando chega uma encomenda depois do plano de produção estar completo).

Na indústria do papel o problema integrado de corte e dimensionamento de lotes foi abordado por Poltroniere et al. [2008], os autores propõem um modelo de programação inteira e duas heurísticas de decomposição. A primeira heurística que propõem baseia-se na relaxação lagrangeana, o problema de dimensionamento de lotes é resolvido primeiro e depois é resolvido o problema de corte. A outra heurística começa por resolver o problema de corte, determinando o número total de jumbos necessários por tipo de papel e período, o planeamento da produção é alcançado resolvendo o problema de dimensionamento de lotes.

O problema *2CS-LSP* foi também tratado por Gramani e França [2006]; o método proposto consiste em determinar as quantidades de itens a produzir em cada período do horizonte de planeamento, de tal modo que o número de placas utilizadas, os custos de armazenamento e de stock sejam minimizados. Um modelo de programação inteira com limite de tempo na capacidade da serra foi resolvido heuristicamente usando uma representação em rede do caminho mais curto. Cada arco da rede está associado com a capacidade do *2CSP* resolvido através do procedimento de Gilmore e Gomory [1965] de geração de colunas. Depois de resolvidos todos os arcos, é resolvido um problema de

caminho mínimo. A abordagem heurística considera a antecipação de toda a produção num dado período de tempo e não a antecipação da produção de apenas alguns itens. O procedimento heurístico foi comparado com o método utilizado tradicionalmente na indústria, que consiste em resolver primeiro o problema de dimensionamento de lotes e depois, num segundo estágio, o problema de corte. Os autores concluíram que resolver a versão combinada do problema permite benefícios acima dos 28%.

Para o mesmo problema, mas considerando a quantidade de produtos finais a produzir em cada período do horizonte de planeamento, em vez da quantidade de itens individuais, Gramani et al. [2009] apresentou um modelo de programação inteira resolvido por um procedimento heurístico baseado em relaxação lagrangeana. Uma extensão deste modelo combinado, através de uma relaxação dos custos de setup, e da frequência inteira dos padrões de corte e a consideração de armazenamento não apenas de produtos finais, mas também de itens individuais foi apresentada por Gramani et al. [2010]. Para esta versão do problema propõem um modelo de programação inteira resolvido através de geração de colunas.

3- Modelo de programação inteira

O modelo aqui proposto é uma extensão do modelo de programação inteira apresentado por Silva et al. [2010] para o problema *2CSP*. Nesse trabalho é proposta uma extensão para o problema de corte a duas dimensões do modelo "*one-cut*" proposto por Dyckhoff [1981] para o problema de corte a uma dimensão. O modelo tem por base a consideração explícita, de como os itens podem ser obtidos das placas, sendo o *corte* o conceito mais importante do modelo, que consiste em usar uma placa e obter um item a partir dela (através de um ou dois cortes em guilhotina).

De um modo genérico, de um corte resulta um item (que não será mais cortado) e duas placas que poderão posteriormente ser cortadas (estas placas denominam-se *placas residuais*). Na figura 1 são apresentados dois exemplos de corte, as linhas a negrito representam os cortes em guilhotina, em ambos os casos do corte obtêm-se um item e duas placas residuais.

São consideradas placas residuais todos os rectângulos com largura maior ou igual que a menor largura dos itens e cuja altura é maior ou igual que a menor altura dos itens. Todos os rectângulos com tamanho menor são considerados desperdício.

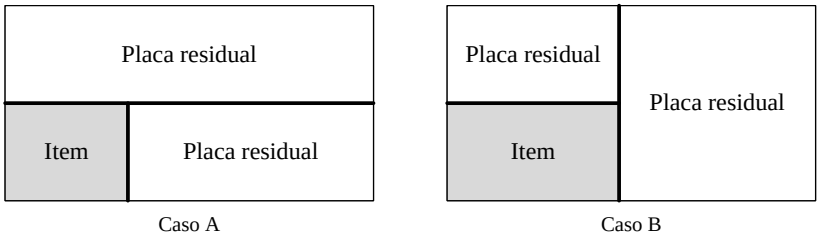


Figura 1 – Exemplos de corte.

O tipo de padrão de corte considerado é o dois-estágios não-exacto (ver Figura 1), significa que os itens são obtidos das placas através de um conjunto de cortes horizontais, que dividem as placas em tiras, sucedidos por um conjunto de cortes verticais, separando os diferentes itens. Um corte horizontal extra pode ser executado para separar os itens dos desperdícios.

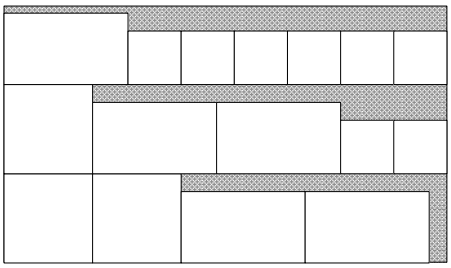


Figura 1 - Padrão de corte dois-estágios não-exacto.

O modelo baseia-se na enumeração de todos os cortes e tipos de placas residuais possíveis, para tal define um conjunto de regras simples para calcular os cortes, a largura, a altura e o estágio das placas residuais resultantes. É importante referir que no tipo de corte apresentado no caso B na figura 1, a placa residual acima do item é considerada desperdício, uma vez que os padrões de corte são do tipo dois-estágios não-exacto. A ideia principal do modelo de programação inteira é associar uma variável de decisão com cada corte, isto é, com o número de vezes que um corte é executado.

Um algoritmo para a geração de todos os cortes e placas residuais, assim como limites superiores para o número de tipos de placas residuais e cortes é também proposto por Silva et al. [2010].

Como referido anteriormente, o modelo de programação inteira proposto para o problema *2CS-LSP* baseia-se no modelo programação inteira de Silva et al. [2010] para o problema *2CSP*; as principais diferenças entre os modelos são a consideração de períodos de tempo, a utilização de variáveis de decisão associadas com o

armazenamento de itens, o armazenamento de placas residuais e também o uso de parâmetros relacionados com os custos.

A partir daqui assume-se que todos os cortes possíveis assim como os tipos de placas residuais são conhecidos para cada período.

No problema *2CS-LSP* as placas rectangulares iniciais têm largura *W* e altura *H* e o item procurado *i* a ser cortado no período *t*, tem largura w_i^t, altura h_i^t e procura d_i^t. O número de tipos de itens é definido como *n*, o número de tipos de placas residuais é definido como *m* e o número de períodos de tempo por *T*. Os tipos de placas são indexados por *j*, *j=0,...,m*, com *j=0* correspondendo à placa inicial e *j=1,...,m* correspondendo aos tipos de placas residuais, os itens são indexados por *i=1,…,n* e os períodos de tempo por *t=1,…,T*. Os parâmetros e variáveis do modelo são definidos como:

Parâmetros:

d_i^t: procura do item *i* no período *t*

c_i^t: custo de cortar o item *i* no período *t*

g_i^t: custo de guardar em stock o item *i* no período *t*

h_j^t: custo de guardar em stock a placa residual j no período *t*

c_0^t: custo da placa inicial no período *t*

a_{ijk}: igual a 1 se a placa *k* resulta de cortar o item *i* da placa *j*

Variáveis de decisão:

x_{ij}^t: número de vezes que o item *i* é cortado da placa *j* no período *t*

I_i^t: número de itens do tipo *i* em stock no final do período *t*

s_j^t: número de placas do tipo *j* em stock no período *t*

Pretende-se definir a quantidade de itens a cortar em cada período de tempo, decidindo se o corte de alguns itens deve ou não ser antecipado, de modo que os custos do processo de corte e de armazenamento e o número de placas iniciais utilizadas sejam minimizados, respeitando a procura de cada item em cada período do horizonte de planeamento. Utilizando a notação previamente apresentada, o modelo de programação inteira para o problema *2CS-LSP* é o seguinte:

$$\text{Minimizar} \quad \sum_{t=1}^{T}\sum_{i=1}^{n}(c_0^t + c_i^t)x_{i0}^t + \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m} c_i^t x_{ij}^t + \sum_{t=1}^{T}\sum_{j=1}^{m} h_j^t s_j^t + \sum_{t=1}^{T}\sum_{i=1}^{n} g_i^t I_i^t \quad (1)$$

Sujeita a: $$\sum_{j=0}^{m} x_{ij}^t + I_i^{t-1} - I_i^t = d_i^t \qquad i = 1, \dots, n; t = 1, \dots, T \qquad (2)$$

$$\sum_{j=0}^{m}\sum_{i=1}^{n} a_{ijk} x_{ij}^t - \sum_{i=1}^{n} x_{ik}^t + s_k^{t-1} - s_k^t = 0 \quad \text{k} = 1, \dots, m;\ t = 1, \dots, T \qquad (3)$$

$$x_{ij}^t \geq 0 \ e \ inteiro \ i = 1, \dots, n; j = 1, \dots, m;\ t = 1, \dots, T \qquad (4)$$

$$s_j^t \geq 0 \ e \ inteiro \ j = 1, \dots, m;\ t = 1, \dots, T \qquad (5)$$

$$I_i^t \geq 0 \ e \ inteiro \ i = 1, \dots, n;\ t = 1, \dots, T \qquad (6)$$

Na função objectivo (1) são considerados os custos de produção dos itens, o custo das placas inicias $c_0{}^t$ que podem variar em cada período do horizonte de planeamento, dependendo da flutuação de preços da matéria-prima nos mercados ($\sum_{t=1}^{T}\sum_{i=1}^{n}(c_0^t + c_i^t)x_{i0}^t + \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m} c_i^t x_{ij}^t$), os custos de armazenamento das placas residuais ($\sum_{t=1}^{T}\sum_{j=1}^{m} h_j^t s_j^t$) e os custos de armazenar os itens, cuja produção foi antecipada ($\sum_{t=1}^{T}\sum_{i=1}^{n} g_i^t I_i^t$).

As restrições (2) garantem que a procura de cada tipo de item em cada período é satisfeita, balanceando os itens que são cortados no respectivo período, de placas iniciais ou residuais, os itens que estão em stock de períodos anteriores e os itens que serão armazenados em stock no final do período de tempo considerado.

Finalmente, as restrições (3) garantem que para cada tipo de placa residual, em cada período, o número de cortes que produz esse tipo de placa residual no período considerado mais o número de placas residuais do mesmo tipo armazenadas em stock no período anterior, $\sum_{j=0}^{m}\sum_{i=1}^{n} a_{ijk} x_{ij}^t + s_k^{t-1}$, é igual ao número de cortes efectuados nessas placas residuais mais o número de placas residuais desse tipo que serão guardadas em stock no final do período, $\sum_{i=1}^{n} x_{ik}^t + s_k^t$; desta forma garante-se que serão apenas realizados cortes em placas residuais existentes.

4- Resultados Computacionais

A construção e optimização do modelo proposto foram implementadas em C++, utilizando a C callable library do *Solver* IBM ILOG Cplex 12.1. Todos os testes computacionais foram executados num processador 1.87 GHz Intel Core Duo com 2GB

Tabela 1 - Características das instâncias e resultados computacionais considerando diferentes horizontes de planeamento

Instância							Número de períodos do horizonte de planeamento													
							t=2				*t*=5				*t*=7				*t*=10			
Nome	*n*	*l*	*W*	*H*	*w'*	*h'*	*t*	*t2*	*nr*	*nv*	*t*	*t2*	*nr*	*nv*	*t*	*t2*	*nr*	*nv*	*t*	*t2*	*nr*	*nv*
10W30_6	3	42	4250	1860	1400	925	0,4	0,3	16	40	0,0	0,0	40	100	0,0	0,0	56	140	0,0	0,0	80	200
146O18_17	6	6570	3760	1860	423	320	0,1	0,1	368	1500	1,8	1,8	920	3750	2,5	2,5	1288	5250	2,6	2,6	1840	7500
15B30_20	5	106	4250	2200	808	418	0,4	0,4	122	486	0,6	0,6	305	1215	2,0	2,0	427	1701	3,0	3,0	610	2430
15J30_22	2	57	4250	2120	2069	484	0,1	0,0	12	30	539,1	539,1	30	75	2038,0	2038,0	42	105	[a]	[a]	60	150
15X18_25	6	1080	4250	1860	458	460	0,0	0,0	98	352	0,0	0,0	245	880	0,1	0,0	343	1232	0,0	0,0	490	1760
15X22_26	7	3000	4250	1860	423	422	0,2	0,2	1044	3246	6,9	6,9	2610	8115	2,7	2,7	3654	11361	44,3	44,2	5220	16230
174N25_31	2	165	4250	1860	434	400	0,5	0,5	52	120	94,1	94,1	130	300	100,2	100,2	182	420	[a]	[a]	260	600
26K19_58	2	365	2820	1870	577	572	0,0	0,0	24	56	0,1	0,1	60	140	0,2	0,2	84	196	0,3	0,2	120	280
26R16_59	2	315	2820	1870	730	595	0,0	0,0	20	50	0,0	0,0	50	125	0,0	0,0	70	175	0,0	0,0	100	250
28C38_63	2	58	4250	1860	694	608	0,1	0,1	28	68	0,3	0,3	70	170	0,7	0,6	98	238	8,4	8,4	140	340
28R18_65	2	156	4250	1860	708	457	0,1	0,1	26	66	[a]	[a]	65	165	[a]	[a]	91	231	[a]	[a]	130	330
30R1_70	1	985	4200	1860	540	900	0,0	0,0	16	32	0,0	0,0	40	80	0,1	0,1	56	112	0,2	0,2	80	160
37A18_73	1	2390	4250	2070	760	490	0,0	0,0	16	32	0,0	0,0	40	80	0,0	0,0	56	112	0,0	0,0	80	160
40C2_82	2	242	4250	1860	450	415	0,2	0,2	70	160	[a]	[a]	175	400	25,0	25,0	245	560	3,0	3,0	350	800
42K18_90	3	209	4250	1860	439	331	0,0	0,0	66	178	0,9	0,9	165	445	0,4	0,4	231	623	0,8	0,8	330	890
69W18_111	10	9090	3760	1860	422	290	0,6	0,5	770	4624	51,3	51,3	1925	11560	98,1	98,1	2695	16184	181,1	181,1	3850	23120
69Y18_112	7	9033	3760	1860	490	320	0,2	0,2	370	1628	7,1	7,1	925	4070	33,1	33,1	1295	5698	27,9	27,9	1850	8140
76R25_121	2	1260	3760	1860	522	388	0,0	0,0	52	124	0,0	0,0	130	310	0,0	0,0	182	434	0,0	0,0	260	620
94H18_123	4	4935	2700	1860	905	540	0,0	0,0	46	178	0,0	0,0	115	445	0,0	0,0	161	623	0,0	0,0	230	890
ARKEN25F_130	1	35	4250	2200	1211	432	0,0	0,0	14	28	0,0	0,0	35	70	0,2	0,2	49	98	0,2	0,2	70	140
ARKEN25P_131	1	400	3760	1860	1211	432	0,0	0,0	12	24	0,0	0,0	30	60	0,0	0,0	42	84	0,3	0,2	60	120
MAC_136	1	32	4250	2120	808	808	0,0	0,0	12	24	0,0	0,0	30	60	0,2	0,1	42	84	0,2	0,2	60	120
RIO25E_143	5	53	4250	1860	845	845	0,0	0,0	12	24	0,1	0,1	30	60	0,2	0,1	42	84	0,2	0,2	60	120
STE1R_147	2	33	4200	2200	730	538	0,1	0,1	42	104	1,6	1,5	105	260	9,6	9,6	147	364	145,1	145,1	210	520

[a] Cplex 12.1 devolve “falta de memória”

de memória RAM. Em todos os casos não foi efectuada nenhuma afinação do *solver* e foram utilizados os parâmetros por omissão.

Na Tabela 1 são apresentadas as principais características de um conjunto de instâncias provenientes de uma empresa da indústria do mobiliário: o número de tipos de itens (coluna *n*), o número total de itens (coluna *l*), o tamanho das placas grandes (colunas *W* e *H*), a menor largura e a menor altura do item (coluna *w'* e *h'*)

O custo do processo de corte de cada item (c_i^t) é dado por 0,01% da área do respectivo item e o custo de armazenar o item em stock (g_i^t) é dado por 2% do custo de produção do item. O custo de armazenar as placas residuais em stock é calculado de forma semelhante ao cálculo do custo de armazenar os itens em stock, isto é, o custo de armazenar a placa do tipo *j* (h_j^t) é igual a 2% do custo da placa residual (0,01% da área da placa residual). O custo das placas iniciais (c_0^t) é 5 e foram considerados 2, 5, 7 e 10 períodos de planeamento. Os stocks iniciais são todos nulos. Os valores utilizados para os parâmetros são semelhantes aos utilizados por Gramani e França [2006].

Para os diferentes períodos de planeamento é apresentado o tempo total necessário para encontrar a solução óptima, o que inclui o tempo despendido na construção do modelo e o tempo de optimização (coluna *t*), o tempo gasto apenas pelo solver (coluna *t2*), o número de restrições (*nr*) e também o número de variáveis (*nv*).

Para um horizonte de planeamento com dois períodos foi possível encontrar a solução óptima para todas as instâncias em menos de 1 segundo. De um modo geral, à medida que o número de períodos do horizonte de planeamento aumenta, aumenta também o tempo computacional. É de salientar que para a maioria das instâncias o tempo aumenta muito pouco, à excepção das instâncias (15J30_22, 174N25_31 e 69W18_111). O tempo de construção do modelo é insignificante, uma vez que *t-t2* é quase sempre zero ou muito próximo de zero.

Em 68% das combinações testadas foi encontrada a solução óptima em menos de 1 segundo e 90% em menos de 2 minutos. Apenas em cinco casos não foi possível encontrar a solução óptima, porque a optimização foi interrompida por não existir memória suficiente disponível.

5- Conclusões

Neste trabalho foi proposto um modelo de programação inteira para o problema integrado de corte bidimensional e dimensionamento de lotes. O modelo é uma extensão do modelo Silva et al. [2010] para o problema de corte bidimensional. No modelo proposto pondera-se a antecipação ou não da produção de cada item em cada período, com o propósito de minimizar os custos de produção, de stock e o número de placas utilizadas, satisfazendo a procura em cada período do horizonte de planeamento.

Embora seja de tamanho pseudo-polinomial, o modelo é optimizado por um *solver* geral de programação inteira (Cplex12.1) que resolveu um conjunto de instâncias reais em tempos computacionais aceitáveis.

Agradecimentos

Este trabalho foi parcialmente financiado pela FCT (Fundação para a Ciência e Tecnologia) através do projecto "SearchCol: Metaheuristic search by column generation" (PTDC/EIA-EIA/100645/2008) e pela bolsa de doutoramento da FCT do primeiro autor SFRH/BD/42259/2007.

6- Referências

Arbib, C., e Marinelli, F. (2005). Integrating process optimization and inventory planning in cutting-stock with skiving option: An optimization model and its application. *European Journal of Operational Research,* 617-630.

Dyckhoff, H. (1981). A new linear programming approach to the cutting stock problem. *Instiute for operations research and the management sciences* 29, 1092-1104.

Erjavec, J., Gradisar, M., e Trkman, P. (2009). Renovation of the cutting process. *International Journal of Production Research* 47, 3979-3996.

Farley, A. (1988). Mathematical programming models for cutting-stock problems in the clothing industry. *The Journal of the Operational Research Society* 39, 41-53.

Gilmore, P. C., e Gomory, R. E. (1965). Multistage Cutting Stock Problems of Two and More Dimensions. *INFORMS, Operations Research* 13, 94-120.

Gramani, M. C., França, P. M., e Arenales, M. N. (2010). A linear optimization approach to the combined production planning model. *Journal of the Franklin Institute* doi:10.1016/j.franklin.2010.05.010.

Gramani, M. C. N., e França, P. M. (2006). The combined cutting stock and lot-sizing problem in industrial processes. *European Journal of Operational Research* 174, 509-521.

Gramani, M. C. N., França, P. M., e Arenales, M. N. (2009). A Lagrangian relaxation approach to a coupled lot-sizing and cutting stock problem. *Int. J. Production Economics* 119, 219-227.

Haessler, R. W. (1971). A heuristic programming solution to a nonlinear cutting stock problem. *MANAGEMENT SCIENCE* 17, 793-802.

Hendry, L. C., Fok, K. K., e Shek, K. W. (1996). A cutting stock and scheduling Problem in the Copper Industry. *The Journal of the Operational Research Society* 47, 38-47.

Nonas, S. L., e Thorstenson, A. (2000). A combined cutting-stock and lot-sizing problem. *European Journal of Operational Research* 120, 327-342.

Nonas, S. L., e Thorstenson, A. (2008). Solving a combined cutting-stock and lot-sizing problem with a column generating procedure. *Computers & Operations Research* 35, 3371-3392.

Poltroniere, S. C., Poldi, K. C., Toledo, F. M. B., e Arenales, M. N. (2008). A coupling cutting stock-lot sizing problem in the paper industry. *Annals of Operations Research* 157, 91-104.

Silva, E., Alvelos, F., e Carvalho, J. M. V. d. (2010). A new integer programming model for two- and three-stage two dimensional cutting stock problems. *European Journal of Operational Research* 205, 699-708.

Optimization of a Distribution Network using Electric Vehicles: A VRP Problem

Frederico Gonçalves[a], *Sónia R. Cardoso*[a], *Susana Relvas*[a], *Ana Paula F. D. Barbosa-Póvoa*[a]

[a]*CEG-IST, UTL, Av. Rovisco Pais, 1049-001 Lisboa, Portugal*
frederico.m.goncalves@ist.utl.pt, sonia.cardoso@ist.utl.pt, susanaicr@ist.utl.pt, apovoa@ist.utl.pt

Abstract

The highly competitive environment that every day companies face forces them to improve the relationships with their clients. The key to success is to provide the right amount of a certain product, in the right place and at the right time in order to achieve an excellent service level at a competitive cost. In this sense, logistics plays a very important role.

This work presents a mixed integer linear programming approach which is developed with the main objective of studying the route planning process and evaluating the feasibility of implementing electric vehicles in this operation. An example based on a Portuguese batteries' distributer is studied and three different scenarios are analyzed. The first corresponds to the application of a vehicle routing problem with pickups and deliveries to the company's current fleet. In the second scenario, the fleet is divided into two types of vehicles, current fleet and electric vehicles without cargo capacity. Finally, the third scenario corresponds to the exclusive use of electric vehicles with transportation capacity.

Keywords: Logistics, vehicle routing problem, pickups and deliveries, fleet optimization, electric vehicles.

1 – Introduction

The highly competitiveness that companies face every day in the market force them to improve their efficiency in order to achieve clients' expectations. Logistics plays a key role in improving companies' efficiency since it is the activity in the supply chain management that is responsible to store and deliver the products when desired. Its main focus is on the transportation activity that appears as the most relevant in terms of logistics total costs. In order to reduce these costs it is essential to optimize distribution vehicles routes so as to achieve an efficient service for costumers [Lau et al., 2009].

Vehicle routing problems have been widely studied in order to provide solutions for real life problems. As an example, Hu et al., [2009] studied a solution for food wholesalers' distribution operation in cities with a circular transportation infrastructure, as well as

Tarantilis and Kiranoudis [2007] analyzed a way to improve fresh milk and concrete distribution. In this way VRP has a huge role by providing solutions to different kinds of real life problems. Given a fleet of identical capacity vehicles, a distribution centre and a set of different customers, that represent demand points, the implementation of a VRP provides a set of routes that minimizes distribution total cost while guaranteeing costumers' demand. One of the most important restrictions is that vehicles have a limited capacity that can not be exceeded [Baldacci et al., 2010].

The vehicle routing problem with pickups and deliveries (VRPPD) is one of the VRP variants. This problem was introduced by Min [1989]. Since then VRPPD has gained strength and its application has been diverse, for example, Gendreau et al. [1999] studied the travelling salesman problem with pickups and deliveries, Blanc et al. [2006] developed a multi-depot pickup and delivery model for optimizing the collection of containers that are used to transport end-of-life materials from dismantled vehicles and to deliver them to recycling centers as well as Liu et al. [2010] that applied a VRP model considering time windows and pickups and deliveries to a barreled drinking water enterprise in order to improve service level and reduce operating costs. Regarding solution approaches, VRP is considered an NP–hard problem and several methods have been developed attempting to find the best possible solution. For example, Laporte et al. [1986] developed the Branch and Bound algorithm (exact method), Clark and Wright [1964] developed the Savings Algorithm, which is a heuristic procedure, Thompson and Psaraftis [1993] developed the Multi-Route Improvement Heuristics.

In this article, a VRPPD model is applied to a real case of a Portuguese company that distributes batteries. The main objective of this work is to study the feasibility of the implementation of electric vehicles on their fleet, given the company's expressed interest in this alternative, as well as to improve the model performance in order to provide a better solution to the company's distribution plan. The problem addressed decomposes the original set of customers using the p-median approach and for each new cluster the VRPPD is applied.

2 – Problem description

In this work, an optimization model for the distribution planning of a Portuguese company is proposed. The main objective of this work is to study the feasibility of implementing electric vehicles on their fleet, as well as to improve model performance

in order to provide a better solution to the company's distribution plan. Thus, two models where developed, one for regular vehicles and another for the electric vehicles. The problem representation of the VRPPD is extended to account for specific constraints for electric vehicles. The problem under study can then be described as follows:

Given the number of clients to visit each day, their location and the reason of the visit, the quantity of batteries to deliver and to collect, each day and at each client, the number of available vehicles and respective transportation capacity, the unitary transportation cost, the variable time of the visit in each site, the fixed time of the distribution operation, the distance and time to travel between all points of the network, the available daily time for the distribution operation, and the autonomy and charging time for each electric vehicle (used only in electric vehicles model);

Determine the optimum routes that guarantee the satisfaction of all clients, obeying the maximum total distance allowed for each vehicle, the maximum total time allowed to carry out the distribution plan, and using the minimum number of vehicles necessary to accomplish the distribution operation;

So as to minimize the total distribution cost.

3 – Model Formulation

The model developed is a Mixed Integer Linear Programming model and uses the following notation, based on sets, parameters and variables. The model constraints and objective function are described subsequently.

Indexes and Sets

i,j – nodes $\in I,\ i,j = 0,1,\dots,N$, where 0 is the the depot and remaining nodes are customers

v – vehicles $\in V, v = 1,\dots,K$

Parameters

e_i - Quantity of batteries to deliver in place i.

r_i - Quantity of batteries to pickup in place i.

max_i - Maximum value between batteries to deliver and pickup in place i.

$cmax_v$ - Vehicle maximum load capacity, in the amount of batteries that can carry.

$cdist_v$ - Cost of fuel per vehicle, in € / km.

cfixo – Fixed costs for the daily use of an electric vehicle, in €.

tv_i - Variable time to deliver and pickup batteries, in place i, in minutes.

$dviag_{ij}$ – Distance between place i and place j, in km.

$tviag_{ij}$ – Travel time between place i and place j, in minutes.

tf – Fixed time to handle administrative issues, in minutes.

tdisp – Daily total time available for distribution operation, in minutes.

autonomia – Value of electric vehicle range with the battery fully charged, in km.

paragem – Time required to recharge the battery of the electric vehicle, in minutes.

Variables

$distot_v$ – Positive variable necessary to obtain the total distance covered by each vehicle.

$Numparagens_v$ – Positive variable necessary to obtain the number of times that each electric vehicle *v* needs to stop to refill its battery.

u_i – Integer variable associated to each costumer *i* related with the sequence of visit and used the sub-tour elimination technique proposed by Miller, Tuckler and Zemlin (1960).

x_{ijv} – Binary variable that assumes the value of 1 if vehicle *v* visits cluster *j* after cluster *i*, otherwise it assumes the value 0.

vf_v – Binary variable that assumes value of 1 if vehicle *v* executes a route, otherwise it assumes value 0.

Note: The parameters *autonomia* and *paragem* and the variable $Numparagens_v$ will only be used in the mathematical model with electric vehicles.

$$\text{Min} \quad \sum_{v}^{m} \sum_{i}^{n} \sum_{j}^{n} (x_{ijv} \cdot dviag_{ij} \cdot cdist_v) + cfixo \qquad (1)$$

Subject to

$$\sum_{i>1} x_{jiv} = vf_v,\ \text{j=0} \ \wedge \forall v \tag{2}$$

$$\sum_{i>1} x_{ijv} = vf_v, \text{j=0} \ \wedge \forall v \tag{3}$$

$$\sum_{i\neq j}\sum_{v} x_{ijv} = 1, \ \ \forall j > 1 \tag{4}$$

$$\sum_{i\neq j}\sum_{v} x_{jiv} = 1 \ \ , \ \ \forall j > 1 \tag{5}$$

$$\sum_{j} x_{jiv} = \sum_{j} x_{ijv} \ \ , \ \ \forall i \wedge \forall v \tag{6}$$

$$\sum_{i>1}\sum_{j\neq i} \max_i \cdot x_{ijv} \leq c \max_v \ \ , \ \ \forall v \tag{7}$$

$$x_{ijv} \leq vf_v \ \ , \ \ \forall i \wedge \forall j \wedge \forall v \tag{8}$$

$$\sum_{i}\sum_{j} (tviag_{ij} + tv_i + tf\ ).x_{ijv} \leq tdisp \ \ , \ \ \forall v \tag{9}$$

$$u_i - u_j + n.x_{ijk} \leq n-1, \qquad 1 \leq i \neq j \leq N \wedge \forall (i,j) \in I \tag{10}$$

$$distot_v = \sum_{i}\sum_{j}\left(x_{ijv} \cdot dviag_{ij}\right) \ \ , \ \forall v \tag{11}$$

$$numparagens_v = \frac{disttot_v}{autonomia} \ \ , \ \forall v \tag{12}$$

$$\sum_{i}\sum_{j} (tviag_{ij} + tv_i + tf\ ).x_{ijv} + (numparagens_v * paragem) \leq tdisp\ , \ \forall v \tag{13}$$

The objective function (1) minimizes the total cost of the operation. Equations (2) and (3) guarantee that all routes begin and end at the distribution centre, respectively, whenever vehicle *v* is used. Equations (4) and (5) ensure that each client *j* is only served by one vehicle. Equation (6) ensures route continuity. Equation (7) ensures that vehicle *v* can handle all the requests from its customers on a given route. Equation (8) defines

that each customer is only visited by a vehicle that has been allocated to a route. Equation (9) ensures that the sum of time spent in the distribution operation does not exceed the daily time available for distribution. Equation (10) eliminates the creation of sub-tours, equation proposed by Miller et al. [1960]. Equation (11) allows to calculate the total distance travelled by each vehicle. Equation (12) calculates the number of stops that an electric vehicle *v* needs to perform the distribution plan. Equation (13) ensures that the total time of a route performed by an electric vehicle *v* does not exceed the maximum time available.

4 – Case Study

A Portuguese batteries' distributor company wants to optimize its distribution operation while studying the feasibility of implementing electric vehicles in their fleet. The distribution network comprises 13 warehouses scattered throughout the country. The batteries distribution structure can be divided in two distinct parts. The primary distribution consists on the transport of goods from the central warehouse (Lisboa) to the regional warehouses, as well as to the customers with direct deliveries. The exchange of products, although not very common, among the regional warehouses (transshipment) is also included here. The secondary distribution includes the distribution of regional warehouses to all sorts of customers. The recovery of batteries to recycle or to recharge is also part of this company service. As well as the distribution, the recovery process is also divided into primary and secondary recoveries.

This work will only focus on Lisboa region since it appears as the most problematic. The study will use typical distribution data based on January of 2008.

In order to study the optimization of the current distribution in Lisbon and also the feasibility of implementing electric vehicles in the company's fleet three scenarios are considered. The first scenario describes the current company distribution characteristics while the second and third scenario evaluates the introduction of electric vehicles. In scenario 1, the actual company's fleet that is heterogeneous and is composed by eight vehicles with different transport capacities is considered. There are five vehicles of type one (V-type 1) with a capacity of 38 batteries and three vehicles of type two (V-Type 2) with a capacity for 82 batteries. For the second scenario, the fleet is divided into two types of vehicles, current fleet and electric vehicles of type one (EV - type 1) that do not have cargo capacity and are only used to visit clients that do not need the deliver or

collection of batteries. These vehicles require 60 minutes to perform charging and have autonomy for 50 km. The third scenario considers the exclusive use of electric vehicles of type two (EV - type 2) with a transportation capacity for 73 batteries, autonomy for 100 km and 120 minutes charging time. The costs with vehicles are divided into two components, a fixed cost, which regards the mensal cost with leasing for the current fleet, and the cost for acquisition and conversion of the vehicle for the electric vehicles. The variable cost component is related to fuel or energy consumption per km. These values are expressed in table 1.

Table 1 – Costs with vehicles

Vehicle	Fixed Cost (€/month)	Variable Cost (€/km)
V-Type 1	365	0.14
V-Type 2	465	0.18
EV-Type 1	678	0.01
EV-Type 2	1626	0.03

The complexity of this problem increases exponentially with the number of customers or the quantity of products to deliver, requiring some simplifications so as to solve the model in a reasonable computational time. Although there are several types of products in the real distribution, these do not have major differences in their transportation and can be treated as a single product. Additionally, the generic battery dimensions and weight are determined by a weighted mean value representative of company's demands. The weight is the limiting parameter on vehicle capacity occupation.

In 2008, the company served 2332 clients, a fact that increases the problem complexity. In order to deal with this complexity, the customers are initially aggregated by geographic area (boroughs) to which they belong, and then, as it is still too complex, the problem is divided into smaller problems by applying the p-median model in the Lisbon area to finally apply the optimization models to each sub-problem.

Customers are visited not only for delivery and collection of batteries, but also to check the conference of the stock on consignment, the collection of claims, for auto sales and to provide technical assistance. The time spent in each costumer can be divided into two components, one fixed and another variable. The fix component is related to the estimated time the driver spends on each route for administrative reasons. The variable component is related to the delivery and collection of batteries. It is estimated that the

salesperson spends a minimum of 20 minutes in each cluster and then it is added 1 minute for every two batteries delivered or collected. Each distribution day has 8 hours duration and 22 working days per month are assumed.

4.1 – Application of p-median model

In the p-median problem it is assumed a given number of *p* of facilities to locate and a number *n* of customers whose demand should be fulfilled. Given the distance between every point and the number of medians to consider, the model determines which points should be aggregated to each median in order to minimize the sum of the total distance between each median and all the associated points. It is necessary to ensure that each point is connected to a single median point and that the number of medians to consider is respected. The demand points are only connected to points that are considered medians. In this problem, median points will be selected customers, to whom other customers will be associated in a cluster.

The number of clusters to be considered should be sufficiently large, so as to be representative of the problem, and at the same time it cannot be too large, because it may become computationally infeasible. To understand how many clusters to consider, the two models, the first (M1) for the current fleet and the second (M2) with electric vehicles were tested for different numbers of clusters and the computational results are summarized in table 2. It was imposed a computational effort limit of 7200 seconds.

Table 2 – Computational results for both models

	Nº of variables		Nº of restrictions		Gap (%)		CPU (s)	
# Clients/ clusters	M1	M2	M1	M2	M1	M2	M1	M2
2	99	107	165	173	0	0	0.11	0.04
5	318	326	555	563	0	0	0.18	0.15
10	1003	1011	1845	1853	0	0	6	1.3
13	1606	1614	3003	3011	0	0	202	649
15	2088	2096	3935	3943	0	7	3404	7200
16	2353	2361	4449	4457	0	11	5590	7200
17	2634	2642	4995	5003	0	12	7200	7200
18	2931	2939	5573	5581	14	13	7200	7200
19	3244	-	6183	-	17	-	7200	-

The optimal number of clusters to consider for M1 is 17 customers. Above this value the solutions present a high relative gap since they were not solved to optimality, due to the time limit. For M2, the optimal number of clusters is 13 clients, for the same reason.

4.2 – VRPPD Model Results

The VRPPD mathematical model was implemented in GAMS 22.8, using CPLEX 11.0, in an Intel 2.20 GHz computer with 4 GB RAM. The three scenarios are compared between themselves and with the results achieved by the company.

Figure 1 shows the distribution costs for all scenarios and for the actual distribution operation at the company.

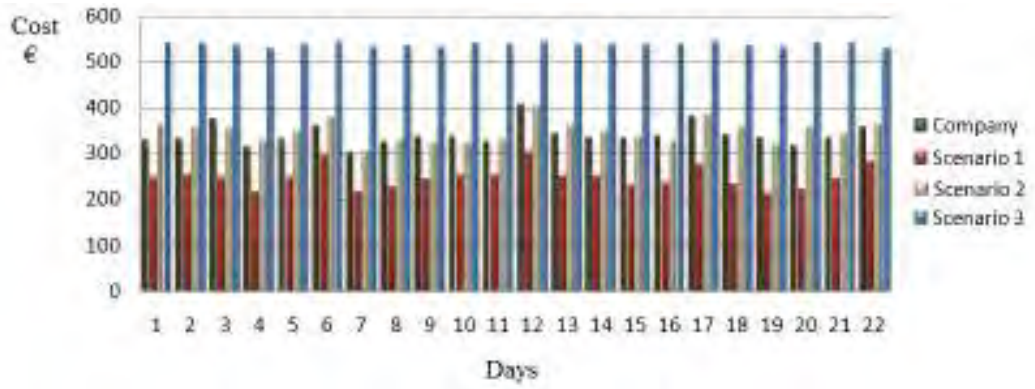


Figure 1 – Distribution costs.

Table 3 – Results for all scenarios

Scenario	Total Costs (€)	Distance Travelled (km)	Time needed (h)	Nº of vehicles
Company	7503	3370,2	176	167
1	5424	2444,8	126,2	130
2	7670	2445,1	145,1	148
3	11814	2440	184,4	130

The total costs for the first scenario are 5424€ for the complete month which is substantially lower than those the company actually had, 7503€. The second scenario implies a total cost of 7670€, which represents an increase of 167€ per month and an average of 8€ per day, when compared with the actual situation (figure 1 and table 3). However, all parameters such as distance travelled, number of vehicles and time needed to complete the distribution operation represent a significant improvement (table 3). The major problem with this scenario is the need of a substantial initial investment in acquiring and converting the vehicles. Scenario 3 appears as more costly fact again

explained by the high vehicle conversion costs. The fixed costs of vehicles represent 96% of total monthly costs that were 11814€ (table 3).

In order to conduct a comparison of the different proposed scenarios, the day that had the largest number of clusters to be visited, which was the 18th of January, was chosen, and a detailed analysis is given.

The computational results for the three scenarios are described in table 4.

Table 4 – Computational results for each scenario.

Scenario	Cluster	Number of Customers	Number of Variables	Number of Restrictions	Gap (%)	Execution Time (s)
1	1	17	2634	4995	0	3688
	2	17	2634	4995	8	7200
	3	16	2353	4449	6	7200
2	1	17	2634	4995	0	397
	2	17	2634	4995	1	7200
	3	8	681	1233	0	11
	4	8	689	1233	0	3
3	1	13	1614	3011	3	7200
	2	13	1614	3011	0	253
	3	13	1614	3011	0	230
	4	11	1196	2207	0	393

It can be seen that some clusters have a gap in the solution, which means that it was not possible to provide the optimal solution within the maximum computation time allowed (2 hours). This is due to the problem complexity, which increases exponentially with the number of customers to serve.

In table 5, the operational results of the three scenarios for the 18th of January are presented.

Table 5 – Operational results for each scenario

Scenario	Total Cost (€/day)	Total distance (km)	Total duration of routes (min.)	Average duration of routes (min.)	Number of vehicles used	Number of stops
Company	343	1298,4	3840	480	8	N/A
1	250	679,4	2537	362	7	N/A
2	364	817,5	2876	360	8	4
3	537	647,5	3245	464	7	6

Similarly to the results presented for the entire month, there is the same trend for 18th of January. Scenario 1 is the best solution for the company, allowing the reduction of all parameters studied. The second scenario presents interesting results, since it reduces all parameters under study. However, the costs are slightly higher than those obtained by the company, motivated by the investment in new vehicles. Finally, scenario 3 represents a significant increase in economic terms because of the need to invest in electric vehicles. Additionally, it is verified that the vehicles have to stop for charging. Nevertheless, the improvement in the routing reduces the effect of charging periods.

4 - Conclusions

In this work, an optimization model for the route planning operation is proposed, and used to evaluate the feasibility of implementing electric vehicles in this operation. The model was applied to a real case study of a Portuguese batteries' distributor. The complexity of the case study required the reduction of the number of demand points which were aggregated by boroughs, allowing a 54% reduction in demand points while preserving the case study characteristics. Secondly, it was applied the p-median model in order to divide total customers into smaller clusters so it would be possible to the VRPPD model to provide optimal results in a reasonable computational time.

Conducting a comparative analysis between the scenarios tested with the results obtained by the company, major improvements in all scenarios were observed namely related to the distance travelled and the number of vehicles used. However, the major objective was to assess the impact that these alternatives would have on the total costs of the operation. Thus, scenario 1 has proved to be the best alternative to the company. Scenarios 2 and 3 involve an investment in the conversion of electric vehicles, so their fixed costs have a strong impact on the cost structure.

As future work, a better aggregation can be studied so that the solution presented can be as close as possible to reality. Since this is a new technology and is subject of continuous development, it is important to study results sensitivity to different parameters related with autonomy and charging time. It would be also interesting to plan the daily distribution using time windows. The time frame of implementation could be extended to analyze how the seasonality may affect the vehicles' needs. Electric vehicles autonomy can be further studied in a deep analysis trough the proposed model

so as to conclude which is the turnover point that allows this alternative to be more attractive than the one with regular vehicles.

References

Baldacci, R., Toth, P. and Vigo, D. (2010). Exact algorithms for routing problems under vehicle capacity constraints. *Annals of Operations Research*. 175, 213–245.

Blanc, I., van Krieken, M., Krikke, H. and Fleuren, H. (2006). Vehicle routing concepts in the closed-loop container network of ARN – a case study. *OR Spectrum*. 28, 53-71.

Clark, G. and Wright, W. (1964). Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*. 12, 568.

Gendreau, M., Laporte, G. and Vigo, D. (1999). Heuristics for the travelling salesman problem with pickup and delivery. *Computers and Operations Research.* 26, 699-714.

Hu, X., Zheng, W., Minfang, H. and Amy, Z. (2009). A computer-enabled solution procedure for food wholesalers' distribution decision in cities with a circular transportation infrastructure. *Computers and Operations Research*. 36, 2201 – 2209.

Laporte, G., Mercure, H. and Nobert, Y. (1986). An Exact algorithm for the asymmetrical capacitated vehicle routing problem. *Networks.* 16, 33-46.

Lau, H., Chan, T., Tsui, W., Chan, F., Ho, G. and Choy, K. (2009). A fuzzy guided multi-objective evolutionary algorithm model for solving transportation problem. *Expert Systems with Applications.* 36 , 8255–8268.

Liu, Z., Li, N., Mi, X., Zhang, B. and Ma, H. (2010). Improvement research on Vehicle Routing Problem with Simultaneous Delivery and Pickup with time windows for Barreled Water. *IEEE 17th International Conference.* 1347.

Miller, C., Tucker, A. and Zemlin, R. (1960). Integer programming formulation of travelling salesman problems. *J. Ass. Comput. Mach.* 7, 326-332.

Min, H. (1989). The multiple vehicle routing problem with simultaneous delivery and pick-up points. *Transportation Research Part A*. 23 (5), 377.

Tarantilis,C. and Kiranoudis, C. (2007). A flexible adaptive memory-based algorithm for real-life transportation operations: Two case studies from dairy and construction sector. *European Journal of Operational Research.* 179, 806–822.

Thompson, M. and Psaraftis, H. (1993). Cyclic Transfer Algorithms for the Multivehicle Routing and Scheduling Problems. *Operations Research.* 41, 935-946.

Assessment of performance and innovation of Portuguese construction companies

Isabel M. Horta[a*], *Ana S. Camanho*[b], *Jorge M. Costa* [c]

[a]*Faculdade de Engenharia da Universidade do Porto, Porto, Portugal*
imhorta@fe.up.pt

[b]*Faculdade de Engenharia da Universidade do Porto, Porto, Portugal*
acamanho@fe.up.pt

[c]*Faculdade de Engenharia da Universidade do Porto, Porto, Portugal*
jmfcosta@fe.up.pt

[*]*Corresponding Author*

Abstract

This paper evaluates the performance of Portuguese construction companies between 1996 and 2007. The purpose of the paper is to explore the trends in construction industry performance, and identify the factors that promote excellence and innovation in the sector. From a methodological perspective, this study enhances the construction of composite indicators by using Data Envelopment Analysis to aggregate the key performance indicators of the construction companies. The statistical significance of the results obtained is ensured by the use of bootstrapping. This paper also proposes an enhanced approach to assess innovation within the industry, enabling the identification of innovative companies and the extent of innovation.

Keywords: Composite indicators, Innovation, Construction Industry.

1- Introduction

In recent years, it is well-known that the competitive environment of the Construction Industry (CI) has become increasingly fierce. The construction companies are aware of the challenges imposed by this context and attempt to implement systematic methods of performance assessment in order to achieve competitive advantage. The research described in this paper intends to promote sustainable performance improvements for CI companies, and to encourage excellence in the sector.

Performance assessment in the CI has been subject to a considerable amount of research over the past 15 years (see Bassioni et al., 2004 for a literature review). Following the publication of the seminal reports, such as the Egan (1998), construction companies

have been mostly concerned with benchmarking systems based on key performance indicators (KPIs). These systems are usually available in the internet, and enable collecting data and produce real-time results concerning performance levels. Costa et al. (2006) describe the scope of the four most well known web-based benchmarking programs focused on construction company performance measurement, carried out in Brazil, Chile, United Kingdom and United States.

More recently, the literature on performance measurement has described successful applications of the Data Envelopment Analysis (DEA) technique to the CI. El-Mashaleh et al. (2007) considered DEA a tool that can offer significant improvements over web-based benchmarking systems for the CI. According to the authors, the existing benchmarking systems have limitations in their ability to guide the industry towards more efficient and effective performance. The DEA models proposed allow construction firms to be evaluated on a company-wide basis and to identify specific areas of improvement. Horta et al. (2010) developed a methodology for assessing company performance combining the use of KPIs and DEA. The models proposed evaluate the relative efficiency of the companies and provide performance improvement targets for all companies, even for the best-practice companies. This is achieved by the specification of virtual companies, whose activity reflects the decision maker preferences.

The purpose of this paper is to evaluate the financial performance and competitive environment of the construction industry in Portugal in the recent past, between 1996 and 2007. In particular, we explored the trends concerning performance improvement and decline, and identified the companies that were able to innovate during this period. This was followed by a study of the factors promoting excellence in the sector. We explored the relationship between performance and company size, location of the headquarters, engagement in Research and Development (R&D) and the evolution of the national economic environment.

The remainder of this paper is organized as follows. Section 2 describes the methods used. Section 3 includes the data used in the study, and Section 4 discusses the results obtained. The last section concludes and suggests recommendations for future research.

2- The methodology

2.1 Performance Assessment

DEA, first introduced by Charnes et al. (1978), is a linear programming technique for comparing the efficiency of a relatively homogeneous set of organizational decision making units (DMUs) in their use of multiple resources (inputs) to produce multiple outcomes (outputs). DEA derives a single summary measure of efficiency for each DMU, which is based on the comparison with other DMUs in the sample. This assessment is relative, as the DMUs are compared to others within a sample. DEA identifies a subset of efficient DMUs, considered as examples of best practices or benchmarks. For the inefficient DMUs, the magnitude of their inefficiency is derived by the distance to the frontier constructed from the benchmark DMUs. The efficient DMUs obtain a performance score equal to 100%, and the inefficient ones obtain a score lower than 100%. The comparison with the benchmarks also allows determining the input and output targets corresponding to efficient operation.

In this paper, the DEA model used followed the research line proposed by Cherchye et al. (2004), which popularized the concept of composite indicators estimated recurring to optimization techniques. The main difference between a traditional DEA efficiency analysis and the construction of a composite indicator as proposed by Cherchye et al. (2004) is that the latter only looks at achievements, without explicitly taking into account the resources used. The rationale of using a DEA model to obtain a composite indicator is to aggregate a set of KPIs into a single summary measure of performance. Nevertheless, these KPIs must be carefully specified, as they must reflect the outputs achieved normalised by the resources used, as otherwise the comparison between companies would be impaired. The advantage of using DEA is to allow each company to select its own weighting system for the performance evaluation, recurring to optimization techniques that emphasize the company strengths. The linear programming (LP) model for deriving the composite indicator C_{j_0} for a company j_0 is shown in (1).

$$C_{j_0} = max \sum_{r=1}^{s} u_r y_{rj_0}$$
$$\text{s.t.} \sum_{r=1}^{s} u_r y_{rj} \leq 1 \qquad j = 1, \ldots, n \tag{1}$$
$$u_r \geq 0 \qquad r = 1, \ldots, s$$

In model (1), y_{rj_0} corresponds to the value of the output indicator r (r=1,...,s) in company j (j=1,...,n), with higher values corresponding to better performance. Model (1) is equivalent to the original DEA constant returns to scale model of Charnes et al. (1978), with all indicators considered as outputs and a “dummy input”' with the same value for all companies. We run model (1) including a pooled sample of the Portuguese CI companies, comparing the financial indicators of each company in each year with a frontier representing the best practices of the period analyzed.

The rationale basis of this procedure to evaluate the construction companies is easy to explain: since it is difficult to identify a priori a set of weights that all companies would agree that reflects adequately the relative importance of each indicator y_r, we let each company select its own weights, such that its composite indicator is as high as possible compared to the composite indicator of other companies evaluated with similar weights. If we impose an upper bound of one to the highest composite indicator obtained across all companies, a value of the composite indicator equal to one signals best performance (i.e., benchmarks). Under these conditions, if a company does not achieve the maximum score, even when evaluated with a set of weights that intends to maximize its performance score, it provides irrefutable evidence that other companies performed better during the period considered.

After computing the composite indicators, we used bootstrapping to obtain unbiased performance estimates. Bootstrapping is based on the idea of resampling from an original sample of data to create replicate datasets from which we can make statistical inference. Performance estimates computed by model (1) are truncated, with an upper value equal to one, and there may exist several estimates close to unity. Consequently, resampling directly from the original data (naive bootstrap) provides a poor estimate of the data generating process. The common approach is to nonparametrically estimate the original densities of the performance scores using kernel smoothing methods, combined with a reflection method. Simar and Wilson (1998} proposed the smoothed bootstrap

method which deals with all the particular features of DEA-based performance scores when mimicking the data generating process. This was the procedure adopted in this paper, which was implemented using the statistical package R including the FEAR library, developed by Wilson (2008).

2.2 Explaining Performance

To explore the factors that may be associated with good performance levels of the construction companies, we used a panel data truncated model controlling both for company and time effects. The model was specified as follows

$$C_{jt} = \alpha_0 + \delta_j + \mu_t + z_{jt}\beta + \varepsilon_{jt} \quad (2)$$

Subscript j represents the jth company (j=1,...,n), subscript t represents the time period (t=1,...,l), α_0 is an intercept, δ_j is a vector of dummy variables for each company, μ_t is a vector of dummy variables for each year, z_{jt} represents the set of regressors: company size, geographic location, R&D engagement, and national Gross Domestic Product (GDP) per capita, β denotes the regression coefficients and ε_{jt} ~ N(0,σ^2_ε) is the error term. Note that C_{jt} corresponds to the performance level of company j in year t, estimated using model (1) and corrected by bootstrapping.

2.3 Assessment of Innovation

The final stage of the performance assessment is related to the evaluation of innovation. In this study, a company is considered an innovator if it shifts the best-practice frontier of a given year to better levels than those observed in previous time periods. This means that an innovator company is able to introduce better practices in the industry. The other companies (i.e. the followers, hereafter) can improve their performance by copying the practices observed in the innovators.

The assessment of innovation using DEA models was originally proposed by Fare et al. (1994), and served as the basis for the analysis described. In this study, the concept of innovation was adapted to an evaluation of performance based on composite indicators. Another enhancement of our approach concerns the comparison of the achievements of

a company in year t with the practices observed in all previous years, up to year t, whereas the concept of innovation proposed by Fare et al. (1994) only involved a comparison of practices between two consecutive years.

To be able to identify the innovators, we run a LP model comparing the performance of each company, in a given year t, with the performance of all other companies, including observations from the same year and from previous years. As proposed by Tulkens and Vandeneeckaut (1995), a frontier that results from an assessment whose production possibility set is defined including observations from a given year as well as from previous periods is called a sequential frontier. This frontier is based on the idea that what was feasible in the past remains achievable in any later period. We consider that a necessary condition to be innovator in year t is to be located in the best practice sequential frontier, which implies having a composite indicator equal to one in a sequential frontier assessment.

To quantify the magnitude of the performance enhancement of innovative companies in relation to previous years, we run a second LP model comparing the performance of the innovators in year t, with all other companies in previous years, up to t-1. The estimate obtained using this procedure is called innovation score hereafter. Note that this second LP model differs from the previous assessment due to the exclusion of observations from year t. The innovation score can either be equal to one, meaning that the achievement of company j in year t is identical to what was observed in previous years, or greater than one, meaning that the company has actually moved the frontier to more productive levels than those previously observed.

The final stage of the analysis of innovation consisted on the study of the factors that promote innovation in the CI. For this purpose, we constructed a logistic regression model, where the dependent variable characterizes the companies as innovators or followers. In particular, a value of one for the dependent variable represents the companies that innovated in at least one of the years analyzed, and zero represents the remaining companies - the followers. The logistic model predicts the odds of being innovator, given known values of the independent variables. The odds is defined as the ratio of the probability of being innovator to the probability of being follower. The logistic model can be specified as follows:

$$ln\left(\frac{P_j}{1-P_j}\right) = \alpha_0 + z_j\beta \quad (3)$$

In the expressions above, P_j is the predicted probability of being innovator for the jth company, and z_j represents the set of regressors: company size, geographic location, R&D engagement, liquidity, financial autonomy, value added and profitability. Companies with P_j values above 0.5 are classified in the innovators group, and companies with P_j values below 0.5 in the followers group. The classification accuracy of the model is obtained by comparing the actual with the predicted status.

3- The data

This paper used a sample of 90 companies corresponding to major Portuguese contractors laboring on public works. The longitudinal assessment reported covers the time period between 1996 and 2007. The data used came from the database of the 500 larger Portuguese companies, in terms of value of sales, published every year in the *Exame* magazine. The construction companies sampled vary from one year to the other, which is explained by the volatility of the market. In particular, new companies enter the market, and others may cease activity, merge or not provide data in some years. This implies that the assessment explores an unbalanced panel data sample.

The performance assessment model specified in this paper characterizes companies mainly from a financial perspective. The indicators used are: i) Profitability - to measure the profit of the company before tax and interest; ii) Value added - to measure the value-added per employee; iii) Financial Autonomy - to measure the contribution of equity on company funding; iv) Liquidity - to measure the company capacity to face short-term commitments. The value added indicator was deflated using the Gross Domestic Product (GDP) deflator, available in the World Bank's World Development Indicators database (http://data.worldbank.org/), considering 1996 as the base year. The last two indicators have minimum obligatory requirements to allow contractor activity in Portugal (at least 110% on Liquidity and 15% on Financial Autonomy).

4- Empirical Results

4.1 Performance Assessment

The first stage of the assessment was intended to explore whether the performance level of Portuguese construction companies improved over time. This required the estimation of a composite performance indicator for each company in each year, which was based on a comparison with a pooled frontier representing the best practices observed in the 12 years analyzed.

Table 1 reports the summary results for the composite indicator obtained using model (1). It also indicates the number of companies with a composite indicator equal to one in each of the years. These observations can be considered the benchmarks, as they correspond to the best financial results for the years studied. The bootstrap results are also shown in the middle panel of the table, including the average of the bias-corrected estimates, bias, standard deviation, and 95% confidence intervals for the bias-corrected composite indicator in the corresponding year.

Table 1 – Results of original and bootstrapped performance estimates

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Original Sample												
Performance score	0.555	0.563	0.591	0.618	0.636	0.617	0.650	0.694	0.682	0.697	0.705	0.678
No. of benchmarks	0	0	0	0	0	0	0	2	1	1	6	1
Bootstrap Sample	0.530	0.533	0.560	0.588	0.605	0.588	0.617	0.653	0.642	0.648	0.650	0.639
Bias-corrected score	0.025	0.030	0.031	0.030	0.031	0.029	0.033	0.041	0.039	0.049	0.055	0.039
Bias	0.025	0.030	0.031	0.030	0.031	0.029	0.033	0.041	0.039	0.049	0.055	0.039
St. Dev.	0.011	0.013	0.014	0.013	0.014	0.013	0.014	0.018	0.017	0.021	0.022	0.016
Lower Bound	0.514	0.516	0.543	0.571	0.587	0.570	0.598	0.633	0.623	0.629	0.633	0.619
Upper Bound	0.549	0.555	0.584	0.611	0.628	0.610	0.641	0.684	0.673	0.686	0.694	0.668

From the bootstrapped results, we can observe that the bias-corrected estimate is within relatively narrow confidence intervals, i.e. the lower and upper bounds of the intervals are relatively close. This provides statistical confidence for the bias-corrected estimate. Moreover, as the estimated bias is much larger than the standard deviation for all the companies, we conclude that the bias-corrected estimates should be preferred to the original estimates, as explained in Fried et al. (2008, p.467). Thus, the bias-corrected estimates were used in all analysis reported hereafter.

From Table 1 we can conclude that the performance estimates exhibited an increasing trend across years, although in 2001, 2004 and 2007 the performance levels suffered a decline in relation to the previous year. Comparing the average bias-corrected performance score of 2007 with the value of 1996, we can conclude that the financial performance of the Portuguese construction companies increased approximately 20% during this period, which is a remarkable improvement. It is also interesting to note that although the confidence intervals from consecutive years overlap, it is possible to observe a significant performance improvement between the first years analyzed (from 1996-1999) and recent years (from 2003-2007). This leads us to the conclusion that the Portuguese construction companies improved their financial performance during the period analyzed, at a statistically significant degree.

4.2 Drivers of good performance

The purpose of this section is to explore the determinants of good performance. In particular, we explored the impact of the economic context, geographic location, company size and engagement on R&D projects on company financial performance. The national economic context was characterized by GDP per capita, measured in US dollars. The data was taken from the World Bank's World Development Indicators database. This variable was deflated using the GDP deflator, considering 1995 as the base year. The regression model included two factors related to GDP: the GDP per capita of the same year as the composite performance indicator used as dependent variable, as well as the GDP per capita of the previous year. The one year lag was included in the regression model to test the hypothesis that the economic context of a given year affects the industry performance in the following year. The company size variable was measured by the value of sales, in real terms. The square of the value of sales was also included in the model to allow for the existence of variable returns to scale. The geographic location of companies corresponds to the location of headquarters. To explore if companies with headquarters located in the capital or its vicinity outperform the other companies, we considered a dummy variable distinguishing this region from the others. The engagement on R&D projects corresponds to a dummy variable distinguishing between “engaged” or “not engaged”

on R&D projects. This qualitative information was provided by CI academic experts routinely involved in conducting R&D projects in many CI companies, with a deep knowledge of the Portuguese CI and its players. Table 2 reports the estimates from the panel data truncated model, the coefficients, standard errors, and p-values. The total number of observations included in the model was 494, corresponding to all observations analyzed in all years. Note that we also used time and company dummies to control for time and company effects. The overall regression model was found to be statistically significant (χ^2 test with p-value of 0.000).

Table 2 – Regression results for the drivers of performance

Variable	Coefficient	Standard error	*p*-value
Size	-0.114	0.022	0.000
Size squared	0.014	0.003	0.000
GDP	-0.014	0.011	0.206
GDP (1 year lag)	0.039	0.014	0.005
Lisbon area	0.035	0.085	0.682
R&D engagement	0.018	0.111	0.873
Constant	0.403	0.097	0.000

From the results of Table 2, we can conclude that the national economic context has an impact on the performance of CI companies with a lag of one year. This impact is positive, meaning that a higher GDP promotes better performance of the CI one year latter. The current economic context has no significant impact on the financial performance of companies. Company size is related to performance nonlinearly, as both the first and the second order coefficients are statistically significant. Performance is U-shaped in relation to size. In particular, performance first decreases as company size increases, and for a value of sales greater than 421 million euros, performance starts to increase as size further increases. This means that the best performance levels tend to be achieved by small specialized companies, as well as by large companies. We reject the hypothesis that the engagement on R&D projects has a direct impact on financial performance. However, the long-term impact of R&D engagement on performance still remains an open question. The geographic location is not influential as well. This means that companies with their headquarters located in the capital or its surrounding area did not outperform the other companies. This could be expected due to the small dimension of Portugal, enabling a nationwide company activity.

4.3 Companies that innovated

This section explores innovation among CI companies in the period 1996 to 2007. Innovation in the context of this study reflects the ability of the CI companies to shift the frontier of financial achievements to better levels than those observed in previous years.

This analysis revealed that 22 companies from the 90 included in the sample were considered innovators in at least one year. Several companies (10) were considered innovative just in one year. Only 6 companies were considered innovators in more than half the years for which data was available for their evaluation. This suggests that keeping in the cutting edge of innovation for long periods is a difficult task. Nevertheless, it is interesting to note that there were innovators in all years studied. The years with the lowest innovation score were 2002 and 2007, both with a score of 1.05. These years coincide with some of the years that occurred a performance decline in the industry. Concerning the years with the highest innovation score (1997 and 2004), both with a score of 1.44, it can be concluded that a good innovation score for a few companies does not necessarily coincide with years of particularly good performance for the whole industry. Analyzing the performance pattern of innovator companies over time, we can conclude that high levels of performance typically precede the innovative status of companies. In particular, the companies that were innovative more than once during the period studied (12) were able to keep high levels of performance in most years for which data was available.

Next, we explored the factors that promote innovation. We used a logistic regression, where the dependent binomial variable represented the company status: innovators (1) or followers (0). The regressors were company size, location, R&D engagement, as well as the average value of the KPIs related to the financial status of the company in the period analyzed (i.e., the average value of liquidity, financial autonomy, value added and profitability). Table 3 reports the estimates from the logistic model, the coefficients, standard errors, Wald statistics, and p-values.

Table 3 shows that company size, R&D engagement, and location are not statistically significant at a 5% level. Concerning the financial status, the most critical factor to promote innovation is profitability, followed by value added. The overall model is statistically significant (χ^2 test with p-value of 0.000), with a Nagelkerke R^2 of 39.9%. The overall percentage of observations classified correctly was 82.2% ((10+64)/90). This value is considerably larger that the proportional chance criterion of 39.8% ($(22/90)^2+(68/90)^2$), which shows that the logistic regression is a good option to model the factors that promote innovation.

Table 3 – Logistic regression estimates

Variable	Coefficient	Standard error	Wald statistic	*p*-value
Size	-0.269	0.468	0.329	0.566
R&D engagement	0.151	0.668	0.051	0.822
Lisbon area	0.403	0.636	0.402	0.526
Profitability	0.390	0.177	4.850	0.028
Value added	0.063	0.035	3.198	0.074
Financial autonomy	0.025	0.042	0.370	0.543
Liquidity	0.012	0.009	1.875	0.171
Constant	-6.928	2.071	11.194	0.001

5- Conclusions

The purpose of this paper was to develop a quantitative approach to evaluate the financial performance of construction companies, and to identify the factors that promote performance improvements and innovation. To assess contractors performance we used a composite indicator calculated using DEA, and complemented with bootstrapping to obtain a robust estimate of performance. The study revealed that the performance of Portuguese construction companies increased during the years analyzed. Regarding the factors considered more influential, we concluded that company performance is affected by the national economic context with a one year lag, and that small specialized companies and large companies tend to achieve the best performance levels.

Concerning the assessment of innovation, this paper proposed an enhanced approach to identify innovative companies, to quantify the extent of innovation, and to explore the factors that drive innovation. It was concluded that keeping the innovative status for long periods is difficult, although a few companies were able to keep their innovative status in consecutive years. It was also found that high levels of performance typically precede the innovative status of companies. Concerning the factors that promote innovation, the results of the assessment suggest that the innovators are typically companies with high profitability.

5- References

Bassioni, H.A., Price A. D., and Hassan T. M. (2004). "Performance Measurement in Construction," *Journal of Management in Engineering* 20 (2): 42-50.

Charnes, A. Cooper W. W., and Rhodes E. (1978). "Measuring Efficiency of Decision Making Units," *European Journal of Operational Research* 2 (6): 429-444.

Cherchye L., Moesen W., and Puyenbroeck T.V. (2004). "Legitimately Diverse, yet Comparable: On Synthesizing Social Inclusion Performance in the EU," *Journal of Common Market Studies* 42(5): 919-955.

Costa D., Formoso C., Kagioglou M., Alarcon F. and Caldas H. (2006). "Benchmarking Initiatives in the Construction Industry: Lessons Learned and Improvement Opportunities.," *Journal of Management in Engineering* 22 (4): 158-167.

Egan J. (1998). *Rethinking Construction.* Technical report, Dept. of the Environment, Transport and the Regions, London.

El-Mashaleh M., Minchin R., and O'Brie W. (2007)., "Management of Construction Firm Performance Using Benchmarking," *Journal of Management in Engineering* 23 (1): 10-17.

Fare R., Grosskopf S., Norris M., and Zhang Z. (1994). "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries," *The American Economic Review* 87(5): 1040-1043.

Fried H.O., Lovell C.A.K., and Schmidt S.S. (2008), *The measurement of productive efficiency and productivity growth.* Oxford University Press US, New York.

Horta I.M., Camanho A.S., and Moreira Costa J. (2010). "Performance Assessment of Construction Companies Integrating Key Performance Indicators and Data Envelopment Analysis," *Journal of Construction Engineering and Management* 136 (5): 581-594.

Simar L. and Wilson P.W. (1998). "Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models," *Management Science* 44(1): 49-61.

Tulkens H. and Vandeneeckaut P. (1995). "Nonparametric efficiency, progress and regress measures for panel-data - methodological aspects," *European Journal of Operational Research* 80(3): 474-499.

Wilson P.W. (2008). "FEAR: A package for frontier efficiency analysis with R," *Socio-Economic Planning Sciences* 42(4): 247-254.

Optimal Network Design of Petroleum Supply Chains

Leão José Fernandes[ab], *Susana Relvas* [b*], *Ana Paula Barbosa-Póvoa* [b]

[a] *CLC-Companhia Logística de Combustíveis, EN 366, Km 18, 2050 Aveiras de Cima, Portugal*
leao.fernandes@clc.pt

[b] *CEG-IST Technical University of Lisbon, Av.Rovisco Pais, 1049-001Lisboa, Portugal*
susanaicr@ist.utl.pt; apovoa@ist.utl.pt

[*] *Corresponding Author*

Abstract

The Petroleum Supply Chain (PSC) being a strategic industry involves substantial planning of huge investments. Also specialized infrastructures are needed to fulfil the logistic requirements of the dynamic market of petroleum products. This research focuses at the design level and presents a mixed integer linear programming (MILP) model to design an optimal downstream PSC network. A multi-echelon system is represented and optimized through a profit objective function that incorporates the PSC member revenues into entity profits thus engaging collaborative operation between petroleum companies, refineries, bulk transport, storage, transporters and retailers. Test results based on representative data of Portugal are presented, and research directions proposed.

Keywords: petroleum supply chain, network design, strategic planning, collaborative.

1- Introduction

The downstream PSC is a major economic sector that surmises production and delivery of petroleum products to consumers in a cost-efficient manner. Here, a key issue is to determine the optimal network design using geographical information and admitting supply, demand, transport and cost uncertainty. Until recent times, PSC networks structures were fairly static and scarcely investigated due to the lack of detailed information and market stability. However, fresh competition and market uncertainty have put the stress on strategic PSC planning and re-design for cost reduction and profit maximization of PSC networks. This has motivated the development of the present study where the prime objective is to help companies design and evaluate PSC networks

so as to improve overall efficiency at the strategic planning level. The main literature on PSC and Process Industry network planning is analysed in this section so as to identify opportunities for the present work.

A two-stage modeling framework is proposed in [Escudero et al., 1999] for scheduling of supply, transformation and distribution in a multi-operator multi-product and multi-period downstream PSC with product demand, spot supply cost and spot selling price uncertainty. [Más and Pinto, 2003] present a two-stage model for continuous-time scheduling for supply problems in petroleum distribution complexes to minimize operating costs. The first MILP solves the assignment problem of oil tankers to piers, and tanker to pipeline operations where the pipeline and intermediate tankage are aggregately represented. The second MILP schedules the tanks and pipelines operations. A multi-period MINLP model is developed in [Neiro and Pinto, 2004] for production planning of a large scale PSC that includes crude oil supply, refinery operations and product transportation. The PSC network is built by connecting PSC nodes representing refineries, terminals and pipeline networks which are structured as groups of elementary entities like storage tanks and pipelines. The MINLP decisions include stream flow rates, properties, operational variables, inventory and facilities assignment.

A multi-period stochastic planning model is proposed in [Al-Othman et al., 2008] for a PSC network, consisting of all activities related to crude oil production, processing and distribution. The two-stage deterministic optimization/ stochastic formulation approach with finite number of realizations is seen to be effective in developing resilient production plans maximum resource utilization and maximum profitability with market demands and price uncertainty. [MirHassani, 2008] considers a capacitated network for the PSC and proposes a linear programming model for operational planning of the transportation network between the existing refineries and depots to meet the uncertain demand minimizing total inventory and transportation costs.

An integrated model is presented in [Kim et al., 2008] for collaborative production and network planning among multi-site refineries. A non-linear model determines the optimal production rates at the refineries, and a MILP model selects locations of distribution centers and transport routes for maximizing profit. More recently, [Al-Qahtani and Elkamel, 2010] propose a two-stage stochastic MILP model for the multi-

site refinery network coordination under uncertainty that allows planning of capacity expansions. They employ the sample average approximation (SAA) method with a statistical bounding and validation technique. The resulting refinery scheduling plans are seen to be stable with crude oil and product prices and product demand variations.

Although these models focus the PSC, they are mainly concerned with the refinery operations planning and scheduling, identifying relatively few references in the downstream PSC and more so for the PSC network design. This led to our development of research related to PSC network structures. The novel MILP presented in this paper, considers a general multi-echelon formulation for optimal design of downstream PSC networks. The model includes company participations in refineries, transportation and depots and other specificities, so as to represent a novel PSC dynamics through the representation of different actors. Thus, the strategic MILP models a PSC network with multiple country, entities, products, and transportation modes for collaborative design and operation aimed at maximizing the total profits for all entities. The following sections present the problem description, the MILP formulation, the PSC network in Portugal, and test results providing insights and future research directions.

2- Problem Description

A PSC network is defined as a network of petroleum companies, refineries, distribution depots, transportation modes and routes that satisfy the retail customer demands for petroleum products in various regions. Fig. 1 presents the downstream PSC network, responsible for the refining, distribution and retail of the petroleum products.

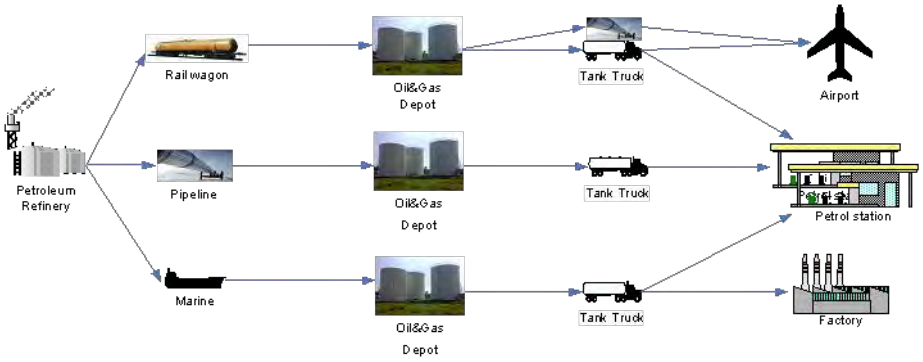


Figure 1: The downstream Petroleum Supply Chain

The strategic planning problem of the downstream PSC design is defined as follows:

Given a set of petroleum companies that produce or import a set of petroleum products at a known set of refineries, and commercialize to a set of customer regions;

Determine the optimal multi-echelon, multi-company and product PSC network consisting of installation or closure of storage depot locations, connecting transportation routes and modes, and the allocation of quantities for refinery production, trading and export, depot transfers and imports, and customer fulfilment;

Subject to customer demand and production, storage and transportation capacity restrictions;

In order to maximize the total network profit, comprising the oil companies' profits and their shared revenue in the participated infra-structures along the network.

A further characterization and assumptions of the problem is performed in order to understand the basis of the model. Specifically, the downstream PSC problem begins with the refineries that buy crude oil, use refining processes to produce petroleum products and trade them to the petroleum companies. The crude flow and trade that occur before the refinery are not considered within the scope of this problem.

The PSC downstream network considered consists of various activities comprising the primary distribution, depot storage, secondary distribution and the retail stages. The problem considers that the prices, tariffs and cost structures are defined for each one of these stages, in order to obtain the optimal network design for the network's total profit maximization. Entity profits can be calculated, from the computation of individual revenues, fixed installation costs, operating costs and unitary volume dependent costs while considering entity share on the installation or activities.

The refineries produce the petroleum products from crude oil. Here the known ex-refinery prices and the refining margins are used to determine the production and crude oil cost for each product. The refinery may produce up to its production capacity per product, and either sell the product to the petroleum companies for national consumption at the free on board (FOB) price. Also they may export the product at the wholesale market price, in which case the refinery must support the cost, insurance and freight (CIF) charges to a wholesale distribution hub. The product exiting the refinery is under the ownership of a petroleum company and will stay so until the retail stage.

The petroleum companies buy the products from the refineries at the FOB wholesale market price and transport the product up to a storage depot of convenience in the region considered. This activity is known as the primary distribution which involves the refinery to depot and inter-depot transportation, and is generally undertaken using the pipeline, maritime or railway transport mode. The petroleum companies may also import the products from distant refineries where again the FOB prices are considered and the companies must support the CIF charges up to the depots.

The storage depots thus receive the product from the refineries on behalf of the companies and will hold the product at their storage facilities until the product is expedited by the petroleum companies to their retail clients. Depot location and capacity sizing is critical in order to reduce relative fixed costs, by maximizing the number of stock rotations, thereby resulting in higher efficiency. These existing infra-structures may be operated or closed. In either case fixed costs must be considered if the infra-structure is not yet amortized. Once the product sale is confirmed by the petroleum companies, the product is expedited by the secondary distribution of the PSC.

The secondary distribution transports the products from the storage depot to final customers. This transport is normally undertaken by road using tank trucks, where the actual travelling distance and travelling time is determinant to the route profitability. Occasionally, this distribution may also be undertaken using pipeline or railway transportation, when huge volumes are to be transported from the depot to a specific destination, such as the transport of JET A1 to an airport terminal.

The retail stage comprises of petrol stations, factories, airports, and hospitals which supply the final consumer. The retail segment receives the product by tank trucks, stores the product in local tanks and sells the products to the end-user, thereby recovering a profit margin from their commercial activity.

3- MILP Model

This novel MILP model performs the Strategic Planning for grassroots and retrofit design of multi-echelon, company and product PSC networks. The MILP maximizes the multi-echelon entity profits by aggregating the net profit margins along the PSC. All crude supply, refining, primary transport, storage, secondary transport, exportation,

retail and unsatisfied demand prices and costs are considered. The model's parametric data, decision variables, objective function and constraints are briefly presented.

The parameter data includes the actual infrastructure and capacity data for refineries, depots and transport routes; information on price, tariff, fixed, operation and unitary costs; capital and demand distributions; and data relating to node locations and distances. Typical capacity of storage tanks and the number of tank rotations (tank refills per year) are considered per product.

The model includes binary variables for strategic installation and operation decisions of the primary and secondary transportation and depot infra-structures. Integer variables decide the number of installed and operating tanks at installed depots. Positive variables determine the produced, traded, exported, imported, retailed and unsatisfied quantities at each location and the transported quantity on primary and secondary distribution routes.

The objective function maximizes the sum of all entity profits. This is the aggregation of sub-functions that calculate the refinery profits, exportation costs, the entities' primary, secondary distribution and depot net profits as part of gross margins, the retail sector profits and finally the entities' return of investment in dealer profits. Each of these objective function terms are calculated per PSC echelon node, entity and product. The total refinery profit is the product of the total production quantity (traded and exported), the refinery margin and the entity percentage participation in the refinery capital. Exported product, however, incur product transportation costs from the refinery to the central distribution hub. The primary and secondary distribution gross margins are calculated as the product of each route travelling distance, the sum of the fixed installation and operating costs and the product of transported quantities on this route and the difference between the volume dependent price and unitary transport mode cost. The same applies to the storage depots' gross margins, however here the costs are modularized by considering depot costs without tank capacity and then adding modules of typical capacity storage as required. The total retail margin is provided as the product of the difference between the retail price and tariff, and the difference between customer demand and the unmet demand quantities. The cost of unmet demand is incurred as lost revenue. Finally, at each activity stage, a proportion of the dealer margin is incorporated into entity profits as per their participation in the infrastructure capital. Company

margins and the dealer margins are calculated by defining a dealer price and a dealer tariff or in the case where the dealer tariff is not known, these can be obtained by defining a parameter for the dealer profit percentage, which is multiplied to the gross margins to compute the dealer and company net margins.

The model is governed by a set of constraints to secure the model feasibility. Each echelon verifies the specific mass balance for the period considered. The storage depot and tank capacity restrictions assume typical capacity tanks, fill factor and tank rotations per product. The primary and secondary distribution route constraints consider existing and allowed routes/distances per product.

Transportation restrictions play a big role in the selection of transportation modes for the primary and secondary distribution. These restrictions are modelled using distance restrictions. For instance, if road transportation is not permitted for primary distribution, than the allowable distance for primary distribution by road could be defined as zero. In the case of fuel oils, as there is a limitation to the distance that can be used for pipeline transportation, this distance restriction can again be used.

4- Case Study

A real case scenario based on the Portuguese PSC network was considered to test the novel MILP model. Fig.2 shows the PSC under study comprising refineries, distribution centres and customer zones in Portugal and includes supply from Spanish refineries.

Figure 2: The Portuguese PSC in an Iberian context

Two national refineries currently supply the petroleum products' consumption in Portugal. One is located on the north coast at Matosinhos and the other at the south coast at Sines. Portugal may also be supplied from eight regional refineries in Spain or from the Amsterdam-Rotterdam-Antwerp (ARA) refinery cluster.

The PSC market in Portugal is currently operated by five major petroleum companies that include multinational and national companies and other hypermarket chains. These companies purchase the oil products from the local refineries or import them from external refineries. The product is first transported using the primary distribution to nine currently operating storage depots and later on using the secondary distribution channels to the retail segment. The retail segment is organized as customer zones, represented in this case study as the eighteen districts in Portugal. The district capital Euclidian coordinates are used for the localization, land route and distance computation.

The PSC network considers the distribution of seven oil products in Portugal, including liquefied petroleum gases (LPG). However, particularities relating to the LPG bottle-filling and the archipelago of Madeira and Azores are not included in this study. Nevertheless, much of the local specifics are considered in this real case example. Pipeline, maritime or railway transportation is considered for the primary distribution while road transportation is assumed for the secondary distribution. Two exceptions to this rule are the existing pipeline and railway transportation for JET A1 to two of the airports, which are also considered in the test data. Maximum production rate is assumed and inventory holding is not considered, hence overproduction above the market requirements is exported while underproduction is imported.

The data includes refineries, which are owned by different companies. The primary distribution and storage depot infra-structures used are rented, partially-owned or fully-owned by single or multi-companies. These may be operated by the company or by external entities. Companies may therefore privilege the utilization of self-owned than partially-owned infra-structures in order to maximize their overall profits. Retail stations can be categorised in a real world setting as company owned and company operated, company owned and dealer operated, and dealer owned and dealer operated and finally hypermarket establishments. However, the model does not yet consider individual stations, and as such currently cannot include this feature to take advantage

of station-level characteristics. Nevertheless, company quota distribution per entity is known and is used to analyse the network under different settings.

In order to perform this study, a statistical database is constructed from the commercial and quantitative data available at official organizations, namely Direcção Geral de Energia e Geologia, Instituto Nacional de Estatística and European Commission Energy, geographical data from Google maps, and logistics data from international, company and regulatory authority sites and publications. All data was gathered for a period comprising the years from 2007 to 2010.

The MILP is tested for two scenarios, namely the individual (A) and collaborative (B) strategies under the Portuguese PSC configured as a multi-company competitive network. For each scenario three situations are analyzed: (1) current network, (2) grassroots network, (3) retrofit network. The individual strategy models the PSC without incorporating participated infra-structure profits, while the collaborative strategy incorporates the profits from company owned distribution infra-structures. The current network situation assumes currently existing transport structures and depot installations as in (A) and (B); the grassroots situation designs the network from scratch without considering the existing transportation and depot infrastructures; and finally the retrofit situation designs an optimized network considering existing installations which may be operated, closed or reconfigured. A five year planning horizon considers all volume dependent revenue, unitary and operating costs however the fixed costs are incurred only as long as the transport or depot infrastructure is not fully amortized.

The MILP is modelled using GAMS 23.5 and solved with MILP commercial solver CPLEX 11.2 using 4 parallel threads on an Intel Core i7 2.67 Ghz, 4 GB RAM, 64-bit Windows 7 computer.

Table 1 presents the connections obtained in each model run for the six situations under study, showing the selected primary distribution connections. D1-D11 are depot while R1-R3 are refineries, being R1 and R2 Portuguese refineries. The primary distribution network in Table 1 selects mostly supply from the Portuguese refineries. Additionally, one product is imported from a Spanish refinery in scenarios A1, A2 and A3 and also for scenario B3 however for a smaller quantity. This is observed for the product P5

where the refinery capacity is short. Most transport is undertaken by sea and pipeline, and rail transport is not selected in the grassroots and retrofit design.

Table 1. Primary distribution network

SCENARIO	INDIVIDUAL									COLLABORATIVE						
REFINERY	CURRENT			GRASSROOTS			RETROFIT			CURRENT		GRASSROOTS		RETROFIT		
/ DEPOT	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R1	R2	R1	R2	R3
D1	PIPE		SEA	PIPE		SEA	PIPE		SEA	PIPE		PIPE		PIPE		SEA
D2	PIPE						PIPE			PIPE				PIPE		
D3	SEA	SEA									SEA					
D4					PIPE											
D5		PIPE					PIPE				PIPE				PIPE	
D6													PIPE			
D7		SEA									SEA					
D8		SEA														
D9		SEA		SEA	SEA PIPE		SEA	SEA			SEA					
D10		PIPE			PIPE			PIPE			PIPE		PIPE		PIPE	
D11		RAIL									RAIL					

The number of depots and tanks to install and operate per depot/product combination are shown in Table2. Both depots and tanks are significantly reduced from current (9 depots and 336 tanks) to retrofit networks (5 depots and 149 tanks) to grassroots (4 depots and 149 tanks). Further reduction of tanks can also be seen from individual to collaborative strategy. The model identifies the capacity needed for the downstream PSC. However this reduction should be observed critically in the case of refinery depots, D1 and D10, as the liberated storage capacity will include that available to refinery operations. The PSC design capacities would be higher if the inventory holding and transport durations were to be modelled. Both these features are identified as opportunities for future research.

Table 2. Depot/tank configurations

Scenarios		INDIVIDUAL STRATEGY							COLLABORATIVE STRATEGY						
	Dep/Prod	P1	P2	P3	P4	P5	P6	P7	P1	P2	P3	P4	P5	P6	P7
Current	D3			2	2	2					2	2	2		
	D11	5	5						5	5					
	D5	8	6	4	3	8		4	8	6	4	3	8		4
	D1			22	11	16	33	11			22	11	16	33	11
	D8			2	2	1					2	2	1		
	D9			2	1	3					2	1	3		
	D10	18	11	26	15	21	47	15	18	11	26	15	21	47	15
	D7	2	2			3			2	2			3		
	D2	15	8						15	8					
Grassroots	D4	3	7	3	1	9									
	D6									1	8	1	16		7
	D1	10	17	8	2	8	6	2	12	19	9	2	8	6	3
	D9	15	10	7	1	8	10	8							
	D10	2	4	2	1	2		3	18	18	2	1	2	10	3
Retrofit	D5	8	8	7	1	12		7	1	6	7	1	13		7
	D1		5	5	1	5	3	2		12	9	2	8	6	3
	D9			2	1	3	10	1							
	D10	11	11	2	1	2	1	3	18	12	4	1	5	10	3
	D2	11	14	4	1	4	3		11	8					

The above identifies excess capacity in the current PSC, mainly related to smaller depots. The more efficient designs, specifically the grassroots and retrofit, as verified later, show a tendency for consolidation to current high volume depots and the closure of low volume depots due to inefficiencies resulting from economies of scale.

Table 3 provides the cost structure and the multi-echelon profits (MILP solution) of the resulting networks given in million monetary units (m.m.u.). The grassroots networks are characterized by huge fixed costs as they need to be newly constructed, whereas the unitary costs are lower in the grassroots and retrofit networks. As in the collaborative networks, entities incorporate revenues and costs, thereby profits from the distribution networks, hence higher unitary costs are observed, however these are compensated by higher revenues as can be observed on the profit line.

Table 3. Scenario total profits and distribution costs

	Scenarios	Individual			Collaborative		
Distribution	Costs	Current	Grassroots	Retrofit	Current	Grassroots	Retrofit
Primary	Fixed	3	172	3	3	61	3
	Unitary	270	131	194	91	52	60
Depots	Fixed	175	140	35	175	105	0
	Operation	63	28	35	63	21	28
	Unitary	84	86	86	83	86	85
Tanks	Fixed Exist	92	0	92	92	0	92
	Fixed new	0	396	102	0	391	77
	Operation	141	79	79	141	78	78
Secondary	Fixed	0	0	0	0	0	0
	Unitary	1069	956	963	1298	1260	1168
Distribution (- fixed) costs		1626	1280	1356	1676	1497	1420
Total Distribution costs		1896	1988	1588	1946	2053	1591
Multi-echelon profits		2230	2288	2489	2458	2481	2881

If only distribution costs were to be accounted, hence without considering fixed installation costs, the grassroots network would be the most efficient under individual strategy and retrofit under collaborative strategy. However considering the multi-echelon profits, the retrofit is the most efficient network under both strategies with the highest profits earned under the collaborative strategy. The above results show a general cost reduction from the individual strategy to the collaborative strategy, only to be reversed in the secondary distribution. This is believed to be due to the model considering a static parameter for the company to dealer profit distribution independent of the PSC activity. A stage dependent parameter estimation is expected to improve the revenue and cost imputation accuracy. Model improvements will consider the insights and difficulties observed in the test results of the current MILP.

6- Conclusions

This paper presents a novel MILP model for strategic planning of downstream PSC networks that determines optimal routes, depot locations, capacities, transport modes and network affectations. A critical novelty in this model is the representation of the collaborative revenue sharing, accomplished by specifying company prices, dealer tariff and costs for the refining, distribution, storage depot and retail PSC activities, thereby incorporating stage profits, formulated per entity, product, location and transport route.

The model was tested using the data for the Portuguese PSC to design the current, grassroots and retrofit networks for individual and collaborative strategies. The results are compared and new insights obtained. Future investigation will focus on modelling of multi-period system evolution, long-term net present value and uncertainty.

Acknowledgements

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

7- References

Al-Othman, W.B.E., Lababidi, H.M.S., Alatiqi, I.M., Al-Shayji, K. (2008). Supply chain optimization of petroleum organization under uncertainty in market demands and prices. *European Journal of Operational Research*, 189, 3, 822-840.

Al-Qahtani, K., & Elkamel, A. (2010). Robust planning of multisite refinery networks: Optimization under uncertainty, *Computers and Chemical Engineering*, 34, 985-995.

Escudero, L.F., Quintana, F.J., Salmerón, J. (1999). CORO, a modeling and algorithmic framework for oil supply, transformation and distribution optimization under uncertainty. *European Journal of Operational Research*, 114, 638-656.

Kim, Y., Yun, C., Park, S.B., Park, S., Fan, L.T. (2008). An integrated model of supply network and production planning for multiple fuel products of multi-site refineries, *Computers and Chemical Engineering*, 32, 11, 2529-2535.

Más, R., Pinto, J.M. (2003). A Mixed-Integer Optimization Strategy for Oil Supply in Distribution Complexes, *Optimization and Engineering*, 4, 23–64.

MirHassani, S.A. (2008). An operational planning model for petroleum products logistics under uncertainty, *Applied Mathematics and Computation*, Vol. 196, 744-751.

Neiro, S.M.S., Pinto, J.M. (2004). A general modeling framework for the operational planning of petroleum supply chains. *Computers & Chemical Engineering*, 28, 871-896.

Developing Methods to Evaluate Benefits and Costs of Arterial Spin Labeling in Alzheimer's Disease

Manuel Rodrigues[a,*], *Mónica Oliveira*[a], *Teresa Peña*[a].

[a]*Instituto Superior Técnico - UTL, Lisboa, Portugal*
manuel.t.rodrigues@ist.utl.pt; monica.oliveira@ist.utl.pt; teresapena@ist.utl.pt

*- *Corresponding author*

Abstract

In this study we developed methods to carry out an economic evaluation of the Arterial Spin Labeling (ASL) MRI technology in the diagnosis of Alzheimer's Disease (AD). Methods were designed to inform about the potential benefits and costs related to the adoption of this new technology. After surveying all relevant literature and data, a decision model based on a Markov cycle tree was built and applied to compare costs and benefits related with four alternative diagnostic strategies: "PET", "MRI-DSC", "MRI-ASL" and "No Exam". By using available input data from diverse studies and by taking, when it existed, Portuguese data, the model allowed us to calculate incremental cost-effectiveness ratios between the different diagnosis strategies. Although the model is sensitive to MRI-ASL cost, to MRI-ASL performance values or to the treatment time span, we found evidence for MRI-ASL to be the dominating strategy, being cost effective in comparison to all concurrent strategies, even when the uncertainty for several parameters is taken into account by using a Monte Carlo simulation method.

Keywords: Arterial Spin Labeling, Alzheimer's Disease Diagnosis, Economic Evaluation.

1- Introduction

In recent years, imaging technology based on Arterial Spin Labeling (ASL) in Magnetic Resonance Imaging (MRI) brought a new window on perfusion physiology and became a promising tool to accurately assess brain hemodynamics. Although ASL has been clinically validated, its adoption as a new technology in health care systems requires economic evaluation [Drummond et al., 2005]. This study aims to understand how the clinical use of ASL might generate added benefits and costs in the diagnosis of Alzheimer's disease (AD) and to compare those net gains with the ones from alternative

technologies. We chose the AD case for our study since it is expected to be a major health concern in the next decades, due to the expected ageing of population. Moreover, it was clinically demonstrated that AD diagnosis can be performed with complementary information provided by imaging technologies such as MRI-ASL [Asllani et al., 2008] or MRI-DSC (Dynamic Susceptibility Contrast MRI) [Harris et al., 1998].

2- Literature Review

ASL involves tagging of inflowing blood before it reaches the imaging plane [Jezzard et al., 2002]. Its main advantage is the absence of injection, while the major drawback is a low signal-to-noise ratio. ASL can provide perfusion images interpretable in a wide range of clinical settings in the central nervous system: from tumors to degenerative diseases as AD. The ASL technology in Portugal has been commercially brought into play by SIEMENS S.A and is now available, for example, in "Hospital da Luz". Other imaging techniques which measure brain perfusion differ on tracers used, on technical requirements, in the quantitative accuracy, etc [Wintermark et al., 2005]. For example, Positron Emission Tomography (PET) demands an intravenous injection. Its main benefit is quantitative accuracy, despite entailing a high cost. Dynamic Susceptibility Contrast-enhanced (DSC) MRI also uses an endovascular tracer, having as a setback the lack of standardization in its results interpretation.

MRI techniques, as DSC or ASL, have the great advantage of avoiding exposure to nuclear radiation. Another key feature is the quantitative accuracy available only in PET and in ASL. The latter technique also has the unique feature of avoiding an intravenous injection, by performing contrast with a pulse rather than with contrast injected.

Dementia is described as a global progressive decline in memory and other brain abilities as thinking or understanding language. It consequently impairs the successful performance of activities of daily living [Fauci, 1998]. Within dementia, Alzheimer's Disease (AD) is the most frequent case. It is a neurodegenerative disease characterized by amyloid plaques, neurofibrillary tangles and progressive loss of neurons.

Although not all AD patients do suffer the same sequence of symptoms, it is common to identify a global disease evolution pattern with three different states: Mild, Moderate and Severe AD. At early stages, patients lose memory of recent events, while in a

Moderate AD state there is an escalation of symptoms as patients can feel disoriented and become unable to make proper judgments. One loses autonomy, depending on others to speaking, walking or eating. A patient with Severe AD has lost nearly all cognitive skills, becoming incapable to use or understand language. No cure for this disease has been discovered, but treatment, as *Donepezil*, may increase quality of life. Nevertheless, whichever the treatment is, it is difficult to slow disease's progression.

AD diagnosis is essentially a diagnosis of exclusion. Usually it includes taking a history of mental and behavioral symptoms by speaking with patient and his family; physical and laboratory tests; neuropsychological tests as the Mini-Mental State Examination and some imaging exams [Geldmacher et al., 1997]. These exams as MRI or computed tomography (CT) provide structural information that can, for example, rule out tumors. The National Institute of Neurological Disorders and Stroke–Alzheimer Disease and Related Disorders Association outlined AD diagnosis criteria, reporting a sensitivity of 81% and a specificity of 70% for a diagnosis of ~~"p~~robable AD" [Knopman et al., 2001].

Various studies describe the impact of ASL on AD diagnosis, even if focusing more on benefits than on costs. They typically report a hypoperfusion pattern in areas as parietal, cingulated and frontal cortices. [Asllani et al., 2008] identified such pattern that might help AD detection. In fact, the sensitivity and specificity obtained were remarkable (100% and 95% respectively). Conversely, a study focusing on AD early detection [Raji et al., 2010], achieved lower values for sensitivity - 85% – and for specificity – 54%.

Analogously PET's quantitative accuracy also makes it suitable for AD diagnosis. The accuracy of standard AD diagnosis is improved as showed in [Silverman et al., 2001], which reported a sensitivity of 94% and a specificity of 78%, not differentiating results by AD severity. PET has though a high cost associated. MRI-DSC may also support AD diagnosis, achieving a sensitivity of 88% for Mild AD and 95% for Moderate AD cases, whereas the specificity stood on 96% [Harris et al., 1998]. Some stress DSC dominance over PET [Harris et al., 1998, McMahon et al., 2003], describing DSC as the current most cost-effective imaging technique for AD diagnosis. However, due to controversy about this evidence, we included PET in this analysis. Still in terms of DSC, [McMahon et al., 2000] in an economic evaluation considered its total costs as being on the limit of what is acceptable to pay for the correspondent improvement in health.

A review paper on existing decision analytic models for AD [Cohen et al., 2008] acknowledged three studies evaluating screening interventions [Kulasingam et al., 2003, McMahon et al., 2000, McMahon et al., 2003] . Nevertheless, up to our knowledge no economic evaluation studies have assessed the ASL technology and compared it with older technologies, and our study aims to make a start on this.

3- Methods

Our goal is to understand how the AD diagnosis process is influenced by different strategies. We aim to establish a method that can, as far as possible, represent current and common practice at hospital units. The framework adopted is represented in Figure 1, which shows all the stages of the model building process, and a Cost-Utility Analysis (CUA) was selected, as it uses a generic measure of outcome potentially comparable across studies, while it is adequate when the program affects both morbidity and mortality [Drummond et al., 2005].

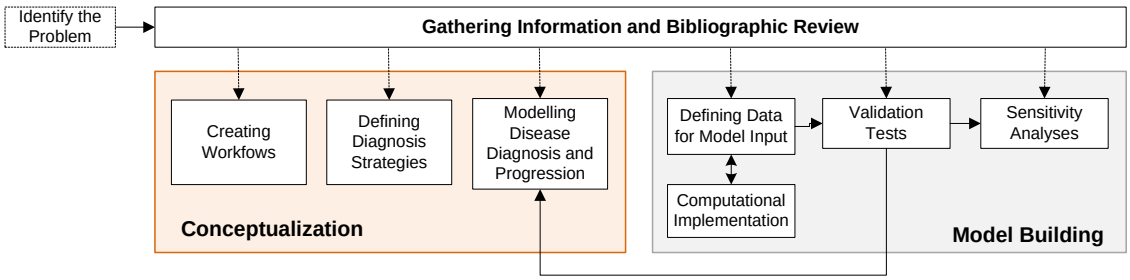


Figure 1- Model Building Process

3.1 – Conceptualization

To define our model, through which the economic analysis should be performed, we have to become familiar with the disease's progression and to survey resources involved. For this purpose workflows are used (e.g. in Figure 2), *i.e.* diagrams based on the literature review that highlight patients' pathways inside and outside the clinical setting. Patients move within these states, which have costs associated and resources utilized. A diagnosis strategy results in one of four options (No AD, Possible or probable AD) that do not include "Certain Alzheimer", as it represents a diagnosis only available *post mortem*. When a patient is diagnosed with "Probable AD", physicians perform a severity assessment of AD: the patient is classified as having "Mild AD", "Moderate AD" or "Severe AD"— the last being highly improbable in a first diagnosis.

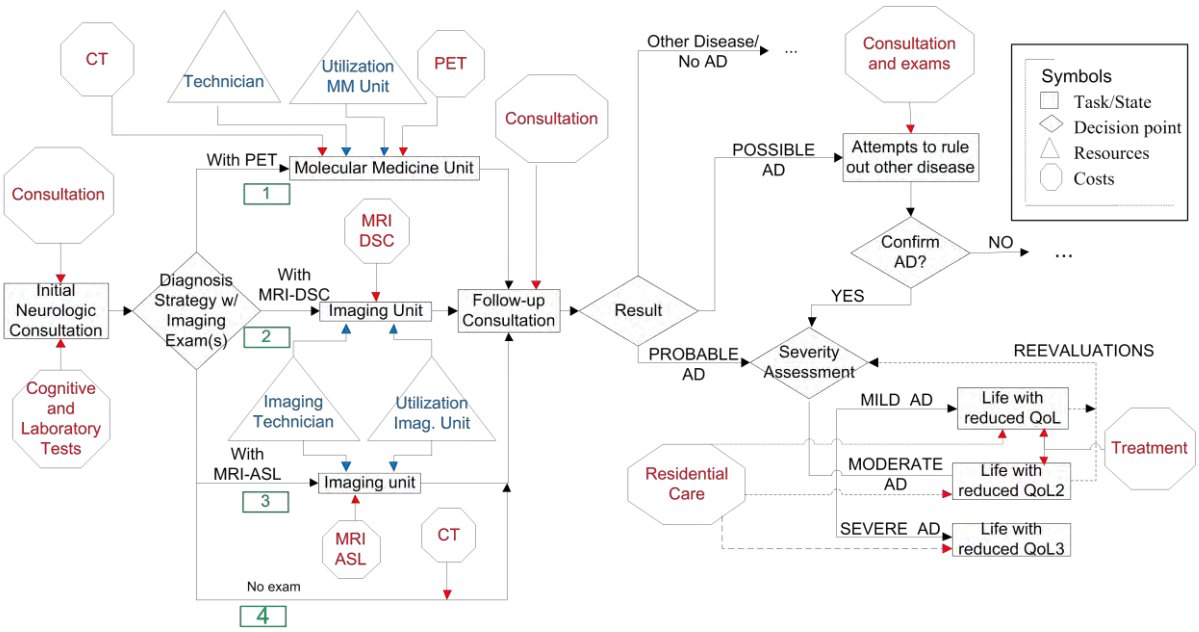


Figure 2 - Clinical workflow

Workflows were built for different diagnosis strategies. However, all strategies embraced taking a medical history of mental and behavioral symptoms, cognitive tests as MMSE, laboratory tests and structural brain imaging that could be either CT or MRI. This was labeled as standard examination. This analysis considered four concurrent strategies, having in common a standard examination and diverging on the functional imaging exam performed (in the fourth option this feature is absent):

1- "PET": Standard examination with a PET scan.
2- "MRI-DSC": Standard examination with MRI-DSC.
3- "MRI-ASL": Standard examination with MRI-ASL.
4- "No Exam": Standard examination.

A Markov cycle tree is used as a core model for this analysis, combining a Decision Tree (DT) and a Markov model [Briggs et al., 2006]. Such model is represented in Figure 3 and was developed from [McMahon et al., 2000, 2003]. A DT is a model that flows from left to right, showing individuals' possible prognoses in alternative branches [Drummond et al., 2005]. DTs are suitable for short-term events, whereas Markov models, which consist of a series of states patients can occupy, are more appropriate to a long-term range. Here, each Markov cycle is set to one month and patients move between states at a speed given by global transition probabilities. Patients have AD in a certain stage, but are also in a specific clinical setting: community or nursing homes. Disease stage-to-stage transition probabilities define the pace of progression through disease stages while setting-to-setting transition probabilities characterize changes

between different clinical settings. This last set of probabilities is conditional on disease stage, but stage-to-stage transitions occur independently of the setting. The target population contains patients in the community with mild AD moderate AD or with a non-AD related dementia that present themselves to a neurology unit.

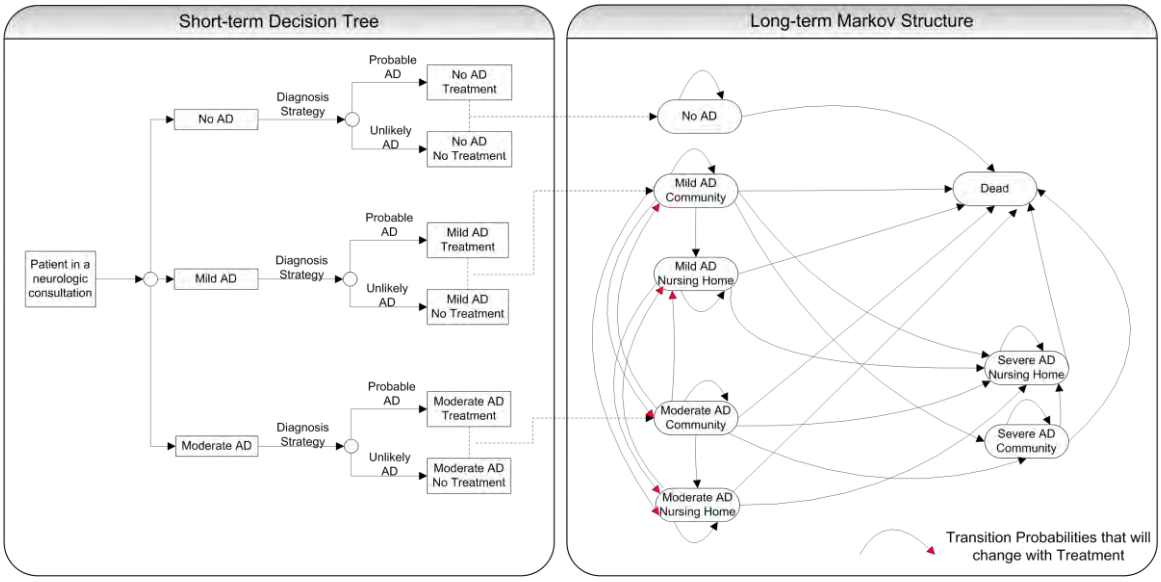


Figure 3 – Markov Cycle Tree

The analysis is performed through a societal perspective: it includes all costs, which are in Euros, and benefits (health effects) affecting every intervenient on this process. The health effects are measured in Quality Adjusted Life Years (QALYs) defined as life-years weighted by their quality that can vary between 0 (death) and 1 (perfect health). The model considers a subject moving through states with a utility value attached (the relative desirability of that state). Thus, it simultaneously assesses gains from reduced morbidity and from reduced mortality.

Diagnosis strategies are associated to a final expected cost and a final health outcome. By comparing two strategies we obtain incremental costs (△C) and incremental QALYs (△QALYs), parameters combined into a single output designated as ICER (Incremental Cost-effectiveness Ratio), given by:

$$ICER = \frac{\Delta C}{\Delta \text{QALY}} \quad (1)$$

ICERs can be compared with λ, a threshold value, to evaluate if a strategy should be recommended. This λ is a critical variable defined by policy makers above which the additional costs generated would not balance the eventual gain in the health outcome.

Patients undergo one of the four diagnosis strategies, receiving a diagnosis that can be "Probable AD" or "Unlikely AD". If one is flagged as probably having AD, one receives a certain treatment. Within AD treatments, *Donepezil* has been largely studied in terms of cost-effectiveness [Neumann et al., 1999], and it is included as treatment for mild and moderate cases. It influences outcomes differently depending if one is an AD patient or if he has been misdiagnosed. In the first case, effects are considered through a slower disease progression with a relative risk factor applied to the original transition probabilities, enhancing for some AD states the hypothesis of a "backward transition". Otherwise, when a subject does not have AD and undergoes an unneeded treatment, the corresponding decrease in QALYs is included. In terms of settings in the model, once a patient has moved to a nursing home, there is no way back.

All strategies included costs related with the "Standard Examination". The long-term Markov Structure also contains, for each state, a quality of life weight and an associated cost. Each state's cost will include direct costs, as expenses on follow-up visits, or medical tests and indirect costs, while *Donepezil* costs are separately accounted for.

The model was implemented in the Microsoft Excel® and TreeAgePro® software which have interconnected and complementing functions. The "base-case" analysis was taken as a deterministic one, using our best estimates for all parameters.

3.2 – Data in use

In terms of performance (sensitivity and specificity) we used PET data from [Silverman et al., 2001]. For DSC values from [Harris et al., 1998] were used. Information on the performance of MRI-ASL for AD was used from [Asllani et al., 2008]. Regarding the No Exam strategy performance values were taken from [Knopman et al., 2001].

The model also requires the prevalence of AD in the patient population who presented itself to the neurology unit and the ratio of mild to moderate AD cases. The first one was an intermediate value (60%) based on both Portuguese and US estimates, whilst the ratio of mild to moderate AD cases was estimated at 4:3. Relative risks for *Donepezil* were extracted from [Neumann et al., 1999]. It confirmed beneficial effects of the treatment with a 50% drop in the transition probability from Mild to Moderate AD and a 2,36 increase in the probability of regression from Moderate to Mild AD. The analysis

time was 18 months based on a survey of experts also performed by [Neumann et al., 1999]. At last, the transition probabilities used were also driven from that same study.

The cost estimate of "No AD" State was calculated from a Portuguese study [Inquérito às Despesas das Famílias 2005/2006, 2008] that surveyed family expenditures. Costs for remaining states of the Markov model were estimated from a Spanish study [López-Bastida et al., 2006]. QALY weights for AD were based on the Mark III version of the Health Utility Index (HUI3) and were reported by [McMahon et al., 2003] from a cross-sectional study of 679 AD patients in the USA, stratified by disease stage and care setting (community/nursing home). As, up to our knowledge, similar estimates for these measures are not available for Portugal, our study is meant to be exploratory to guide possible decisions on this matter. Patients with non-AD related dementia that underwent treatment had an estimated decrease in QALYs of 0,05.

All strategies include costs associated with their common section (one specialized consultation, diverse tests and a structural imaging exam). These values and the cost of functional imaging exam were extracted from a regulation published by the Portuguese Ministry of Health [Portaria nº 132/2009]. Another item is the treatment cost calculated from the price of a generic drug of *Donepezil* found in an online Infarmed database.

4- Results

It is common to represent strategies in the cost-effectiveness plane. This plane includes differences in costs (△Cost) to a common base-line per patient against a similar difference in utility (△QALY). If we consider the base-line as the "No exam" strategy, results from the base-case scenario (deterministic analysis) are shown in Figure 4.

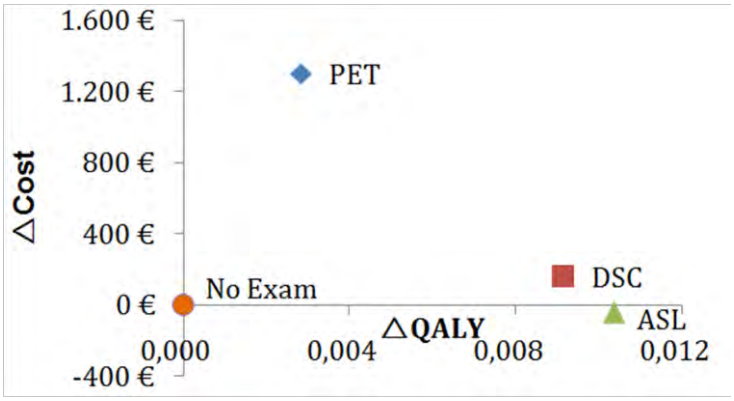


Figure 4 – Diagnosis Strategies on the cost-effectiveness plane

It is clear that ASL strategy provides the intervention with higher effectiveness, while achieving lower cost than any other strategy. As this happens simultaneously, the ASL strategy dominates all strategies. Since it also avoids an extra exam, one could expect the "No Exam" strategy final cost to be lower than ASL strategy cost, yet it is slightly superior which is explained by differences on each strategy performance values. AD patients benefit more often from the treatment that slows disease's progression by changing specific transitions probabilities. This ultimately decreases total related costs.

"PET" and "DSC" strategies are located on the first quadrant of Figure 4 and have positive slopes (the ICER values according to (1)). In contrast, a line uniting plot origin with the ASL point will have a negative slope. If a decision maker does not want to invest in ASL, he also finds in this study information on whether to adopt DSC or PET strategies *vs.* "No exam" strategy. In that case, it is undoubtedly advisable to adopt DSC strategy as it has a much lower ICER: with DSC strategy one will pay €17 803 for each QALY gained instead of the €461 227 for an extra QALY that PET strategy demands. These results are valid in case all technologies are implemented in the clinical setting.

If one wants to implement ASL, it is worth knowing how many exams are required to compensate the investment, considering for each ASL exam cost the ASL equipment price. It should be interpreted as a scenario where AD is the only condition for which a neurologic unit is using ASL. Thus, depending on λ (the maximum cost per QALY the society is willing to pay), the minimum number of exams can be calculated and it is visible in Table 1. If society is not paying any euro for a QALY increase, the number of exams is 1172. However, if society pays a suggested threshold cited in a Spanish study [Sacristán et al., 2002] (30 000€/QALY) and DSC is not implemented, it would be necessary at least 142 exams for ASL to be considered cost-effective in AD diagnosis.

Table 1 - Minimum Number of Exams for specific values that the society is willing to pay.

λ	0	5 000	10 000	17 803	30 000	50 000	100 000
Minimum Number of Exams	1172	530	342	221	142	90	47

We performed sensitivity analyses (ICER values vs. "No exam" strategy are visible in Table 2) and occasionally dominances present in the base-case scenario have changed. The most sensitive parameters were ASL cost, ASL performance and the time span. Less relevant parameters that might also change relations among strategies under certain ranges were the "Ratio of Mild to Moderate AD Cases" or the "Non-invasiveness

Benefic Effect". It is clear that ASL strategy does not dominate the "No exam" strategy for short time spans (in which DSC and PET strategies are extremely unadvisable).

Table 2 - Sensitivity Analyses

Types Of Scenarios	Scenarios		ASL	DSC	PET
Base-case			-4 126	17 804*	461 227*
Exam Characteristics and Treatment Rules	***ASL Cost identical to:***	***"No exam" Cost***	-17 017	17 804*	461 227*
		"MRI-DSC" Cost	10 376	17 804	461 227
	ASL Performance	***Perfect Examination***	-5 299	17 804*	461 227*
		Raji et al. values	-52 665†	17 804	461 227*
Drug Effects and Duration	***Other Treatment with a Possible New Drug***		-18 841	2 524*	225 832*
	Model Time Span	***12 months***	16 210	46 664*	1 133 614*
		24 months	-15 225	2651*	2 398 333*
		48 months	-31 093	-19 568*	32 996*
	Ratio of Mild to Moderate AD Cases	***1:1***	-2 415	19 419*	517 245*
		1:2	2 102	23 502*	713 769*
QALYs	***Non-invasiveness Benefic Effect***		-2 812	161483*	-250 960*

Strategies with * are dominated by ASL while those with † are dominated by DSC. Items in light grey dominate "No exam" strategy while those in darker grey are dominated by it.

To transform the model into a probabilistic one, appropriate distributions were applied to parameters for which literature provides reliable data about uncertainty. We included gamma distributions for the cost of each strategy, since as highly skewed distributions they model correctly the upper tail of costs distributions, while beta distributions were applied to relative risks. A Monte Carlo simulation with 10000 draws was carried out, leading to the acceptability curve displayed in Fig. 5. Independently of the λ value society is willing to pay, it is always more probable that "ASL" strategy is the best strategy (this probability even increases with λ). The model was also partially run with data from a private health care provider in Lisbon corroborating results previously described, which provides evidence on the robustness of the model.

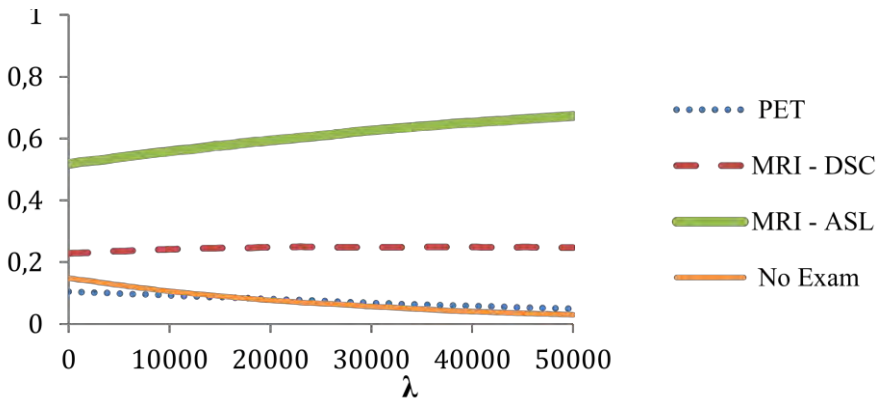


Figure 5 - Acceptability Curve for a MonteCarlo Simulation

5- Discussion and Conclusions

This study proposes a model to assess the economic impact of ASL on AD diagnosis. Incremental cost-effectiveness ratios between the different diagnosis strategies of AD were calculated and even if the model is sensitive to MRI-ASL cost and performance values, as well as to the treatment time span, we found evidence for MRI-ASL to be generally cost effective in comparison to all concurrent strategies.

Our results have shown that the model is robust and current evidence suggests that ASL might be a cost-effective option for diagnosing AD patients. The economic evaluation conducted here might be regarded as a first reference in the assessment of ASL, suggesting that ASL adoption may add value in comparison to alternative technologies.

Some limitations of the model and its application should be discussed: since there is little information about the cost of the ASL exam, another sensitivity analysis on the cost of ASL Strategy was performed between the costs of "No exam" and "MRI-DSC" strategies with ASL remaining cost-effective. ASL performance values are probably excessively close to values from a theoretical "perfect examination"; and patients' age does not influence transitions, despite AD being an age-related disease. Mild Cognitive Impairment was also not considered as a previous state to AD. At last, the wide range of data sources in use should also be taken into account when interpreting the results. It is clear the need for further studies in this area where resource scarcity is always present and where there are demands for a careful selection of which technologies to adopt. These limitations can be surpassed when an improved set of data is available.

6- References

Asllani, I. et al. (2008). Multivariate and univariate analysis of continuous arterial spin labeling perfusion MRI in Alzheimer's disease. Journal of Cerebral Blood Flow & Metabolism. 28. 725-736

Briggs, A.H. et al. (2006).Decision modelling for health economic evaluation. Oxford University Press Oxford, UK: ISBN: 978-0-19-852662-9

Cohen, J.T. et al. (2008). Decision analytic models for Alzheimer's disease: State of the art and future directions. Alzheimer's and Dementia. 4. 212-222

Drummond, M.F. et al. (2005).Methods for the economic evaluation of health care programmes. Oxford University Press, USA. ISBN: 019-852944-9

Fauci, A.S. (1998).Harrison's principles of internal medicine. McGraw-Hill New York.

Geldmacher, D.S. et al. (1997). Differential diagnosis of Alzheimer's disease. Neurology. 48. 2

Harris, G.J. et al. (1998). Dynamic susceptibility contrast MR imaging of regional cerebral blood volume in Alzheimer disease: a promising alternative to nuclear medicine. American Journal of Neuroradiology. 19. 1727

Instituto Nacional de Estatística. (2008). Inquérito às Despesas das Famílias 2005/2006. Lisboa - PORTUGAL.

Jezzard, P. et al. (2002).Functional MRI: an introduction to methods. Oxford University Press Oxford. ISBN: 019263071

Knopman, D.S. et al. (2001). Practice parameter: diagnosis of dementia (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology. 56. 1143

Kulasingam, S.L. et al. (2003). When should functional neuroimaging techniques be used in the diagnosis and management of Alzheimer's dementia? A decision analysis. Value in Health. 6. 542-550

López-Bastida, J. et al. (2006). Social-economic costs and quality of life of Alzheimer disease in the Canary Islands, Spain. Neurology. 67. 2186

McMahon, P.M. et al. (2000). Cost-effectiveness of Functional Imaging Tests in the Diagnosis of Alzheimer Disease1. Radiology. 217. 58

McMahon, P.M. et al. (2003). Cost-Effectiveness of PET in the Diagnosis of Alzheimer Disease1. Radiology. 228. 515

Neumann, P.J. et al. (1999). Cost-effectiveness of donepezil in the treatment of mild or moderate Alzheimer's disease. Neurology. 52. 1138

Portuguese Health Ministry. (2009). Portaria nº 132/2009 - Regulamento Das Tabelas De Preços Das Instituições E Serviços Integrados No Sistema Nacional De Saúde.

Raji, C.A. et al. (2010). Initial Experience in Using Continuous Arterial Spin-Labeled MRI for Early Detection of Alzheimer Disease. American Journal of Neuroradiology.

Sacristán, J.A. et al. (2002). What is an efficient health technology in Spain? Gaceta Sanitaria. 16. 334-343

Silverman, D.H.S. et al. (2001). Positron emission tomography in evaluation of dementia: regional brain metabolism and long-term outcome. Jama. 286. 2120

Wintermark, M. et al. (2005). Comparative overview of brain perfusion imaging techniques. Stroke 2005 36; e83-e89

Uma variação da função Lagrangiana Aumentada Barreira Modificada para resolução de problemas de Programação Não Linear

Maria C. S. Masiero[a*], *Edméa C. Baptista*[b], *Aguinaldo A. Pereira*[c], *Vanusa A. Sousa*[d]

[a]*Universidade Estadual Paulista – Unesp - Programa de Pós Graduação em Engenharia Elétrica -FEB*
Bauru – SP - Brasil
mariacsm@feb.unesp.br

[b]*Universidade Estadual Paulista- Unesp - Departamento de Matemática - FC*
Bauru – SP - Brasil
baptista@fc.unesp.br

[c]*Universidade Estadual Paulista – Unesp – Grupo de Pesquisa Modelagem e Otimização de Sistemas*
Bauru – SP – Brasil
agnld@hotmail.com

[d]*Universidade Federal de São de Carlos – UFSCar – Engenharia Elétrica – CCET*
São Carlos – SP - Brasil
vanusa@ufscar.br

[*]*Corresponding Author*

Resumo

Nesse trabalho utilizamos uma abordagem da Função Lagrangiana Aumentada Barreira Modificada para a resolução de problemas de programação não linear, restritos e gerais. Nessa abordagem, as restrições de desigualdade e as variáveis canalizadas são transformadas em igualdades através do uso de variáveis de folga e excesso, positivas. Desta forma, as restrições de igualdade passam a ser tratadas pela função Lagrangiana Aumentada e as variáveis auxiliares pela função Barreira Modificada resultando em uma nova Função Lagrangiana Aumentada Barreira Modificada. O problema restrito é então, transformado em um problema irrestrito através dessa função auxiliar e as condições necessárias de primeira ordem são aplicadas resultando num sistema não-linear. Este sistema é linearizado e resolvido através do Método de Newton. O parâmetro de barreira é atualizado segundo Melman e Polyak. O parâmetro de penalidade é atualizado por um fator de correção. Os multiplicadores de Lagrange associados à Lagrangiana aumentada são atualizados pela regra de Goldfarb et al. Os multiplicadores de Lagrange associados à Barreira modificada são atualizados pela regra de Polyak. Foi realizada uma aplicação desta abordagem à um exemplo numérico e a um Problema de Fluxo de Potência Ótimo.

Palavras Chaves : função Barreira Modificada; função Lagrangiana Aumentada; Fluxo de Potência Ótimo; Programação não linear.

1- Introducão

Antes de 1940, relativamente muito pouco havia sido desenvolvido sobre métodos para otimização numérica de funções de muitas variáveis. Vários métodos de otimização foram desenvolvidos após o surgimento do computador, entre eles damos ênfase aos métodos da Lagrangiana Aumentada e os de pontos interiores.

O método da função Lagrangiana aumentada pode ser visto como uma combinação dos métodos da função penalidade e o Lagrangiano, sendo aproveitadas as melhores qualidades de ambos. Ela foi introduzida por [Hestenes, 1969] e [Powell, 1969]. Em seu trabalho, Hestenes se preocupou em apresentar a 'idéia' da função Lagrangiana Aumentada e da atualização dos multiplicadores, porém, não forneceu um algoritmo para o método. Em [Powell, 1969], foi apresentado um algoritmo que utilizava uma regra para aumentar o fator de penalidade e demonstrou-se que, se as condições suficientes de segunda ordem para a otimalidade são satisfeitas, o algoritmo converge localmente com uma taxa linear, sem que o fator de penalidade precise tender ao infinito. Após a apresentação da função Lagrangiana Aumentada por Hestenes e Powell, muitos pesquisadores passaram a trabalhar nessa área, a fim de aprimorá-la. Entre eles, citamos Rockafellar [Rockafellar, 1973] que apresentou uma função Lagrangiana Aumentada para problemas convexos com restrições de desigualdade e a relação dessa função com a teoria da dualidade. O autor não se preocupou, neste trabalho, em fornecer um algoritmo particular, mas sim em dar uma forma geral de funcionamento do método. Em [Boukari and Fiacco, 1995] foi apresentado um levantamento histórico dos métodos de programação não linear, o qual contempla artigos que tratam do assunto função Lagrangiana Aumentada no período de 1968 a 1993 e a observação que foi feito muito pouco sobre tal assunto, em virtude da atenção despertada pelos métodos de pontos interiores. Em [Baptista et al, 2006], a aplicação da função Lagrangiana Aumentada Barreira Logaritmica ao problema de Fluxo de Potência Ótimo (FPO) é realizada.

Para a resolução de problemas com restrições de desigualdade, cujo interior é não-vazio, é utilizado o método da função barreira modificada. A função barreira logarítmica foi estudada por [Frisch, 1955] e a função barreira inversa foi proposta por [Carrol, 1961]. O método de barreira foi realmente popularizado por [Fiacco e McCormick, 1968], que realizaram um estudo teórico mais detalhado do método e

desenvolveram um novo; "misturando"a função barreira e a função penalidade em uma mesma função auxiliar. Muitos outros autores que trabalharam com o método da função barreira perceberam que ele apresenta alguns problemas, tais como: o mal condicionamento da matriz Hessiana quando seu fator de barreira tende a zero; a dificuldade na escolha do fator de barreira, entre outros. O interesse pelo método da função barreira reapareceu somente após a apresentação feita por Karmarkar [Karmarkar, 1984], de seu método projetivo para programação linear.

Após o trabalho de Karmarkar, surgiram vários artigos de pontos interiores. Entre eles citamos Polyak [Polyak, 1992], que desenvolveu uma teoria de métodos da função Barreira Modificada. Estes métodos combinam as melhores propriedades da função Lagrangiana Clássica e da função Barreira Clássica, evitando os problemas que ambas enfrentam e buscando explorar as melhores propriedades de cada uma dessas funções. Em contraste com a função Barreira Clássica, as funções Barreira Modificadas são definidas na solução; são suaves na vizinhança do ótimo; não vão para o infinito quando se aproximam do ótimo e representam explicitamente os Multiplicadores de Lagrange. Em [Baptista et al, 2009] uma nova abordagem da Função Barreira Modificada é apresentada.

Com o objetivo de aproveitar as características dos métodos citados anteriormente, alguns autores apresentaram abordagens misturadas. Aqui destacamos o método de [Goldfarb et al, 1999] em que as restrições de igualdade são tratadas através da função Lagrangiana Aumentada e as de desigualdade através da função da Barreira modificada de Frisch.

Neste artigo, propomos uma variação do método apresentado em [Goldfarb et al,1999] usando variáveis auxiliares. A aplicação consiste em transformar as restrições de desigualdade e as variáveis canalizadas em igualdades através do uso de variáveis de folga e excesso, positivas. Desta forma, as restrições de igualdade passam a ser tratadas pela função Lagrangiana Aumentada e as variáveis auxiliares pela função Barreira Modificada resultando em uma nova Função Lagrangiana Aumentada Barreira Modificada (NFLABM). O problema restrito é então transformado em irrestrito e as condições necessárias de primeira ordem são aplicadas resultando num sistema não-linear. Este sistema é linearizado e resolvido através do Método de Newton. O parâmetro de barreira é atualizado segundo [Melman e Polyak, 1996]. O parâmetro de penalidade é atualizado por um fator de correção. Os multiplicadores de Lagrange associados à Lagrangiana aumentada são atualizados pela regra de [Goldfarb et al,

1999]. Os multiplicadores de Lagrange associados à Barreira modificada são atualizados pela regra de [Polyak,1992]. Para a validação do método, apresentamos um exemplo numérico detalhado e a sua aplicação a um problema de Fluxo de Potência Ótimo, área de estudo na Engenharia Elétrica, de sistemas Elétricos de Potência. O problema de FPO foi proposto por Carpentier [Carpentier, 1962] e resolvido por diversos métodos matemáticos.

O artigo está organizado como segue: na seção 2 um problema de programação não linear é apresentado. A variação da função Lagrangiana Aumentada Barreira Modificada é apresentada na seção 3, seguida do algoritmo na seção 4. Os resultados obtidos usando um exemplo numérico e um problema de FPO são mostrados na seção 5. São feitos alguns comentários conclusivos na seção 6, seguido de agradecimentos e referências.

2- Formulação do Problema

Seja um problema de Programação Não Linear (PNL) com restrições de igualdade e desigualdade:

$$\begin{aligned} &\text{Minimizar} \quad f(x) \\ &\text{sujeito a}: \quad g_i(x) = 0 \qquad\qquad i = 1,2,...,m \\ &\qquad\qquad h_{min_j} \le h_j(x) \le h_{max_j} \qquad j = 1,2,...,p \\ &\qquad\qquad x_{min} \le x \le x_{max} \end{aligned} \tag{1}$$

em que : $x \in R^n$; f(x): função objetivo; g_i (x): equações de igualdade; h_j(x): restrições canalizadas; $hmin_j$ e $hmax_j$: limites inferiores e superiores das restrições h_j(x), respectivamente e x_{min} e x_{max}: vetor dos limites inferiores e superiores das variáveis x, respectivamente.

3- A Variação da Função Lagrangiana Aumentada Barreira Modificada

Para resolução do problema (1), desdobramos as restrições canalizadas em duas restrições de desigualdade e, posteriormente, em igualdades através da adição de variáveis de folga e excesso estritamente positivas. Logo, o problema (1) modificado pode ser apresentado como:

$$
\begin{aligned}
&\text{Minimizar} \quad f(x) \\
&\text{sujeito a:} \quad g_i(x) = 0, \qquad i = 1,\ldots,m \\
&\qquad\qquad h_j(x) - s_{1j} = h_{min_j}, \quad j = 1,\ldots,p \\
&\qquad\qquad h_j(x) + s_{2j} = h_{max_j}, \quad j = 1,...,p \\
&\qquad\qquad x - s_3 = x_{min} \\
&\qquad\qquad x + s_4 = x_{max} \\
&\qquad\qquad s_{1j} \geq 0 \,,\; s_{2j} \geq 0,\; s_3 \geq 0 \,,\; s_4 \geq 0
\end{aligned}
\tag{2}
$$

em que: $s_1 \in R^p$, $s_2 \in R^p$, $s_3 \in R^n$ e $s_4 \in R^n$.

As variáveis de folga e de excesso vistas em (2) são incorporadas à função objetivo através da função barreira modificada e o seguinte problema é originado:

$$
\begin{aligned}
&\text{Minimizar} \quad F(x,s) \\
&\text{sujeito a:} \quad g_i(x) = 0, \qquad i = 1,\ldots,m \\
&\qquad\qquad h_j(x) - s_{1j} = h_{min_j}, \quad j = 1,\ldots,p \\
&\qquad\qquad h_j(x) + s_{2j} = h_{max_j}, \quad j = 1,...,p \\
&\qquad\qquad x - s_3 = x_{min} \\
&\qquad\qquad x + s_4 = x_{max}
\end{aligned}
\tag{3}
$$

em que:

$$
F(x,s) = f(x) - k_1\left(\sum_{j=1}^{p} u_{1j} \ln\left(\frac{s_{1j}}{k_1}+1\right) + \sum_{j=1}^{p} u_{2j} \ln\left(\frac{s_{2j}}{k_1}+1\right) + \sum_{r=1}^{n} u_{3r} \ln\left(\frac{s_{3r}}{k_1}+1\right) + \sum_{r=1}^{n} u_{4r} \ln\left(\frac{s_{4r}}{k_1}+1\right)\right)
$$

e u_{q1}, $q1 = 1,...,4$ são as estimativas dos multiplicadores de Lagrange associados à função Barreira Modificada, segundo [Polyak, 1992].

As restrições de igualdade do problema (3) são tratadas através da função Lagrangiana Aumentada. Desta forma, a nova Função Lagrangiana Aumentada Barreira Modificada (NFLABM) associada ao problema (3) é dada por:

$$
\begin{aligned}
\text{NFLABM} = {} & f(x) - k_1\left(\sum_{j=1}^{p} u_{1j} \ln\left(\frac{s_{1j}}{k_1}+1\right) + \sum_{j=1}^{p} u_{2j} \ln\left(\frac{s_{2j}}{k_1}+1\right) + \sum_{r=1}^{n} u_{3r} \ln\left(\frac{s_{3r}}{k_1}+1\right) + \sum_{r=1}^{n} u_{4r} \ln\left(\frac{s_{4r}}{k_1}+1\right)\right) + \\
& - \sum_{i=1}^{m} v_{1i}(g_i(x)) - \sum_{j=1}^{p} v_{2j}(h_j(x) - s_{1j} - h_{min_j}) - \sum_{j=1}^{p} v_{3j}(h_j(x) + s_{2j} - h_{max_j}) - \sum_{l=1}^{n} v_{4l}(x_1 - s_{3l} - x_{min}) + \\
& - \sum_{l=1}^{n} v_{5l}(x_1 + s_{4l} - x_{max}) + \frac{k_2}{2}\left(\sum_{i=1}^{m} (g_i(x))^2 + \sum_{j=1}^{p} (h_j(x) - s_{1j} - h_{min_j})^2 + \sum_{j=1}^{p} (h_j(x) + s_{2j} - h_{max_j})^2 + \right. \\
& \left. + \sum_{l=1}^{n} (x_1 - s_{3l} - x_{min})^2 + \sum_{l=1}^{n} (x_1 + s_{4l} - x_{max})^2\right)
\end{aligned}
\tag{4}
$$

em que:

k_1: parâmetro de barreira; k_2: parâmetro de penalidade; u_{q1}, $q1 = 1,...,4$: multiplicadores de Lagrange associados à função Barreira Modificada e v_i, $i = 1,2,...,m$: multiplicadores de Lagrange associados à função Lagrangiana Aumentada.

A solução do problema (1) consiste em encontrar os valores do vetor $d = (x, s_1, s_2, s_3, s_4)$ que satisfaça as condições de otimalidade de primeira ordem aplicados à (4), isto é:

$$\nabla NFLABM = 0 \tag{5}$$

em que:

$$\nabla NFLABM = \begin{bmatrix} \nabla_x f(x) - J(x)^T v_1 - J_1(x)^T v_2 - J_1(x)^T (v_3) - (v_4) - (v_5) + \frac{k_2}{2}\varpi \\ \frac{-k_1 u_{1j}}{s_{ij} + k_1} + v_{2j} - k_2(h_j(x) - s_{1j} - h_{min_j}),\ j = 1,...,p \\ \frac{-k_1 u_{2j}}{s_{2j} + k_1} - v_{3j} + k_2(h_j(x) + s_{2j} - h_{max_j}),\ r = 1,...,n \\ \frac{-k_1 u_{3r}}{s_{3r} + k_1} + v_{4r} - k_2(x_r - s_{3r} - x_{min}),\ r = 1,...,n \\ \frac{-k_1 u_{4r}}{s_{4r} + k_1} - v_{5r} + k_2(x_r + s_{4r} - x_{max}),\ l = 1,...,n \end{bmatrix}, \tag{6}$$

$\varpi = 2g(x)J(x) + 2(h(x) - s_1 - h_{min_j})J_1(x) + 2(h(x) + s_2 - h_{max_j})J_1(x) + 2(x - s_3 - x_{min}) + 2(x + s_4 - x_{max})$ e $J(x)^T = (\nabla_x g_1(x),...,\nabla_x g_m(x))$ e $J_1(x)^T = (\nabla_x h_1(x),...,\nabla_x h_p(x))$ são denominadas matrizes Jacobianas.

O sistema (5) é resolvido através do método de Newton, o que resulta na determinação dos fatores de correção das variáveis:

$$W\Delta d = -\nabla NFLABM \tag{7}$$

em que:

$$W = \begin{pmatrix} \nabla^2_{xx} NFLABM & -k_2 J_1(x)^T & k_2 J_1(x)^T & -k_2 I_n & k_2 I_n \\ -k_2 J_1(x)^T & S_1 & 0 & 0 & 0 \\ k_2 J_1(x) & 0 & S_2 & 0 & 0 \\ -k_2 I_n & 0 & 0 & S_3 & 0 \\ k_2 I_n & 0 & 0 & 0 & S_4 \end{pmatrix}$$

é a matriz Hessiana da função Lagrangiana Aumentada Barreira Modificada, sendo:

$$S_1 = \begin{pmatrix} \frac{k_1 u_{11}}{(s_{11}+k_1)^2}+k_2 & & 0 \\ & \ldots & \\ 0 & & \frac{k_1 u_{1p}}{(s_{1p}+k_1)^2}+k_2 \end{pmatrix}, \quad S_2 = \begin{pmatrix} \frac{k_1 u_{21}}{(s_{21}+k_1)^2}+k_2 & & 0 \\ & \ldots & \\ 0 & & \frac{k_1 u_{2p}}{(s_{2p}+k_1)^2}+k_2 \end{pmatrix},$$

$$S_3 = \begin{pmatrix} \frac{k_1 u_{31}}{(s_{31}+k_1)^2}+k_2 & & 0 \\ & \ldots & \\ 0 & & \frac{k_1 u_{3n}}{(s_{3n}+k_1)^2}+k_2 \end{pmatrix}, \quad S_4 = \begin{pmatrix} \frac{k_1 u_{41}}{(s_{41}+k_1)^2}+k_2 & & 0 \\ & \ldots & \\ 0 & & \frac{k_1 u_{4n}}{(s_{4n}+k_1)^2}+k_2 \end{pmatrix};$$

$$\nabla^2_{xx} NFLABM = \nabla^2_{xx} f(x) - \sum_{i=1}^{m} v_1 \nabla^2_{xx} g_i(x) - \sum_{j=1}^{p} v_2 \nabla^2_{xx} h_j(x) - \sum_{j=1}^{p} v_3 \nabla^2_{xx} h_j(x) + \frac{k_2}{2} \nabla^2_{xx} \varpi \ ;$$

$\Delta d^T = (\Delta x, \Delta s_1, \Delta s_2, \Delta s_3, \Delta s_4)$; e o $\nabla NFLABM$ é dado em (6).

Os multiplicadores de Lagrange u_{q1}^{t+1} são atualizados de acordo com a regra de [Polyak, 1992]:

$$u_{q1}^{t+1} = \frac{u_{q1}^t k_1^{t+1}}{s_{q1}^{t+1} + k_1^{t+1}}, \ q1 = 1,...,4. \tag{8}$$

Os multiplicadores de Lagrange v_{q2}^{t+1} são atualizados de acordo com a regra de [Hestenes e Powell, 1969]:

$$v_{q2}^{t+1} = v_{q2}^t - k_2 b_{q2}(x^{t+1}), \ q2 = 1,...,5. \tag{9}$$

em que $b_{q2}(x^{t+1})$ representa as equações de igualdade do problema (3).

O parâmetro de barreira k_1 é atualizado conforme [Melman and Polyak, 1996]:

$$k_1^{t+1} = k_1^t \left(1 - \frac{\alpha}{\sqrt{2p+2n}}\right), \tag{10}$$

em que 2p+2n é o número de restrições de desigualdade do problema (2) tratadas pela função Barreira Modificada e $\alpha = \max\left(\frac{1}{k_1^{-1} s_{q1} + 1}\right), q1 = 1,...,4$, para $s_{q1} > 0$.

O parâmetro de penalidade k_2 é atualizado por:

$$k_2^{t+1} = c k_2^t, \tag{11}$$

em que $c > 1$.

Os vetores x, s_1, s_2, s_3 e s_4 são atualizados da seguinte maneira:

$$\begin{aligned} &x^{t+1} = x^t + \alpha \Delta x^t; \qquad s_1^{t+1} = s_1^t + \alpha \Delta s_1^t; \\ &s_2^{t+1} = s_2^t + \alpha \Delta s_2^t; \qquad s_3^{t+1} = s_3^t + \alpha \Delta s_3^t; \qquad s_4^{t+1} = s_4^t + \alpha \Delta s_4^t, \end{aligned} \tag{12}$$

em que o tamanho do passo $\alpha \in (0,1]$ é escolhido para preservar a positividade do vetor s. Isto se traduz por:

$$\alpha_p^{max} = \min\left\{\frac{-s_1}{\Delta s_1} : \Delta s_1 < 0,\ \frac{-s_2}{\Delta s_2} : \Delta s_2 < 0,\ \frac{-s_3}{\Delta s_3} : \Delta s_3 < 0,\ \frac{-s_4}{\Delta s_4} : \Delta s_4 < 0\right\}$$

$$\alpha = \min\left\{\tau\alpha_p^{max}, 1\right\}$$

onde o escalar $\tau \in (0,1)$ é valor determinado empiricamente e dado por $\tau = 0,9995$. De acordo com [Wright, 1995], pode ser derivado da fórmula $1 - 1/9\sqrt{z}$ onde z é o número de restrições do problema.

Um ponto inicial estritamente admissível não é obrigatório, mas a condição $s > -k_1$ deve ser satisfeita em todos os pontos. O processo de otimização termina quando as condições de Karush-Kuhn-Tuker (KKT) são satisfeitas.

4- Algoritmo

1) Dado o problema (1), construa a função (4). Faça t=0, calcule uma estimativa inicial para $d^t = (x^t, s_1^t, s_2^t, s_3^t, s_4^t),\ u_1^t,\ u_2^t,\ u_3^t,\ u_4^t,\ v_1^t,\ v_2^t,\ v_3^t,\ v_4^t,\ v_5^t,\ k_1^t$ e k_2^t que satisfaça as condições propostas e assuma $\varepsilon = 0.0001$;

2) Determine o sistema (7) e resolva-o;

3) Atualize d^t utilizando (12);

4) Se a norma de (6) for menor que uma precisão ε vá para o passo **5**. Caso contrário volte para o passo **2**;

5) Se as condições de KKT são satisfeitas então **pare**. Caso contrário vá para o passo **6**;

6) Atualize k_1 utilizando (10) , k_2 utilizando (11), os multiplicadores de Lagrange u_{q1}^{t+1} usando (8) e v_{q2}^{t+1} usando (9). Faça t=t+1 e retorne ao passo **2**.

5- Resultados Computacionais

Teste 1: Exemplo Numérico

Seja o problema de otimização, proposto por Pereira (2007):

$$\begin{aligned} \text{Minimizar} \quad & (x_1-2)^4+(x_1-2x_2)^2 \\ \text{sujeito a} \quad & x_1+x_2-3=0 \\ & -1\le x_1^2-x_2\le 0 \\ & 1{,}5\le x_2\le 2 \end{aligned} \tag{13}$$

A nova Função Lagrangiana Aumentada Barreira Modificada associada ao problema (13) é dada por:

$$\begin{aligned} \text{NFLABM}= & (x_1-2)^4+(x_1-2x_2)^2-k_1u_1\ln\left(k_1^{-1}s_1+1\right)-k_1u_2\ln\left(k_1^{-1}s_2+1\right)+ \\ & -k_1u_3\ln\left(k_1^{-1}s_3+1\right)-k_1u_4\ln\left(k_1^{-1}s_4+1\right)-v_1(x_1+x_2-3)-v_2(x_1^2-x_2-s_1+1)+ \\ & -v_3(x_1^2-x_2+s_2)-v_4(x_2-s_3-1{,}5)-v_5(x_2+s_4-2)+\frac{k_2}{2}(x_1+x_2-3)^2+ \\ & +\frac{k_2}{2}(x_1^2-x_2-s_1+1)^2+\frac{k_2}{2}(x_1^2-x_2+s_2)^2+\frac{k_2}{2}(x_2-s_3-1{,}5)^2+\frac{k_2}{2}(x_2+s_4-2)^2 \end{aligned} \tag{14}$$

As condições necessárias de primeira-ordem são aplicadas à nova Função Lagrangiana Aumentada Barreira Modificada (14), gerando um sistema de equações não-lineares, como segue:

$$\nabla\text{NFLABM}\ (x1,x2,s1,s2,s3,s4)=0, \tag{15}$$

sendo:

$$\nabla NFLABM=\begin{bmatrix} 4(x_1-2)^3+2(x_1-2x_2)-v_1-2x_1v_2-2x_1v_3+k_2(x_1+x_2-3)+2x_1k_2((x_1^2-x_2-s_1+1)+(x_1^2-x_2+s_2) \\ -4(x_1-2x_2)-v_1+v_2+v_3-v_4-v_5+k_2(x_1+x_2-3)-k_2(x_1^2-x_2-s_1+1)-k_2(x_1^2-x_2+s_2)+k_2(x_2-s_3-1{,}5)+k_2(x_2+s_4-2) \\ -\frac{k_1u_1}{s_1+k_1}+v_2-k_2(x_1^2-x_2-s_1+1) \\ -\frac{k_1u_2}{s_2+k_1}-v_3+k_2(x_1^2-x_2+s_2) \\ -\frac{k_1u_3}{s_3+k_1}+v_4-k_2(x_2-s_3-1{,}5) \\ -\frac{k_1u_4}{s_4+k_1}-v_5+k_2(x_2+s_4-2) \end{bmatrix}$$

O método de Newton é aplicado ao sistema de equações não-lineares (15) para determinar as direções de busca das variáveis, isto é, $\Delta x1$, $\Delta x2$, $\Delta s1$, $\Delta s2$, $\Delta s3$, $\Delta s4$, o que resulta em:

$$W\Delta d = -\nabla\text{NFLABM}\ ,$$

em que: $\Delta d = (\Delta x1,\Delta x2,\Delta s1,\Delta s2,\Delta s3,\ \Delta s4)^T$ e W a matriz Hessiana da função Lagrangiana, dada por:

$$W=\begin{pmatrix} \nabla^2 x_1x_1 NFLABM & \nabla^2 x_1x_2 NFLABM & -2x_1k_2 & 2x_1k_2 & 0 & 0 \\ \nabla^2 x_2x_1 NFLABM & \nabla^2 x_2x_2 NFLABM & k_2 & -k_2 & -k_2 & k_2 \\ -2x_1k_2 & k_2 & \frac{k_1u_{11}}{(s_{11}+k_1)^2}+k_2 & 0 & 0 & 0 \\ 2x_1k_2 & -k_2 & 0 & \frac{k_1u_{21}}{(s_{21}+k_1)^2}+k_2 & 0 & 0 \\ 0 & -k_2 & 0 & 0 & \frac{k_1u_{31}}{(s_{31}+k_1)^2}+k_2 & 0 \\ 0 & k_2 & 0 & 0 & 0 & \frac{k_1u_{41}}{(s_{41}+k_1)^2}+k_2 \end{pmatrix}$$

$$\nabla^2 x_1x_1 NFLABM = 12(x_1-2)^2+2-2v_2-2v_3+k_2+8k_2x_1 \text{;}\quad \nabla^2 x_2x_2 NFLABM = 8+5k_2$$

$$\nabla^2 x_1x_2 NFLABM = \nabla^2 x_2x_1 NFLABM = -4+k_2-4x_1k_2 \text{;}$$

O algoritmo foi implementado utilizando o software Matlab Starter Application versão R2010a (http://www.mathworks.com) em um microcomputador com processador Intel i5 M430 de 2.27GHz, 4 Gb de memória RAM e sistema operacional Windows 7 de 64 bits, do Laboratório de Estudos Econômicos em Sistemas de Potência (LEESP) da Faculdade de Engenharia da Unesp de Bauru.

A Tabela 1 apresenta o processo de convergência do método da nova Função Lagrangiana Aumentada Barreira Modificada aplicado ao problema (13). Na Tabela 1, It. é o número da iteração e F. Ob. é o valor da função objetivo. Os valores iniciais de todas as variáveis foram tomadas iguais a 1.

Tabela 1 – Convergência do método da NFLABM aplicado ao problema (13)

It.	F. Ob.	x1	x2	s1	s2	s3	s4
1	6,5991	1,2759	1,1938	1,2897	0,0000	0,0458	0,8789
2	5,7093	1,2916	1,5089	1,1058	0,0000	0,1373	0,5183
3	4,5820	1,3059	1,6960	0,9989	0,0000	0,1946	0,3037

Teste 2: Sistema elétrico teste 3 barras (FPO)

Este exemplo foi proposto por [Dommel e Tinney, 1968] a fim de testar a formulação do FPO. A função objetivo, a ser minimizada, representa as perdas de potência ativa na transmissão. Essa função é não-linear, não convexa tornando o problema em um dos mais difíceis de resolver [Monticelli e Liu, 1992]. O critério de parada foi satisfazer as condições de KKT e garantir a convergência na região admissível a um erro máximo para as restrições de igualdade de 10^{-4} p.u.

$$\text{min imizar } 4[(V_1^2 + V_3^2 - 2V_1V_3\cos\theta_{13}) + (V_2^2 + V_3^2 - 2V_2V_3\cos\theta_{23})]$$

s.a $P_2 - 1{,}7 = 0$

$P_3 + 2{,}0 = 0$

$Q_3 + 1{,}0 = 0$

$0{,}1 \le Q_2 \le 2{,}0$

$0{,}8 \le V_1 \le 1{,}2$

$0{,}8 \le V_2 \le 1{,}2$

$0{,}99 \le V_3 \le 1{,}01$

$\theta_1 = 0$

em que:

$$P_2 = 4V_2^2 - 4V_2V_3\cos\theta_{23} + 10V_2V_3\text{sen}\theta_{23};$$

$$P_3 = (4V_3^2 - 4V_3V_2\cos\theta_{32} + 10V_3V_2\text{sen}\theta_{32}) + (4V_3^2 - 4V_3V_1\cos\theta_{31} + 5V_3V_1\text{sen}\theta_{31});$$

$$Q_3 = (10V_3^2 - 10V_3V_2\cos\theta_{32} - 4V_3V_2\text{sen}\theta_{32}) + (5V_3^2 - 5V_3V_1\cos\theta_{31} - 4V_3V_1\text{sen}\theta_{31});$$

$$Q_2 = V_2(10V_2 - 10V_3\cos\theta_{23} - 4V_3\text{sen}\theta_{23});$$

$$\cos\theta_{ij} = \cos(\theta_i - \theta_j);$$

$$\text{sen}\theta_{ij} = \text{sen}(\theta_i - \theta_j).$$

O teste foi executado com as seguintes condições iniciais: $x_1 = 1$, $x_2 = 1$, $x_3 = 1$, $\theta_1 = 0^\circ$, $\theta_2 = -2^\circ$, $\theta_3 = -5^\circ$. As variáveis x representam a magnitude da tensão nas barras. Observamos que a magnitude da tensão na barra 3 é fixa e igual a 1.0 p.u. Os multiplicadores de Lagrange relacionados com as restrições de igualdade foram iniciados por: $v_1 = -0{,}0001$, $v_2 = 0{,}01$, $v_3 = 0$, $v_4 = 0$, $v_5 = 0$, $v_6 = -0{,}04$, $v_7 = 0{,}07$, $v_8 = 0{,}02$, $v_9 = 0{,}07$, $v_{10} = -1$, $v_{11} = 1$ e os multiplicadores de Lagrange relacionados com as restrições de desigualdade : $u_1 = 1$, $u_2 = 1$, $u_3 = 1$, $u_4 = 1$, $u_5 = 1$, $u_6 = 1$. O valor do parâmetro de barreira inicial foi definido como $k_1 = 0{,}01$ e o de penalidade $k_2 = 1$. O processo convergiu em 3 iterações e o total de perdas de potência ativa foi de 12,92 MW, conforme Tabela 2.

Tabela 2. Processo de otimização do sistema teste 3 barras.

Iterações	Perdas (MW)	ΔP (MW)	ΔQ (MVAr)
1	13,48	14,32	4,62
2	12,89	0,14	0,19
3	12,92	0,00	0,00

6- Conclusões

Neste artigo apresentamos uma variação de uma abordagem que utiliza a função Lagrangiana Aumentada e a função Barreira Modificada para resolução de problemas de programação não linear. Esta função combina as melhores características das funções citadas ao tratar as restrições de igualdade e de desigualdade. Os testes numéricos demonstram a eficiência desta variação.

Agradecimentos

Este projeto foi financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e pela Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior (CAPES).

Referências

Baptista, E.C., Belati, E., Souza, V., Costa, G.R.M. (2006). Primal-Dual Logarithmic Barrier and Augmented Lagrangian Function to the Loss Minimization in Power Systems. *Electric Power Components and Systems Lexington* – USA v. 34, n. 7, 775-784.

Baptista, E.C., Belati, E., Souza, V., Costa, G.R.M. (2009). Loss minimization by the predictor corrector modified barrier approach. *Electric Power Systems Research (Print)* v. 79, 803-808.

Boukari, D., Fiacco, A.V. (1995). Survey of Penalty, Exact – Penalty and Multiplier Methods From 1968 to 1993. *Optimization.* v. 32, 301-334.

Carpentier, J. L. (1962). Contribution à l'étude du dispatching économique. *Bulletin de la Société Française des Electricien.* ser. 8, 3:431-447.

Carrol, C. W. (1961). The Created Response Surface Technique for Optimizing Nonlinear Restrained Systems. *Operations Research*, v. 9, p. 169 – 185.

Dommel, H. W., Tinney, W. F. (1968). *Optimal Power Flow Solutions.* IEEE Transactions on Power Apparatus and Systems. v.87, n.10, 1866-1876.

Fiacco, A.V., McCormick, G. P. (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. New York, John Wiley & Sons.

Frisch, K. R. (1955). The logarithmic Potential Method of Convex Programming, *Memorandum*,University Institute of Economics, Oslo, Norway.

Goldfarb, D., Polyak, R., Scheinberg, K., Yuzefovich, I. (1999) A modified Barrier-Augmented Lagrangian Method for Constrained Minimization. *Computacional Optimization and Applications*. v. 14, 55 – 74.

Hestenes, M. R. (1969). Multiplier and Gradient Methods. *J.O.T.A* v.4, n.5, 303-320.

Karmarkar, N. (1984). A New Polynomial – Time Algorithm for Linear Programming. *Combinatorica,* v. 4, p. 373 - 395.

Melman, A., Polyak, R. A. (1996). The Newton Modified Barrier Method for QP Problems. *Annals of Operations Research*. v. 62, 465-519.

Monticelli, A., Liu, W.H.E. (1992). Adaptive Movement Penalty Method For The Newton Optimal Power Flow. IEEE Transactions on Power Systems. v. 7, n.1, 334-341.

Pereira, A. A. (2007). *O Método da Função Lagrangiana Barreira Modificada/Penalidade.* São Carlos. 114p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.

Polyak, R. (1992). Modified barrier functions. *Mathematical Programming*. 54(2):177-222.

Powell, M.J.D. (1969). A Method for Nonlinear Constraints in Minimization Problems. In: R. Fletcher, *Optimization*, New York, Academic Press 283-298.

Rockafellar, R.T. (1973). The Multiplier Method of Hestenes and Powell Applied to Convex Programming. *Journal of Optimization Theory and Applications*, v. 12, n. 6, p. 555 - 562.

Wright, M. H. (1995). Why a Pure Primal Newton Barrier Step May Be Infeasible. *SIAM Journal on Optimization*. 5(1):1-12.

Rect-TOPOS: Uma heurística construtiva para a minimização da área da envolvente rectangular no posicionamento de formas com componentes rectangulares

Marisa Oliveira[a*], *Eduarda Pinto Ferreira*[b], *A. Miguel Gomes*[c]

[a]*ISEP – Instituto Superior de Engenharia do Porto*
Rua Dr. António Bernardino de Almeida, 431 4200-072 Porto, Portugal
mjo@isep.ipp.pt

[b]*ISEP – Instituto Superior de Engenharia do Porto*
GECAD - Grupo de Investigação em Engenharia do Conhecimento e Apoio a Decisão
Rua Dr. António Bernardino de Almeida, 431 4200-072 Porto, Portugal
epf@isep.ipp.pt

[c]*FEUP – Faculdade de Engenharia da Universidade do Porto*
INESC - Instituto de Engenharia de Sistemas e Computadores do Porto.
Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal
agomes@fe.up.pt

[*]*Corresponding Author*

Resumo

Neste artigo propomos uma heurística construtiva, Rect-TOPOS, para resolver o problema da minimização da área da envolvente rectangular que contém, sem sobreposições, um conjunto de peças com componentes rectangulares (peças com formas em L, T, etc.). Este problema pertence a uma classe mais alargada de problemas de optimização combinatória: os problemas de corte e empacotamento. A complexidade computacional deste problema é conhecida por ser NP-difícil. Para avaliar a heurística Rect-TOPOS foram realizados testes computacionais que permitiram validar a heurística na resolução do problema apresentado. Nestes testes foram usadas instâncias com diferentes características, nomeadamente no número total de peças por instância e na diversidade das formas das peças. Os resultados obtidos mostram que se trata de uma heurística rápida e que é capaz de lidar facilmente com todo o tipo de formas com componentes rectangulares.

Palavras-chave: Optimização combinatória. Corte e empacotamento. Heurística construtiva. Minimização de área.

1- Introdução

No problema de minimização da área no posicionamento de formas com componentes rectangulares (MAPFCRE) pretende-se posicionar um conjunto de peças sem sobreposições de forma a minimizar a área da envolvente rectangular que contém todas as peças (Figura 1). Do ponto de vista científico é um problema NP-difícil [Murata et al, 1995]. Este problema pertence a uma classe mais abrangente de problemas de optimização combinatória: os problemas de corte e empacotamento (C&E), que são problemas combinatórios com uma forte componente geométrica. O objectivo nestes problemas poderá ser a maximização de itens pequenos a serem posicionados num objecto grande ou a minimização de objectos grandes para posicionar os itens pequenos. A componente geométrica dos problemas de C&E está presente no posicionamento dos itens, uma vez que é necessário assegurar que estes não se sobrepõem, enquanto a componente combinatória surge na selecção dos itens, onde os posicionar, na selecção do objecto grande, etc. Os problemas de C&E podem ser resolvidos por qualquer uma das abordagens e técnicas habituais de resolução de problemas de optimização combinatória: programação mista, heurísticas, meta-heurísticas, etc.

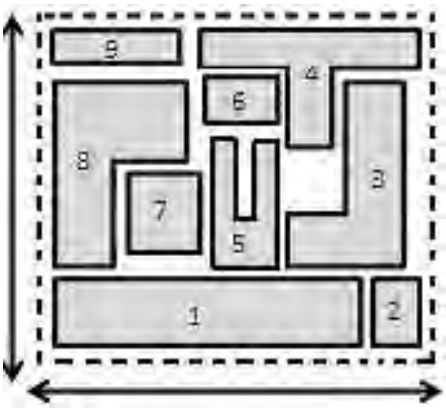


Figura 1 Objectivo do problema MAPFCRE

Para resolver o problema MAPFCRE propomos a aplicação de uma variante da heurística construtiva TOPOS [Oliveira et al, 2000]. As principais diferenças entre a variante proposta, Rect-TOPOS, e o TOPOS advêm da forma das peças usadas, da função objectivo e dos critérios usados para a escolha da próxima peça a posicionar, a sua orientação e o ponto de posicionamento.

Na secção 2 é feita uma descrição detalhada do problema MAPFCRE, seguindo-se na secção 3 a descrição da heurística proposta para resolver este problema. Na 4 secção são

apresentados os testes computacionais realizados e, finalmente, na última secção as apresentam-se as conclusões.

2- Descrição do Problema

O objectivo do problema MAPFCRE é minimizar a área da envolvente rectangular que contém um conjunto de peças com componentes rectangulares (i. e. peças cujos ângulos internos são 90^o ou 270^o, por exemplo peças com formas em L, T, U, peças 8, 4 e 5 da Figura 1, respectivamente) sem que estas se sobreponham. As dimensões das peças são conhecidas, o posicionamento deve ser ortogonal (i.e., com os lados paralelos aos eixos horizontal e vertical) e as peças podem sofrer rotações de 90^o, 180^o ou 270^o. É um problema bidimensional e de acordo com a recente tipologia de problemas de C&E apresentada por Wäscher et al [2007] está classificado como sendo um problema de dimensões variáveis (ODP), uma vez que as dimensões do rectângulo exterior envolvente não são conhecidas. Tal como já foi referido do ponto de vista científico é um problema de Optimização Combinatória NP-difícil.

O problema MAPFCRE surge em vários contextos reais: tais como o posicionamento de módulos em circuitos VLSI, (*Very Large Scale Integration*), que tem como objectivo encontrar o melhor posicionamento dos módulos para obter circuitos compactos, e no projecto do *layout* de instalações, no qual um dos objectivos passa por encontrar o arranjo mais eficiente de departamentos dentro de uma instalação. Por exemplo, no posicionamento de peças em circuitos VLSI módulos com componentes rectangulares surgiram para facilitar a utilização da área do circuito e melhorar a ligação entre as peças, contribuindo para o aumento da performance do circuito.

Os primeiros trabalhos que surgiram na literatura para resolver problemas de minimização de área apenas lidam com rectângulos e a principal preocupação é encontrar estruturas de dados eficientes para representar os padrões. Estas representações codificam soluções em sequências, grafos ou representações em árvore. Ao longo do tempo, foram surgindo novas representações que são justificadas por melhorias na eficiência da avaliação das soluções, o tipo de esquemas de codificação, a quantidade de redundância que existe na codificação e o número total de configurações. Um dos primeiros trabalhos foi proposto por Wong et al [1986] que apresenta uma

estrutura em árvore para representar padrões guilhotináveis[1]. Um marco importante foi o aparecimento da primeira representação para padrões gerais (não guilhotináveis) proposta por Murata et al [1995] que se baseia num par de sequências, *sequence pair*, para especificar posições relativas entre rectângulos. Muitas outras representações surgiram após o *sequence pair*: *bounded slicing grid* [Nakatake et al, 1996], *O-tree* [Guo et al, 1999], *B*-tree* [Chang e tal, 2000], *transitive closure graph* [Lin e tal, 2001], etc. As representações existentes para o posicionamento de rectângulos foram sendo adaptadas de forma a permitir a sua aplicabilidade a problemas com formas com componentes rectangulares.

Ao contrário do que acontece na maioria das abordagens existentes na literatura para o problema MAPFCRE a abordagem proposta não lida com representações do padrão mas opera directamente sobre o padrão. Na próxima secção apresenta-se uma descrição da heurística proposta para a resolução deste problema.

3- Rect-TOPOS para o problema MAPFCRE

Para a resolução do problema MAPFCRE propomos uma heurística construtiva, Rect-TOPOS. Esta heurística é uma variante da heurística TOPOS que foi originalmente desenvolvida para resolver problemas com peças cujas formas são irregulares[2]. A ideia principal da heurística TOPOS é adicionar sucessivamente uma nova peça à volta de uma solução parcial, i.e., um conjunto de peças já posicionadas (Figura 2). A escolha da peça a posicionar e do respectivo ponto de posicionamento é efectuada em dois níveis. Num primeiro nível são testadas todas as peças a posicionar e todas as orientações admissíveis, num conjunto finito de pontos de posicionamento possíveis, e escolhido um ponto de posicionamento para cada par peça/orientação. Esta escolha baseia-se num dos seguintes critérios:

- minimização da envolvente rectangular da nova solução parcial;
- minimização do comprimento da nova solução parcial;

[1]Um padrão diz-se guilhotinável se puder ser obtido por sucessivos cortes horizontais ou verticais, de um lado ao outro do padrão, que o dividem em dois rectângulos.

[2] Peças com ângulos interiores arbitrários.

- maximização da sobreposição entre as envolventes rectangulares da nova peça e da solução parcial.

A escolha do par peça/orientação efectivamente a posicionar é realizada num segundo nível com base nos critérios:

- diferença entre a área da envolvente rectangular da solução parcial e a área de todas as peças já posicionadas (desperdício);
- sobreposição entre a envolvente rectangular da peça em avaliação e a envolvente rectangular de cada peça já posicionada;
- distância euclidiana entre o centro da envolvente rectangular da peça em avaliação e o centro da envolvente rectangular da solução parcial.

O objectivo é minimizar o comprimento do padrão já que nestes problemas a largura é fixa.

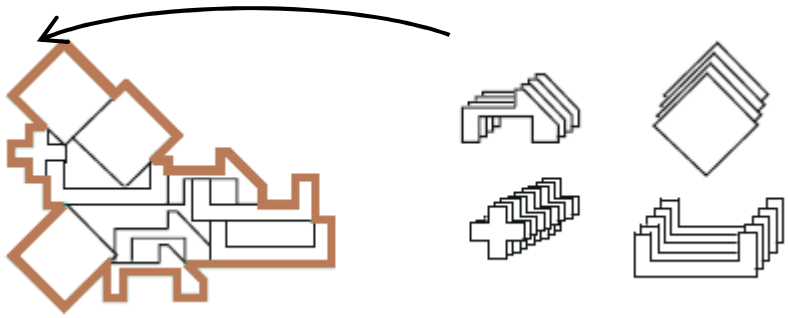
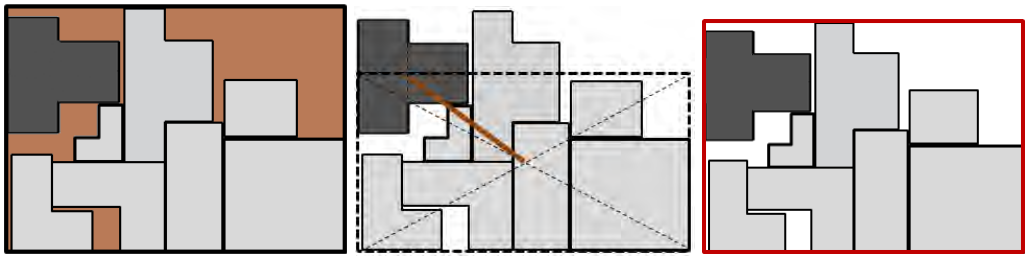


Figura 2 - Construção da solução parcial [Oliveira et al, 2000]

Na variante Rect-TOPOS usamos a mesma ideia geral da heurística TOPOS seleccionando e acrescentando, sucessivamente, uma nova peça com componentes rectangulares, à solução parcial, com o objectivo de obter o rectângulo exterior envolvente de área mínima que contém todas as peças. Optamos por utilizar um único nível para encontrar a próxima peça a posicionar, a sua orientação e o ponto de posicionamento. A existência de um só nível permite que se escolha a melhor peça para posicionar de entre todas as possibilidades contrariamente ao que acontece quando há dois níveis em que existe uma selecção inicial do ponto de posicionamento para cada peça a posicionar. Os critérios usados para efectuar esta escolha foram:

- minimização do *desperdício* da envolvente rectangular resultante (Figura 3(a));

- minimização da *distância* do centro da envolvente rectangular da peça a posicionar ao centro da envolvente rectangular das peças já posicionadas (Figura 3(b));
- minimização do *perímetro* da envolvente rectangular resultante (Figura 3(c)).



(a) ***Desperdício*** **(b)** ***Distância*** **(c)** ***Perímetro***

Figura 3 Critérios para escolha da peça a posicionar e respectivo ponto de posicionamento

Os dois primeiros critérios são retirados directamente dos critérios utilizados no segundo nível do TOPOS e foi introduzido um novo critério - a minimização do perímetro da envolvente rectangular que contém a peça em avaliação e a solução parcial, Figura 3(c).

O 3º critério utilizado no TOPOS (minimização da sobreposição entre a envolvente rectangular da nova peça a posicionar e a envolvente rectangular de cada peça já posicionada) não foi considerado uma vez que não é adequado a situações em que nas peças a posicionar existe um grande número de rectângulos. Nestas situações, como a envolvente rectangular de um rectângulo é o próprio rectângulo, não faz sentido tentar maximizar a sobreposição de dois rectângulos pois não é permitida a sobreposição de peças (Figura 4).

Figura 4 Critério da sobreposição com rectângulos

Tal como acontece no TOPOS o processo iterativo necessita de seleccionar uma primeira peça, para iniciar a solução parcial. Para esta selecção optamos por usar critérios que privilegiam a selecção de peças “grandes”:

- peça com maior *área*;

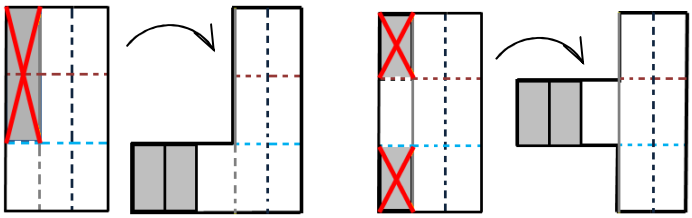
- peça com maior *perímetro*;
- peça com maior *largura*.

4- Testes Computacionais

Nesta secção serão apresentados os resultados computacionais obtidos com a heurística Rect-TOPOS. Os testes computacionais foram realizados numa estação de trabalho Linux equipada com duplo processador Intel XEON Dual Core 5160, 3GHz. Apesar de a estação de trabalho possuir dois CPUs, foi utilizado apenas um processador nos testes.

Para avaliar a heurística Rect-TOPOS foram utilizados 4 conjuntos de instâncias de teste, tendo as instâncias de um deles sido geradas artificialmente e as dos 3 restantes retiradas de publicações. Foram utilizados os seguintes 4 conjuntos de instâncias:

- instâncias do conjunto de referência MCNC (http://vlsicad.eecs.umich.edu/BK/MCNCbench/HARD/) com origem no projecto de circuitos VLSI, no qual todas as peças têm forma rectangular e onde o número total de peças a posicionar não ultrapassa as 50 (APTE, XEROX, HP, AMI33, AMI49);
- instâncias que contém também apenas rectângulos, no entanto, diferem das anteriores por apresentarem um número total de peças entre 100 e 500 (http://www.simplex.t.u-tokyo.ac.jp/~imahori/packing/), (RP100, RP200, PCB146, PCB500);
- instâncias de [Nakatake et al,1996] (NAKATAKE1, NAKATAKE2), [Lin et al, 2002] (LIN) e [Xu et al, 1998] (AMI49L, AMI49LT) que contêm um misto de peças que são rectângulos, peças com forma em L e/ou T e outras peças com componentes rectangulares (U, +, H, em escada);
- instâncias geradas a partir das instâncias AMI33 e AMI49 do conjunto de referência MCNC (AMI33LTa, AMI49LTa). Nestas instâncias usou-se como regra alterar aproximadamente 10% do número total de rectângulos para peças com formas em L e/ou T. Cada peça é construída de forma a ter dimensões inteiras e a sua área seja aproximada da área do rectângulo que lhe deu origem, Figura 5.



(a) **(b)**

Figura 5 Construção de peças com componentes rectangulares nas instâncias geradas

As instâncias de teste utilizadas apresentam características variadas, nomeadamente no número total de peças, no número de peças com formas diferentes (número de tipos de peças), na forma das peças (rectangular, em L, em T e outras formas com componentes rectangulares). Estas características encontram-se resumidas na Tabela 1.

Instância	# Peças		Peças com componentes rectangulares			
	Total	Tipos	# Rect.	# L	# T	# Outras
APTE	9	3	9	—	—	—
XEROX	10	10	10	—	—	—
HP	11	6	11	—	—	—
AMI33	33	31	33	—	—	—
AMI49	49	46	49	—	—	—
RP100	100	99	100	—	—	—
PCB146	146	22	200	—	—	—
RP200	200	99	146	—	—	—
PCB500	500	417	500	—	—	—
AMI49L	28	28	7	21	0	0
AMI49LT	27	27	6	20	1	0
NAKATAKE1	40	35	30	10	0	0
NAKATAKE2	19	19	5	3	2	9
LIN	29	21	22	1	1	5
AMI33LTa	33	31	30	2	1	0
AMI49LTa	33	46	41	3	2	0

Tabela 1 – Características das instâncias usadas

A Tabela 2 resume os testes computacionais efectuados para avaliar a heurística Rect-Topos. Foram testados os 3 critérios, para escolha da próxima peça a posicionar, a sua orientação e o ponto de posicionamento, anteriormente apresentados (*desperdício*,

distância e *perímetro*) e, para cada um deles, foram consideradas três possibilidades, *área*, *perímetro* e *distância*, para a escolha da peça que inicia a solução parcial. Os valores indicados na tabela são os aproveitamentos obtidos, medidos como o quociente entre a área das peças posicionadas e a área da envolvente rectangular obtida. A negrito indica-se o melhor resultado obtido para cada instância. Na Tabela 2 apresenta-se também, para cada instância, o tempo médio de resolução (Tm), Refira-se que os tempos de resolução, para cada uma das combinações dos critérios, aplicadas a cada instância, não apresentam grande variabilidade. Finalmente apresentam-se ainda, para cada instância, o melhor resultado encontrado na literatura: aproveitamento (Aprov.), tempo de resolução (T) e a indicação da referência à publicação.

	Desperdício			***Distância***			***Perímetro***			**Tm**	**Melhor Resultado**		
Instância	***Área***	***Per.***	***Larg.***	***Área***	***Per.***	***Larg.***	***Área***	***Per.***	***Larg.***	**(s)**	**Aprov.**	**T (s)**	
APTE	**0.917**	**0.917**	**0.917**	0.893	0.893	0.893	0.894	0.894	0.894	0.01	0.992	2.38	[1]
XEROX	0.801	0.801	0.801	**0.804**	**0.804**	**0.804**	0.788	0.788	0.788	0.09	0.977	9812	[1]
HP	0.848	0.848	0.695	0.834	0.834	0.695	0.924	0.924	**0.936**	0.03	0.987	891	[1]
AMI33	0.813	0.813	0.875	0.712	0.712	0.745	0.832	0.832	**0.863**	0.84	0.986	2.01	[2]
AMI49	0.807	0.807	0.807	0.792	0.792	0.792	**0.843**	**0.843**	**0.843**	1.97	0.983	56.61	[2]
RP100	0.819	0.819	0.857	0.721	0.721	0.773	**0.924**	**0.924**	0.905	9.35	0.968	200	[3]
PCB146	0.622	0.622	0.622	0.786	0.786	0.786	**0.881**	**0.881**	**0.881**	0.95	0.977	300	[3]
RP200	0.876	0.876	0.878	0.746	0.746	0.754	**0.929**	**0.929**	0.913	13.2	0.963	400	[3]
PCB500	0.865	0.865	0.865	0.781	0.781	0.781	**0.967**	**0.967**	**0.967**	221.0	0.963	1000	[3]
AMI49L	0.625	0.625	0.667	0.761	0.761	0.761	**0.829**	**0.829**	0.792	1.11	0.956	2728	[4]
AMI49LT	0.731	0.731	0.663	0.787	0.787	0.753	0.793	0.793	**0.823**	1.08	0.951	2843	[4]
NAKATAKE1	0.825	0.825	0.763	0.807	0.807	0.784	0.852	0.852	**0.867**	1.35	0.969	10.24	[4]
NAKATAKE2	0.590	0.619	0.590	0.669	0.614	0.669	0.731	**0.770**	0.731	2.14	0.802	93.32	[4]
L**IN**	0.789	0.740	0.780	0.780	0.780	0.740	0.840	0.840	**0.910**	0.82	1.000	2.72	[4]
AMI33LTa	0.764	0.764	0.711	0.752	0.752	0.73147	0.832	0.832	**0.844**	1.05	—	—	—
AMI49LTa	0.875	0.875	0.875	0.761	0.761	0.76072	**0.881**	**0.881**	**0.881**	2.41	—	—	—

[1] [Chan et al, 2004] - 1.2 GHz Linux Athlon workstation
[2] [Chen et al, 2007] - 2.4 GHz PC with 512 MB memory
[3] [Imahori et al, 2005] - Intel Pentium III 1 GHz, 1 GB of memory
[4] [Chen et al, 2010] - IBM portable PC with 2.0 GHz processor and 512MB memory.

Tabela 2 – Comparação dos resultados obtidos pelo Rect-Topos vs. resultados da literatura

Da Tabela 2 podemos constatar que os melhores resultados foram obtidos quando se usa para a escolha da próxima peça a posicionar e respectivo ponto de posicionamento o critério *perímetro*, à excepção das instâncias APTE e XEROX. Estas duas instâncias são muito sensíveis à escolha da primeira peça a posicionar, já que têm um reduzido número

de peças a posicionar, 9 e 10 respectivamente. Relativamente à escolha da primeira peça, os resultados mostram um maior equilíbrio entre os três critérios. Na comparação dos resultados obtidos com os melhores resultados publicados deverá ter-se em consideração que o Rect-TOPOS é apenas uma heurística construtiva, enquanto os melhores resultados publicados foram obtidos com abordagens baseadas em pesquisa local e em algoritmos de pesquisa em árvore. Em relação à comparação dos tempos de execução, a informação insuficiente, acerca das características do processador usado, disponível em [Chan et al, 2004], [Chen et al, 2007] e [Chen et al, 2010], não nos permite fazer uma comparação rigorosa. No entanto, uma simples análise permite concluir que os nossos tempos são claramente inferiores. Na Tabela 3 apresenta-se a comparação com os tempos de execução publicados em [Imahori et al, 2005]. Nesta tabela, os tempos da coluna T' resultaram da aplicação de um factor correctivo que reflecte as velocidades relativas dos computadores utilizados nos dois conjuntos de testes. O factor correctivo foi calculado com base nos dados disponíveis em www.cpubenchmark.net. Da comparação apresentada podemos concluir que os tempos de execução da heurística Rect-TOPOS são bastante competitivos.

	Rect-TOPOS	**[Imahori et al, 2005]**	
Instância	**Tm (s)**	**T (s)**	**T' (s)**
RP100	9.35	200	50.9
PCB146	0.95	300	76.4
RP200	13.2	400	101.8
PCB500	221.0	1000	254.7

Tabela 3 – Comparação dos tempos de execução entre Rect-Topos e [Imahori et al, 2005]

Assim, como seria de esperar, os resultados obtidos com o Rect-TOPOS ficam aquém dos publicados, mas em contrapartida os tempos de execução são muito inferiores. É de realçar que para a instância PCB500 o resultado obtido pelo Rect-TOPOS foi melhor do que o melhor resultado encontrado na literatura [Imahori et al, 2005].

A Tabela 3 permite também mostrar o grande impacto que o número de tipos de peças tem nos tempos de execução da heurística Rect-TOPOS. Por exemplo, apesar de a instância PCB146 ter mais 46 peças que a instância RP100 o seu tempo de execução é cerca de 10 vezes inferior, já que tem apenas 22 tipos de peças diferentes enquanto a instância RP100 tem 99 tipos diferentes.

Na Figura 6, a título de exemplo, apresenta-se a solução obtida para a instância PCB500.

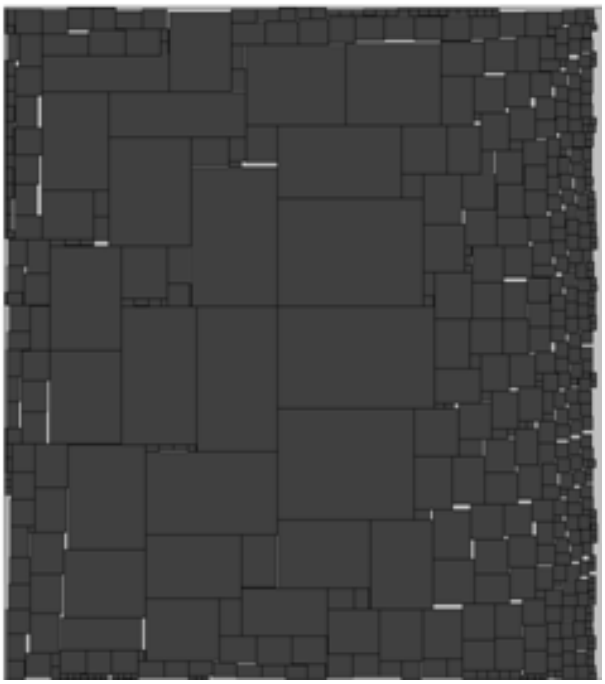


Figura 6 Solução obtida para a instância PCB500

5- Conclusões

Neste artigo apresentamos uma heurística construtiva, Rect-TOPOS, para o problema MAPFCRE. Trata-se de uma heurística rápida e que é capaz de lidar facilmente com todo o tipo de formas com componentes rectangulares. Esta heurística usa vários critérios para escolher a próxima peça a posicionar, a sua orientação e o melhor ponto de posicionamento. A qualidade das soluções obtidas revelou-se bastante satisfatória dado tratar-se de uma simples heurística com reduzidos tempos de execução. Estas características sugerem, como desenvolvimentos futuros, a incorporação da heurística Rect-TOPOS numa abordagem baseada em pesquisa local. Nesta abordagem seria possível, à custa de um aumento do tempo de execução, melhorar os já bons resultados alcançados pelo Rect-TOPOS em situações onde tal fosse necessário.

6- Referências

Chan H., Markov. L. (2004). Practical slicing and non-slicing block-packing without simulated annealing. *Proceedings of the 14th ACM Great Lakes symposium on VLSI, GLSVLSI '04 (ACM, New York, NY, USA)*. 282-287.

Chang Y. C., Chang Y. W., Wu G. M., Wu S. W. (2000). B*-trees: A new representation for nonslicing foorplans. *Proceedings of the 37th Design Automation Conference, DAC '2000.* 458-463.

Chang Y.W, Lin J. M. (2000). TCG: A Transitive Closure Graph-Based Representation for Non-Slicing Floorplans. *Proceedings of the Design Automation Conference, DAC '2001*. 764-769.

Chen M., Huang. W. (2007). A two-level search algorithm for 2D rectangular packing problem. *Computers and Industrial Engineering.* 53:1123-136.

Chen D., Liu J., Fu Y., Shang. M. (2010). An efficient heuristic algorithm for arbitrary shaped rectilinear block packing problem. *Computers and Operations Research.* 37-6, 1068-1074.

Guo P. N., Cheng C. K., Yoshimura T. (1999). An O-tree Representation of Non-Slicing Floorplan and Its Applications. *Proceedings of the36th annual ACM/IEEE Design Automation Conference, DAC '99 (New York, NY, USA), ACM.* 268-273.

Imahori S., Yagiura M., Ibaraki T. (2005). Improved local search algorithms for the rectangle packing problem with general spatial costs. *European Journal of Operational Research, Elsevier*. 167(1), 48-67.

Lin J.-M., Chen H.-L., Chang Y.-W. (2002). Arbitrarily shaped rectilinear module placement using the transitive closure graph representation. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*. 10:6, 886- 901.

Murata H., Fujiyoshi K., Nakatake S., Kajitani Y. (1995). Rectangle packing module placement. *Proceedings of the 1995 IEEE/ACM international conference on Computer-aided design, ICCAD' 95 (Washington, DC, USA), IEEE Computer Society*. 472-479.

Nakatake S., Fujiyoshi K., Murata H., Kajitani Y. (1996). Module placement on BSG-structure and IC layout applications. *Proceedings of the 1996 IEEE/ACM international conference on Computer-aided design, ICCAD' 96 (Washington, DC, USA), IEEE Computer Society.* 484-491.

Oliveira J. F., Gomes A. M., Ferreira J. S. (2000). TOPOS: A new constructive algorithm for nesting problems. *OR Spectrum.* 22, 263-284.

Wäscher G., Haußner H., Schumann H. (2007). An Improved Typology of Cutting and Packing Problems. *European Journal of Operational Research*. 183, 1109-1130.

Wong D. F., Liu C. L. (1986). A new algorithm for floorplan design. *Proceedings of the 23rd ACM/IEEE Design Automation Conference, DAC '86, (Piscataway, NJ, USA), IEEE Press.* 101-107.

Xu J., Guo P.-n., and Cheng C.-K. (1998). Rectilinear block placement using sequence-pair. *Proceedings of the 1998 international symposium on Physical design (ISPD '98). ACM, (New York, NY, USA)*. 173-178.

Prioritizing Health Care Interventions: A Multicriteria Resource Allocation Model to Inform the Choice of Community Care Programmes

Mónica D. Oliveira [a,*], *Teresa C. Rodrigues*[a], *Carlos A. Bana e Costa*[a], *Armando Brito de Sá*[b]

[a] *Centre for Management Studies of IST, Instituto Superior Técnico, Technical University of Lisbon, Lisbon, Portugal*

[b] *Faculdade de Medicina da Universidade de Lisboa, Lisbon, Portugal*

*- *Corresponding author:* monica.oliveira@ist.utl.pt

Abstract

Many countries, and Portugal in particular, are currently dealing with budget cuts and a shortage of resources in the health sector, while the demand for health care services is increasing. Within this context, the Group of Health Centres (GHC) of Northern Lisbon faces the challenge of prioritizing community care programmes so as to decide which programmes should be funded. This article describes the development of a Multi-criteria ***M***odel to ***A***llocate Human ***R***esources in ***C***ommunity *C*are Pr***O***grammes (MARCCO) with the GHC to appraise the costs and the benefits of the proposed programmes within a multicriteria resource allocation model structure. Building MARCCO followed a socio-technical approach using Multi-Criteria Decision Analysis (MCDA) and decision conferencing. The GHC used the results obtained by MARCCO in the selection of programmes and in the re-design of its information system. MARCCO contributes to the literature by showing how a constructive approach using MCDA methods and decision conferencing is an alternative to conventional approaches used in the prioritization of interventions in the health care sector.

Keywords: Resource Allocation, Portfolio Decision Analysis, Multicriteria Analysis, Primary Care, Portugal.

1. INTRODUCTION

Many countries, and Portugal in particular, are currently dealing with budget cuts and a shortage of resources in the health sector, while the demand for health care services is increasing. This calls for transparent prioritization of health care interventions by carefully balancing their multiple costs and benefits [Mitton, 2002]. Clearly, there is a "need for multi-criteria decision analysis" [Baltussen and Niessen, 2006] (MCDA) and decision conferencing [Phillips, 2007] to support resource allocation in health care organizations. In fact, the use of MCDA and decision conferencing within a "socio-technical process improves communication

within the organisation, develops shared understanding of the portfolio and generates a sense of common purpose about those projects that will best realise the organisation"s objectives" [Phillips and Bana e Costa, 2007]. Previous studies have already shown the usefulness of MCDA to help prioritizing health care interventions [Baltussen and Niessen, 2006] [Peacock, Mitton et al., 2009].

Based on a National Health Service (NHS) structure created in 1979, the Portuguese health care system is defined as universal, nearly free at the point of use and funded by general taxation [Assembleia da República, 1992]. The first level of contact in the NHS is primary care, which is undergoing major reforms since 2005. Groups of Health Centres (GHC) were instituted in 2008, to promote quality and equity of access and reorganize the existing Health Centres (HC) into operative divisions with complementary responsibilities [Pisco, 2007]; followed by the creation of a community care division (CCD) in each HC, responsible for providing health care services and educating citizens, offering both social and psychological services to individuals, families and vulnerable groups at higher risk and with physical and functional dependence [Missão para os Cuidados de Saúde Primários, 2008]. In 2010, each of the four HC of the Northern Lisbon GHC started the process of creating its own CCD. The first step was the development of an action plan composed of several community care programmes, which should be approved by the GHC board. A first analysis of the four action plans showed that it was not realistic to implement all programmes proposed, because of shortage of financial and human resources (in particular, nurses). Consequently, a selection of programmes should be made within each action plan. This called for the development of a common basis to appraising the costs and the benefits of the programmes. For this purpose, the authors suggested the GHC board to follow a socio-technical approach using the MCDA approach to construct a Multi-criteria ***M**odel to **A**llocate Human **R**esources in **C**ommunity **C**are Pr**O**grammes* (MARCCO).

2. BUILDING MARCCO

MARCCO was the result of a socio-technical process framed within a multicriteria resource allocation model [Phillips and Bana e Costa, 2007] structured and built with the decision-making group (DM) – the members of the Clinical Board and the Executive Director of the Northern Lisbon GHC – in a sequence of decision conferences during which the group was involved in the model building activities depicted in Figure 1.

The group started by discussing the specificities of the GHC/CCD decision context and concluded that the key resource constraining the selection of all proposed programmes was the scarce number of hours of nursing available in each HC. A careful selection should therefore be made and it was consensual that although the main objective is to maximize health gains, the selection should also consider other potential benefits of the programmes, as shown in the value tree of Figure 2.

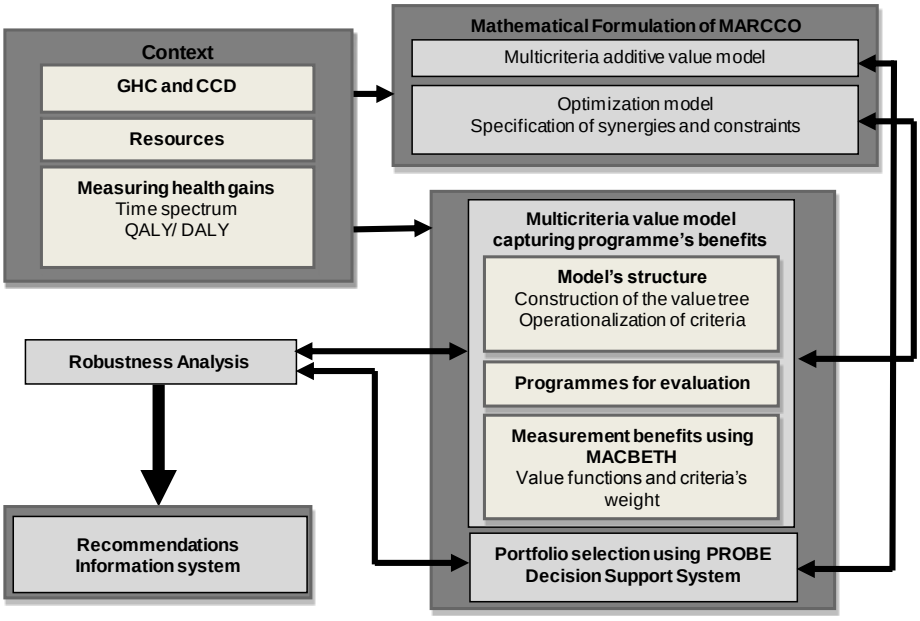


Figure 1 - Activities developed to build MARCCO

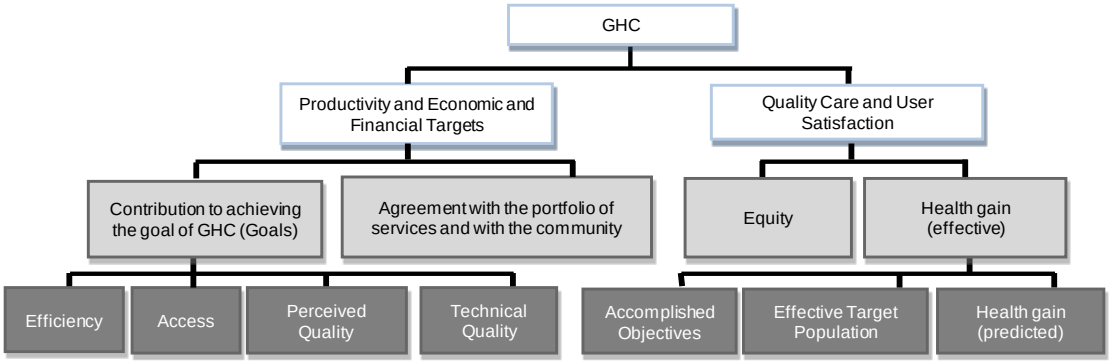


Figure 2 - Value tree - strategic objectives in white; evaluation criteria in light-grey and elementary aspects in dark-grey

The multicriteria method MACBETH (*Measuring Attractiveness by a Categorical Based Evaluation Technique* [Bana e Costa, De Corte et al., 2010] [Bana e Costa and Chagas, 2004]) and its software (M-MACBETH [Bana Consulting, 2005]) were used to construct the multicriteria evaluation model. A distinctive characteristic of MACBETH is that it requires only qualitative judgements from the DM about differences in attractiveness (or added value) in order to score programmes on each criterion and to weight criteria. The multicriteria model allows to measure the overall benefit of each programme by a simple weighted sum procedure. These overall benefits of the programmes and the "costs" in nursing hours required

to achieve them are then used to perform a portfolio decision analysis and select the portfolio of programmes that offers the best global benefit value for the limited number of nursing hours available. The portfolio selection process was supported by the PROBE approach and software (*Portfolio Robustness Evaluation* [Lourenço and Bana e Costa, 2009]). A distinctive characteristic of PROBE is that it allows the DM to take into consideration, in the selection of a portfolio, different forms of interdependences between programmes – for example, synergies and/or incompatibilities - but also to analyse the robustness of a selection in face of uncertainty phenomena (like imprecise performance estimations or doubts about values and weights).

Each CCD proposes to the GHC an action plan with m programmes j (j = 1,..,m). The benefit values of the programmes are to be evaluated, by the additive value model, on the n benefit evaluation criteria i (i=1,...,n). The benefit performance of each programme j on each criterion i is appraised by associating it with one level of performance of the descriptor (or attribute) X_i of criterion i (i=1,…,n). For example, the performances on the criterion "agreement with the portfolio of services and with the community" are described by the three-level qualitative scale displayed in Figure 3. Two reference levels of performance, $neutral_i$ and $good_i$ (those shaded in Figure 3), were identified in each descriptor X_i. The performance of j on each criterion i (i=1,…,n) is converted in added value by a value function $v_i(.)$ constructed with MACBETH in such a way that a score of 0 is assigned to $neutral_i$ and a score of 100 is assigned to $good_i$ and the difference between two scores measures the difference in attractiveness between the respective programmes, with respect to the same difference between $good_i$ and $neutral_i$. The MACBETH pairwise judgements expressed by the DM between performance levels of the descriptor in Figure 3 are shown in the matrix in Figure 4, for which M-MACBETH suggested the numerical values depicted in Figure 5a (for details, see [Bana e Costa, De Corte et al., 2010]). The DM was then invited to analyse the differences between the MACBETH scores and if necessary adjust the score of the intermediate level so that the differences reflect the relative differences in attractiveness between the respective performance levels. The final scale validated by the DM is shown in Figure 5b. Moreover, the comparison, with MACBETH, of the differences between the two references on the several criteria allowed the DM to assign a weight k_i to each criterion i (i=1,...,n), with $\sum_{i=1}^{n} k_i = 1$ and $k_i > 0$). In this way, the common critical mistake of weighting criteria based on judgements of importance could be avoided [Keeney, 1992].

Levels	Description
CCD+Community (*good*)	Programme consistent with the CCD's portfolio of services and a priority to the community
CCD	Programme consistent with the CCD's portfolio of services but not a priority to the community
No_CCD (*neutral*)	Programme not consistent with the CCD's portfolio of services and not a priority to the community

Figure 3 - Qualitative levels of 'Agreement with the portfolio of services and with the community'

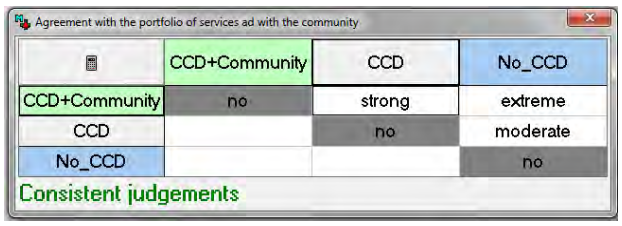
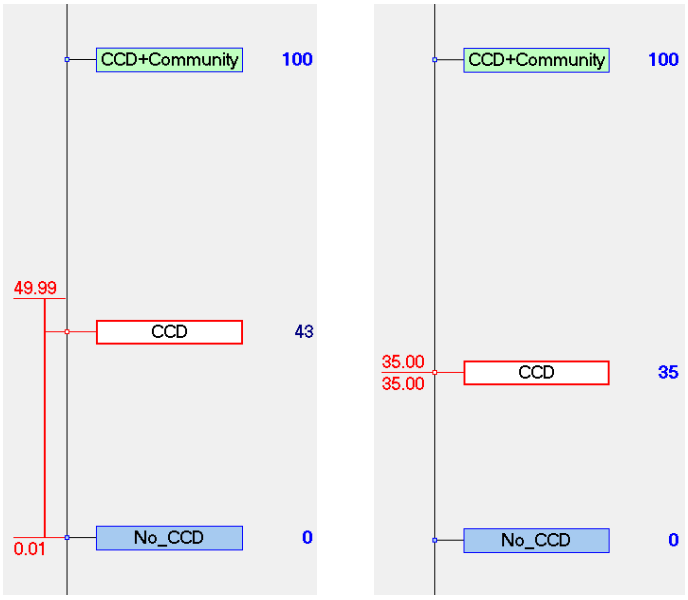


Figure 4 - M-MACBETH judgements for the 'Agreement with the portfolio of services and with the community' criterion



(a) (b)

Figure 5 – MACBETH thermometers for the 'Agreement with the portfolio of services and with the community' criterion (a) MACBETH scale; (b) Interval Scale

Consider a specific programme j with benefits appraised by the performance profile $x_{1j}, \ldots, x_{nj}$. The overall benefit value for that programme ($v_j(.)$) is determined by the multicriteria additive value model defined in Equation (1).

$$v_j = v_j(x_{1j}, \ldots, x_{nj}) = \sum_{i=1}^{n} k_i v_i (x_{ij}) \quad (1)$$

The use of the additive model (1) should acknowledge that the difference independence condition is respected, i.e., the difference in attractiveness between performance levels on each evaluation criterion can be measured independently of the performance levels on the rest of the criteria, being this considered a working hypothesis in the model building process [Bana e Costa, Carnero et al., 2010]. This is the reason why several elementary evaluation

aspects (shown in dark-grey in Figure 2) were clustered into two of the evaluation criteria (shown in light-grey in Figure 2).

If the overall value of a proposed programme results negative from model (1), it means that the programme is globally unattractive and therefore it should be rejected. Consequently, the baseline in the MARCCO resource allocation model is a programme that adds no value [Clemen and Smith, 2009] [Lourenço, Bana e Costa et al., 2010].

Consider now that programme j "costs" c_j nursing hours and B is the total number of nursing hours available. Assume that B is not fixed and can vary within a certain range. The portfolio of programmes offering the highest global benefit under these conditions can be found by solving the basic optimization problem (2) (where l_j is the decision variable that assumes the value of 1 if programme j is included in the portfolio and zero otherwise).

$$\text{maximize: } \sum_{j=1}^{n} v_j \, l_j \qquad (2)$$

$$\text{subject to: } \sum_{j=1}^{n} c_j \, l_j \leq B$$

$$l_j \in \{0,1\}, j = 1, \ldots, m$$

Other (linear) constraints enabling to model interdependencies between programmes can be added to model (2), for example: programme j must be chosen given past commitments ($l_j = 1$); j cannot be carried out this year ($l_j = 0$); j can only be chosen if programme h is also selected ($l_j - l_h \leq 0$); j and h need to be implemented simultaneously ($l_j - l_h = 0$), or the opposite ($l_j + l_h \leq 1$). It is also possible to take into account restrictions in other type of resources, for instances, other human resources such as nutritionists, physiatrists, etc. Let h_{kj} represent the hours of a human resource k consumed by programme j and H_k the total number of hours of k that are available. This restriction can be considered in the selection by simple adding to the basic model (2) the linear constraint $\sum_{j=1}^{n} h_{kj} \, l_j \leq H_k$.

3. APPLYING MARCCO AT THE GHC

Results from applying MARCCO to the selection of a portfolio of community care programmes for the Alvalade CCD are here presented. Table 1 shows the overall benefit values, calculated by the multicriteria model, and „costs" (in a total of 9589,5 nursing hours required) of the programmes, which are presented by decreasing order of their benefit/cost ratios (a measure of their efficiency). A first finding for the group was that there was a significant percentage of low efficient programmes and wadding little value. Anyway, as the Alvalade budget is 8820 nursing hours, some programmes should be excluded.

Table 1. Data for Alvalade community care programmes.

Programmes	Overall Benefit	Nursing hours required	Benefit/Cost ratio
Social integration income	51,72	0	+∞
Domiciliary visits: healthy lives	53,28	5,5	9,69
Physical exercise for the elderly	19,04	4	4,76
Prevention of domestic accidents for the elderly	44,35	64	0,69
Reproductive health and family planning	51,55	96	0,54
Group support for teenagers	45,19	182	0,25
Support to the child and youth vulnerable groups	68,58	1002	0,068
Preparation for maternity/paternity	23,68	464	0,051
Preparation for post-delivery	23,07	464	0,050
Integrated long-term Care	143,74	2940	0,049
Health at home	4,90	204	0,024
Promoting family parenting for vulnerable families	53,32	4164	0,013
TOTAL	582,42	9589,5	-

The overall benefits and the nursing hours costs of the programmes were the inputs for a portfolio decision analysis with the PROBE software, in which one linear constraint was included, to ensure that the „Preparation for post-delivery" programme is chosen together with „Preparation for maternity/paternity." Supposing that the available nursing hours varies from 0 to 9589,5 hours, Figure 6 presents the main screen of PROBE, in which the dots represent the efficient portfolios (a portfolio is efficient when any other portfolio with higher global benefit require more nursing hours); they form the efficient frontier. Figure 7 helps to make clear that when moving up in the efficient frontier, from one portfolio (A) to the next one (B), what matters is the ratio between added benefit and added cost, while maximizing the global benefit.

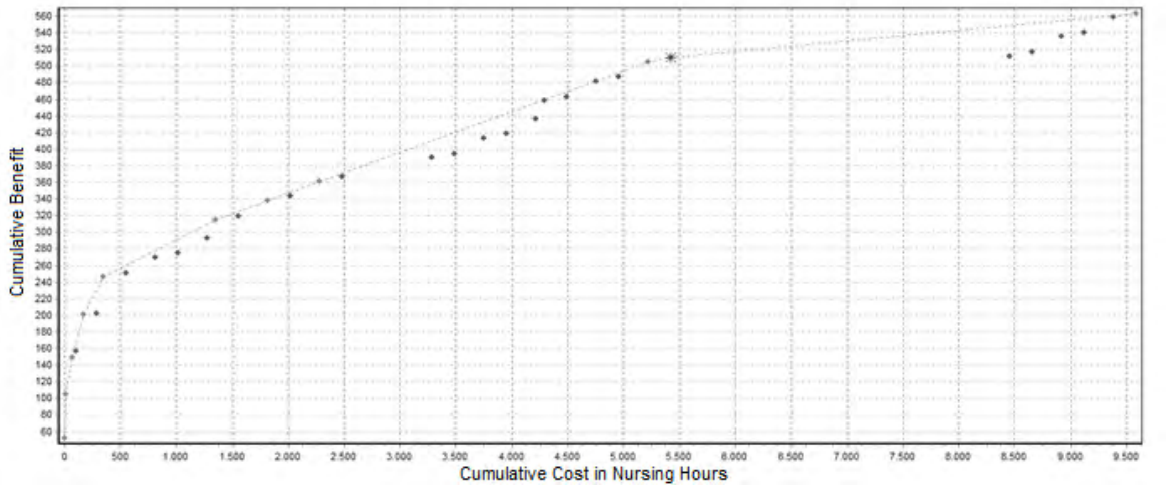


Figure 6 - Efficient portfolios for Alvalade's CCD analyzed using the PROBE DSS

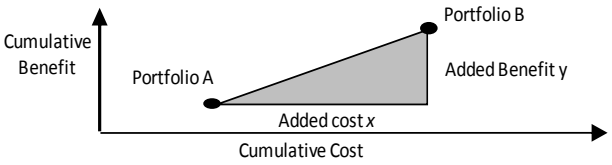


Figure 7 - Cost-benefit triangle

Three zones of the graph in Figure 6 were observed by the group: a first zone (corresponding to portfolios summing less than 500 nursing hours), almost vertical, where the added global benefit per added nursing hour is very high; a second zone (between 500 and 5500 hours) with smaller but still high efficiency; and a third zone (demanding more than 5500 hours) in which an additional benefit unit „costs" a lot of hours. This observation made the DM decide that it is not worthwhile to spend more than 5500 nursing hours and therefore the portfolio represented by a stat dot in Figure 6 was selected. A robustness analysis confirmed that this portfolio is highly stable to uncertainty in criteria weights and/or in programmes. Moreover, the Alvalade CCD was asked to reconfigure current programmes or propose alternative ones, in order to spend more efficiently the remaining nursing hours available (more than 3000 hours). MARCCO has also helped to improve the GHC's information system by building a template for future action plans, so that CCDs could propose programmes to be compared on a common basis and so that, when defining programmes, CCDs could better understand which programme features are valued by the GHC.

4. DISCUSSION

MARCCO allowed for transparent prioritization of health care interventions by balancing programmes' costs and benefits for the GHC within a multicriteria resource allocation model structure, and decision conferencing enabled achieving an agreement on the way forward. The members of the Clinic Board and the Executive Director of the Northern Lisbon GHC have identified themselves with the results obtained by MARCCO and used them in the selection of programmes and in the re-design of the information system. MARCCO contributes to the literature by showing how a constructive approach using MCDA methods and decision conferencing is an alternative to conventional approaches used in the prioritization of interventions in the health care sector, highlighting the usefulness of MCDA in priority setting in the health sector suggested in previous studies [Peacock, Mitton et al., 2009] [Baltussen and Niessen, 2006].

REFERENCES

Assembleia da República, *Constituição da República Portuguesa*. 1992, Lisboa: Assembeia da República.

Baltussen, R. and L. Niessen, *Priority setting of health interventions: the need for multi-criteria decision analysis.* Cost Effectiveness and Resource Allocation, 2006. **4**(14).

Bana Consulting, *"M-MACBETH Version 1.1: User Manual."*. 2005: Lisbon.

Bana e Costa, C.A., M.C. Carnero, and M.D. Oliveira, *A multi-criteria model for auditing a Predictive Maintenance Program.* Working Paper Centre for Management Studies of Instituto Superior Técnico (CEG-IST), 2010. **4/2010**.

Bana e Costa, C.A. and M.P. Chagas, *A career choice problem: an example of how to use MACBETH to build a quantitative value model based on qualitative value judgments.* European Journal of Operational Research, 2004. **153**(2): p. 323-331.

Bana e Costa, C.A., J.-M. De Corte, and J.-C. Vansnick, *MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique)*, in *Wiley Encyclopedia of Operations Research and Management Science*, J.J. Cochran, Editor. 2010, Wiley.

Clemen, R.T. and J.E. Smith, *On the Choice of Baselines in Multiattribute Portfolio Analysis: A cautionary Note.* Decision Analysis, 2009: p. 1-7.

Keeney, R.L., *Value-Focused Thinking: A Path to Creative Decision Making*. 1992: Havard University Press.

Lourenço, J.C. and C.A. Bana e Costa, *PROBE - A Multicriteria Decision Support System for Portfolio Robustness Evaluation, Working Paper LSEOR 09.108, London School of Economics, London.* 2009.

Lourenço, J.C., C.A. Bana e Costa, and J.O. Soares. *Portfolio Decision Analysis with PROBE: Addressing Costs of not Financing Projects* in *Advances in Mathematical and Computational Methods - 12th WSEAS International Conference on Mathematical and Computational Methods in Science and Engineering*. 2010. University of Algarve, Faro,Portugal: WSEAS Press.

Missão para os Cuidados de Saúde Primários, *Unidade de Cuidados na Comunidade*. 2008, Available in the internet www. URL: http://www.mcsp.min-saude.pt/engine.php?cat=95. Consulted in May 2010.

Mitton, C., *Priority setting for decision makers: using health economics in practice.* European Journal of Health Economics, 2002. **4**: p. 240-243.

Peacock, S., C. Mitton, A. Bate, et al., *Overcoming barriers to priority setting using interdisciplinary methods.* Health Policy, 2009. **92**: p. 124-132.

Phillips, L.D., *Decision Conferencing*, in *Advances in Decision Analysis: From Foundations to Applications*, Edwards, Miles, and v.W. (eds), Editors. 2007, Cambridge University Press.

Phillips, L.D. and C.A. Bana e Costa, *Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing.* Annals of Operations Research, 2007. **154**(1): p. 51-68.

Pisco, L., *A Reforma dos Cuidados de Saúde Primários*, in *Cadernos de Economia*. 2007. p. 60-66.

Índice Técnico para Pentatlo Moderno baseado em DEA

Nissia Carvalho Rosa Bergiante, Mariana Vieira Rangel Nunes, João Carlos Soares de Mello*

Universidade Federal Fluminense, Niterói, Brasil
nissia.rosa@gmail.com

Universidade Federal Fluminense, Niterói, Brasil
mariana.vr.nunes@gmail.com

Universidade Federal Fluminense, Niterói, Brasil
jcsmello@pq.cnpq.br
* *Corresponding Author*

Resumo

Esse artigo tem como objetivo desenvolver um Índice Técnico baseado nos melhores resultados de cada modalidade do Pentatlo Moderno. O Índice técnico será útil para auxiliar na formação de eficientes programas de treinamento para os atletas do esporte. Foram utilizados os resultados da Copa do Mundo do Pentatlo Moderno, realizado em Medway (Reino Unido) em Abril de 2010. A Copa do Mundo é uma séria anual de competições internacionais e uma preparação para os jogos Olímpicos oficiais de 2012. Para o desenvolvimento do índice técnico foi utilizado o modelo avançado de DEA com inputs unitários. Os resultados indicam que alguns países devem priorizar as estratégias de maximização de desempenho em determinadas modalidades de modo a que potencializem os seus resultados finais.

Palavras-chave: Esporte, Pentatlo Moderno, Análise de Regras, Índice Técnico.

1- Introdução

O Pentatlo Moderno é um esporte que inclui cinco modalidades sendo elas esgrima, natação, hipismo e combinado de corrida e tiro. Foi introduzido oficialmente por Pierre de Coubertin nas Olimpíadas de 1912, em Estocolmo, mas seu início data de séculos atrás [Garniner (1925)]; [Langdon (1989)]; [Kile (1990)]. Sua versão primitiva está relacionada às Olimpíadas de 708 a.C quando era usada como método de treinamento de soldados, tamanha a exigência física e de outras habilidades tais como flexibilidade e coordenação consideradas fundamentais para a construção de um atleta ideal. Essa visão de treinamento continuou ainda nas Olimpíadas atuais, quando, em sua introdução era usado como parte dos exames finais da Academia Militar Européia [Clearlead (2009)].

Desde seu início, em 1912 até 1980, o Pentatlo Moderno era disputado ao longo de cinco dias de competição. Esse período foi alterado nas Olimpíadas de 1984, 1988 e 1992 onde se testou uma disputa em quatro dias, onde se unia em um só dia, a natação e o tiro, ou a corrida e o tiro. Esse modelo, porém, não se perpetuou, sendo substituído, em 1996, por apenas um dia de provas. [Kile (1990)]; [Lee (1993)].

Em janeiro de 2009, com a necessidade de tornar o Pentatlo um esporte mais dinâmico e atraente para mídia, foi criado o "combinado", que une as modalidades de Tiro e Corrida [Nunes et al (2010)].

Esse artigo tem como objetivo desenvolver um índice técnico, baseado nos melhores resultados em cada modalidade do esporte. Este índice será construído como uma alternativa de medição de desempenho dos atletas, que poderá auxiliar na elaboração de programas e estratégias de treinamentos que permitam o aumento da eficiência dos países participantes dos jogos.

Para a formulação deste índice será usado como ferramenta um modelo de análise envoltória de dados – DEA (Data Envelopment Analysis). Este modelo possibilita a comparação de eficiência utilizando para tanto, programação linear. Todo o estudo de DEA baseia-se em avaliação de unidades produtivas, que para o caso deste artigo será formada de duas maneiras: A primeira considerará o melhor resultado de cada um dos países participantes das competições e a segunda os resultados dos melhores atletas de cada um dos países. Mesmo no segundo caso, a unidade produtiva terá a denominação do país que o atleta representa.

A obtenção dos índices DEA em um modelo tal como o utilizado neste trabalho fornece um critério de avaliação através da diferença entre a pontuação obtida por cada país/atleta em cada uma das modalidades que disputa e a pontuação que deveria ser obtida para maximizar o desempenho do país tornando-o apto a competir por medalhas. Neste sentido, este índice é importante para analisar se cada país prioriza as estratégias de treinamento nas modalidades corretas que potencializam os seus resultados finais.

2- Descrição das regras do pentatlo

2.1 - Esgrima

No pentatlo, a modalidade de esgrima disputada é a espada, na qual o toque no adversário é valido em qualquer parte da superfície do corpo.

Nesta modalidade os atletas enfrentam uns aos outros e todos devem se enfrentar. Cada um dos embates tem a duração total de 1 minuto ou o tempo que seja suficiente para que um atleta toque o outro. Não havendo qualquer toque por parte dos adversários, ambos devem assumir a derrota. As regras de pontuação na Esgrima determinam que 70% dos toques ganhos correspondem a 1000 pontos. Esse número de toques ganhos e sua pontuação são calculados de acordo com o número de atletas participantes. Cada vitória acima ou abaixo dessa pontuação é calculada de acordo com o numero de toques da competição.

2.2- Natação

Para a natação os atletas devem nadar um total de 200m livre. O importante nesta modalidade é o tempo de prova. Completando a distância em 2 minutos e 30 segundos, o pentatleta soma 1000 pontos. Cada segundo, abaixo ou acima de 2:30 minutos, representa um acréscimo ou perda de 12 pontos.

2.3 - Hipismo

No hipismo, o percurso tem entre 350 e 450m de comprimento, com 12 obstáculos, sendo 1 duplo e 1 triplo. A altura da vara varia entre 1m e 1,20m. Esse percurso desempenhado sem faltas e dentro do tempo previsto soma 1200 pontos. O tempo determinado para o percurso de menor distância é de 1 minuto e para o de maior distância, 1 minuto e 17 segundos. Cada ponto perdido é deduzido do total:

- Cada falta no percurso representa 24 pontos perdidos;
- Tempo excedido;
- Cada obstáculo não saltado representa a perda de 100 pontos;
- Duas quedas levam à interrupção da participação do atleta na prova.

2.4- Combinado (Corrida e Tiro)

Em janeiro de 2009 houve uma alteração nas regras do Tiro e da Corrida que determinou que os atletas devem correr uma distância de 3000 metros. O atleta inicia a prova, também com a saída em handcap, corre aproximadamente 20 m e faz a primeira série de Tiro e, em seguida, inicia a corrida, que é interrompida duas vezes (a cada 1000

metros) para a execução da prova de tiro. A prova de tiro, consiste em acertar 5 alvos, a uma distância de 10 metros, durante um tempo máximo de 1 minuto e 10 segundos.

Neste combinado a pontuação é de 2000 pontos que são obtidos completando o percurso em 14 minutos. Cada segundo mais rápido ou mais lento, representa em média, 4 pontos.

3- Modelamento

A base de dados que será utilizada para as análises que este artigo se propõe a fazer envolve os resultados da Copa do Mundo de Pentatlo Moderno, realizada em Medway (Reino Unido), em Abril de 2010.

A Copa do Mundo de Pentatlo Moderno é realizada anualmente e o evento de 2010 é parte intrínseca de preparação para os jogos oficiais Olímpicos de 2012.

3.1- Modelo DEA

A análise envoltória de dados (DEA – *Data Envelopment Analysis*) foi desenvolvida por Charnes et al [1978] e utiliza programação linear para uma avaliação comparativa da eficiência de unidades, chamadas de Unidades tomadoras de decisão (DMU – *Decision maker unit*).

O propósito principal do modelo DEA é comparar certo número de DMUs que realizam atividades similares, mas que se distinguem pela quantidade de recursos de entradas e saídas utilizada.

Neste trabalho será utilizado o modelo DEA com *inputs* unitários. A interpretação deste modelo, pode ser vista em Caporaleti [1999] e Soares de Mello et al [2009] Esta abordagem é equivalente ao modelo DEA vote-ranking proposto por Doyle et al [1996]. O modelo de *input* unitário é um caso particular do modelo CCR proposto por Charnes et al [1978]. Para o caso particular de uma constante unitária (entrada), o modelo CCR torna-se um modelo no qual h_0 é medida de eficiência da DMU o, y_{jk} é o output j (j = 1, . . . , s) da k^{th} DMU (k = 1, . . . , n), x_{ik} é o input i (i = 1, . . . , r) da k^{th} DMU e μ_j e o v_i são os pesos dos outputs e os inputs respectivamente.

Como o input é uma constante unitária não aparecerá na formulação. O modelo Primal (1) e o Dual (2) são descritos a seguir:

Modelo Primal:

$$\begin{aligned} &\textit{Maximizar } h_o = \sum_{j=1}^{s} \mu_j y_{jo} \\ &\textit{sujeito a} \\ &\sum_{j=1}^{s} \mu_j y_{jk} \le 1, \quad k = 1,\ldots,n \\ &\mu_j \ge 0 \quad j = 1,\ldots,s \end{aligned} \tag{1}$$

Modelo Dual:

$$\begin{aligned} &\textit{Minimizar} \sum_{k=1}^{n} \lambda_k \\ &\textit{sujeito a} \\ &\sum_{k=1}^{n} y_{jk} \lambda_k \le y_{jo}, \quad j = 1,\ldots,s \\ &\lambda_k \ge 0, \quad \forall k. \end{aligned} \tag{2}$$

É importante ressaltar que um modelo de input unitário baseado em na formulação com retornos variáveis de escala não teria sentido. Com efeito, a restrição adicional dessa formulação igualaria a função objetivo do modelo dual à unidade e todas as DMUs teriam a mesma avaliação.

Outra abordagem desse trabalho será a consideração de um modelo de avaliação cruzada proposta por Doyle e Green [1994] como um modo de aumentar a discriminação entre as DMUs consideradas. A idéia principal é avaliar todas as DMUs como um grupo (avaliação pareada). Enquanto no DEA clássico cada DMU é avaliada sob o seu próprio ponto de vista, na avaliação cruzada cada uma das DMUs passa a ser avaliada sob o ponto de vista das demais unidades [Lins e Angulo-Meza (2000)]. Matematicamente, a avaliação cruzada utiliza os pesos atribuídos a cada DMU k, u_{ik} e v_{jk} para avaliar as demais unidades, conforme mostrado em (3), onde Eks é a eficiência da DMUs avaliada segundo os pesos da DMU k. O modelo a ser utilizado é descrito a seguir em (4):

$$E_{ks} = \frac{\sum_i u_{ik} y_{is}}{\sum_j v_{jk} x_{js}} \tag{3}$$

$$\begin{aligned} &\textit{Minimizar} \quad \sum_i \left(u_{ik} \sum_{s \ne k} y_{is} \right) - \left(\sum_j v_{kj} \sum_{s \ne k} x_{js} \right) \\ &\textit{sujeito a} \\ &\sum_j v_{jk} x_{jk} = 1 \\ &\sum_i u_{ik} y_{ik} - E_{kk} \sum_j v_{jk} x_{jk} = 0 \\ &\sum_i u_{ik} y_{is} - \sum_j v_{jk} x_{js} \le 0, \quad \forall s \ne k \\ &u_{ik}, y_{is} \ge 0 \end{aligned} \tag{4}$$

Também será utilizada a fronteira invertida de DEA [Entani et al.,2002] que permite uma melhor discriminação das DMUs que atribuem peso zero a muitas variáveis. Este modelo pode ser considerado como um modo otimista de avaliar a eficiência dos países. Neste modelo os inputs e outputs são invertidos. Assim uma alta eficiência neste modelo representa um baixo desempenho da DMU. Como a fronteira clássica de DEA

representa uma avaliação otimista e a fronteira invertida representa uma avaliação pessimista, um índice composto [Angulo-Meza et al., 2005] pode ser descrito em (5) como:

$$Índice\ \ de\ eficiência\ \ composta = \frac{eficiência\ clássica - eficiência\ invertida + 1}{2} \qquad (5)$$

3.2- Modelamento

A partir dos modelos discutidos na seção anterior a modelagem do problema considerou inputs unitários e outputs gerados a partir da análise de duas classes de dados: (1) o melhor resultado do país em cada esporte (não necessariamente os mesmos atletas) e (2) os resultados do atleta com a melhor posição no ranking para cada um dos países.

Assim, resumidamente, as variáveis utilizadas foram:

- 27 DMUs (todos os países que participaram dos jogos);
- Inputs unitários;
- 4 Outputs (as quatro modalidades de competição)

E, para cada país (DMU) foi definido:

- k DMUs (k=1…27)
- xik inputs (i=1; k=1,...,n)
- yjk output (j=1…4; k=1,...,n)

Tabela 1 - Caso 1: Dados considerando o melhor resultado de cada um dos países

País	Input	Esgrima	Natação	Hipismo	Combinado
ARG	1	748	1232	0	1848
AUT	1	904	1256	1160	2272
BLR	1	904	1320	1160	2244
BRA	1	568	1224	0	1992
BUL	1	748	1324	0	1784
CAN	1	832	1280	1140	2352
CHN	1	784	1376	1172	2168
CZE	1	952	1344	1120	2352
EGY	1	676	1248	0	1876
ESP	1	880	1248	1100	1988
FRA	1	928	1344	904	2232
GBR	1	928	1448	1060	2388
GER	1	856	1276	0	2300
GRE	1	784	1396	400	0
HUN	1	1000	1384	1200	2400
IRL	1	676	1332	0	1936
ITA	1	904	1320	964	2424
JPN	1	880	1400	1040	2236
LAT	1	904	1304	1140	2208
LTU	1	856	1280	724	2228
NED	1	784	1304	1160	2276
POL	1	784	1436	1140	2340
RUS	1	952	1364	1180	2272
SWE	1	964	1252	0	2092
TPE	1	712	1248	0	2044
UKR	1	880	1300	1180	2284
USA	1	856	1336	1144	2256

Tabela 2 – Caso 2: Dados considerando o atleta melhor posicionado no ranking

País	Input	Esgrima	Natação	Hipismo	Combinado
ARG	1	748	1232	0	1848
AUT	1	904	1256	1160	2272
BLR	1	904	1280	1160	2244
BRA	1	496	1204	0	1992
BUL	1	748	1324	0	1784
CAN	1	832	1280	1140	2352
CHN	1	712	1320	1172	2108
CZE	1	952	1344	1120	2344
EGY	1	676	1248	0	1876
ESP	1	880	1248	1100	1988
FRA	1	808	1344	904	2232
GBR	1	712	1424	1060	2356
GER	1	856	1196	0	2300
GRE	1	784	1396	400	0
HUN	1	1000	1372	1200	2188
IRL	1	640	1332	0	1936
ITA	1	904	1240	964	2424
JPN	1	688	1400	1040	2236
LAT	1	904	1304	1140	2208
LTU	1	856	1280	724	2228
NED	1	784	1304	1160	2276
POL	1	784	1436	1140	2300
RUS	1	952	1344	1180	2236
SWE	1	964	1252	0	2092
TPE	1	712	1248	0	2044
UKR	1	856	1276	1180	2284
USA	1	760	1312	1100	2256

4- Resultados

Utilizando o software SIAD (Sistema Integrado de Apoio a Decisão) [Angulo-Meza (2002)] para fazer esta análise comparou-se o os resultados encontrados para cada um dos países em termos da eficiência clássica e da eficiência invertida utilizando a eficiência composta e a eficiência composta normalizada (sinalizada por composta*) parcialmente apresentada nas **Error! Reference source not found.** e **Error! Reference source not found.**, para os casos 1 e 2 respectivamente. A eficiência composta foi calculada a partir da expressão (5). Além disso, considerou-se também a eficiência composta normalizada, denotada por composta* que fornece a mesma ordenação da composta, mas os valores são divididos pelo maior, para ficarem entre 0 e 1.

Os resultados da avaliação cruzada, calculados a partir de (3) e (4), são exemplificados nas Tabela 4 e Tabela 6.

Tabela 3 – Excerto dos resultados para o Caso 1

DMU	Padrão	Invertida	Composta	Composta*
ARG	0,850829	1	0,425414	0,736842
AUT	0,966667	0,974522	0,496072	0,859226
BLR	0,966667	0,927273	0,519697	0,900145
BRA	0,845304	1	0,422652	0,732057
BUL	0,914365	1	0,457182	0,791866
CAN	0,978548	0,95625	0,511149	0,88534
CHN	0,986689	0,891892	0,547398	0,948125
CZE	0,978226	0,910714	0,533756	0,924496
EGY	0,861878	1	0,430939	0,746411

Tabela 4 - Excerto dos resultados da Avaliação Cruzada para o Caso 1

	ARG	AUT	BLR	BRA
ARG	0,85	0,87	0,91	0,85
AUT	0,00	0,97	0,97	0,00
BLR	0,00	0,97	0,97	0,00
BRA	0,85	0,87	0,91	0,85
⋮				
Média	0,55	0,92	0,94	0,54
Média*	0,53	0,92	0,94	0,53

Para o caso 1, no qual se considerou os melhores resultados do país em cada uma das modalidades (e não necessariamente os mesmos atletas), observou-se que alguns países, tais como Bulgária e Grécia que eram quase eficientes na fronteira clássica tornaram-se ineficientes quando a fronteira invertida foi utilizada.

Estes resultados indicam que estes países só eram considerados eficientes por causa de seus resultados em uma ou duas disciplinas do Pentatlo.

A Grã-Bretanha é sempre eficiente, não importando qual a fronteira analisada. Isso ocorre porque os seus resultados, para o caso 1, estão bem perto dos melhores resultados de cada uma das modalidades. Situação parecida ocorre com a Polônia.

Os resultados permitem demonstrar que alguns países atribuíram peso zero a algumas modalidades como meio de obter a máxima eficiência o que é contra o espírito competitivo do Pentatlo, justificando o uso da avaliação cruzada e da fronteira invertida.

Outra razão para estes resultados é que devido a construção matemática do modelo DEA, os países com o melhor resultado em uma modalidade serão sempre considerados eficientes. Entretanto, como apresentado aqui, eles não são, de fato, eficientes.

A análise dos resultados da avaliação cruzada permitiu interessantes observações. A Hungria é sempre eficiente, mesmo considerando outras políticas de atribuição de pesos para as modalidades. Isso indica que este país é um importante *benchmarking* (não no sentido apropriado de DEA, mas no conceito real do termo) na competição.

Para a Alemanha, Bulgária, Irlanda e outros, altos Mavericks indicam que são países quase eficientes apenas porque os pesos usados na avaliação individual são construídos de forma muito peculiar. A Suécia possui o pior resultado dentre os demais indicando que ela não considerou aspectos avaliados por outros países como relevantes para se obter melhores resultados.

Os resultados encontrados para o Caso 2, no qual se considerou as pontuações dos melhores atletas de cada um dos países podem ser observados, de forma resumida nas **Error! Reference source not found.** e Tabela 6.

Tabela 5 – Excerto dos resultados para o Caso 2

DMU	Padrão	Invertida	Composta*
HUN	1	0,882	0,966
CZE	1	0,894	0,955
RUS	0,999	0,896	0,953
POL	1	0,842	1
CAN	1	0,936	0,919
UKR	1	0,94	0,915
AUT	0,995	0,955	0,898
BLR	0,989	0,938	0,907
LAT	0,975	0,922	0,909

Tabela 6 - Excerto dos resultados da Avaliação Cruzada para o Caso 2

	HUN	CZE	RUS	POL
HUN	1,00	0,93	0,98	0,95
CZE	1,00	1,00	1,00	0,98
RUS	1,00	1,00	1,00	1,00
POL	1,00	0,99	1,00	1,00
⋮				
Media	0,98	0,98	0,98	0,98
Media*	0,98	0,98	0,98	0,98

Para este caso foi possível avaliar a eficiência de alguns atletas. Os resultados mostraram que alguns deles empenham maior esforço para obter um melhor resultado em uma modalidade específica.

Como o Pentatlo não atribui pesos aos seus eventos, uma estratégia mais interessante seria obter resultados médios em todas as disciplinas do que ser melhor em apenas uma. Alguns países, tais como a Itália, que é eficiente na fronteira padrão (ou clássica) tornam-se ineficiente quando a fronteira invertida é usada. Este resultado demonstra que esses países são considerados eficientes apenas devido a um melhor resultado em alguma modalidade. No caso da Itália, o país possui o melhor resultado na disciplina Combinado. Isso também pode ser explicado pela formulação matemática de DEA que beneficia DMUs com melhores resultados em alguns de seus *outputs*.

Entretanto, como exceção a este caso, a Polônia é eficiente em ambas as análises mesmo tendo o melhor resultado em uma das modalidades Isto significa que seus resultados nas outras modalidades não são tão ruins quando comparados aos demais atletas.

Na avaliação cruzada dois resultados interessantes merecem comentários. A Suécia apresentou os mesmos resultados da análise do Caso 1. Isto ocorreu porque o país levou apenas um atleta à competição. A pontuação do país é a pior de todas indicando que sua estratégia para os jogos não é a melhor na obtenção dos melhores resultados.

O Brasil, a Bulgária, o Egito e a Irlanda, assim como outros, obtiveram altos Mavericks, o que indica que eles são eficientes apenas porque a política de pesos atribuída às modalidades foi favorável a eles.

5- Considerações Finais

Os resultados apontados por este trabalho indicam que alguns países devem reconsiderar suas políticas de treinamento e suas estratégias para os atletas. Países que estabeleceram uma estrutura de pesos e atribuição de importância média a todos os eventos parecem obter melhores resultados na competição.

Este trabalho também pode ser útil aos atletas e seus treinadores na definição de quais modalidades investir para que suas pontuações finais sejam mais competitivas. Em trabalhos futuros este ponto será aprofundado.

Como proposta para outros estudos futuros, uma interessante contribuição residiria no uso do número de atletas de cada um dos países como um input na modelagem além de estudar a consideração do uso da restrição aos pesos para comparar o comportamento de cada DMU em uma base comum de sistema de pesos. Também se intenciona utilizar os atletas como unidades de análise no modelo proposto.

Agradecimentos

À FAPERJ pelo apoio financeiro.

6- Referências

Angulo Meza, L.; Biondi Neto, L.; Soares de Mello, J.C.C.B.; Gomes, E. G. ISYDS– Integrated System for Decision Support (SIAD – Sistema Integrado de Apoio a Decisão): a software package for data envelopment analysis model. Pesquisa Operacional, v.25, n.3, p 493-503. 2005

Angulo Meza, L.; Lins, M.P.E. Review of Methods for Increasing Discrimination in Data Envelopment Analysis. Annals of Operations Research, v. 116, p. 225-242, 2002.

Caporaletti, L. E., Dul´a , J. H. & Womer, N. K. (1999) Performance evaluation based on multiple attributes with nonparametric frontiers. Omega, 27, 637–645.

Charnes, A., Cooper, W. W. & Rhodes, E. (1978) Measuring the efficiency of decision-making units. Eur. J. Oper. Res., 2, 429–444.

Clearlead. Modern Pentathlon - Pentathlon Event - Olympic Pentathlon. Disponível em http://www.clearleadinc.com/site/pentathlon.html. Acesso em 20 de outubro de 2009.

Doyle, J. R., Cook, W. D. & Green, R. H. (1996) Preference voting and project ranking using DEA and cross evaluation. European. Journal of Operational Research., 90, 461–472.

Doyle J, Green R (1994) Efficiency and cross-effciency in DEA – Derivations, meanings and uses. Journal of the Operational Research Society, 45(5), 567-578.

Entani, T., Maeda, Y., Tanaka, H. (2002). Dual Models of Interval DEA and its extensions to interval data. European Journal of Operational Research, 136, 32-45

Garniner, N.E. The System of the Pentathlon. The Journal of Hellenic Studies, Vol. 45, Part 1 , pp. 132-134. (1925).

Harris, H.A. The Method of Deciding Victory in the Pentathlon. Greece & Rome, Second Series, Vol. 19, No. 1, pp. 60-64 . (1972).

Hillman, DJ. How to Win Pentathlon – Unifying Theories of Information Science. American Documentation. Vol. 20. Ed. 4. Pg: 335-336. (1969)

Kile, DG. Winning and Watching the Greek Pentathlon. Journal of Sport History. Vol. 17. Ed. 3. Pg 291-305. (1990)

Kile, DG. Games, prizes and athletes in Greek sport: Patterns and perspectives (1975-1997). Classical Bulletin. Vol. 74. Ed. 2. Pg. 103-127. (1998).

Ladany, SP. Optimization of Pentathlon Training Plans. Management Science Series B – Application. Vol. 21. Ed. 10. Pg: 1144-1155. (1975)

Langdon, MK. Scoring the Ancient Pentathlon: Final Solution?. Zeitschrift für Papyrologie und Epigraphik, Bd. 78, pp. 117-118. (1989).

Lee, HM. Wrestling in the Repechage of the Ancient Pentathlon. Journal of Sport History. Vol. 20. Ed. 3. Pg: 277-283. (1993)

Nunes, M.; Bergiante, N.C.R.; Soares de Mello, J.C.C.B. Algumas Considerações sobre a Mudança de Regras do Pentatlo Moderno. Relatórios de Pesquisa em Engenharia de Produção V. 10 N. 12, Universidade Federal Fluminense (2010)

Pentathlon GB. The History of Modern Pentathlon. Disponível em: http://www.pentathlongb.org/aboutpentathlon/aboutmp_history.php. Acesso em 20 de outubro de 2009.

Percy, D.F. A mathematical Analysis of Badminton Scoring Systems. Journal of the Operational Research Society (2009) 60, 63--71

Sexton, T. R., Silkman, R. H. & Logan, A. J. (1986) Data envelopment analysis: critique and extensions. Measuring Efficiency: An Assessment of Data Envelopment Analysis (H. Silkman ed.). San Francisco, CA: Jossey-Bass, pp. 73–105.

Soares de Mello, J.C.C.B.; Angulo-Meza, L.; Branco da Silva, B.P. A ranking for the Olympic Games with unitary input DEA models. IMA Journal of Management Mathematics, v. 20, n 2, p. 201-211, 2009.

Soares de Mello, J.C.C.B., Gomes, E.G., Angulo Meza, L., Biondi Neto, L. Cross evaluation using weight restriction in unitary input DEA models: Theoretical aspects and application to Olympic Games ranking. WSEAS Transactions on Systems, v7, n. 1, p. 31-39, 2008.

Supply Chain Optimization: Application to a real case

Nuno Miguel Santana[a], *Maria Isabel Gomes*[b], *Ana Paula Barbosa-Póvoa*[c]

[a]*IST, Lisboa, Portugal*
nuno.santana@ist.utl.pt

[b]*CMA, FCT, Universidade Nova de Lisboa, Monte da Caparica, Portugal*
mirg@fct.unl.pt

[c]*Centro de Estudos de Gestão do IST,UTL, Lisboa, Portugal*
apovoa@ist.utl.pt

Abstract

The increase focus of companies in clients' satisfaction, independently of the market they operate in, results from a growing degree of consumers expectations. To fulfill such increase, companies have to arm themselves with the correct internal capabilities, achieving not only efficacy in the satisfaction of each customer, but also efficiency, by operating at the lower possible cost. One of those internal capabilities is the logistics network, which is representing, more and more, a great challenge for companies, given the complexity of the variables that compose it. In this paper, an optimization model is developed that contemplates simultaneously the design and planning of supply chains with reversed flows, through a multi-period and a multi-product network. The proposed model acts as a support tool for the decisions to be made in a real case of a Portuguese telecommunications company, in order to optimize its logistics network.

Keywords: Forward supply chain, closed loop supply chain, logistics providers, supply chain modeling, supply chain integration.

1- Introduction

Whatever the business sector, customer satisfaction has become the central focus of any company that wants to win in an increasingly competitive market. The concept of customer satisfaction no longer refers only to the performance of goods and repairs a company offers, it also considers its interconnection with the internal capabilities that enable the company to respond adequately to the needs of increasingly demanding customers.

Supply chains management, and in particular the associated logistics networks, can be one of the most relevant internal capabilities to meet customers' needs. Companies must

deliver products to the right places and quantities, at the appropriate time and at the lowest possible cost. However companies have customers spread across several geographical regions, national or worldwide, with demands for different types of products. This makes the management of the forward flow of products a challenge when efficiency and effectiveness are at stake. Adding to this the different life cycle times products have and a good quality after sales service (product breakdowns, repairs or replacements, i.e., the reverse flow of products), the challenge becomes even greater.

Such problem is addressed along the present paper where the case of a national telecommunications company is studied. In order to deal with this complexity, the company in study outsourced most of its logistics operations namely transportation, warehousing and repair. Different third-party logistic (3PL) providers manage forward and reverse flows. Due to this decision, the company currently faces the challenge (not less complex) of coordinating and managing all the involved 3PLs. If, outsourcing on one hand, allows the company to focus on their core business, leaving such operations for those who are supposed to run it in an effective and efficient manner, on the other hand, the company may lose control over such activities if not properly supervised, which can create a negative image in end customers.

Within this context, the objective of this study is to investigate a possible integration of product flows. Particularly in terms of transportation and sharing of facilities to support the logistics processes in order to minimize total costs, without diminishing the after sales quality service provided to the consumer. To achieve this goal a Mixed Integer Linear Program (MILP) model is developed and different scenarios based on the logistics network main decisions are studied. This paper is organized as follows. In Section 2 an overview of the existing literature related with this work is presented. In Section 3 the case study is characterized including forward and reverse supply chains. The problem's description is given in Section 4. The characterization of the model formulation follows in Section 5, where some model aspects are explained. In Section 6 are presented the main results obtained for the case study under study. Finally, some conclusions and future research directions are drawn.

2- Literature Review

Each industry has very particular characteristics respecting its supply chains, differing even between companies in the same field. Despite these differences, and independently from the type of industry in which companies operate, supply chains models are quite similar in the approach they follow. Besides that, there is a need to have in mind that no model can capture all the aspects of the processes of a supply chain, reflecting only the essential key aspects to its study, while being capable enough of being mathematically solved (Min and Zhou, 2002).

The key components for the construction of supply chains models as suggested by Min and Zhou (2002) are essentially three: 1) drivers, which consist in the establishment of the aims that arise from the identification of the main drivers of the supply chain; 2) constraints, that represent the real restrictions of the logistics network, like production capacity constraints, etc.; 3) decision variables, that represent the answers that we want to extracted from the construction of the supply chain model, being these answers that, in a last analysis, will contribute to the final decision. Everything mentioned above applies to all supply chain models, whether they are strategic, tactical, operational nature, or even of a mix nature, like strategic and tactical models. The strategic models generally involve temporal horizons of four to five years, depending on the industry in question. According to Goetschalckx and Vidal (1997), the strategic level contemplates long-term decisions, for example, the number, location, capacity and type of facilities to use. Most of the case studies in supply chain management published in the literature are strategic in nature, mostly because of its modeling easiness. A good review on the subject is provided by Melo et al. (2009). At the tactical level, a smaller temporal horizon is considered when compared to the strategic one, generally between a quarter and a year. Tactical decisions are more detailed than the strategic ones. Being associated with the supply chain's planning in itself, they establish performance levels related to inventory, transport and sales policies. According to Swaminathon et al. (2003), these levels allow for an orientation so that daily tasks might be executed. However, few are the purely tactical models found in the literature, being mostly strategic or operational (daily decisions). According to Dogan et al. (2002), to achieve a global and integrated system, the strategic decisions must be considered simultaneously with the tactical

aspects, since the integration of decisions with different time units might imply high savings for the company.

It is based on this premise that the present work is developed. Within the context of closed-loop supply chains, a case study is presented together with a model that simultaneously integrated strategic and tactical decisions. The supply chain structure is defined, in terms of logistics centers to install. Furthermore, different tactical decisions on how much capacity is required and on how operate the structure chosen are also evaluated.

3- Case study

3.1 General description

The company in study has, in the recent past, restructured its logistics network by outsourcing most of its logistical operations, with recourse to third-party logistics (3PL) providers. Currently, this company operates with two logistics operators and four different distributors. The two logistics operators are individually responsible for the forward and reverse logistics. Distributors can operate with flows of both supply chains. Overall the company's network consists of one warehouse and two logistics centers, with different characteristics in terms of dimensions, locations and performed functions. The company logistics department is in charge of the management of all logistics processes involved in both forward and reverse flows. This management is done independently for the two flows, i.e., the forward and reverse flows are not integrated. Nonetheless some reverse flow products are reused through the forward network. In order to evaluate the quality of the collected products, the company has a Repair Center where equipments are tested and its housing is exchanged for a new one.

3.2 Forward supply chain description

The forward supply chain has four different levels: suppliers, Logistics Center/warehouse, Distribution Channels/Service Providers (SPs) and clients (see Figure 1). The first level represents all suppliers, which provide the various components and products. Transportation between the suppliers and the next level in the chain is their responsibility. All components and products are transported to the logistics center that, along with a warehouse (which operates in cross-docking) composes the second

level of the forward supply chain. The logistics center also operates as a warehouse and is responsible for all the forward logistics operations (includes reception, dispatch, inventory, kitting and serial number reading). The third level of the forward supply chain is composed by the Service Providers (SP), which main function is to supply the final customer (last level of the forward supply chain). The SPs are split in two major groups according to their geographical location: north and south. The southern SP are supplied by the Logistic Center, while the SPs of the northern area are supplied from the warehouse which in turn, is supplied by the Logistics Center (transhipment). As mentioned above, customer supplying is an outsourced operation (performed by the SPs). These distribution channels work on consignment, i.e., products flowing in these channels belong to the company. This forward supply chain can be characterized as a push-pull system. The service providers place their orders and are responsible for the transportation. After SP orders have been placed, the production planning as well as the placement of orders to suppliers is based on a make-to-stock strategy. Production planning, as well as orders to suppliers, is based on monthly and weekly forecasts (respectively), with a security inventory level between 7-8 weeks of expected demand.

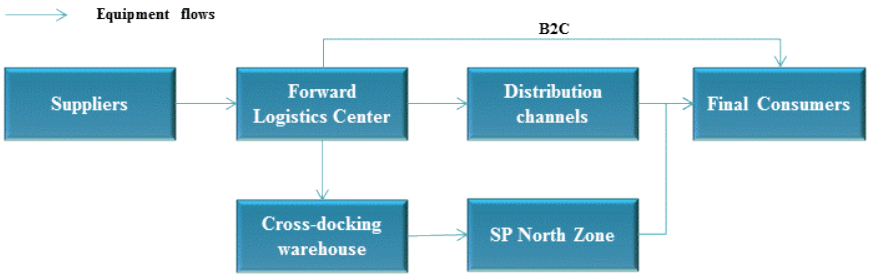


Figure 1- Simplified forward supply chain scheme

3.3 Reverse supply chain description

The complexity of processes and flows of the reverse supply chain is much higher when compared to forward supply chain. This is explained by the multiple origins and destinations of the returning equipment. However, the underlying principle of the two chains is similar. That is, the reverse chain operations are also centralized in a single logistics center (different than the one of the forward flow). To this center returns all sorts of products included in the company's portfolio. Main reasons for this reverse flow are faulty equipment, the service cessation and upgrade equipment. The intrinsic uncertainty of the quantity and quality of equipment returned is difficult, if not

impossible, to predict. This uncertainty is mainly reflected in terms of transport planning and operations at the logistics center. In terms of structure, the reverse logistics network can be also decomposed into four levels (see Figure 2): distribution channel, reverse logistics center, support entities and consumers. The first level consists in some distribution channels, which can stem the equipment returned. The second level is composed by the Logistics Center. While forward products are transported directly to the respective distribution channels, in the reverse this is not the case. Depending on the equipment processed at the Logistics Center, they may be referred to the repairers, for the slaughter (which includes resale and scrap dealers), or circuits for maintenance and / or post-sale (third level). The latter feeds the last level which consists of consumers. The Repair Centre integrates the reverse supply chain, being the operations split between two entities: logistics operator performs equipments test operations and an external repairer performs the equipment housing exchange operation.

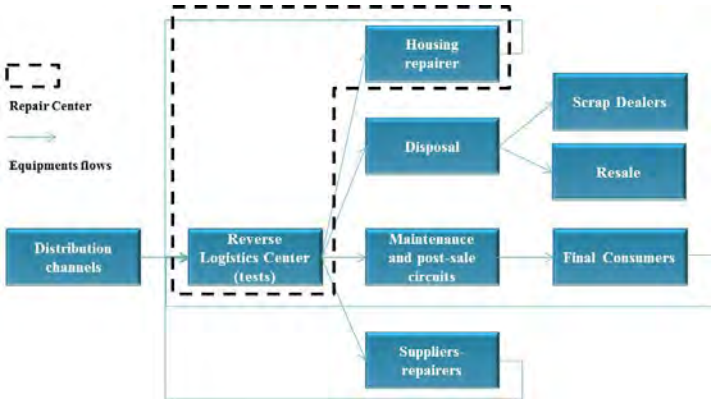


Figure 2 - Simplified reverse supply chain scheme

4 - Problem's Description

The company wants to restructure its logistics network where the location and operation of the logistics center, Repair Center and cross-docking warehouse are to be evaluated. Two alternatives to the Logistics Center are studied: the actual one (see Section 3) and a future one where the Logistics Center common to both flows with outsourcing. The operation of the Repair Centre and the cross-docking warehouse, and consequently the monthly payment to SPs, is studied and results compared to the direct distribution to this channel by the company. So, the objective of this work is to evaluate different logistics network structure and planning schemes having the best structure as a reference base. The studied problem can be defined as:

Given:

- The super-structure composed by the potential entities in the supply chain;
- The demand volumes of the commercial circuit, the maintenance circuit and the distribution channel;
- Bill of materials;
- The relation between distributed products and the returned ones, given by the return fraction;
- The initial stock levels at the facilities;
- The relation between returned products considered as new and the ones considered obsolete, given by an historical fraction;
- The relation between tested products and possible results of those tests, given again by historical fractions;
- Transportation cost between each pair of entities;
- Cost of the products bought from the suppliers;
- Cost structure of the logistics operators;
- Logistics Center's investment cost;
- Repair Center's investment cost;
- Operation costs of the Repair Centers;
- Labor costs;
- Maximum warehousing capacity of the logistics operators and the Repair Center;
- Maximum and minimum limits of the flows between facilities;
- Maximum transportation capacity of the transportation companies vehicles;
- Physical characteristics of the products such as weight and volume;
- The time for kitting, rework, testing and housing by product.

To determine:

- The number and type Logistics Centers;
- The type of Repair Center;
- Distribution strategy for the Repair Provider (SP) distribution channel;
- Products stock levels, in each micro time unit;
- Levels of kits production, housing and tests in each micro time unit;
- Number of necessary trips between logistics operators and SP or vice-versa, in each micro time unit;

- Number of people necessary to execute the tasks of testing equipment, kitting, housing and rework without the occurrence of backlog of the Logistics Center and Repair Center;
- The stock amounts, in each micro time unit;
- Amount of products transported between each pair of facilities and in which micro period they happen, including amount of equipment bought from the suppliers.

So as to minimize the total cost of the supply chain.

5 – Model Formulation

The mathematical model used to solve the problem under study was based on the model presented by Salema et al. (2010). In this work, authors propose a MILP model that integrates the strategic decisions with the tactical ones through the integration of two interrelated temporal scales. The network representation is achieved through a graph approach, where nodes represent entities (e.g. suppliers, logistics centers, etc.) and the arcs represent the flow of material moving between entities. For further details the reader is referred to the work of Salema et al. (2010).

The model involves the definition of the network super-structure. Starting from the defined structure, a set linear equations are able to represent the constraints of the problem to be satisfied. Considering the indices *i, j, k* for the entities, and *m, n* for the products the following variables (continuous, binary and integer) were defined:

X_{mijt} quantity of product m dislocated from entity i to entity j, during micro period t';

S_{mit} quantity of product m in stock in entity i, during micro period t'.

$Y_i = 1$ if entity $i \in I_{inst}$ is open, otherwise $Y_i = 0$. About the remaining entities, it is imposed that they are open ($Y_i = 1$), since the suppliers, the clients and the junkyard are always included in the model;

$E_{ijt'} = 1$ if the flow between entity i and entity j occurs during micro period t', otherwise $E_{ijt'} = 0$. This variable is an auxiliary variable that allows the modeling of the flows existing in the model.

$Z_{ijt'}$ number of trips from entity i to entity j, during micro period t';

$P_{mit'}$ number of pallets of product m in entity i, during micro period t';

PES_{it} number of workers in entity i, during micro period t'.

Using the above variables and considering the characteristics of the problem the model was developed. An objective function that minimizes the total logistics costs is considered. This includes investment and operational costs. In terms of restrictions, the model includes a diverse set of constraints, namely: (i) material balance in each entity, thus ensuring that everything that comes out is equal to anything that enter, (ii) the demand of each distribution channel in each time unit is fully satisfied in the same period, (iii) the amount of returned products (which depend on the amount demanded) must be collected in the shortest possible time, (iv) storage capacity in the entities, (v) safety stock levels, (vi) consider the minimum and maximum capacity constraints flows between entities, (vi) ensure that a flow exists if the respective entities also exist, (vii) the number of each type of entities opened, (viii) number of transportation trips between entities, (ix) the number of human resources to carry out operations on entities.

The above model was built in a generic form and allows the solution of the case-study previously described. The obtained results are presented next.

6 – Results

In this section are presented the results obtained from the execution of the previously described model to the study case described in section 3. In this analysis only aspects related to the logistics network and respective costs are considered although planning results are also obtained. The results for the actual logistics network (Current case - section 6.1) and for the optimal network which is being implemented by the company (Future Case - section 6.2) are presented and a brief comparison between both is made.

The MILP model was solved by GAMS/CPLEX (version 23.1.2), an Intel ® Core ™ 2 Duo CPU, 4, 2.66 GHz.

6.1 - Current case (Case 1)

The current logistics network is composed by two logistics centers dedicated to each flow (forward and reverse) with outsourcing, the Repair Center in charge of the reverse chain logistics operator and is of the responsibility of a third entity. In this case, the payment of settlements to the SP distribution channel is used. After applying the model

to this case, the costs obtained are shown in Table 1. These are the base of comparison for the analysis developed along this article. The cost of the current logistics structure is of 107.674.641 monetary units (m.u.).

Table 1 – Costs of Current Case

Rubrics	Cost (m.u.)	Rubrics	Cost (m.u.)	Rubrics	Cost (m.u.)
Suppliers	103 315 800	Repair Center	256 787	Rework	1 031 086
Transport	1 208 488	Logistics Operator	1 868 476	Junkyard	(5 723)

Total: 107 674 641m.u.

When analyzing the costs it can be seen that the costs related to new products acquisition represent 97% of the total costs due to the high prices associated. Follows the logistics operators costs related to the operations performed in the logistics centers representing almost 1,9M m.u.. Transports costs 1,2M m.u./year while Rework operation, which represents the operation of kitting used and functional equipments, costs 1M m.u.. Finally, the costs related to equipment tests and housing exchange (Repair Center) are 256 787 m.u. while the sale of damaged products to scrap dealers (Junkyard) contributes to a cost reduction of 5 723 m.u..

6.2 - Future Case (Case 2)

The optimal case is composed by a common Logistics Center that manages all the outsourced forward and reverse chain operations. The logistics operator responsible for the Logistics Center is also responsible for the Repair Center. After running the model the costs associated with such option are shown in Table 2. The total cost of this scenario is 106 845 214 m.u..

Table 2 - Costs associated with the Future Case

Rubrics	Cost (m.u.)	Rubrics	Cost (m.u.)	Rubrics	Cost (m.u.)
Suppliers	103 315 800	Repair Center	179 061	Rework	1 031 086
Transport	1 125 957	Logistics Operator	1 199 982	Junkyard	(5 723)

Total: 106 845 214 m.u.

6.3 - Comparison between the actual and the future case

The comparison of the results of the two previous cases aims at analyzing the differences between the logistics network in operation with the one that is under analysis by the company, allowing the conclusion of what network configuration is better suited from a financial perspective. Comparing both scenarios Figure 3 shows the

cost differences between the future case and the current one. The future network presents better results since it allows for larger annual savings than the current network, a total of 826.908 m.u.. These savings appear at the level of transportation, of Logistics Center operations and Repair Center operations (see Figure 3). When analyzing the transportation costs, three factors contribute to the savings of 82.530 m.u./year of the future case. Those factors are the consolidation of orders, the elimination of flows between different logistics centers and the fact that the operations of housing of equipment are now located in the Logistics Center own facilities, opposing to what happens in the actual case in which the Repair Center is split between two entities in different locations. About the Logistics Center operations, the results show that the billing imposed by the logistics operator of Case 2 is more advantageous for the company, since it allows the nonoccurrence of fixed costs related to information systems maintenance (less 245.000 m.u./year) and the reduction of reception and expedition costs of 145.000 m.u. and 176.000 m.u./year respectively. Additionally, the 40.000 m.u. saved in the kitting operation indicates that a method of cost based on work hours is more advantageous than the one based on processed unit.

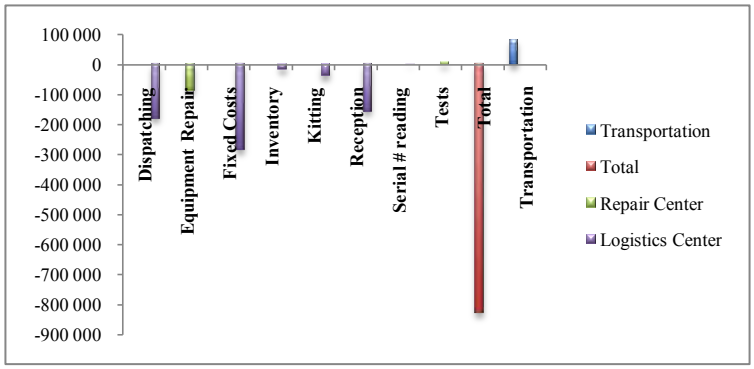


Figure 3 - Cost difference (€) between Case 2 and 1

The new network structure has no impact on acquisition costs (suppliers), which means that no renegotiation has to be performed with the current suppliers.

Finally, Figure 4 shows the operational costs within the Repair Centre for both scenarios. Concerning the testing operation, the cost is higher in the future case which reveals that, for

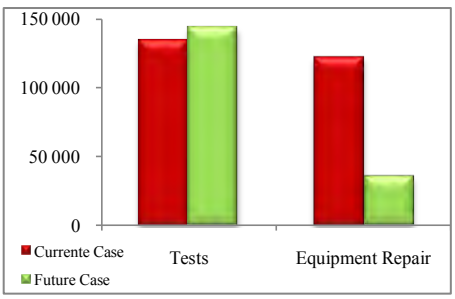


Figure 4: Repair Center operations costs (€)

this operation, it is not advantageous a method of cost based on work hours but rather on processed units, as it happens in the actual case. However, for the housing operation the costs are inferior for the future case 86.000 m.u./year, being of advantage the cost based on the hour worked. Comparing to the actual case, the total savings of the future case are of 826.908 m.u./year.

7– Conclusions

This work studies the restructuring of a Portuguese company's logistics network. To that end, an optimization model was developed, based on the work of Salema et al. (2010).

The developed model is applied to design and access the cost structure of two different logistics networks: the current operating network and a new one that the management board intent to put in place. In a yearlong analysis, the structure that presents the least total costs is the new one, where the Logistics Center and the repair centre become part of a single logistics operator. The annual savings with this new network structure are estimated to be more than €1 m.u..

As future work, some limitations of this work are to be overcome. The first one is related the temporal aggregation. The present study was conducted assuming a week basis (mostly due to computational problems). A smaller time unit such as a day would allows a more detailed planning of kit production and equipment testing (for instance), which would be very useful for the company.

8 – References

Dogan, K., Goetschalckx, M. e Vidal, C. (2002), "Modeling and design of global logistics systems: A review of integrated strategic and tactical models and design algorithms", *European Journal of Operational Research*, 143, 1-18.

Goetschalckx, M. e Vidal, C. (1997), "Strategic production-distribution models: A critical review with emphasis on global supply chain models", *European Journal of Operational Research*, 98, 1-18.

Melo, M.T., Nickel, S. e Saldanha da Gama, F. (2009), "Facility Location and Supply Chain Management – A review", *European Journal of Operational Research*, 196, 401-412.

Min, H. e G. Zhou (2002), "Supply chain modeling: past, present and future", *Computers and Industrial Engineering,* 43, 231-249.

Salema, I., A. P. F. D. Barbosa-Póvoa, e A. Q. Novais (2010), "Simultaneous design and planning of supply chains with reverse flows: a generic modelling framework", *European Journal of Operational Research*, 203, 336-349.

Operação Ótima de Sistemas Hidrotérmicos com Geração de Cenários Hidrológicos Através de Modelos Auto-Regressivos Periódicos identificados via Bootstrap

Reinaldo Castro Souza[a*], *André Luís Marques Marcato*[b], *Bruno Henriques Dias*[b], *Fernando Luiz Cyrino Oliveira*[a]

[a]*Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brasil*
reinaldo@ele.puc-rio.br; fcyrino@ele.puc-rio.br

[b]*Universidade Federal de Juiz de Fora, Juiz de Fora, Brasil*
andre.marcato@ufjf.br; bhdias@ieee.org
[*]*Corresponding Author*

Abstract

This paper deals with the development of models to generate synthetic series for the medium and long term planning of the hydrothermal Brazilian integrated system. The proposed model is based on the so called "periodic autoregressive structure", denoted by PAR(p), where the order p of the model is identified based on the estimated PACF (Partial Autocorrelation Function) whose significance is tested using the bootstrap technique. In particular, systems where the hydro plants dominates the thermal ones, make a thorough use of synthetic series to generate streamflow scenarios which are considered the stochastic input of the operational procedure to produce the minimal costs operation. Such optimization uses a Stochastic Dynamic Programming (SDP) based on convex hull and energy equivalent reservoirs approach, which has produced some promising results. The SDP and convex hull techniques take into account the expected cost-to-go functions which are quite important in the choice of the best long term operation planning strategy carried out by the model.

Keywords: Periodic Model, Bootstrap, Long Term Operation Planning, Convex Hull, Stochastic Dynamic Programming.

1- Introdução

O objetivo do planejamento ótimo da operação de sistemas hidrotérmicos de grande porte é determinar os montantes a serem gerados por cada unidade hídrica e térmica que minimizem o custo esperado de operação do sistema ao longo do horizonte considerado [Pereira, 1989].

O planejamento de médio/longo prazo, considerado de 5 a 20 anos no Setor Elétrico Brasileiro, apresenta como principais características: (i) o acoplamento espacial, devido ao fato de se construir diversas usinas ao longo da mesma bacia [Ni, 1999]; (ii) acoplamento temporal: a decisão de geração em um determinado mês impacta no custo dos meses subseqüentes; (iii) Estocasticidade: que aparece devido à incerteza com relação às afluências futuras e são mais significativas quanto maior o horizonte de estudo.

Desta forma, a incerteza com relação às afluências é de grande importância, tornando o problema mais complexo do ponto de vista estatístico. Modelar esta incerteza não é uma tarefa trivial. Para o ajuste dos modelos estatísticos, tem-se disponível o histórico de afluências, uma série temporal que, na verdade, consiste em apenas uma das possíveis realizações do processo estocástico gerador. O objetivo é, portanto, dado um modelo ajustado à série original, aproximar o comportamento estocástico e, sinteticamente, gerar tantas novas séries temporais quanto se queira, diferentes do histórico original, mas igualmente possíveis do ponto de vista estatístico. Os cenários gerados são utilizados no cálculo da política e simulação da operação.

A técnica de Programação Dinâmica Estocástica (PDE) com o algoritmo de fechos convexos, proposto por Dias *et al.* [2010], é utilizada neste trabalho. Para reduzir o problema do crescimento exponencial do tempo computacional em função do número de reservatórios considerados, característico da PDE, é adotada a agregação de reservatórios de bacias com características hidrológicas semelhantes em reservatórios equivalentes de energia [Arvanitidis, 1970]. Por exemplo, no Sistema Elétrico Brasileiro (SEB), são considerados quatro subsistemas: Sudeste/Centro-Oeste, Sul, Norte e Nordeste, que representam a agregação de usinas destas regiões em um único reservatório equivalente de energia.

O presente trabalho faz uso de um modelo gerador de cenários hidrológicos conhecido como "autorregressivo periódico", PAR(p), cujos parâmetros apresentam um comportamento periódico em função das secas e chuvas, para a geração de séries sintéticas para o problema do planejamento de médio prazo (5 anos) do Sistema Elétrico Brasileiro (SEB). A utilização do modelo PAR(p) apresenta uma fase crítica: a identificação das ordens p. Neste trabalho, uma nova abordagem para esta fase é

considerada, conforme proposto por Oliveira [2010]. A geração de cenários é realizada e o desempenho do modelo é comparado com as séries históricas de afluências aos reservatórios no planejamento da operação do SEB por meio de uma série de testes estatísticos.

2- O Modelo de Geração de Cenários Hidrológicos

De acordo com Hipel & McLeod [1994], algumas séries históricas, dentre estas as hidrológicas sazonais, exibem uma estrutura de autocorrelação que depende não somente do intervalo de tempo entre as observações, mas também do período observado. Ainda, segundo Salas & Obeysekera [1982], os processos estocásticos naturais são, em geral, estacionários em sentido lato, isto é, os momentos de primeira e segunda ordem da distribuição de probabilidades não são afetados por variações devido à escolha da origem dos tempos, um dos pressupostos para a aplicação da metodologia Box & Jenkins. Na classe dos processos periódicos, dois modelos se destacam: PAR (periodic autoregressive) e PARMA (periodic *autoregressive*-moving average). O modelo PAR (*p*) ajusta para cada período da série um modelo AR (*p*). De maneira similar, um PARMA (*p,q*) consiste num modelo ARMA (*p,q*) para cada período em estudo. Em hidrologia, a modelagem PAR (*p*) surgiu a partir das pesquisas de Thomas & Fiering ([1962], de acordo com Hipel & McLeod [1994].

O modelo PAR (*p*) é descrito matematicamente por:

$$\left(\frac{Z_t - \mu_m}{\sigma_m}\right) = \varphi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right) + \varphi_2^m \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}}\right) + \ldots + \varphi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right) + a_t \quad (1)$$

Onde Z_t representa a série sazonal de período *s*; *S* o Número de períodos (*s* = 12 para séries mensais); *T* o índice do tempo, *t* = 1, 2, ..., *sN*, função do ano *T* (*T* = 1, 2, ..., *N*) e do período *m* (*m* = 1,2, ..., *s*); N o número de anos; μ_m a média sazonal do período *m*; σ_m o desvio-padrão sazonal do período *m*; φ_i^m o *i*-ésimo coeficiente autorregressivo do período *m*; p_m a ordem do operador autorregressivo do período *m*; a_t a série de ruídos independentes com média zero e variância $\sigma_a^{2(m)}$.

Isto posto, o modelo PAR (*p*) é adotado na modelagem e simulação de afluências no setor elétrico brasileiro e ajusta, portanto, um modelo autorregressivo de ordem *p* para

cada um dos períodos (meses) das séries históricas de vazões de cada um dos quatro subsistemas que compõem o Sistema Interligado Nacional (Sudeste/Centro-Oeste, Sul, Nordeste e Norte). Para a geração de cenários hidrológicos ajusta-se uma distribuição Lognormal aos resíduos, Penna [2009].

Atualmente, segundo Maceira [1989], a identificação tradicional das ordens p dos modelos PAR (p) é realizada com base na avaliação da significância dos coeficientes da função de autocorrelação parcial (FACP), baseados na aproximação assintótica de Quenouille. Para cada período m procura-se a maior ordem i tal que todas as estimativas dos i coeficientes da FACP para $k > i$ não sejam mais significativas (Critério 1 de identificação).

Em seu trabalho, Stedinger [2001] faz críticas à forma tradicional de identificação, haja vista que, com base no critério considerado, *lags* intermediários não significantes são considerados e os podemos podem se tornar pouco parcimoniosos. O autor propõe que a identificação seja baseada no seguinte critério: para cada período m procura-se a maior ordem i tal que todas as estimativas dos k coeficientes da FACP para $k < i$ sejam significativas (Critério 2 de identificação).

Isto posto, com vistas à construção de intervalos de confiança para cada um dos lags da FACP, estimando a real significância dos coeficientes da mesma, Oliveira [2010] propõe a utilização da técnica de computação intensiva Bootstrap, introduzida por Effron & Tibshirani [1993].

Os resultados obtidos mostram que a identificação via Bootstrap é sensivelmente mais parcimoniosa, levando à identificação de ordens inferiores na maioria dos casos e corroborando alguns pontos levantados em estudos anteriores sobre a abordagem tradicional.

3- Programação Dinâmica Estocástica Utilizando Algoritmo de Fechos Convexos

Técnicas de Programação Dinâmica têm sido utilizadas na resolução de problemas solucionáveis de forma seqüencial, onde uma decisão ótima global pode ser obtida através da otimização de subproblemas (ou ótimos locais) individuais [Bellman, 1957]. No planejamento da operação de médio prazo de sistemas hidrotérmicos o horizonte de

estudos, considerado de 5 anos no Setor Elétrico Brasileiro, é subdividido em estágios mensais.

Desta forma, a decisão ótima atual está ligada a um conjunto de acontecimentos futuros. O problema é resolvido em sentido contrário, ou seja, tomando como início o último estágio de decisão e realizando a recursão no tempo, onde em cada estágio são consideradas as decisões do estágio em análise e as conseqüências futuras [Silva, 2001].

Adicionalmente, o problema de planejamento da operação de sistemas hidrotérmicos depende das afluências às usinas hidrelétricas do sistema sendo, portanto, um problema estocástico. Desta forma pode-se utilizar a Programação Dinâmica Estocástica para a solução do problema.

Os algoritmos da classe "Convex Hull" (CH) apresentam, a partir de um conjunto de pontos fornecidos, o menor polígono convexo que contém os pontos dados [Cormen,2001], conhecido como fecho convexo.

No presente trabalho, utiliza-se o algoritmo "QuickHull" [Barber, 1996], onde dados os pontos correspondentes à solução ótima de cada um dos estados dos problemas de coordenação hidrotérmica, o algoritmo indica o conjunto de hiperplanos que correspondem a um fecho convexo, que são utilizados como uma aproximação linear das funções de custo futuro para cada estágio do problema.

Estes hiperplanos formam uma aproximação linear por partes das funções de custo futuro (FCFs), que são calculadas para cada estágio do horizonte de planejamento. Detalhes da obtenção destas funções pode ser visto em [Dias *et al.*, 2010a]. A partir das FCFs obtidas pode-se realizar o cálculo da política operativa, que consiste na avaliação de um cenário específico ao longo do horizonte de estudo, utilizando as informações das FCFs em cada estágio. Desta maneira obtendo-se o custo operacional ótimo esperado.

4- Resultados

4.1- Modelo de Geração de Cenários Hidrológicos

A seguir os resultados obtidos para as ordens de cada um dos períodos dos subsistemas brasileiros. Para a forma tradicional de identificação, são mostrados os valores de p

encontrados segundo os critérios estabelecidos. O objetivo é, portanto, comparar os resultados obtidos por meio das duas modelagens e verificar a aplicabilidade da técnica proposta.

Nas tabelas seguintes os valores das ordens selecionadas.

Tabela 1: Resultados da estimação das ordens para os subsistemas Sudeste/Centro-Oeste e Sul

	SUDESTE/CENTRO-OESTE			
	Clássica		Bootstrap	
	Critério 1	Critério 2	Critério 1	Critério 2
Jan	5	1	1	1
Fev	6	1	1	1
Mar	1	1	1	1
Abr	2	2	2	2
Maio	3	1	3	1
Jun	1	1	1	1
Jul	3	3	3	3
Ago	1	1	1	1
Set	1	1	1	1
Out	3	3	1	1
Nov	1	1	1	1
Dez	4	1	1	1

	SUL			
	Clássica		Bootstrap	
	Critério 1	Critério 2	Critério 1	Critério 2
Jan	1	1	1	1
Fev	1	1	1	1
Mar	4	1	1	1
Abr	5	1	1	1
Maio	1	1	1	1
Jun	5	1	1	1
Jul	4	2	1	1
Ago	1	1	1	1
Set	1	1	1	1
Out	1	1	1	1
Nov	6	1	1	1
Dez	1	1	1	1

Tabela 2: Resultados da estimação das ordens para os subsistemas Nordeste e Norte

	NORDESTE			
	Clássica		Bootstrap	
	Critério 1	Critério 2	Critério 1	Critério 2
Jan	1	1	1	1
Fev	4	2	2	2
Mar	1	1	1	1
Abr	1	1	1	1
Maio	1	1	3	1
Jun	1	1	5	1
Jul	1	1	1	1
Ago	1	1	1	1
Set	1	1	3	1
Out	3	1	1	1
Nov	2	2	2	2
Dez	5	1	1	1

	NORTE			
	Clássica		Bootstrap	
	Critério 1	Critério 2	Critério 1	Critério 2
Jan	1	1	1	1
Fev	4	2	2	2
Mar	1	1	1	1
Abr	1	1	1	1
Maio	1	1	1	1
Jun	1	1	1	1
Jul	3	3	3	3
Ago	1	1	2	2
Set	5	1	5	2
Out	2	2	1	1
Nov	5	1	1	1
Dez	1	1	1	1

Por meio da análise dos resultados mostrados nas Tabelas 1 e 2, é possível notar que dos 48 períodos considerados (12 meses em cada um dos subsistemas), em apenas nove casos as ordens apontadas pelo Critério 2 da identificação tradicional e os Critérios 1 e 2 do Boostrap apresentam alguma diferença. Para cada um destes casos, foram simulados, também via procedimentos de Bootstrap, o intervalo de confiança dos parâmetros do possível modelo ajustado.

Nos casos em que as ordens apontadas pelo Critério 2 da identificação Tradicional foram maiores que aquelas via Bootstrap (meses de outubro do subsistema

Sudeste/Centro-Oeste e Norte e julho no Sul), as simulações mostraram que os parâmetros dos modelos que utilizaram o Critério 2 da identificação Tradicional não foram significantes, a menos do primeiro parâmetro autorregressivo, corroborando a ordem identificada pelo procedimento Bootstrap nestes períodos.

Nos casos dos meses de maio do subsistema Sudeste/Centro-Oeste e maio, junho e setembro do subsistema Nordeste, em que as ordens apontadas pelo Critério 1 do Boostrap foram maiores que as do Critério 2 da identificação Tradicional, as simulações mostraram que ao nível de confiança de 95%, somente são significantes o primeiro e o último parâmetros dos modelos, isto é, os parâmetros intermediários não o são. Em casos desta natureza, é comum a utilização de modelos mais parcimoniosos, optando-se pelas ordens inferiores, apontadas pelo Critério 2 da identificação via Bootstrap.

No caso do mês de agosto do subsistema Norte, nos Critérios 1 e 2 da identificação via Bootstrap, as ordens selecionadas foram maiores que o Critério 2 da Tradicional. As simulações realizadas apontaram que ao nível de confiança de 95%, os dois parâmetros do modelo são significantes, porém ao nível de 99%, somente o primeiro parâmetro permaneceria no modelo.

Finalmente, para o período de setembro do subsistema Norte, as simulações mostraram que para o modelo de ordem igual a cinco, somente o primeiro, segundo e quinto parâmetros são significantes. Pelo princípio da parcimônia, opta-se pelo modelo de ordem 2, que é aquela apontada pelo Critério 2 da identificação via Bootstrap.

Portanto, a técnica de Bootstrap apresentou-se eficiente na identificação das ordens dos modelos de cada período (sobretudo com relação ao Critério 2), aproximando-se do Critério 2 do método clássico de identificação em 90% dos casos. Desta forma, recomenda-se como alternativa mais acertada a utilização da técnica de computação intensiva na fase crítica de identificação das ordens p ou, como mudança inicial na metodologia, a utilização do Critério 2 em substituição ao Critério 1, implantado atualmente.

Com relação aos cenários gerados, os mesmos mostraram-se aderentes ao histórico para todos os subsistemas. Foram realizados testes de médias, variância, aderência de distribuição, correlação e sequências negativas, PENNA (2009). Todos se mostraram

adequados e as séries sintéticas geradas conservaram as propriedades estatísticas do histórico original.

4.2- Comparação Entre o Modelo Gerador de Cenários e o Conjunto de Séries Históricas

A partir das séries obtidas no modelo gerador descrito anterior, realizou-se a comparação do desempenho destas séries em relação às séries históricas no que diz respeito à PDE, ou seja, pode-se comparar os custos operativos totais de ambos os conjuntos de séries hidrológicas, para o sistema hidrotérmico brasileiro.

Inicialmente, o conjunto total de séries é utilizado para a política operativa, que consiste na fase onde são calculadas as Funções de Custo Futuro, também conhecida como fase *backward*. Foram realizadas duas simulações, uma utilizando o conjunto de séries históricas e outra o conjunto de séries sintéticas.

A partir destas funções, foram realizadas as simulações da operação, onde cada uma das séries é utilizada para o cálculo dos montantes ótimos a serem gerados pelo conjunto de usinas térmicas e hídricas de forma a minimizar o custo total esperado da operação do sistema hidrotérmico, sendo esta fase conhecida como *forward*. Nesta fase, cada série histórica foi simulada com as funções de custo futuro obtidas para o conjunto de séries históricas, sendo ao final calculado o custo médio para todas as 70 simulações. O mesmo foi realizado para as séries sintéticas.

Para a avaliação inicial das séries geradas, são mostrados na Figura 1 os gráficos da Energia Natural Afluente (ENA) média dos 70 cenários, para o sistema brasileiro. A ENA representa a energia que pode ser gerada com o correspondente volume de água afluente ao sistema. Observa-se a aderência entre as curvas das séries histórica e sintética. Este gráfico mostra que o modelo gerador de cenários de afluências apresenta a robustez estatística esperada e possíveis diferenças nas médias das variáveis de decisão do problema podem estar associadas a períodos críticos menos ou mais severos nas séries sintéticas do que nas séries históricas. Tendo em vista a disponibilidade de séries hidrológicas às usinas hidrelétricas do setor elétrico brasileiro desde 1931, utilizou-se um total de 70 séries históricas. Para a comparação, foram utilizadas também 70 séries sintéticas obtidas com o modelo descrito nas seções anteriores.

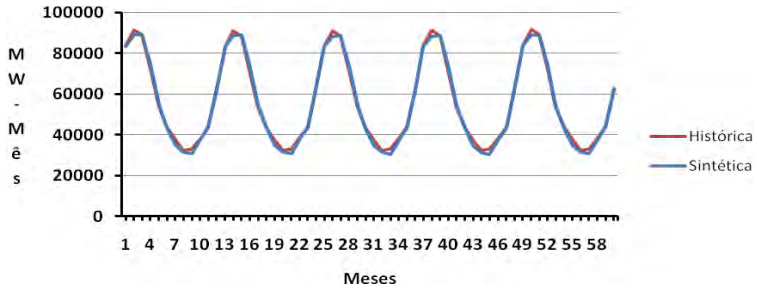


Figura 1: Energia Natural Afluente às usinas do sistema hidrotérmico.

O custo esperado de operação total que é calculado a partir dos gastos com combustíveis e possíveis déficits de energia para os dois casos são mostrados na Tabela 3, onde observa-se que utilizando as séries sintéticas obtém-se um custo aproximadamente 7% inferior às séries históricas. Isto pode indicar que o planejamento da operação utilizando-se séries sintéticas geradas pelo modelo proposto é menos restritivo que o histórico, sinalizando ao operador do sistema que espera-se cenários hidrológicos mais otimistas que o histórico de energias afluentes. Com isto, utilizando-se as séries sintéticas, observa-se que em média o planejamento da operação consegue utilizar melhor os recursos hídricos.

Tabela 3: Resultado dos custos esperados de operação total para os dois conjuntos analisados

	Custo Esperado de Operação
Histórica	2.127933 x 10^7
Sintética	1.977998 x 10^7

Para as variáveis a serem analisadas a seguir, serão mostrados apenas os resultados do subsistema Sudeste, o maior do Sistema Brasileiro. Na Figura 2, são mostrados os resultados de energia armazenada final do subsistema Sudeste, onde se observa que em ambos os casos, para a média das simulações, as curvas apresentam a mesma morfologia, além de apresentarem resultados bem próximos.

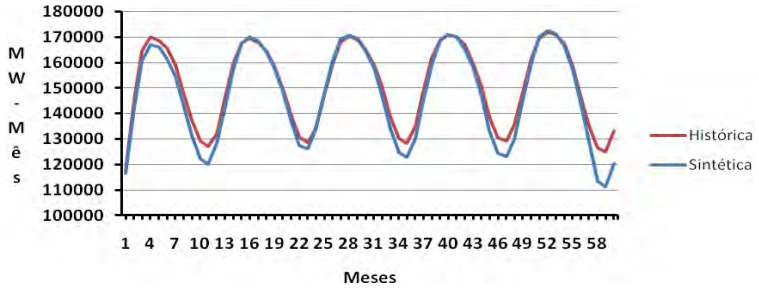


Figura 2: Energia armazenada ao final de cada mês do subsistema Sudeste/Centro-oeste.

Outra variável de importância é o Custo Marginal da Operação (CMO). Na Figura 3, é mostrado o CMO para o subsistema Sudeste onde observa-se, novamente um comportamento semelhante. No entanto, devido ao melhor aproveitamento dos recursos energéticos com os cenários sintéticos, em média, observou-se CMOs menores nas séries sintéticas do que nas séries históricas.

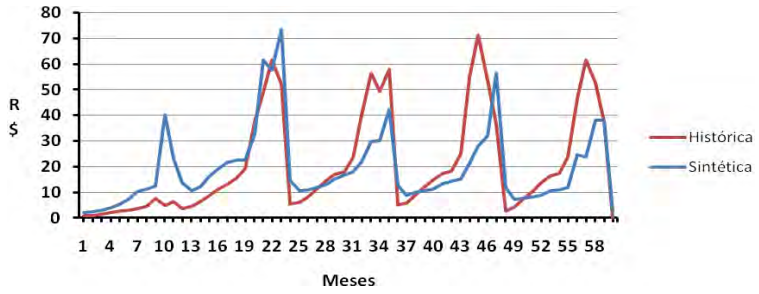


Figura 3: Custo marginal da operação médio do subistema Sudeste/Centro-Oeste.

Na Figura 4, é mostrada a média dos déficits no mesmo subsistema e devido às explicações acima observa-se, na maioria das ocorrências, déficits médios menores nos cenários sintéticos.

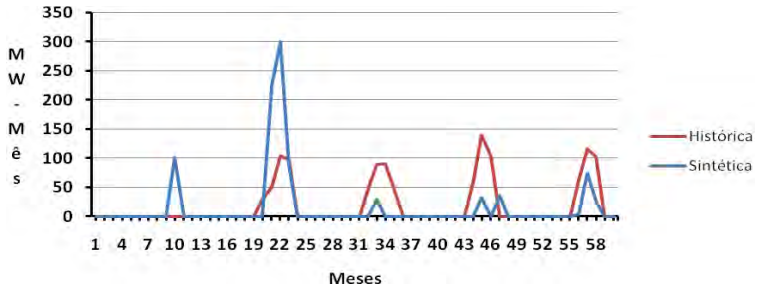


Figura 4: Déficit médio do subsistema Sudeste/Centro-Oeste.

5- Conclusões

Este trabalho apresentou uma nova metodologia de geração de cenários sintéticos de energia natural afluente cujo conceito está associado ao montante de energia que pode ser produzido por um conjunto de usinas hidrelétricas a partir da expectativa de vazões afluentes aos seus respectivos reservatórios. A partir do histórico mensal de energia afluente aos quatro subsistemas brasileiros, Sudeste/Centro-Oeste, Sul, Norte e

Nordeste, foi ajustado um modelo autoregressivo baseado na estrutura PAR(p) com o diferencial de que a identificação da ordem dos modelos é realizada com base na significância dos coeficientes da Função de Autocorrelação Parcial, determinadas através da técnica de Bootstrap. O método mostrou-se adequado, levando à escolha de modelos mais parcimoniosos. A partir dos cenários sintéticos gerados, a técnica de PDE foi utilizada incorporando o conceito de algoritmo de fechos convexos.

As principais variáveis de decisão associadas ao problema, Custo Marginal de Operação (CMO), Energia Armazenada no Final do Mês e Déficit, foram comparadas em estudos de planejamento de cinco anos com base mensal, ora utilizando os cenários históricos, ora utilizando os cenários sintéticos.

Os resultados mostraram que a nova técnica de geração de cenários sintéticos atende às expectativas associadas ao planejamento da operação de sistemas de grande porte como o Brasileiro. As diferenças nas médias das variáveis de decisão estudadas se devem à existência de séries sintéticas mais otimistas do que as séries históricas, indicando que a utilização de séries históricas pode ser muito restritiva para o planejamento da operação do sistema brasileiro. Nos próximos trabalhos espera-se poder utilizar mais cenários sintéticos para aumentar-se a riqueza estatística das análises e, também, avaliar-se com mais detalhes as seqüências negativas das séries sintéticas confrontadas com as séries históricas.

6- Referências

Arvanitidis, N. V & Rosing, J., (1970), "Composite representation of a multireservoir hydroelectric power system", IEEE Transactions on Power Apparatus and Systems PAS-89, pages 319-326.

Barber, C. B.; Dobkin, D. P.; Huhdanpaa, H., (1996) "The Quickhull Algorithm for Convex Hulls". *ACM Trans. Math. Software*, 22, 4.

Bellman, R., (1957). *Dynamic Programming*, Princeton – Princeton University Press.

Cormen, T. H., (2001). *Introduction to Algorithms*. 2nd Edition, Cambridge, The MIT Press.

Dias, B. H.; Marcato, A. L. M.; Souza, R. C., Soares, M. P.; Silva Jr, I. C.; Oliveira, E. J.; Brandi, R. B. S.; and Ramos, T. P.; (2010a), "Stochastic Dynamic Programming

Applied to Hydrothermal Power Systems Operation Planning Based on the Convex Hull Algorithm", *Mathematical Problems in Engineering*, vol.2010.

Dias, B. H.; Marcato, A. L. M.; Souza, R. C., Soares, M. P.; Silva Jr, I. C.; Brandi, R. B. S.; Ramos, T. P.; Tomim, M. A.; e Oliveira, E. J.; (2010b) "Processamento Paralelo e Algoritmo de Fechos Convexos na Programação Dinâmica Estocástica Aplicada ao Planejamento da Operação de Sistemas Hidrotérmicos", XVIII Congresso Brasileiro de Automática (CBA), 12 a 16 de Setembro de 2010, Bonito, MS - Brasil.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Hipel, K. W. & McLeod, A. I. (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier. Amsterdam.

Maceira, M. E. P. (1989). *Operação Ótima de Reservatórios com Previsão de Afluências*. Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, Brasil.

Ni, E.; Guan, X.; Li, R., (1999). "Scheduling Hydrothermal Power System with Cascaded and Head-Dependent Reservoir". *IEEE Trans. Power Systems*. Vol 14, No. 3, pages 1127 - 1132.

Oliveira, F. L. C. (2010). *Nova abordagem para geração de cenários de afluências no planejamento da operação energética de médio prazo*. Dissertação de Mestrado, Departamento de Engenharia Elétrica, PUC-Rio, Brasil.

Penna, D. D. J. (2009). D*efinição da árvore de cenários de afluências para o planejamento da operação energética de médio prazo*. Tese de Doutorado, Departamento de Engenharia Elétrica, PUC-Rio, Brasil.

Pereira, M. V. F. (1989) Optimal stochastic operations scheduling of large hidroelectric systems. *Electrical Power & Energy Systems* 11, 3, 161–169.

Salas, J. D & Obeysekera, J. T. B. (1982). ARMA model identification of hydrologic time series. *Water Resource Research*, 18, 1011-1021.

Silva, E. L., (2001). *Formação de Preço em Mercados de Energia Elétrica*, Editora Sagra Luzzatto.

Stedinger, J. R. (2001). *Report on the evaluation of CEPEL's PAR models*. Cornell University. School of Civil and Environmental Engineering. Ithaca, New York.

Thomas, H. A.; Fiering, M. B. (1962). *Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation*. In: *Design of Water Resource Systems* edited by A. Maass et al.], Harvard Univ. Press, Cambridge, Massachusetts, 459-493.

Avaliação da Gestão de Efluentes em Explorações Leiteiras: uma abordagem com integração de Análise Multicritério num Sistema de Informação Geográfica

Sandra Silva[a], *Luís Alçada-Almeida*[b], *Luís Dias*[c]

[a]*Instituto Politécnico de Viana do Castelo, Viana do Castelo, Portugal e INESC Coimbra*
sandrasilva@esa.ipvc.pt

[b]*Faculdade de Economia da Universidade de Coimbra, Coimbra, Portugal e INESC Coimbra*
alçada@fe.uc.pt

[c]*Faculdade de Economia da Universidade de Coimbra, Coimbra, Portugal e INESC Coimbra*
lmcdias@fe.uc.pt

Resumo

A Gestão Ambiental de Efluentes em Explorações Leiteiras é complexa devido à quantidade de informação envolvida, definição e escolha de critérios e consequente avaliação e ainda a necessidade de interligação com ferramentas de visualização de informação espacial. Contudo, é essencial a construção de um instrumento de apoio à decisão com o objectivo de avaliar e tomar decisões sobre a gestão de efluentes de forma a garantir que a continuidade da actividade leiteira seja compatível com a preservação do ambiente, o bem-estar animal e das populações. As metodologias multicritério de apoio à decisão (MCDA) são ferramentas que, quando conciliadas com Sistemas de Informação Geográfica (SIG), podem dar resposta a esse objectivo. Este trabalho foca-se no desenvolvimento de um interface de integração completa entre SIG-MCDA, através da construção de uma Macro escrita em linguagem de programação VBA no software ARCGIS 9.3, com o objectivo de avaliar a gestão dos efluentes de 1860 explorações da região Entre Douro e Minho, segundo 8 critérios seleccionados de acordo com a legislação portuguesa em vigor, utilizando o método de agregação de soma ponderada para obter diferentes compromissos entre os critérios. Esta ferramenta permite ao utilizador a visualização da informação dos critérios envolvidos, atribuição de pesos, obtenção de uma classificação das explorações e ainda, a visualização do seu mapeamento na região em estudo.

Palavras-chave: Sistemas de Informação Geográfica, Análise Multicritério, Gestão Ambiental de Efluentes, Sistema de Apoio à Decisão Espacial

1- Introdução

Nas últimas décadas, em Portugal, tem-se verificado uma intensificação da actividade leiteira, com o aumento do número de animais, o que leva a um a elevada produção de efluentes cujo armazenamento, tratamento e destino final, levantam problemas ambientais, sociais e económicos. Estes factores associados por vezes a deficientes concepções das explorações leiteiras do ponto de vista ambiental, escassas capacidades de armazenamento e tratamento de efluentes e utilização inadequada de métodos e equipamentos de espalhamento no solo, resultam numa prática disforme levando a uma contaminação de culturas agrícolas e cursos de água, produção de odores desagradáveis, entre outros problemas identificados. Esta problemática é notória na Região Entre Douro e Minho, onde pode estar comprometida a viabilidade de algumas das explorações leiteiras por não cumprirem o novo enquadramento legal relativo às exigências ambientais definidas na Legislação Portuguesa e Europeia.

Estes problemas são notórios na Bacia Leiteira Primária do Entre Douro e Minho, em particular nos dez Concelhos em estudo: Póvoa do Varzim, Vila do Conde, Santo Tirso, Trofa, Maia, Matosinhos, Barcelos, Esposende, Famalicão e Viana do Castelo (POBLPEMD, 2007). Esta região tem assim em mãos novos desafios no âmbito do cumprimento das exigências ambientais exigidas por Legislação Portuguesa e Europeia, podendo ver comprometida a viabilidade de muitas explorações leiteiras por não estarem de acordo com os novos enquadramentos.

Há claramente a necessidade de construir um instrumento de apoio à decisão na gestão ambiental da actividade leiteira, de forma a realizar um conjunto de análises e de obter uma classificação, para uma posterior reflexão e tomada de decisão. Este tipo de problemas possui uma grande quantidade de dados espacialmente distribuídos (agregados em critérios a analisar) contendo informação relativa a cada exploração. Os Sistemas de Informações Geográficas (SIG) auxiliam a resolução deste tipo de problemas, pela sua grande capacidade em armazenar, exibir e manipular dados espacialmente distribuídos. Essa capacidade aumenta consideravelmente quando se faz seu uso combinado com Metodologias Multi critério de Apoio à Decisão (MCDA - Multiple Criteria Decision Aid) pois fornecem um rico conjunto de técnicas e procedimentos para a avaliação das alternativas sob a óptica de múltiplos critérios

(Ascough et al., 2002; Chakhar e Mousseau, 2007a; Densham, 1991). Os métodos multicritério de apoio à decisão permitem ao decisor, ao longo do percurso, avaliar as possíveis implicações de cada curso de acção, de modo a obter uma melhor compreensão das vinculações entre as suas acções e seus objectivos. Neste trabalho é usado o software ArcGis, versão 9.3 e a linguagem de programação VBA.

A integração SIG-MCDA permite a realização de procedimentos simultâneos de análise de componentes de dados geográficos, espaciais e temáticos, proporcionando um apoio à solução de problemas espaciais complexos, mas esta nem sempre pode ser realizada de forma directa através dos comandos disponíveis no software. Por vezes é necessária a implementação de uma interface através de linguagem de programação, neste caso o VBA, através do uso de ArcObjects (mais informações em Burke, 2003; Chang, 2008; Jackson, 2000). No interface desenvolvido esta incorporação do VBA no ARCGIS permite aos decisores introduzir em o seu julgamento de valor relativamente às suas preferências de acordo com os critérios, recebendo de imediato o feedback dessas implicações na avaliação, podendo explorar e obter uma classificação ordinal das alternativas, contendo 4 classes: Má, Média, Boa e Excelente.

A especificidade dos Sistemas de Apoio à Decisão Multicritério Espacial (MC-SDSS – *Multicriteria Spatial Decision Support Systems*) é o apoio a problemas de decisão espacial multicritério com recurso à análise multicritério, ou seja, a decisão multicritério espacial refere-se à aplicação da análise multicritério no contexto espacial onde as alternativas, critérios e outros elementos do problema de decisão têm uma dimensão espacial explícita (Chakhar e Mousseau, 2007b). Os resultados da análise não dependem apenas da distribuição geográfica dos atributos, mas também dos juízos de valor envolvidos no processo decisório (Ascough et al., 2002). O papel dos MC-SDSS no espaço de tomada de decisão é ajudar o decisor na atribuição de pesos para os critérios, para avaliar as alternativas viáveis e visualizar os resultados da sua escolha (Jankowski, 1995).

As ferramentas MC-SDSS permitem a consideração de um grande conjunto de alternativas possíveis e a avaliação de múltiplos critérios conflituantes e incomensuráveis, e oferecem capacidades únicas para gerir, analisar, elaborar cenários e interacção com o decisor, tornando o processo mais consistente e confiável no que

respeita aos resultados obtidos. A integração MCDA e SIG apresenta um grande potencial na melhoria do processo de decisão espacial.

Este trabalho foca-se no desenvolvimento de uma interface de integração completa entre SIG-MCDA, através da construção de uma Macro escrita em VBA no software ARCGIS 9.3, com o objectivo de avaliar a gestão dos efluentes para cada uma das 1860 explorações da região em estudo, segundo 8 critérios seleccionados de acordo com a legislação portuguesa em vigor, usando o método de Soma Ponderada, de forma a permitir ao utilizador a visualização da informação dos critérios envolvidos, atribuição de pesos, obtenção de uma classificação e ainda, a visualização do seu mapeamento na região em estudo. O objectivo é disponibilizar no ambiente SIG, de forma integrada, funcionalidades típicas multicritério ligando informação residente no modelo relacional, informação introduzida pelo Agente de Decisão e visualizações de resultados em tabelas e nos mapas.

2- Metodologia

2.1. Definição dos Critérios

Para a gestão correcta dos efluentes é essencial que as explorações adoptem medidas eficientes a vários níveis, como as estruturas de armazenamento e tratamento de efluentes e a capacidade de armazenamento, que deverá ser dimensionada de forma a ser possível realizar uma gestão adequada e segura dos efluentes que sejam produzidos tendo em consideração a sua utilização, transferência para terceiros ou eliminação.

Com o objectivo de apresentar uma classificação sobre a Gestão dos Efluentes nas 1860 explorações leiteiras considerou-se a seguinte informação de base para definição dos critérios.

- A Portaria n.º 631/2009 de 9 de Junho, que legisla sobre a questão da gestão dos diferentes efluentes pecuários de forma a regulamentar o Regime do Exercício da Actividade Pecuária (REAP), aprovado pelo Decreto-Lei n.º 241/2008, de 10 de Novembro, cria um quadro de licenciamento para o encaminhamento destes efluentes, no qual a prioridade é a valorização agrícola e a minimização dos impactes negativos desses efluentes sobre o ambiente.
- Os resultados do inquérito realizado no âmbito do Plano de Ordenamento da Bacia Leiteira Primária de Entre Douro e Minho (2007), fundamentalmente o tópico

Armazenamento e Tratamento de Dejectos, que contém informação de cada exploração relativa a: forma de armazenamento de efluentes, capacidade, distância às linhas de água, tempo de armazenamento, separação sólido/líquido, recolha individualizada de águas pluviais, localização em relação ao estábulo e tipo.

Após consideração destes dados, foram considerados os seguintes critérios:

Tabela 1 – Definição dos critérios de avaliação e seus objectivos.

Critérios	Objectivo
• Capacidade de armazenamento da exploração (considerando os diferentes tipos)	Maximizar
• Distância de armazenamento às linhas de água a menos de 25 metros	Minimizar
• Tempo de armazenamento superior a 3 meses	Maximizar
• Número de armazenamentos com separação sólido/líquido	Maximizar
• Número de armazenamentos sem águas pluviais (ou seja, com separação das águas pluviais)	Maximizar
• Azoto total produzido	Minimizar
• Número de armazenamentos fora do estábulo	Maximizar
• Número de armazenamentos do tipo fechado	Maximizar

2.2. Método de agregação Soma Ponderada

O método de soma ponderada é um das técnicas MCDA mais aplicadas em várias áreas de estudo, em particular, na área ambiental (Pohekar e Ramachandran, 2004). Trata-se de um método conceptualmente muito simples, fácil de compreender e de fácil integração em ambientes SIG. No entanto, também tem alguns inconvenientes, destacando-se a errada interpretação que pode ser conferida aos seus parâmetros (nomeadamente os pesos utilizados) e o facto de permitir uma compensação total entre desempenhos nos diferentes critérios (p.ex. um desempenho péssimo num critério pode ser compensado por um desempenho excelente noutro critério), o que nem sempre reflecte a vontade do decisor.

Este método requer a normalização de todas as medidas de desempenho em unidades proporcionais. As medidas de desempenho são sujeitas a uma normalização da razão de diferenças, substituindo cada desempenho pela razão entre a diferença para o pior desempenho da escala e a diferença entre o melhor e o pior desempenho (note-se porém que existem outros possíveis métodos de normalização).

A medida de desempenho global é calculada para cada alternativa através da soma da multiplicação do valor normalizado para cada alternativa pelo peso atribuído para cada critério, isto é, usando uma soma ponderada simples. Há que salientar que os pesos usados neste tipo de agregação não reflectem directamente a importância de cada critério (Keeney, 1992). Neste trabalho são tratados meramente como um artifício que permite ao decisor obter diferentes resultados: ao aumentar-se o peso de um critério mantendo os restantes inalterados ir-se-á encontrar, na melhor classe, alternativas em que o critério em causa é melhorado à custa do piorar dos outros.

2.3. Integração SIG e Análise Multicritério

Existem quatro modos possíveis de integração física de SIG e das ferramentas de análise multicritério (Chakhar e Martel, 2003; Jankowski, 1995; Laaribi et al., 1996; Malczewski, 1999 e 2006): (i) sem integração, (ii) integração fraca, (iii) integração forte, e (iv) integração completa.

Neste trabalho, desenvolve-se um sistema de integração completa (Figura 1), que se caracteriza pela existência de um interface e uma base de dados únicos. O método de análise multicritério é activado directamente a partir do interface do SIG, como que de uma função SIG se tratasse. A base de dados SIG é ampliada por forma a suportar os dados geográficos e descritivos, por um lado, e os parâmetros requeridos para a aplicação da técnica MCDA, por outro (Chakhar e Martel, 2003). Tal modo é obtido através de uma linguagem de programação, de forma a criar de raiz, ou reforçar sistemas já existentes, facilitando a visualização de mapas de alternativas e sua avaliação. Nesta abordagem um sistema fornece ao utilizador um interface completo de visualização e controlo do aplicativo.

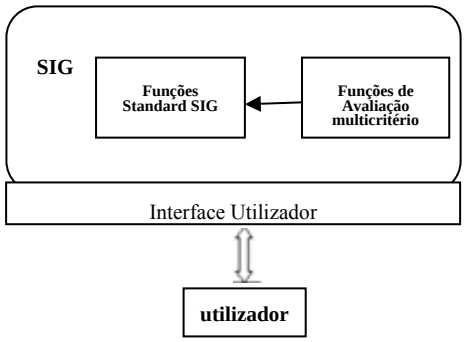


Figura 1: Esquema do modo de integração SIG e MCDA completa, baseado em Chakhar e Mousseau (2007b).

O interface desenvolvido foi escrito em Visual Basic for Aplication (VBA) usando ArcObjects, que é a plataforma de desenvolvimento para ArcGIS (ESRI, 2001). O interface funciona como uma Macro em VBA no ambiente ArcGIS.

As principais razões para o desenvolvimento do interface intitulado *Gestão Ambiental de Efluentes nas Explorações*, foram:

- as opções disponíveis no software não permitem a interacção desejada com o utilizador e a obtenção dos resultados pretendidos neste problema;
- as macros VBA em aplicativos ArcGIS podem usar as funcionalidades do ArcGis na sua plena extensão (Marinoni, 2004).

3. Caso de Estudo

3.1. Procedimento e Resultados

O procedimento implementado na avaliação da gestão ambiental dos efluentes em explorações leiteiras foi realizado tendo em consideração as seguintes três etapas:

Dados: conceptualização da base de dados com a informação em estudo: código de exploração, concelho e freguesia e a matriz com os 8 critérios envolvidos; indicação do objectivo relativo a cada critério, maximização ou minimização, segundo Portaria n.º 631/2009; determinação do máximo e mínimo dos valores em cada um dos critérios. A apresentação dos dados no interface desenvolvido é ilustrada na Figura 2.

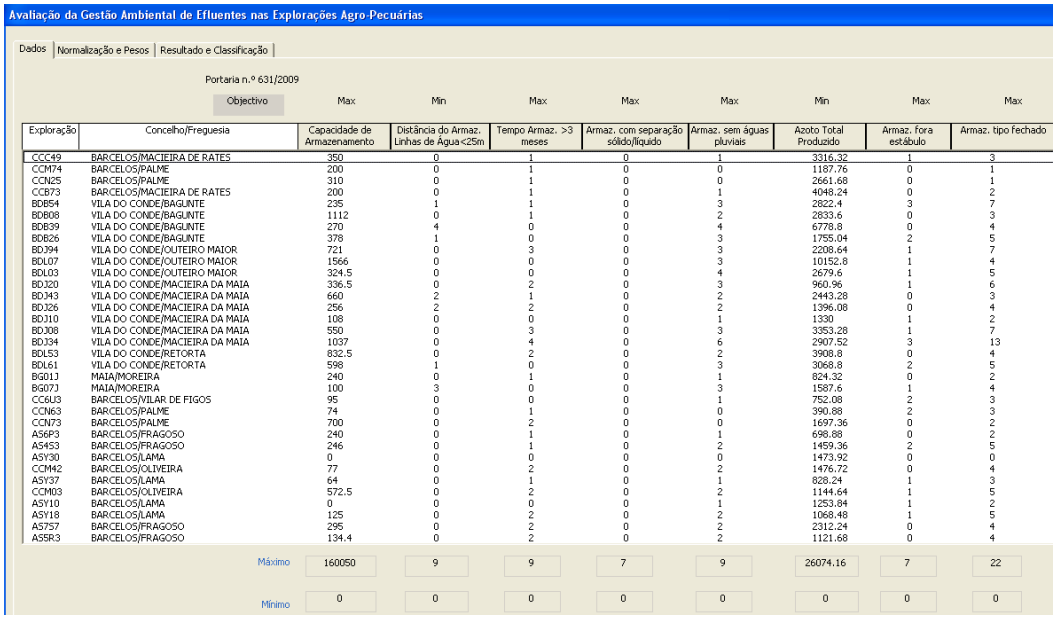


Figura 2. Visualização de Dados.

Normalização e Pesos: determinação da matriz normalizada de acordo com o objectivo de Maximização ou Minimização; estabelecimento de valores de preferência aos critérios envolvidos na matriz; realização de uma soma ponderada de todos os critérios conjuntos. A Figura 3 ilustra o interface para definição dos pesos em função dos valores normalizados.

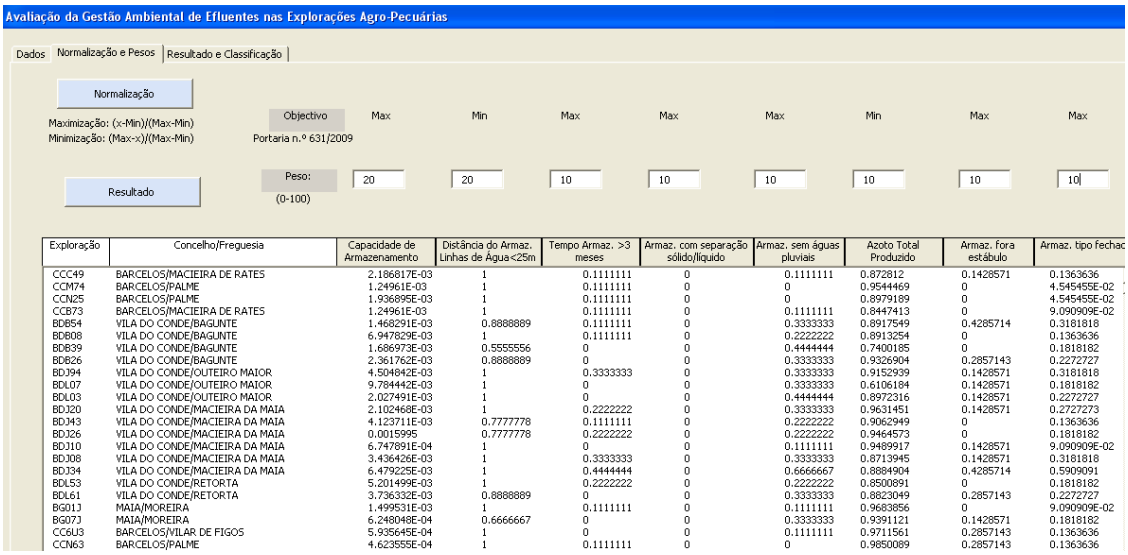


Figura 3: Normalização e Pesos.

Resultado e Classificação: obtenção do resultado e respectiva classificação para cada exploração; determinação do máximo e mínimo do resultado; a classificação é ordinal, contendo 4 classes: Má, Média, Boa e Excelente. Cada uma destas categorias corresponde a um intervalo de valores para a soma ponderada, tendo em consideração a divisão em quatro partes iguais da diferença obtida entre os valores máximo e mínimo resultante.

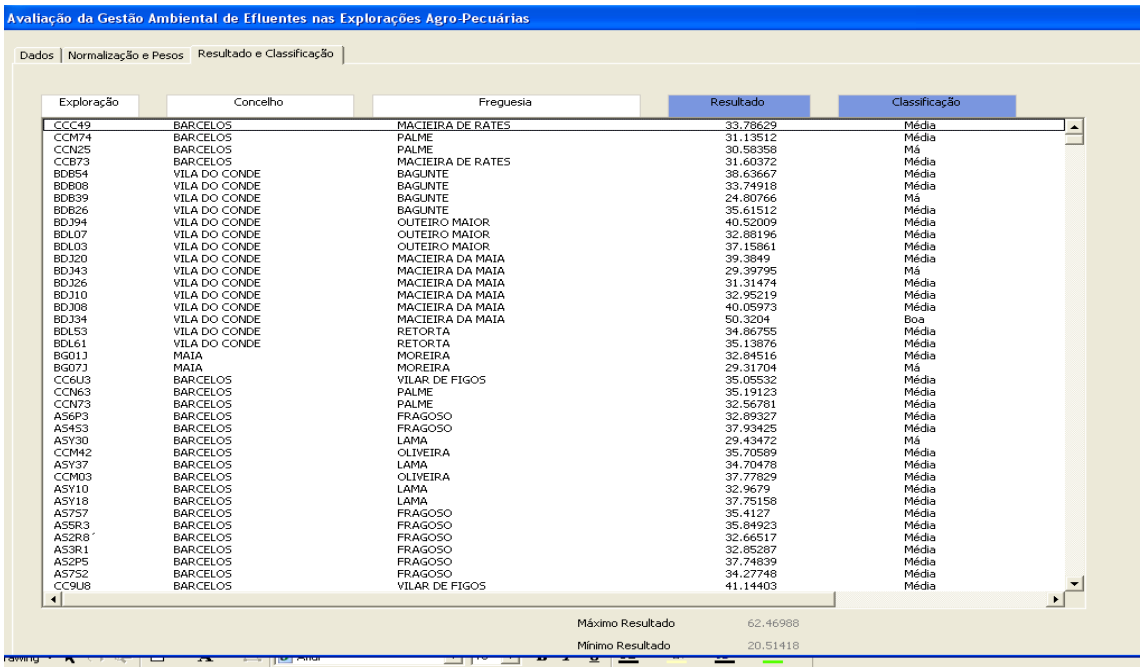


Figura 4: Resultado e Classificação.

O interface (v. exemplo na Figura 4) permite de forma dinâmica que os resultados e classificação obtidos, relativos a cada exploração, sejam introduzidos de forma automática na tabela de atributos do software ArcGis, em duas colunas: Resultados e Classificação e a partir daí o utilizador pode construir uma simbologia adequada à classificação obtida, de forma a visualizar a classificação obtida directamente no mapa.

Por fim, o utilizador pode analisar o mapa obtido (Figura 5), de forma a verificar as classificações obtidas pelas explorações espacialmente distribuídas de acordo com a informação geográfica de base. Estes resultados são apenas um exemplo do interface de visualização, pois as classificações dependem dos valores escolhidos para os pesos da soma ponderada.

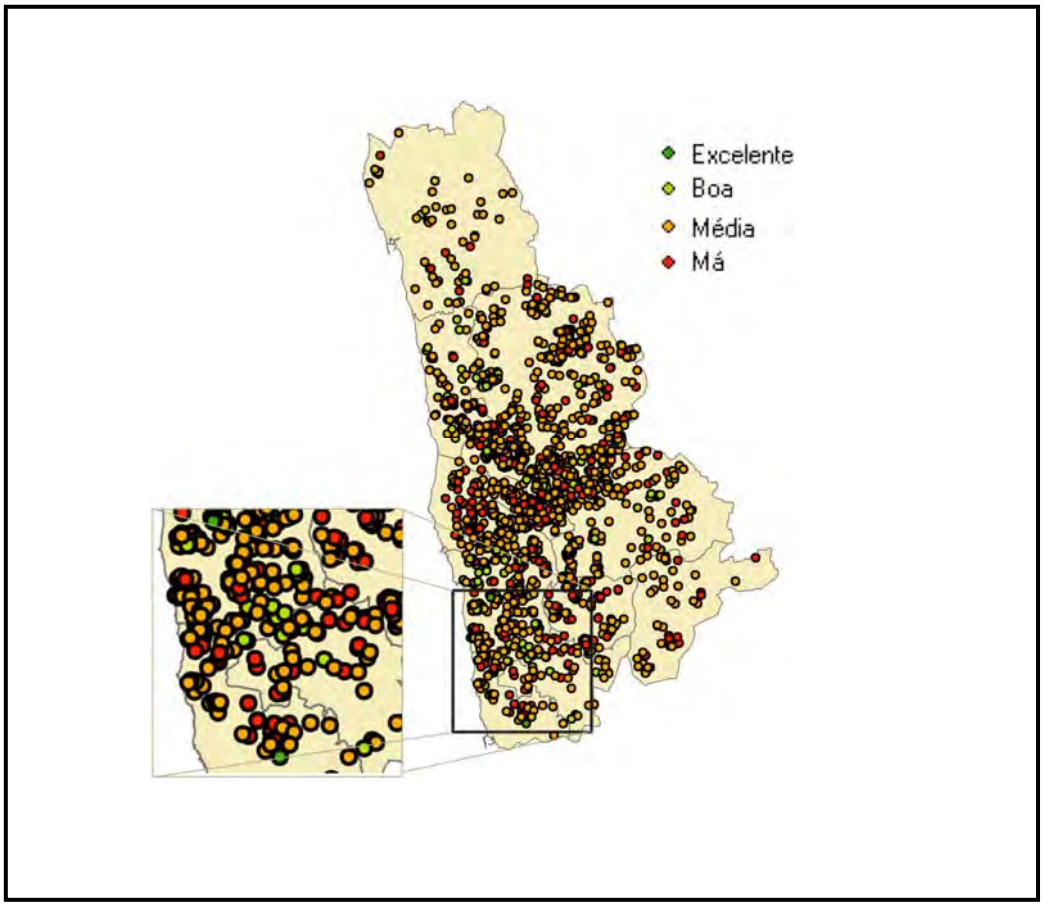


Figura 5: Exemplo de um mapa de resultados de acordo com a classificação.

Quanto à gestão de efluentes, de acordo com os pesos atribuídos, a maioria das explorações classificam-se, como "Média" (cerca de 73%), seguidamente como "Má" (cerca de 20%) e "Boa" (cerca de 7%) e apenas existem 2 explorações com a classificação "Excelente", em dois concelhos vizinhos (realçado no mapa de resultados).

Neste estudo de caso, verifica-se facilmente que a variação nos pesos atribuídos aos critérios provoca a variação na classificação final, ou seja, os dados são sensíveis aos valores dos pesos. Apesar dessa variação, verifica-se que maioritariamente as explorações são classificadas como "Má" ou "Média" e um menor número são classificadas em "Boa" ou "Excelente".

4. Conclusão e vias de trabalho a prosseguir

Admitindo que não há uma solução universal e única para a gestão dos efluentes pecuários, mas que existe antes, um conjunto de objectivos que deve ser considerado no planeamento de um qualquer sistema de gestão dos efluentes (sistema esse que deve constituir parte integrante e essencial de qualquer unidade de produção leiteira), podemos estar convictos que é possível encontrar uma solução que, quando convenientemente aplicada, conduza para que as actividades agro-pecuárias sejam compatíveis com a preservação do ambiente, o bem-estar animal e a qualidade dos produtos.

Este tipo de estudo é um passo fundamental na análise deste tipo de questões, uma vez que faz uma avaliação da situação actual, classificando as explorações de acordo com a Portaria referida e os pesos atribuídos, permitindo verificar o panorama actual. Nesta fase do trabalho debruçámo-nos apenas sobre a definição dos atributos a considerar e alguns aspectos da integração SIG-MCDA, nomeadamente a fase de selecção e estruturação de dados e o desenvolvimento da interface. A definição do modelo ou modelos de agregação multicritério que serão utilizados está ainda em aberto, bem como a escolha dos actores a envolver na definição dos parâmetros da agregação.

Este trabalho define-se como uma primeira abordagem no desenvolvimento da integração plena entre SIG-MCDA na área da gestão ambiental de efluentes, realizado através da linguagem de programação VBA e recorrendo-se a um método simples, que permite uma análise, visualização e reflexão sobre a situação. Existem limitações na abordagem efectuada e novos rumos a explorar em pesquisas futuras, como a integração de outros métodos, nomeadamente métodos que não sejam completamente compensatórios, a identificação de *clusters* regionais de explorações semelhantes e,

ainda, procurar, com base na interacção com os actores relevantes neste sistema, eliciar julgamentos de valor que permitam restringir a gama de pesos aceitável.

O objectivo último do trabalho passaria pela tomada de decisão das autoridades competentes e definição de estratégias para médio e longo prazo. A imposição de boas práticas nas explorações não poderá resultar, apenas de imposição de regulamentos e sanções. A efectividade destas medidas deverá passar pela assunção junto dos produtores (actores) de formação e aconselhamento técnico.

5. Referências

Ascough II. J. C., Rector, H.D.; Hoag, D. L.; McMaster, G.S.; Vandenberg, B.C.; Shaffer, M. J.; Weltz, M. A.; Ahuja, L. R. (2002). Multicriteria spatial decision support systems: overview, applications, and future research directions. In: Conference on Integrated Assessment and Decision Support, 2002, Suíça. International Environmental Modelling and Software Society. vol. 3, 175-180.

Burke, R. (2003). Getting to know ArcObjects: programming Arc Gis with VBA, volume 1- ESRI.

Chang, K-T. (2008) Programming ArcObjects with VBA - A Task-Oriented Approach, Second Edition. MATLAB, CRC Press.

Chakhar, S.; Martel, J.M. (2003). Enhancing geographical information systems capabilities with multi-criteria evaluation functions. Journal of Geographic Information and Decision Analysis, 7, 47-71.

Chakhar, S.; Mousseau, V. (2007a). Spatial Multicriteria Decision Making. In Shashi Shekhar and Hui Xiong, editors,Encyclopedia of Geographical Information Science. Springer.

Chakhar, S.; Mousseau, V. (2007b). Multicriteria Spatial Decision Support Systems. In Shashi Shekhar and Hui Xiong, editors,Encyclopedia of Geographical Information Science. Springer.

Densham, P. J. (1991). Spatial Decision Support Systems. In D. J. Maguire, M. F. Goodchild; D. Rhind (Eds.), Geographical Information Systems: principles and applications (Vol. 1, pp. 403-412). London: Longman.

Jankowski, P. (1995) Integrating geographical information systems and multiple criteria decision making methods. International Journal of Geographical Information Systems, 9, 251-73.

Jackson, J. (2000). Exploring ArcObjects, ESRI Development Lead, June.

Keeney, R.L. (1992). Value-Focused Thinking: A Path to Creative Decision Making. Harvard University Press, Cambridge, Massachusetts.

Laaribi, A.; Chevallier, J.-J.; Martel, J.-M. (1996). A Spatial decision aid: A multicriterion evaluation approach, Computers and Urban Systems, 20(6), 351-366.

Malczewski, J. (1999) GIS and Multicriteria Decision Analysis. New York: John Wiley & Sons.

Malczewski, J. (2006) GIS-based multicriteria decision analysis: a survey of the literature. International Journal of Geographical Information Science, 20, 703-26.

Marinoni, O. (2004). Implementation of the analytical hierarchy process with VBA in ArcGIS. Computers & Geosciences 30, 637–646.

POBLPEMD (2007). Plano de Ordenamento da Bacia Leiteira Primária do Entre Douro e Minho. Direcção Regional de Agricultura de Entre Douro e Minho; Escola Superior Agrária do Instituto Politécnico de Viana do Castelo; Instituto para o Desenvolvimento Agrário da Região Norte; Universidade do Porto – CIBIO. 153 pp.

Pohekar S. D.; M. Ramachandran (2004), Application of multi-criteria decision making to sustainable energy planning — A review, Renewable and Sustainable Energy Reviews 8, 365-381.

Portaria n.º 631/2009. Diário da República, 1.ª série – N.º 111- 9 de Junho de 2009 (3581-3594). Ministério do Ambiente, do Ordenamento do Território e do Desenvolvimento Regional e da Agricultura, do Desenvolvimento Rural e das Pescas.

Optimal Network Design of Supply Chains with Reverse Flows

Sónia R. Cardoso[a], *Ana Paula F. D. Barbosa-Póvoa,*[a] *Susana Relvas*[a]

[a]*CEG-IST, UTL, Av. Rovisco Pais, 1049-001 Lisboa, Portugal*
sonia.cardoso@ist.utl.pt, apovoa@ist.utl.pt, susanaicr@ist.utl.pt

Abstract

Markets increasing competition together with a growing concern with the environment have created a new way of thinking when designing and planning supply chains. A need to augment supply chains' sustainability has been emerging in recent years. To achieve such aim, supply chains should integrate reverse logistics activities such as the transportation and reprocessing of end-of-life collected products supported by different types of reverse flows.

In this paper, a mixed integer linear programming (MILP) formulation is developed for the design and planning of supply chains while considering simultaneously production, distribution and reverse logistics activities with the goal of maximizing the net present value. The developed model is applied to a representative example in order to show its applicability.

Keywords: reverse logistics, optimisation, design, planning, MILP.

1- Introduction

In recent years, there has been a growing concern on the environment preservation. On the other hand, supply chains are important structures in any organization, that if efficiently designed and operated can positively contribute to the minimization of the negative environmental impacts. To achieve this, companies must invest in the design and operation of these systems in order to reduce their ecological footprint [Barbosa-Póvoa *et al.* 2007]. Design and planning decisions need then to be considered simultaneously with reverse logistics activities [Seuring and Muller, 2008]. For this purpose, the supply chain should be now seen as a closed loop system [Guide and Van Wassenhowe, 2002] where reverse logistics activities are included encompassing the transportation and reprocessing of collected products. However, the establishment of a reverse network that is independent of the forward one can increase infrastructure costs and reduce the potential profit associated with remanufacturing [Uster *et al.* 2007]. So, for a better network design and planning it is necessary to consider simultaneously the forward and reverse flows.

Several studies were published in this area, such as Fleischmann *et al.* [1997] who analyzed reverse logistics concluding that the research on this subject focuses only on separate aspects and there is no holistic analysis of the supply chain. Also Savaskan and Van Wassenhowe [2006] highlighted the benefits of a closed loop supply chain. Melo *et al.* [2009] concluded on their review that, despite its importance, few works have addressed the design of supply chains with reverse flows. The same conclusion is corroborated by Salema *et al.* [2010] who state that the number of published works where both forward and reverse flows are taken into account simultaneously is less than the ones that treat them separately. Based on that, it is possible to see the importance of the development of a model that deals with both flows through an integrated approach. Such problem will be explored along the present paper.

The paper is structured as follows. In Section 2, it is made a characterization of the problem of designing and planning supply chains with forward and reverse flows. A representative example is presented in Section 3 that demonstrates the applicability of the model. Finally, in Section 4, some conclusions are made and possible future research directions are given.

2- Problem definition

In this paper, it is proposed an optimization model for the simultaneous design and planning of supply chains with forward and reverse flows, given a certain time horizon. The model has been developed based on the work by Guillén-Gosálbez and Grossmann [2010] where the authors studied the design and planning of supply chains, taking into account economic and environmental aspects. However, they only consider a three echelon supply chain and did not incorporate reverse logistics. Such model is in the current work generalized.

The problem addressed in this work has the objective of determining the configuration of the network along with planning decisions that maximize the net present value (NPV). The supply chain is formed by four echelons: plants with a set of available processes to be installed, warehouses, where final products are assembled and stored before being sent to retailers which, in turn, deliver the final products to markets. It is also considered that factories can exchange all type of products between themselves, including raw materials or intermediate products. In the reverse flow, products are collected by the retailers from markets and sorted. Products that are too damaged are

sent to disposal while products in their end of life are directed to factories to be disassembled and reprocessed. Non-conforming products are sent to warehouses to be repacked. In figure 1 it is possible to see a generic representation of this network.

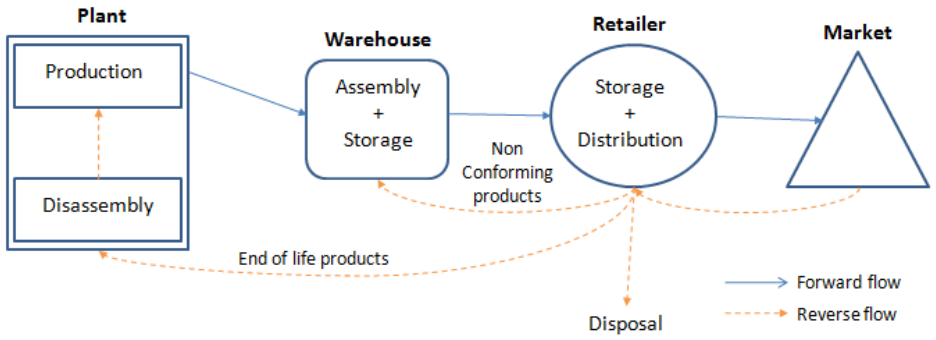


Figure 1 - Network representation.

Having such network in consideration the problem can be generally described as follows:

Given the input data:

- Upper and lower bounds for the infrastructures capacities expansions;
- Inventory costs at the warehouses and retailers;
- Initial capacity and initial inventory level in the existing infrastructures;
- Demand and price of each final product;
- Travel distance between each pair of entities of the network;
- Investment costs, operational costs and unitary transportation costs;
- Price of each raw material;
- Percentage of collected products that are sent to disposal;
- Products' bills of materials;
- Turnover ratio of warehouses and retailers;
- Minimum percentage of demand and returned products that has to be satisfied;
- Upper and lower bounds for purchases and flow of materials;
- Salvage value, tax and interest rates;
- Time horizon and time slots' extent.

Subject to:

- Mass balance satisfaction at each node of the network (plants, warehouses, retailers and markets),
- The capacity of each infrastructure cannot be exceeded,
- Material flows between each pair of entities have to be within the boundaries,
- Products' bill of materials must be complied.

Compute the following results, that maximize the NPV:

- Design variables: number, location and capacity of the processes that have to be installed in each plant, warehouses and retailers,
- Planning decisions: production rates at the plants, forward and reverse flows between the nodes of the network, inventory levels at warehouses and retailers, establishment of transportation links between all entities.

The problem was modeled trough a Mixed Integer Linear Programming (MILP). This model uses a discretization of the time horizon where all the design and planning constraints defined above were modeled. Furthermore, in terms of reverse logistics activities it was assumed that collected products in their end of life are sent by retailers to disassembly lines in plants where they are integrated into the forward supply chain after being reprocessed, along with new products. Also, and for the non-conforming products these are sent by retailers to warehouses to be repacked and thereafter to be introduced again in the network as new products. Therefore, the reverse flow consists of different types of products: (1) end of life products collected in time period t, that were sold to the market in a previous period Tn=t-n, where n accounts for the products' estimated lifetime; (2) non-conforming products, that are a fraction of the forward products supplied to markets; (3) returned products to be sent to disposal whose amount is calculated as a percentage of the total collection of returns at the end costumers.

3- Supply Chain Example

The model developed was applied to a supply chain example where two distinct scenarios are explored: (1) Case A where the expansion of an existent plant is considered and where no reserve flows exist; (2) Case B where the same supply chain is expanded but reverse flows are considered with all the associated possible reverse logistics activities.

The model was implemented in GAMS 22.8, using CPLEX 11.0, in an Intel(R) Core(TM) i7 CPU 2.80 GHz computer with 6 GB RAM.

3.1- Case A

The expansion of an existent supply chain is studied considering a time period of four years. The existent supply chain is characterized by a plant where three operational processes are installed, a warehouse with two assembly lines for packing products, and two retailers that supply four markets. Facing an increase on demand, such supply chain

has the possibility of extending its superstructure by installing a new plant and/or a new warehouse. Figure 2 represents the superstructure of the supply chain under study with the possible links between all nodes and respective distances (D) in Km. In this case, is only considered the supply chain forward flow, so there is no return of collected products.

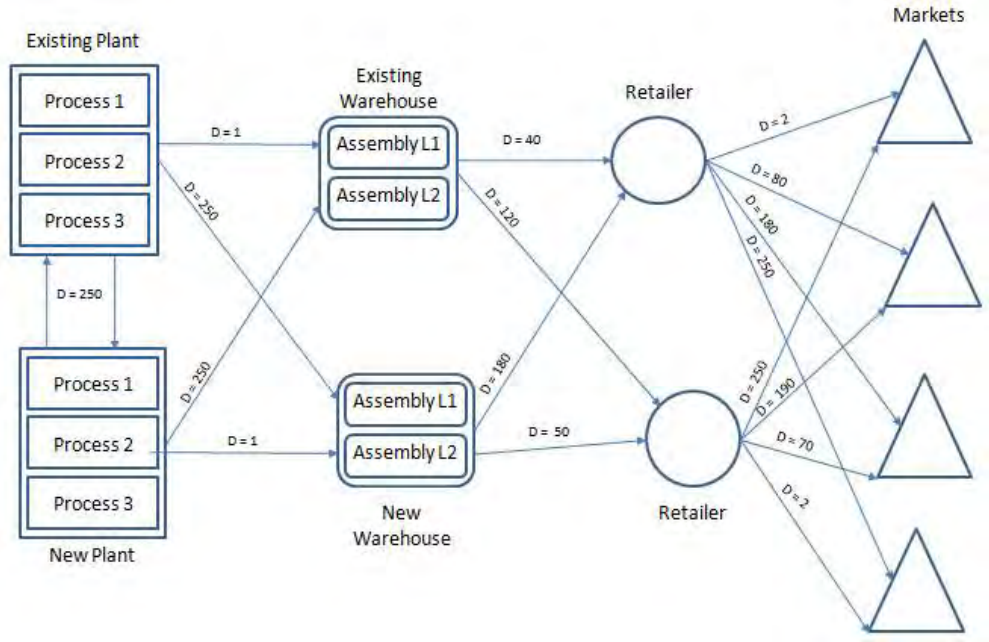


Figure 2 – Superstructure of the forward supply chain.

In each plant there are three different processes to manufacture three products, which are P4, P5 and P6. As represented in Figure 3, there are three distinct raw materials (P1, P2 and P3) and an intermediate product that is P4. This is a final product of process 1 and simultaneously an input in process 3. Figure 3 also includes the quantities of each raw material needed to produce one unit of each final product.

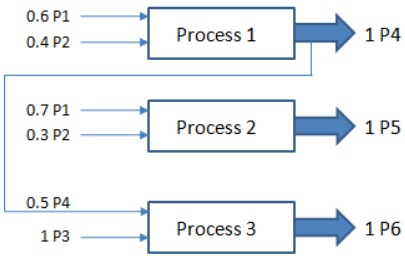


Figure 3 – Production process.

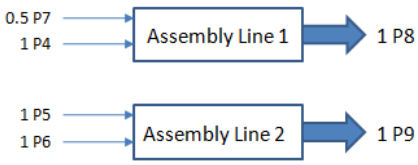


Figure 4 – Packaging and assembling process.

After being manufactured at the plants, the products are sent to warehouses to be packed and assembled in orders. As it is possible to see in Figure 4, in assembly line 1, product P4 is packed (the package is bought by the warehouse and it is called product P7). In assembly line 2, products P5 and P6 are grouped into another type of order, named P9.

A demand satisfaction target level of 90% must be attained in each of the four years in which the total time horizon is divided. All the three technologies of the existing plant have an installed capacity of 100 ton whereas the existing warehouse's assembly lines have an initial capacity of 150 ton. The existing warehouse and the retailers have a storage capacity of 100 ton. The lower and upper limits within of the capacity expansions can vary are 1 and 4000 ton for all the supply chain entities. No bounds on the total number of expansions of plants, warehouses and retailers were imposed and no upper limits on the purchases of raw materials are fixed. However, in order to prevent outsourcing, the upper limits of purchases of intermediate and final products (P4, P5, P6, P8 and P9) are zero. The lower and upper bounds on the flows of materials between all nodes of the supply chain are between 1 and 500 ton, respectively. The turnover ratio for the warehouses is equal to 5 and for the retailers is 4. The initial inventories are assumed to be zero in both cases. No minimum production levels are fixed at the plants.

The interest rate, the salvage value and the tax rate are equal to 10%, 20% and 30%, respectively. In this case it is assumed a value of 0.1 euros/ton/Km for every transportation costs except between plants, which has a cost of 0.01 euros/ton/Km. All the remaining data associated with the problem is given in Tables 1 - 5. It is assumed an increase of 5% per year in all costs and demand.

Table 1 – Variable, fixed investment and operational costs of technologies in plants for t=1.

	Variable (€/ton)		Fixed (€)		Operating Cost (€/ton)	
Tech (I)/Plant (J)	**J1**	**J2**	**J1**	**J2**	**J1**	**J2**
i1	10	8	1000	1000	1.5	1.3
i2	11	9	1000	1500	1.2	1
i3	14	10	1100	1000	1.1	1.4

Table 2 – Price and demand of final products for t=1, in euros and ton

Product (P)/Market (M)	M1		M2		M3		M4	
	€	Ton	€	Ton	€	Ton	€	Ton
P8	40	50	43	50	44	60	41	40
P9	50	50	52	70	51	40	50	60

Table 3 – Cost of raw material for t=1, in euros/ton

Product (P)/ Plant (J) or warehouse (K)	J1	J2	K1	K2
P1	9	8	-	-
P2	8	6	-	-
P3	9	10	-	-
P7	-	-	0.3	0.3

Table 4 – Variable, fixed investment and inventory costs of warehouses and retailers for t=1

Warehouse (K)/ Retailer (R)	Variable (€/ton)	Fixed (€)	Inventory Costs(€/ton)
K1	2	1000	0.25
K2	2.4	1500	0.30
R1	2	1200	0.30
R2	2.5	1900	0.30

Table 5 – Variable, fixed investment and operational costs of assembly lines in warehouses for t=1

	Variable (€/ton)		Fixed (€)		Operating Costs (€/ton)	
A. Line (M)/ Warehouse (K)	**K1**	**K2**	**K1**	**K2**	**K1**	**K2**
M1	1.5	1.5	50	50	0.5	0.4
M2	0.5	0.5	30	30	0.4	0.5

Figure 5 shows the resulting supply chain topology obtained for a maximum NPV. As it can be observed, the solution implies the construction of a new plant, with all of the total available processes installed, and a new warehouse also with the two assembly lines. The initial capacities of the processes of the existing plant are not changed and a new plant is opened with higher capacities to take advantage of reduced operational costs. In this way, the existing plant receives from the new plant, on average, 42% of all P4 that it sends to warehouse 1, 23% of all P5, almost 49% of P6 and also 70 ton of raw material P2 in the last time period considered. The network structure is stable over the

time horizon and there is only one new connection in time periods 3 and 4 between retailer 1 and market 3. The production and storage capacities of all nodes of the network and links that are established between them are represented in Figure 5.

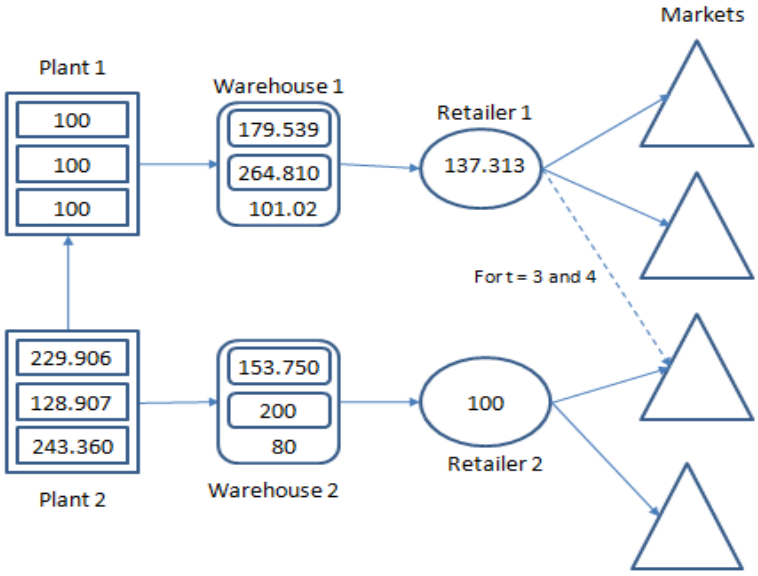


Figure 5 – Case A solution.

The objective value obtained is approximately 12878 € and the computational results of the model are presented in Table 6.

3.2- Case B

As previously referred, case B addresses the design and planning of the supply chain described above, but also incorporates reverse logistics activities. As shown in Figure 6, the markets can now send returned products to retailers where they are sorted. Products that are too damaged are disposed while products in their end of life are reprocessed in the plants. Non-conforming products are sent to warehouses to be repacked. In Figure 6 are also shown all possible links between the nodes of the reverse supply chain and the respective distances (D) in Km.

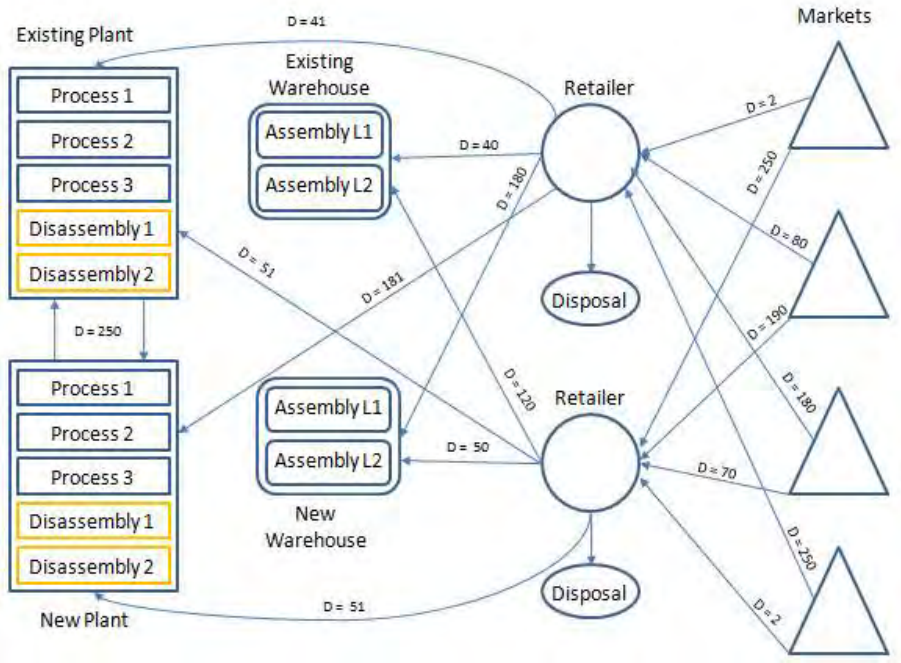


Figure 6 – Superstructure of the reverse supply chain.

At plants it is necessary to install two processes with the necessary capacity to disassembly the returned products that are after reprocessed. The returned products are P8 and P9, at their end of life. For each four units of these products, that enter in the respective disassembly process, one unit of P4, P5 or P6 is originated (Figure 7).

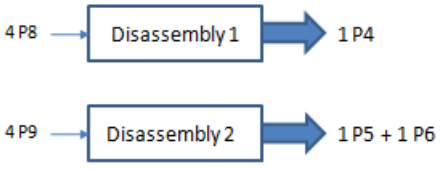


Figure 7 – Disassembly process.

In this case there are two additional disassembly processes in the existing plant with an initial capacity of 10 ton and there is also the possibility to install these processes in the new plant, if it opens, with the necessary capacities. The operational cost of the disassembly processes for both processes is assumed to be 0.01 €/ton in the first period of time. No minimum percentage of products to collect is imposed. The total quantity of non-conforming products per year is 20% of the amount that was sold in each year and

also 20% of the collected products are sent to disposal. The lifetime of the final products is 2 years for P8 and 3 years for P9.

Figure 8 shows the resulting supply chain topology that optimizes the NPV taking into account forward and reverses flows. The blue lines represent the links that are established in the forward flow, the orange lines are the reverse flows and the dashed line represents a link that suffered some alteration through the planning horizon and the periods where it was active are expressed in the figure.

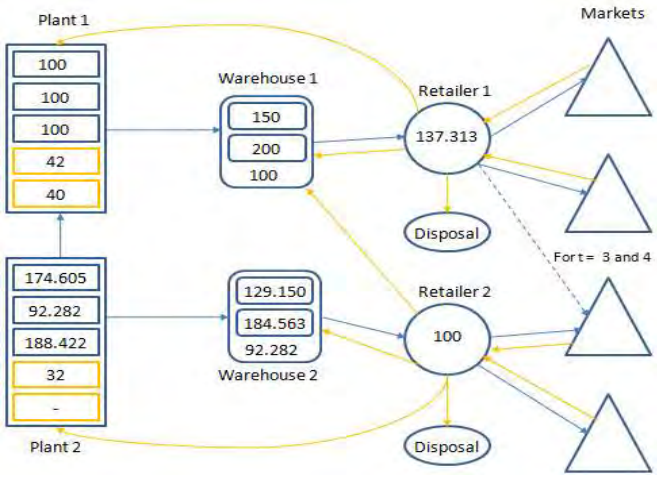


Figure 8 – Case B solution.

As it is possible to observe, it is necessary to build a new plant with all processes of manufacturing and with only one disassembling process installed. The initial manufacturing capacities of the existing plant are not changed and the disassembling capacities are expanded. A new warehouse is also built with the two assembling machines. The capacities of each entity are expressed in Figure 8. The markets are served by the nearest retailer since it has sufficient capacity. As an example, in periods 3 and 4, market 3 had to be supplied by both retailers. Each retailer sends its end of life products to one plant and the non-conforming products are sent to one or to both warehouses, such as in the case of retailer 2.

The net present value obtained with this supply chain is approximately 17075 € and the CPU times for each instance are presented in Table 6.

Table 6 – Computational results of both case studies.

	Case A	Case B
Total No. of variables	2155	3643
Binary Variables	136	216
Total No. of constraints	1650	3198
Gap (%)	0	0
CPU time (s)	0.347	0.417

As expected the increase of complexity by introducing the reverse flow in case B, results in an increase in the number of variables and constraints. However, the gap of the solution obtained is zero in both cases and the computational time takes less than half a second to provide the results of both cases.

Comparing both supply chains, it is possible to conclude that the inclusion of the reverse flow with the associated reverse logistics activities allows the achievement of a better net present value, in this case with an increase of approximately 32.6%.

4 - Conclusions

In this work it is proposed an optimization model for the design and planning of a four echelon supply chain with forward and reverse flows. The model was applied to two examples, one that considers only the forward flow of products and another one that includes forward and reverses flows simultaneously. The comparison of the results highlights the financial and environmental benefits of addressing reverse flows in an integrated approach in the supply chain, which can be seen as a business opportunity.

As future work, it is intended to incorporate environmental issues in the model with the goal of maximizing the NPV and simultaneously minimizing the environmental impacts, using a bicriteria formulation of the problem. In addition, we also intend to apply the developed model to different supply chains in order to show its applicability to real case studies and to infer on model improvement directions.

Acknowledgments

The authors gratefully acknowledge financial support from FCT, grant SFRH/BD/64125/2009 and PeerChain Project PTDC/SEN-ENR/102869/2008.

References

Barbosa-Póvoa, A., Salema, M., and Novais, A., (2007), Design and Planning of Closed-Loop Supply Chains. *In Supply Chain Optimization in Supply Chain Optimization*, Editores L. Papageorgiou and M. C. Georgiadis, Wiley-VCH, Germany, 7, 187-218.

Fleischmann, M., Bloemhof-Ruwaard, J., Dekker, R., Van der Laan, Van Nunen, J. and Van Wassenhove, L., (1997), Quantitative models for reverse logistics: A review, *European Journal of Operational Research*, 103, 1-17.

Guide, D. R. and Van Wassenhowe, L.N., (2002). The Reverse Supply Chain, *Harvard Business Review*, 80, 2, 25-26.

Guillén-Gosálbez, G. and Grossmann, I., (2010), A global optimization strategy for the environmentally conscious design of chemical supply chains under uncertainty in the damage assessment model, *Computers and Chemical Engineering*, 34, 42-58.

Melo, M.T., Nickel, S. and Saldanha da Gama, F., (2009), Facility Location and Supply Chain Management – A review, *EJOR*, 196, 401-412.23

Salema, M., Barbosa-Póvoa, A. and Novais, A., (2010), Simultaneous design and planning of supply chains with reverse flows: A generic modelling formulation, *European Journal of Operational Research*, 203, 336-349.

Savaskan, R. and Van Wassenhove, L., (2006), Reverse channel design: the case of competing retailers, *Management Science*, 52, 1-14.

Seuring, S. and Muller, M. (2008), From a literature review to a conceptual framework for sustainable supply chain management, *Journal of Cleaner Production*, 16, 1699–1710.

Uster, H., Easwaran, G., Akçali, E. and Çetinkaya, S., (2007), Benders Decomposition with Alternative Multiple Cuts for a Multi-Product Closed-Loop Supply Chain Network Design Model, *Naval Research Logistics*, 54, 890-907.

A Goal Programming Approach for the Retrofit of Supply Chain Networks

Tânia Pinto-Varela[a,b], Ana Paula F. D. Barbosa-Póvoa[b] and Augusto Q. Novais[a]*

[a]UMOSE, LNEG-INETI, Lisboa, Portugal

(tania.pinto@lneg.pt; augusto.novais@lneg.pt)

[b]Centro de Estudos de Gestão, IST, Lisboa, Portugal

(apovoa@ist.utl.pt)

**Corresponding Author*

Abstract

In order to achieve sustainability, the design and planning of a supply chain has to fulfil economic, social and environmental objectives. Traditionally the design of supply chains has been based on economic objectives. As societal environment concerns grow, environmental aspects are also emerging, not only at the industry level, but also within the context of supply chain management. The investment towards logistics structures that consider both economic and environmental performance is nowadays an important research topic. However, much is still to be done.

This paper, addresses the retrofit of supply chain networks where planning aspects are also considered. The supply chain network design and planning is modeled through a Resource-Task-Network (RTN) methodology. A mixed integer linear programming (MILP) multi-objective approach is developed, which attempts to simultaneously maximize the annual profit of the supply chain, taking into account the network retrofit, while environmental impacts are minimized. The environmental impacts are accounted for through the Eco-indicator 99 methodology. Profit and environmental impacts are balanced through the use of goal programming. The model applicability is illustrated through the solution of an example.

Keywords: Planning, Design, Supply Chain, Environment, RTN.

1- Introduction

Sustainable Supply Chains can be seen as logistic structures that guarantee the production and distribution of products globally in an environmental friendly manner (Barbosa-Póvoa 2009). To achieve such goal, companies must invest on the design and planning optimization of their logistic structures, while accounting for the trade-off between profits and environment impacts. In spite of a considerable amount of research that has already been carried out on supply chain management, a new area exploring environmental aspects is now emerging.

In recent years there has been a growing awareness for the importance of incorporating environmental aspects along with the traditional economic indicators. This trend has been motivated by several issues, a major one being tighter governments regulations and customers' perception towards more environmentally conscious systems, which may eventually lead to higher product sales (Guillen-Gosalbez et al. 2009).

The simultaneous annual profit maximization and environmental impacts minimization requires a multi-objective approach. A way to deal with such problem is through the use of a goal programming (GP) approach, which can combine the optimization with the decision maker desire to satisfy several goals. Pati et al (2008) applied GP to a recycled paper distribution network. Three goals were pursued: reduction in reverse logistics cost; product quality improvement through increased segregation at the source; and environmental benefits through increased wastepaper recovery. In the following year, Melo et al. (2009) presented a survey, where the majority of cited works feature a cost minimization objective and noticed that very few refer to models using multiple and conflicting objectives, covering both profit and environmental aspects. More recently, the design of a sustainable supply chain network was studied by Chaabane et al. (2010). A multi-objective approach is used that denotes an emission trading scheme to achieve sustainability objectives in a cost-effective manner.

In this work the simultaneous optimal design and planning of a supply chain is investigated. A generic and uniform mathematical framework is developed which accounts for the design and retrofit (of the supply chain, while considering profit maximization together with minimization of environmental impacts. A GP formulation is used to deal with the multiple objectives involved.

The paper is structured along five sections. Section 1, the current one, gives a brief introduction to the work. In section 2 the model framework is characterized. Section 3 presents the problem characterization followed by section 4, which illustrates the model applicability. This paper concludes with final remarks and lines for future work.

2- Modelling framework

As stated above, the current problem uses three main methodologies, the Resource-Task-Network (RTN), the Life Cycle Analysis (LCA) and the Goal Programming (GP) Approach. Briefly the main concepts of these methodologies are described below.

a. Resource-Task-Network Methodology

RTN, presented by Pantelides (1994), is a general and conceptually simple representation methodology. Its main characteristics lie in the uniform description and characterization of the available resources, with no distinction between them, and on the

definition of tasks. A Task is an abstract operation that consumes and/or produces a specific set of Resources. Resources can be classified in: non renewable, which represents raw materials, utilities, manpower, etc., and renewable, which represents all types of equipment associated to the supply chain network (manufacturing, warehouse, distribution centre, transportation, technology, etc). Each task *k* has a fixed duration τ_k and the execution of task *k* starting at time *t* is characterized by the pair of variables (N_{kt}, ξ_{kt}). N_{kt} is an integer variable while ξ_{kt} is continuous. These variables define respectively the number of instances of task *k* starting simultaneously at time *t*, and the total amount of resource being processed by all these instances. The amount of resource type *r* produced at time θ relative to the start of task *k* at time *t* is given by:

$$\mu_{kr\theta} N_{k,t-\theta} + \nu_{kr\theta} \xi_{k,t-\theta} \quad (1)$$

where $\mu_{kr\theta}$ and $\nu_{kr\theta}$, $\theta = 0,..., \tau_k$ are known constants. Negative values for the latter indicate consumption of the resource, while positive values denote production. These coefficients allow the modeling of different type of tasks such as those with time-dependent manpower or utility requirements, as well as those that involve transformations from raw materials to products using a process facility. In terms of resources, changes on their utilization can only occur at interval boundaries. The variable R_{rt} denotes the amount of excess resource *r* over time interval *t* (e.g. the amount that is not involved in the tasks that are active over this time interval). The change in the excess resource level for each resource type from one time interval to the next is given by excess resource balance constraints that will be defined later on.

b. Life Cycle Analysis

LCA can be described as a quantitative framework for considering the environmental impacts associated with every stage in the life cycle of a product, from raw materials production to final disposal. Based on this, the Eco-indicator methodology appears as a powerful tool for designers to aggregate LCA results into easily understandable and user-friendly quantitative units (Ministry of Housing et al. 2000). The Eco-indicator 99 introduces a damage function approach that represents the relation between the impact and the damage to human health, resource and ecosystem. Such methodology involves three main steps: (1) inventory of all relevant emissions, resource extractions and land-use in all processes that form the life cycle of a product, (2) calculation of the damages that these emissions flow cause on the Human Health, Ecosystem Quality and Resources and (3) weighting these damage categories. The set of weights used, noted by Pt (it reads Points), results from the contributions of all types of damage. Those weights

were defined using a social study (PRéConsultants 2001) measuring the social sensibility to each type of damage for population samples representative of three cultural perspectives: equalitarian, hierarchical and individualistic. Furthermore, due to the small dimension of the social panel used in that particular study, all subsequent work uses a set of mean weights, which are: 40% for both Human Health and Ecosystem, with the remaining 20% assigned to Resources.

c. Multi-Objective Optimization

The design and planning of a supply chain taking into account environmental impacts is obtained through the solution of a deterministic MILP formulation (Pinto-Varela et al. 2011), which is extended in this paper with the application of a GP approach.

The model developed considers the following sets, parameters and variables:

Sets:

C= {r: set of all non-renewable resources}
C_r={r: set of material resources storable in dedicated warehouse/distribution centers}
C_p / C_f = {r: set of final products /raw materials}
D={d :set of damages}
E={p:set of pollutants emitted}
F ={f:set of facilities}
L={p:set of pollutants from land use}
N={p:set of pollutants from natural resources extraction}
T_P={k: set of technological processes(tasks) in an resource technology, r}
T_v={k: set of technological processes(tasks) in an dedicated warehouse and distribution centers }
T_T ={k: set of all technological processes associated to connection between two entities}
T_r^f ={r: set of all resources technologies that may be installed in facility f}
U={u: set of utilities}
W_r={r: set of all resources renewable}
W_{fp}/W_{rm}={r ∈ W: set of dedicated warehouse and distribution centers for final products/raw materials}
W_v={r ∈ W: set of dedicated warehouse and distribution centers}
W_c ={r ∈ W: set of all connections between entities}
W_p={r ∈ W: set of resource technologies}

Parameters:

CC_r^0 -fixed cost associated to the supply chain resources
CC_r^1- variable cost associated to the supply chain resource
CC_r^s - variable cost of material storage
CCF - capital charge factor
F^f - maximum amount of resource technologies available in facility f
H - planning horizon per year
$HourYr$ – number of hours per
$NormF_g$ - weighted value of the damages g
$Q_r^{\min}$ / $Q_r^{\max}$ - min/max capacity available for resource r
$R_r^{\min}$ / $R_r^{\max}$ - the minimum/maximum demand of the resource at H
Km_r- distance between two entities
v_r /p_r - price of resource (raw material / product)
α_k^0 / α_k^1 - fixed and variable cost coefficients for techonogical processes
α_{ur}^{WD}/β_{ur}^{WD}- fixed and variable utility cost coefficients for dedicated warehouse and distribution center
$\mu_{kr\theta}$/$\nu_{kr\theta}$ - consumption /production of a renewable(-1,1)/non-renewable(-1,0) resource r, at the start/end of θ
$\phi_r^{\max}$ - resource techonogy size factor
$\phi_r^{\min}$ - resource techonogy size factor
$\Omega_{u,p}$ - quantity of pollutant emitted to generate an unit of utility u consumed
η_u - amount of diesel consumed m^3/km
ς_{dp} - impact factor coeficient
λ_{pr} - defines the quantity of pollutants p, emitted per unit mass of resource r used
λ_{pf} - defines the quantity of pollutants p, emitted to soil occupation/transformation
λ_{pu} - defines the quantity of pollutant p, emitted per unit of utility u consumed

$\alpha^F_{uk}/\beta^F_{uk}$ - *fixed and variable utility coefficients for the techonological process*

Variables:

DR^u- *direct costs associated to the utilities consumption*

Dam^{SC}_d - *set of damages*

Eco99- environmental indicator

P_{rt} - *amount of material delivered from the resource technology r in instant t*

Q_r - *capacity of resource technology r*

QR_p=*quantity of pollutant emitted from natural resources*

QL_p-*quantity of pollutants emitted from the land use*

QE_p- *quantity of pollutants emitted from the emissions*

DR- direct costs

Q^{Utotal}_p - *total amount of pollutants emitted*

R_{rt} - *excess of resource at t*

R_{r0} - *level of resource provisions necessary*

UT_u- *total amount of utility consumed*

ξ_{kt} - *production/storage size of technological process k at time t*

N_{kt} – *technological process selection k at instant t*

y_r -*1 if resource technology r is used; 0 otherwise*

y^f - *1 if the facility is opened; 0 otherwise*

PR-operational revenues

OCI-Indirect cost

The model evaluates two criteria, the environmental impact defined in equation (2) and annual profit in equation (3), followed by a sets of constraints from equation (4) to (25), showed below:

$$Eco99 = \sum_g NormF_g Dam^{SC}_g \tag{2}$$

$$Profit = (PR - DR) \times \frac{HourYr}{H} - OCI \times CCF \tag{3}$$

The Design and planning model of supply chain:

Excess resource technology balance: the mass balance for each resource technology must be satisfied at every instant *t.*

$$R_{rt} = R_{r_0|t=1} + R_{r,t-1|t\geq 2} + \sum_k \sum_{\theta=0}^{\tau_k} (\mu_{kr\theta} N_{k,t-\theta} + \upsilon_{kr\theta} \xi_{k,t-\theta}) + P_{rt} \quad \forall r \in W_p \; t = 1...H+1 \tag{4}$$

Operational Constraints are based on the following rules: at any one time each resource technology is either idle or being used by a technological process (task) that cannot be pre-empted once started. This can be modelled by:

$$\sum_{t'=t-\tau_k+1}^{H} \sum_{k\in T_r} N_{kt'} \leq y_r \qquad \forall\, r \in W_p \tag{5}$$

The amount of material being processed through a technological process, *k*, in resource *r* must always be within the maximum and minimum capacity available of the resource technology:

$$\phi^{\min}_{kr} Q_r N_{kt} \leq \xi_{kt} \leq \phi^{\max}_{kr} Q_r N_{kt} \qquad \forall\, k \in T_p, r \in C, t = 1...H \tag{6}$$

The previous equation is replaced by the equation (8) after the linearization.

$$\phi^{\min}_{kr} \sum_{j=1}^{N^{\max}_k} j \tilde{Q}_{rjkt} \leq \xi_{kt} \leq \phi^{\max}_{kr} \sum_{j=1}^{N^{\max}_k} j \tilde{Q}_{rjkt} \qquad \forall\, k \in T_p, r \in W_p, t=1...H \tag{7}$$

For simplicity, a detailed explanation of the linearization is omitted, but the equations are showed. A more detailed information is presented in Pinto-Varela et al. (2011).

$$N_{kt} = \sum_{j=1}^{N_k^{\max}} j\,\tilde{N}_{jkt} \qquad \forall\, k \in T_p,\ t=1...H \qquad (8)$$

$$\sum_{j=0}^{N_k^{\max}} \tilde{N}_{jkt} \le 1 \qquad \forall\, k \in T_p,\ t=1...H \qquad (9)$$

$$Q_r^{\min}\tilde{N}_{jkt} \le \tilde{Q}_{rjkt} \le Q_r^{\max}\tilde{N}_{jkt} \qquad \forall\, k \in T_p,\ r \in W_p,\ j=1...N_k^{max},\ t=1...H \qquad (10)$$

$$\sum_{j=0}^{N_k^{\max}} \tilde{Q}_{rjkt} = Q_r \qquad \forall\, k \in T_p,\ r \in W_p,\ t=1...H \qquad (11)$$

Resource technology capacity and design constraints:

$$Q_r^{\min} y_r \le Q_r \le Q_r^{\max} y_r \qquad \forall\, r \in W_p \qquad (12)$$

Dedicated warehouse/distribution centre capacity and design constraints:

$$R_{rt} \le \phi_r^{\max} Q_r \qquad \forall\, r \in W_v,\ k \in T_V,\ t=1...H+1 \qquad (13)$$

$$Q_r^{\min} y_r \le Q_r \le Q_r^{\max} y_r \qquad \forall\, r \in W_v \qquad (14)$$

Transport constraints:

$$\xi_{kt} \le \phi_{kr}^{\max} Q_r \qquad \forall\, k \in T_T,\ r \in W_c \qquad (15)$$

$$Q_r^{\min} y_r \le Q_r \le Q_r^{\max} y_r \qquad \forall\, r \in W_c \qquad (16)$$

Market demand constraints: the model allows the possibility, as part of the supply chain design problem, of optimizing its design taking into account the trade-off between its total cost and the added value of satisfying demand over the planning horizon *H*, together with the global environmental impact.

$$R_r^{\min} \le R_{rt} \le R_r^{\max} \qquad \forall\, r \in C_p,\ t=1, ..., H+1 \qquad (17)$$

Supply Chain Facilities Design: The choice of a certain facility is defined by the choice of any of the technological resources associated to it:

$$y^f \ge \sum_{r \in T_r^f} y_r \qquad \forall\, f \in F \qquad (18)$$

Life Cycle Analysis constraints:

The utilities consumptions are defined by equation (20), which is used to obtain the emission inventory (equation (21)), followed by land and resource inventory in equation (22) and (23), respectively.

$$UT_u = \sum_t \sum_{k \in T_p} \sum_{\theta}^{\tau_k} \left(\alpha_{uk}^F N_{kt-\theta} + \beta_{uk}^F \xi_{kt-\theta} \right) + \sum_{r \in W_v} \alpha_{ur}^{WD} y_r + \sum_t \sum_{r \in W_v} \beta_{ur}^{WD} R_{rt} \qquad \forall\, u \in U \qquad (19)$$

$$QE_p = \sum_u \Omega_{up} \left(UT_u + \eta_u \sum_{r \in W_c} Km_r y_r \right) \qquad \forall\, p \in E \qquad (20)$$

$$QL_p = \sum_f \lambda_{pf} y^f \qquad \forall\, p \in L \qquad (21)$$

$$QR_p = \sum_{r \in C_f} \lambda_{pr} \left(R_{r_0} - R_{rt|t=1+H} \right) + \sum_u \lambda_{pu} UT_u \qquad \forall\, p \in N \qquad (22)$$

The total pollutant inventory and the environmental impact are expressed by the equation (24) and equation (25) respectively:

$$Q_p^{Utotal} = QE_p + QL_p + QR_p \qquad \forall\, p \qquad (23)$$

$$Dam_d^{SC} = \sum_p \varsigma_{dp} Q_p^{Utotal} \qquad \forall\, d \in D \qquad (24)$$

Taking into account the previous constraints, the objective functions ((2) and (3)) become soft constraints specifying the requirements that are desirable to verify. The deviational variables d_i are added to the model, because there may be no solution that simultaneously achieves the desire levels of all soft constraints. The model presented can be stated for the GP approach as:

$$\begin{aligned} &Min \quad d_1 + d_2 \\ &s.t. \\ &\quad Profit + d_1 \geq t_1 \quad (z_1 \succ t_1) \\ &\quad Eco - d_2 \leq t_2 \quad (z_2 \prec t_2) \\ &\quad g_j(x) \leq 0 \qquad j = 1, 2, \ldots J; \\ &\quad h_k(x) = 0 \qquad k = 1, 2, \ldots, K; \\ &\quad x_i^{(L)} \leq x_i \leq x_i^{(U)} \quad i = 1, 2, \ldots, n. \end{aligned} \qquad (25)$$

Since the two objectives are defined in different units with different magnitude, the equalization of the criterion value over the efficient set is used. Each objective function is then multiplied by its representative *range equalization factor*.

$$\pi_i = \left(\frac{1}{Z_i} \right) \left[\sum_{j=1}^{k} \frac{1}{Z_j} \right]^{-1} \qquad (26)$$

where Z_i is the range width of the *i th* criterion value over the efficient set.

3- Problem statement

Summarising and assuming a uniform discretization of time, the problem in study can be stated as follows. *Given*: a fixed time horizon, set of products, markets in which products are available to customers and their nominal demand; a set of geographical sites for locating facilities, warehouse and distribution centres; a set of technologies for product manufacturing; the lower and upper bounds for the capacity of facilities, warehouse and distribution centres; the RTN representation for the product recipes; the suppliers capacity; the fixed and variable costs associated to the setting up of facilities, warehouse, distribution centres, the materials transportation and operational costs; price for every product in each market and raw-material costs; diesel and electricity consumptions; all the necessary environmental specifications and parameters.

Determine: the facilities to be opened; the technologies to be selected at each production facility; warehouse and distribution centres capacity; the amount of final products to be sold in different markets and the flow of materials to be transported.

So as to balance the maximization of the supply chain profit, while simultaneously considering the environmental impact minimization.

4- Illustrative Example

An illustrative example based on a real case from a pulp and paper industry is studied. All the presented data were modified due to confidentiality reasons.

Two cases are used to show the applicability of the proposed model. Case a) that corresponds to the optimization of the operating conditions of the current supply chain, where the maximization of the profit is obtained for the available facilities, not taking into account environmental aspects. Case b) employs the GP approach to investigate alternative strategic targets, taking case a) as a reference.

Both cases were solved for an optimality margin of 5%. A Pentium Intel Core 2 Quad, 3.0 GHz and 3.5 GB RAM, and the GAMS 21.3 software package with the CPLEX (v 11.0) solver was employed.

The current supply chain operates with two production sites (A and B). Based on their technological characteristics, a prospective acquisition of more modern multipurpose technologies is to be analysed for site B, aiming at the reduction of the associated environmental impact. However, since new investment is required, a set of conditions were predefined namely, site A facility must produce at least 10 tonnes of each final product per year, maintaining the characteristics already defined related with the

technological resource suitability, variable not only production capacity but also annual production in the facility and two new targets must be evaluated (1) an increase around 20% of profit and (2) a decrease of 10% in the environmental impact. Based on these conditions a viability study analysis was made.

Sites' Facilities characteristics

The paper industry has two facilities respectively installed in Site A and Site B. Each one of these production sites has three parallel multipurpose technological resources (MP1, MP2 and MP3) available to produce 12 types of different products. MP1 produces six final products (P1 to P6) and MP2 and MP3 three final products each (P7 to P9 and P10 to P12, respectively). All the production is exported to the international market, mainly in Central Europe. The technological resources suitability remains unchanged, as well as the annual production range of site A. Site B may install a new technology that is able to increase 50 % of its annual production.

In Table 1 the capacity associated to the facilities technologies is characterized and it is showed that the maximum capacity available in site A remains constant from case a) to case b). In Table 2 is showed the demand for each market, where there is a dedicated distribution centre (DC).

Table 1– Facilities suitability technological resources and capacities.

		Capacity (min:max Tonnes)	
Facilities	Technological Process	Case (a)	Case (b)
Site A	MP1	0:85	0:85
	MP2	0:45	0:45
	MP3	0:25	0:25
Site B	MP1	0:65	0:130
	MP2	0:25	0:50
	MP3	0:25	0:50

Table 2 – Product annual demand by each market.

		Demand for each product (min :max Tonnes)	
Product	Market	Case (a)	Case (b)
P1-P6	M1	0:200	10:300
	M2	0:260	10:390
P7-P9	M3	0:200	10:300
	M4	0:140	10:210
P10-P12	M5	0:100	10:150
	M6	0:80	10:120

In terms of environment, not only the electricity consumption associated to the technological resource (MPi) is quantified, but also the material transportation

environmental impact, namely in terms of CO2, NOx and SOx emissions. The corresponding emitted values per utility consumption (equations 20-24) are given in Table 3.

Table 3- Pollutants emitted per utility consumption (Duque et al. 2010).

Utility	CO	CO2	NOx	SOx	Units
Diesel	14.828	2609.5	34.6	-	Kg/m3
Electricity	4.151e-3	7.306e-1	1.941e-3	3.872e-3	Kg/kwh

The transportation costs are dependent on the geographical distance between the locations involved and quantities transported. It was assumed full truck load freights and an average speed of 80 km/h.

In this work the environmental analysis is focused on the Human Health (HH) damage and Table 4 presents it reflecting the respiratory effects of the inorganic substances emitted by the utilities consumption (equation 25).

Table 4- Damage to Human Health (Geodkoop et al. 2001).

Damage	CO	CO2	NOx	SOx
Human Health (DALYs/kg emission)	-	7.5e-4	8.74e-5	5.35e-5

The results from the two cases are showed in Table 5. In case (a) the site A facility presents a higher production than site B, providing all products for central European markets, while site B does not supply markets M2 and M4. In case (b) and site B, where retrofit is possible, the use of better performing technologies leads to an increase in production, leaving the production for some products at site A, at the level imposed by the board. In both facilities all the technological resources are used. The retrofit of site B allows an increase in production around 36 %.

Table 6 shows the values of profit and environmental impact for each case. Case (b), due to the retrofit in site B, allows an increase of 42% in profit when compared to case (a) and exceeds by 28.9% the board target. On the other hand, it presents a reduction of 21.4% in the environmental impact when compared with case (a). While this represents a 9.3% drop, it is slightly below the board target of 10%. An adjustment to the latter target could be sought, but it is considered satisfactory in terms of the broad goals set by the board.

In case (a) the solution was reached in 1.8 s, with 9784 constraints and 2 679 binary variables in a total of 7 660. Case (b) is characterised by 9787 constraints, 2 679 discrete variables in a total of 7 663 and solution reached in 1.3 s.

Table 5–Warehouse/DC sites, products produced in each facility and respective flows.

		Case a (Tonnes)			Case b (Tonnes)		
	Products	Site A	Site B	Total	Site A	Site B	Total
DC-M1	P1	85	100	185	75	195	270
	P2	85	65	150	75	200	275
	P3	85	100	185	75	185	260
	P4	85	100	185	75	195	270
	P5	85	100	185	75	195	270
	P6	85	100	185	85	195	280
DC-M2	P1	-	-	-	10	190	200
	P2	85	-	85	10	190	200
	P3	85	-	85	10	10	20
	P4	85	-	85	10	190	200
	P5	130	-	130	10	190	200
	P6	-	-	-	10	190	200
DC-M3	P7	100	100	200	10	160	170
	P8	90	75	165	45	160	205
	P9	90	100	190	10	85	95
DC-M4	P7	70	-	70	35	140	175
	P8	70	-	70	45	140	185
	P9	45	-	45	35	140	175
DC-M5	P10	50	50	100	15	100	115
	P11	50	50	100	15	100	115
	P12	50	50	100	15	100	115
DC-M6	P10	25	40	65	10	75	85
	P11	40	20	60	10	75	85
	P12	40	40	80	10	75	85

Table 6- Profit and environmental results.

	Case a)	Target	Case b)	Units
Profit	4e9	4.9e9	6.9 e9	m.u.
Eco 99	86055	77452	70847	mPt

The warehouse and distribution centres design for both networks are showed in Table 7.

Table 7- Warehouse and distribution centre design.

	Warehouse/DC design (Tonnes)	
	Case a	Case b
DC_M1	1075	1625
DC_M2	385	1020
DC_M3	555	470
DC_M4	185	535
DC_M5	300	345
DC_M6	205	255

5- Conclusions

In this paper the design and retrofit of supply chains is studied while accounting simultaneously for profit maximization and environmental impacts minimization. The mathematical formulation incorporates three methodologies, Resource-Task-Network used to define the supply chain characteristics with no ambiguity, Eco-indicator 99,

which quantifies the environmental aspects and Goal Programming approach that deals with the multi-objective characteristics of the problem.

The goal programming approach allows a multi-objective problem to be represented in a very simple way, which is an approach that might be expected to become more attractive as the number of targets increases.

An illustrative example based on a real case study was presented where the retrofit of an existing supply chain is optimized taken into account not only economic and environmental aspects, but also administration board's strategies. These consist of two defined targets employed when analysing the retrofit of the existing supply chain.

As future work a more detailed analysis of the quantification of the environmental impacts will be considered, as well as other methods to deal with the multi-objective characteristics of the problem.

Acknowledgment

The authors gratefully acknowledge the support of the Portuguese National Science Foundation through the project PTDC/SEN-ENR/102869/2008.

References

Barbosa-Póvoa, A. (2009). Sustainable Supply Chains. Computer Aided Chemical Engineeringg. Brasil, Elsevier**:** 127-132.

Chaabane, A., A. Ramudhin, et al. (2010). "Design of sustainable supply chains under the emission trading scheme." International Journal of Production Economics: doi:10.1016/j.ijpe.2010.10.025.

Duque, J., A. Barbosa-Povoa, et al. (2010). "Design and Planning of Sustainable Industrial Networks: Application to a Recovery Network of Residual Products." Industrial & Engineering Chemistry Research **49**(9): 4230-4248.

Geodkoop, M. and R. Spriensma (2001). The Eco-indicator 99. A damage oriented method for Life Cycle Impact Assesment, Pré Consultants B.V. .

Guillen-Gosalbez, G. and I. E. Grossmann (2009). "Optimal Design and Planning of Sustainable Chemical Supply Chains Under Uncertainty." Aiche Journal **55**(1): 99-121.

Melo, M. T., S. Nickel, et al. (2009). "Facility location and supply chain management - A review." European Journal of Operational Research **196**(2): 401-412.

Ministry of Housing, S. Planning, et al. (2000). Eco-Indicator 99, Manual for Designers, A damage oriented method for Life Cycle Impact Assessment 36.

Pantelides, C. C. (1994). Unified frameworks for optimal process planning and scheduling. , D.W.T. Proc. Second Conf. on Foundations of Computer Aided Operations, CACHE Publications

Pati, K. R., P. Vrat, et al. (2008). "A goal programming model for paper recycling system " Omega **36**: 405-417.

Pinto-Varela, T., A. P. Barbosa-Povoa, et al. (2011). "Bi-objective optimization approach to the design and planning of a supply chain: economic versus enviromental performance." Computers & Chemical Engineering: DOI:10.1016/j.compchemeng.2011.03.009

PRéConsultants (2001). Eco-indicator 99 methodology report. PRÉ Consultants b.v. Amersfoort. Netherlands.

Solving a multi-product, multi-depot vehicle routing problem by a hybrid method

Tânia Rodrigues Pereira Ramos[a,c*], *Maria Isabel Gomes*[b], *Ana Paula Barbosa-Póvoa*[c]

[a] *Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal*
tania.ramos@iscte.pt

[b] *CMA – FCT - Universidade Nova de Lisboa, Lisboa, Portugal*
mirg@fct.unl.pt

[c] *CEG-IST – Universidade Técnica de Lisboa, Lisboa, Portugal*
apovoa@ist.utl.pt

[*] *Corresponding Author*

Abstract

This paper is focused on the delimitation of service areas and on vehicle routes definition in recyclable waste collection systems with more than one depot. Three types of materials have to be collected in separated routes, so the problem is modelled as a multi-product, multi-depot vehicle routing problem. A hybrid method is developed where a MIP solver is embedded inside a heuristic framework. The effectiveness of this method is tested by comparing the results obtained for some test instances when solved only by an exact formulation. The hybrid method is then applied to a medium size problem based on a real recyclable waste collection system.

Keywords: multi-depot, vehicle routing, multi-product, hybrid method, recyclable waste collection system.

1- Introduction

Recycling of packaging materials, imposed by the European Union, has forced member states to develop new collection systems. The traditional routes defined for organic waste do not fit the particularities where different vehicles, different collection rates, and different bin locations are required. These circumstances motivated the creation of two different waste collection systems: selective and undifferentiated. The selective collection is intended for recyclable products, while the undifferentiated is intended for non-recyclable products.

A recyclable waste collection system is responsible to collect, within a certain geographic area and on a regular basis, the three types of recyclable materials used in packaging (namely, paper, glass and plastic/metal) which are dropped by the final consumer into special containers. Afterwards, those materials are sorted at sorting stations and delivered to recyclers. Recyclable materials have different collection frequencies, so it is necessary to consider a planning horizon equal to the lowest collection frequency founded.

The present work aims to support tactical decisions for such systems, focusing on the delimitation of service areas of collection systems with more than one depot, and on vehicle routes definition. The vehicle routes definition is considered a tactical decision as the routes are to be maintain for a medium period of time (e.g. 6 to 12 months) due to the stability of the containers fulfilling rates. In this case, we are dealing with static routes and not with dynamic routes. Therefore, the routes defined are to be applied at the operational level, and only need to be revised due to seasonal demands or due to a significant increase in the quantities dropped by the population in the containers. Since we want to establish the service areas, the vehicle routes and we are dealing with three different products to be collected in separated routes, the problem is modelled as a multi-product, multi-depot vehicle routing problem. The problem is solved through a hybrid method where a MIP solver is embedded within a heuristic framework. The effectiveness of this method is tested by comparing the results obtained for some test instances solved only by an exact formulation. The hybrid method is then applied to a medium size problem based on a real recyclable collection system.

This paper is structured as follows. After a brief review of the literature on multi-depot vehicle routing problem (MDVRP) in Section 2, we describe the hybrid method and test its effectiveness in Section 3. The computational results obtained for a medium size problem are presented in Section 4. Finally, we draw conclusions and discuss future work directions.

2- Literature Review

MDVRP is usually defined by a graph $G=(V,A)$, where $V=\{1,...,N+W\}$ is the vertex set and $A=\{(i,j):i,j\in V,i\neq j\}$ is the edge set. The vertex V is partitioned into two subsets $V_c=\{1,...,N\}$ and $V_d=\{N+1,...,N+W\}$, representing the set of cities or customers and the set of depots, respectively. At each depot are based $k\in\{1,...,K\}$ vehicles of capacity Q_k. Each vertex i in V_c has a nonnegative demand p_i and a nonnegative service duration t_i. A distance matrix D = (d_{ij}) is associated with set A.

The MDVRP consists of building a set of vehicle routes in such a way that: (1) each route starts and ends at the same depot, (2) each customer is visited exactly once by a vehicle, (3) the total demand of each route does not exceed the vehicle capacity, (4) the total duration of each route (including travel and service times) does not exceed a preset limit so that (5) the total routing cost is minimized.

For the MDVRP, there are several models developed (exact and approximate approaches). Due to its NP-hard combinatorial nature, the models proposed in the literature are mostly heuristics-based. There are still few exact algorithms in the literature. Laporte et al. (1984), as well as Laporte et al.(1988), developed exact branch and bound algorithms for solving the symmetric and asymmetric version of the MDVRP, respectively. Recently, Baldacci and Mingozzi (2009) developed an exact method for solving the Heterogeneous Vehicle Routing Problem (HVRP) that is capable to solve, among other problems, the MDVRP. This algorithm is based on the set partitioning formulation, where a procedure is applied to generate routes and three bounding procedures are used to reduce the number of variables in the formulation. However, when analysing heuristic algorithms to solve MDVRP several ones have been proposed (Tillman and Cain (1972), Golden, Magnanti and Nguyen (1977), Renaud et al.(1996), Salhi and Sari (1997), Lim and Wang (2005), Crevier et al. (2007), among others).

Therefore we can conclude that few exact models for the multi-depot problems have been proposed, while many heuristic procedures exist for the same problem. The combination of these two methods has also received little attention from the academia. Therefore, this work explores this opportunity and proposes a hybrid method combining

exact formulation with heuristic procedures to solve the multi-product, multi-depot vehicle routing problem.

3- Hybrid Method to Solve the Multi-Product, Multi-Depot VRP

3.1- Description

The main idea of this hybrid method is to relax the original model (Multi-Product, Multi-Depot VRP), and then set decision variables that present feasible values along the procedure.

The first step is then to relax the original Multi-Product, Multi-Depot VRP problem where more than one product is considered and the vehicles are restricted to start and finish at the same depot. This relaxation leads to the Single-Product, Multi-Depot VRP with Multi-Depot Routes, where we have just one product and multi-depot routes are allowed. By solving the first step of the model, we will have some collection sites that belong to a route that start and finish at the same depot and some collection sites that do not. For the "feasible" collection sites we set their assignment to the start and finish depot but not to a particular route or vehicle. After setting these variables the next step of the procedure is to run the Single-Product, Multi-Depot VRP, where all the routes are required to start and end at the same depot.

The first two steps of the algorithm are run for each of the products in study.

After the model Single-Product, Multi-Depot VRP has been run for all the products individually, in the solutions there are some collection sites that belong to routes of the same depot. For those sites, we set their assignment to that depot and then run the Multi-Product, Multi-Depot VRP (which has now the constraint that guarantees that the service areas are defined by depot). This is required because in the original problem, the "Multi-Product" problem, all types of products in each collection site have to be collected from the same depot. In Figure 1 it can be seen a schematic diagram of the hybrid method proposed where the example for three products (paper, glass and plastic/metal) is used.

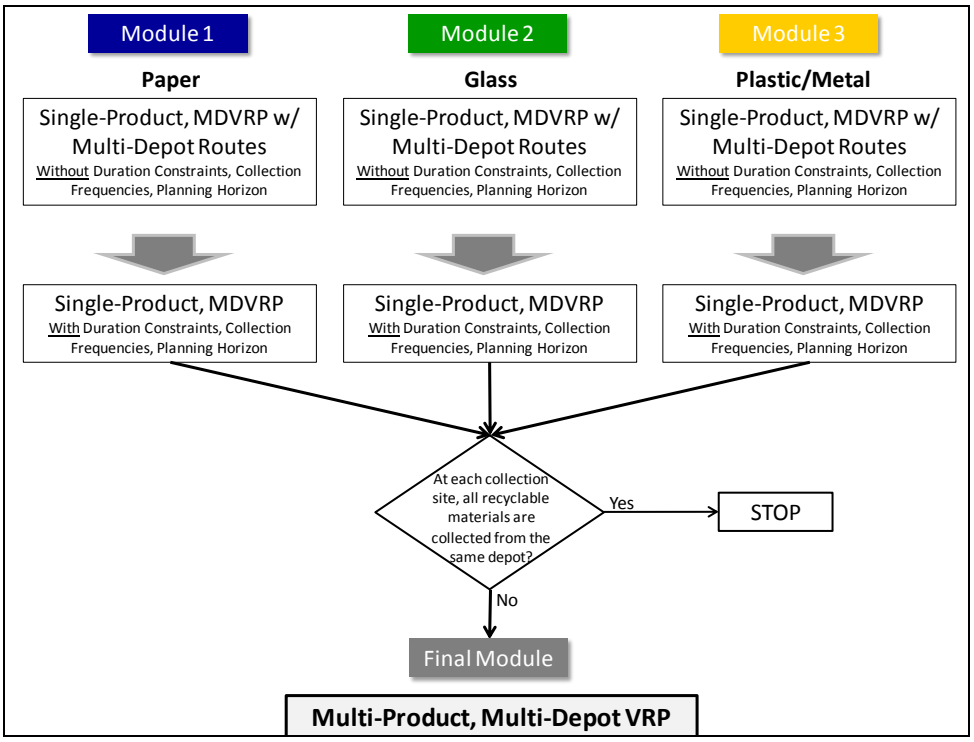


Figure 1 – Hybrid Method Modules

The method proposed is called "hybrid" since it combines exact formulations with the procedure that set the value of some variables. This algorithm solves three mathematical models: 1) Single-Product, Multi-Depot VRP with Multi-Depot Routes, 2) Single-Product, Multi-Depot VRP, 3) Multi-Product, Multi-Depot VRP. The last one is the original problem that has some fixed variables from the results of the first two models (which are relaxations of the original one). All mathematical formulations are based on the two-commodity flow formulation proposed by Baldacci et al. (2004) for the capacitated VRP (CVRP). We adapted Baldacci et al. (2004) formulation to comprise multiple depots and multiple products by adding a binary decision variable z_{ik} that links each collection site i to a vehicle route k, by adding the index m and k to the decision variables x and y, and by adding some constraints. This formulation requires the extend graph $\overline{G} = (\overline{V}, \overline{A})$ obtained from G by adding the vertex set $V_f = \{N+W+1, \ldots, N+2W\}$ which is a replica of the set of depots. Thus, $\overline{V} = V \cup V_f$, $\overline{A} = A \cup \{(i,j) : i \in V_c, j \in V_f\}$ and $d_{i(j+W)} = d_{ji}, i \in V_c, j \in V_d$.

All routes start at one real depot (set V_d) and end in the corresponding replica depot (set V_f). Each route is defined by two flow paths: one path from the real depot to the replica depot, defined by variables y_{ijmk} (representing the vehicle load, which increases along

the route since we are collecting waste); the other path, the reverse one, starts at the replica depot and ends at the real depot, and is defined by variables y_{jimk} (representing the empty space on the vehicle, which decreases along the route). Besides the decision variables y_{ijmk} and y_{jimk}, this formulation has a binary variable x_{ijmk} which represents the routing solution (1, if site j is visited immediately after site i, to collect material m, by vehicle route k; 0, otherwise).

The objective function of the model focuses on minimizing the total distance travelled to collect all recyclable materials at collection sites over the timeframe.

3.2- Effectiveness Testing

In this section, we evaluated the effectiveness of the proposed hybrid method by applying it to a set of test instances and comparing some performance measures (i.e., lower bound, quality of the final solution and computational time) with the optimal solutions from the exact formulation. Due to the complexity of the multi-product, multi-depot vehicle routing problem, the instances that can be solved to optimality will typically be small. Therefore, eight small instances were generated since in the literature there are not available test instances for multi-product problems. The structure of these instances is presented in Table 1.

Table 1- Structure of the Test Instances

Instance	*Nº of Depots*	*Nº of Collection Sites*	*Nº of Recyclable Materials*	*Nº of Vehicles*
1	2	8	3	4
2	2	11	3	4
3	2	11	3	6
4	2	13	3	6
5	2	15	3	4
6	2	16	3	4
7	2	16	3	4
8	2	18	3	4

All the instances are solved by the branch-and-bound algorithm using GAMS/CPLEX Optimizer 12.1.0. An Intel(R) Core(TM) i7 CPU 930 @ 2,80 GHz is used.

The results obtained by applying the branch-and-bound and the hybrid method to the test instances are presented in Table 2.

Table 2 - Comparisons Between the Optimal Value and the Hybrid Method Solution

Instance	Exact Formulation			Hybrid Method			Deviation		
	Lower Bound	Opt. Value	CPU Time (sec)	Lower Bound	Opt. Value	CPU Time (sec)	Lower Bound	Opt.Value	CPU Time
1	589.3	691.1	10	613.7	691.1	2	4.1%	0%	-80%
2	570.1	669.4	104	585.1	682.6	43	2.6%	2%	-59%
3	770.8	872.7	232	788.0	872.7	40	2.2%	0%	-83%
4	802.1	875.1	6115	810.8	875.1	859	1.1%	0%	-86%
5	557.5	656.5	5520	567.2	656.5	487	1.7%	0%	-91%
6	759.2	904.3	987	770.0	904.3	648	1.4%	0%	-34%
7	660.3	750.4	5071	660.3	750.4	726	0%	0%	-86%
8	769.1	946.8	48917	779.2	946.8	2100	1.3%	0%	-96%

Deviation= (Hybrid Method Value/Exact Formulation Value – 1)×100

As mentioned above, three performance measures are compared to assess the effectiveness of the hybrid method: the root node lower bound, the solution quality and the CPU time to compute and prove the optimal solution. We calculate a percentage deviation between the values founded by the exact formulation and by the hybrid method for each of the performance indicators. From Table 2, we can see that, with the exception of instance 7, the root node lower bound provided by the hybrid method is of better quality than the one provided by the exact formulation without compromising the search for the optimal solution (which could happen in other examples). Again, with one exception (instance 2), the hybrid method always finds the optimal solution. For instance 2, the hybrid method finds a solution 2% worse than the optimal one. In terms of computational time, the hybrid method reaches the optimal solution in less time than the exact formulation.

These results show that the hybrid method is an effective method to produce a good or even the optimal solution in less time that the exact formulation.

4- Application to a medium size problem

In this section, the method developed is applied to a medium size problem, with 50 nodes. The nodes location is represented in Figure 2.

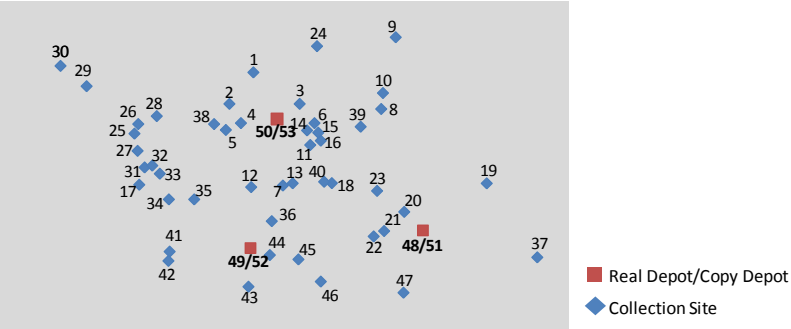


Figure 2 – Location of the Collection Sites and Depots

Since this problem is based on a real case study regarding a recyclable waste collection system, we have three recyclable materials to be considered (Paper, Glass and Plastic/Metal). The recyclable materials have different collection frequencies, so we considered a two-weeks planning horizon, which corresponds to 10 working days. Paper has to be collected four times over the timeframe, Glass has to be collected once and Plastic/Metal twice.

In order to illustrate how the hybrid method works, the detailed results obtained for each module are presented below. As mentioned above, all modules are solved by GAMS/CPLEX Optimizer 12.1.0, being computation time limited to 2 hour for each module.

"Paper" is the first recyclable material for which routes are design by the Single Product, Multi-Depot VRP with Multi-Depot Routes. The results provide five routes, where two are feasible (since they start and end at the same depot) and the other three are infeasible (see Figure 3). Therefore, for sites that are collected by feasible routes their assignment to a depot is fixed. The remaining sites are kept free when the Single-Product, MDVRP is runned. The solution obtained for the Single-Product, MDVRP involves two routes by depot, and the corresponding service areas are showed in Figure 3.

Single-Product, MDVRP with Multi-Depot Routes:

Routes	Feasible for MDVRP?
48-19-37-47-21-22-46-45-44-52	Infeasible
48-20-23-18-40-15-16-14-53	Infeasible
49-34-17-33-32-31-27-26-29-30-28-38-53	Infeasible
49-36-13-7-12-35-41-42-43-52	Feasible
50-3-39-8-10-24-1-2-5-4-53	Feasible

Single-Product, MDVRP:

Routes:

Depot 48/51:

48-19-37-47-51

48-21-22-40-18-23-20-51

Depot 49/52:

49-42-41-17-31-32-33-34-35-12-7-13-36-52

49-43-46-45-44-52

Depot 50/53:

50-1-2-28-30-29-26-27-38-5-4-53

50-14-16-15-39-8-10-24-3-53

Service Areas:

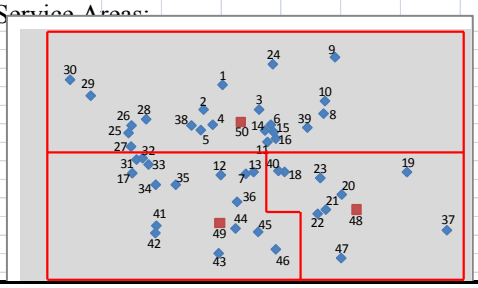


Figure 3 - Results Obtained for Module 1 (Paper)

For material "Glass", three infeasible routes and four feasible routes are the result obtained for the Single-Product, MDVRP with Multi-Depot Routes (Figure 4). When Single-Product, MDVRP is run, one route is defined to depot 48/51, two routes are defined to depot 49/52 and four routes are defined to depot 50/53. The resulting service areas have a different configuration when compared with the service areas defined by the collection of material "Paper".

Single-Product, MDVRP with Multi-Depot Routes:

Routes	Feasible for MDVRP?
49-36-13-7-12-35-34-17-31-32-38-53	Infeasible
49-41-42-43-52	Feasible
49-44-45-46-47-37-19-51	Infeasible
50-1-2-5-4-53	Feasible
50-14-15-40-18-23-22-21-20-51	Infeasible
50-26-25-29-30-28-53	Feasible
50-3-24-9-10-39-16-6-53	Feasible

Single-Product, MDVRP:

Routes:

Depot 48/51:

48-19-37-47-22-21-23-20-51

Depot 49/52:

49-35-34-17-31-32-12-36-44-52

49-45-46-43-42-41-52

Depot 50/53:

50-4-5-38-2-53

50-7-13-40-18-15-16-14-6-3-53

50-26-25-29-30-28-53

50-39-10-9-24-1-53

Service Areas:

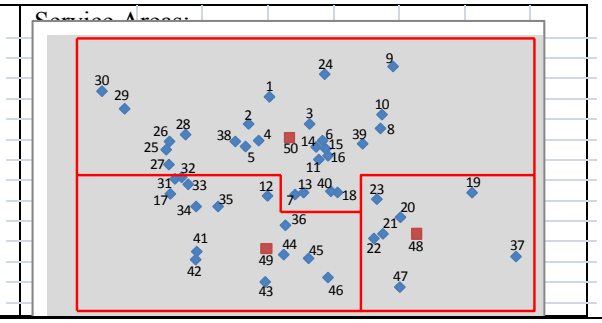


Figure 4 - Results Obtained for Module 2 (Glass)

For material "Plastic/Metal", the Single-Product MDVRP produces only two infeasible routes out of seven routes. When Single-Product, MDVRP is runned, one route is defined to depot 48/51, two routes are defined to depot 49/52 and four routes are

defined to depot 50/53. The resulting service areas have a different configuration when compared to the service areas to collect material "Paper" and "Glass" (Figure 5).

Single-Product, MDVRP with Multi-Depot Routes:

Routes	Feasible for MDVRP?
48-20-23-21-22-47-37-19-51	Feasible
49-35-34-41-42-52	Feasible
49-36-7-13-40-18-15-11-53	Infeasible
49-43-45-44-52	Feasible
50-2-26-29-30-28-1-53	Feasible
50-3-9-10-8-39-6-14-53	Feasible
50-4-38-27-31-32-33-17-52	Infeasible

Single-Product, MDVRP:

Routes:

Depot 48/51:
48-20-23-21-22-47-37-19-51

Depot 49/52:
49-42-41-34-35-36-52
49-43-45-44-52

Depot 50/53:
50-1-28-30-29-26-2-53
50-3-9-10-8-39-6-14-53
50-11-15-18-40-13-7-53
50-33-17-31-32-27-38-4-53

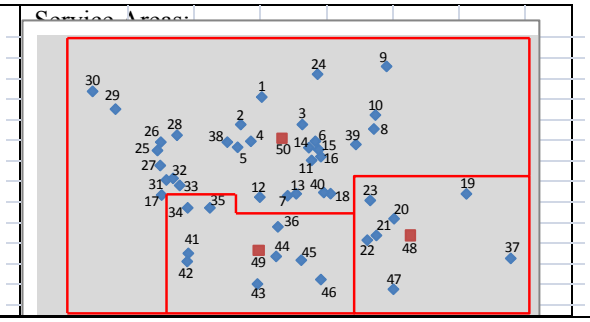


Figure 5 - Results Obtained for Module 3 (Plastic/Metal)

Since the resulting service areas are different among the three recyclable materials, is necessary to run the final module of the hybrid method – the Multi-Product, MDVRP. The collection sites with an infeasible assignment regarding service areas definition are identified – collection sites 7, 12 13, 17, 18, 31, 32, 33 and 40 have not all recyclable materials collected from the same depot. For example, collection site 40 is collected from depot 48 in Paper collection routes, and from depot 50 in Glass and Plastic/Metal collection routes.

The total distance travelled over the timeframe to collect all recyclable materials, where all routes start and end at the same depot and service areas are defined by depot is 2.523 Km. The vehicle routes for each depot are presented on Figure 6 and the corresponding service areas are shown in Figure 7. In this case, the final service areas configuration is equal to the service area configuration provided by module 1. However, this is not always the case in other examples.

Depot 48/51	Depot 49/52	Depot 50/53
Paper 48-19-37-47-51 48-20-23-18-40-22-21-51 Glass 48-19-37-47-21-22-40-18-23-20-51 Plastic/Metal 48-19-37-47-51 48-21-22-40-18-23-20-51	Paper 49-36-13-7-12-35-34-33-32-31-17-41-42-52 49-44-45-46-43-52 Glass 49-35-34-17-31-32-12-7-13-36-44-52 49-45-46-43-42-41-52 Plastic/Metal 49-35-33-32-31-17-34-52 49-36-7-13-45-44-52 49-43-42-41-52	Paper 50-4-5-38-27-26-29-30-28-2-1-53 50-14-16-15-39-8-10-24-3-5 Glass 50-2-38-5-4-53 50-14-15-16-6-3-53 50-26-25-29-30-28-53 50-39-10-9-24-1-53 Plastic/Metal 50-3-9-10-8-39-6-53 50-4-38-1-14-15-11-53 50-28-30-29-26-27-2-53

Figure 6 - Final Routes by Depot

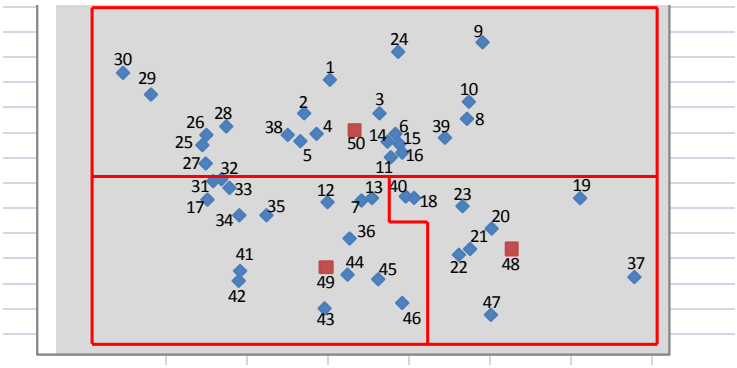


Figure 7 - Final Service Areas

The computational results for the hybrid method are presented in Table 3. We compare the performance of the hybrid method with the performance of the exact formulation for this problem. Within the same CPU time, the hybrid method has found a better solution than the exact formulation for the original problem (2.523 *vs.* 2.638). If the original problem is used to find the same solution provided by the hybrid method, the exact formulation needs 70 hours of CPU time, presenting a gap of 9,8%.

Table 3 – Computational Results Obtained for Hybrid Method and Exact Formulation

Methods	Opt. Value	GAP	CPU (hours)
Hybrid Method			
Module 1 (Paper)			
• Single-Product, MDVRP w/ Multi-Depot Routes	332	0%	0,5
• Single-Product, MDVRP	1350	6,6%	1
Module 2 (Glass)			
• Single-Product, MDVRP w/ Multi-Depot Routes	378	2,9%	1
• Single-Product, MDVRP	388	7,5%	1
Module 3 (Plastic/metal)			
• Single-Product, MDVRP w/ Multi-Depot Routes	366	2,6%	1
• Single-Product, MDVRP	747	5,7%	1
Final Module			
• Multi-Product, MDVRP	2.523	8,9%	2
Exact formulation for the Multi-Product, MDVRP	2.638	14,1%	7,5

5- Conclusions

The present work addresses the multi-product, multi-depot vehicle routing problem often present in a recyclable waste collection system. For medium and large scale problems, exact formulations are not capable to solve the problem since this is a hard combinatorial problem. This triggered off the development of a hybrid method that combines exact formulations with some heuristic procedures. The main idea of this method is to relax some constraints of the problem, and set the variables that do not violate the relaxed constraints. The exact formulations embedded in the hybrid algorithm are based on the two-commodity flow formulation for the Vehicle Routing Problem. The effectiveness of the proposed procedure is tested and shows that it is capable of producing the optimal solution in less time than the exact formulation in seven out of the eight test instances. Since the hybrid method proves to be effective, it is applied to a medium scale problem based on a real recyclable waste collection system. The detailed results for each module of the hybrid method are showed.

As further work, we intend to apply the proposed method to a larger problem, as well as incorporate a procedure to schedule the vehicle routes within the timeframe. Efforts to improve exact formulations performance will also be developed.

References

Baldacci, R., Hadjiconstantinou, E. and Mingozzi, A. (2004). An exact algorithm for the capacitated vehicle routing problem based on a two-commodity network flow formulation. *Operations Research,* 52**,** 723-738.

Baldacci, R., Mingozzi, A. (2009). A unified exact method for solving different classes of vehicle routing problems. *Mathematical Programming,* 120**,** 347-380.

Crevier, B., Cordeau, J. F. and Laporte, G. (2007). The multi-depot vehicle routing problem with inter-depot routes. *European Journal of Operational Research,* 176**,** 756-773.

Golden, B. L., Magnanti, T. L. and Nguyen, H. Q. (1977). Implementing Vehicle Routing Algorithms. *Networks,* 7**,** 113-148.

Laporte, G., Nobert, Y. and Arpin, D. (1984). Optimal solutions to capacitated multi-depot vehicle routing problems. *Congressus Numerantium,* 44**,** 283-292.

Laporte, G., Nobert, Y. and Taillefer, S. (1988). Solving a Family of Multi-Depot Vehicle-Routing and Location-Routing Problems. *Transportation Science,* 22**,** 161-172.

Lim, A., Wang, F. (2005). Multi-depot vehicle routing problem: A one-stage approach. *Ieee Transactions on Automation Science and Engineering,* 2**,** 397-402.

Renaud, J., Laporte, G. and Boctor, F. F. (1996). A tabu search heuristic for the multi-depot vehicle routing problem. *Computers & Operations Research,* 23**,** 229-235.

Salhi, S., Sari, M. (1997). A multi-level composite heuristic for the multi-depot vehicle fleet mix problem. *European Journal of Operational Research,* 103**,** 95-112.

Tillman, F. A., Cain, T. M. (1972). Upperbound Algorithem for Single and Multiple Terminal Delivery Problem. *Management Science Series a-Theory,* 18**,** 664-682.

Modelação e optimização de cadeias de abastecimento de biomassa florestal

Tiago Costa Gomes[a*], *Filipe Pereira e Alvelos*[b], *Maria Sameiro Carvalho*[b]

[a]*Universidade do Minho, Braga, Portugal*
tiago.gomes@dps.uminho.pt

[b] *Departamento de Produção e Sistemas, Universidade do Minho, Braga, Portugal*
falvelos@dps.uminho.pt
sameiro@dps.uminho.pt

[*]*Autor para correspondência*

Resumo

Neste artigo, é proposto um modelo de programação inteira mista que integra decisões tácticas e operacionais para problemas de cadeias de abastecimento de biomassa. Considera-se uma empresa que faz a recolha de recursos florestais e os entrega a clientes, após a sua transformação em estilha. O objectivo da empresa é entregar a quantidade de produtos desejada no momento pretendido aos clientes, minimizando o custo total da cadeia. São apresentados resultados de testes computacionais em instâncias baseadas em dados da literatura. O modelo proposto permite a obtenção de soluções de qualidade num tempo aceitável em instâncias de pequena e média dimensão e em algumas de grande dimensão.

Palavras-chave: programação inteira mista, cadeias de abastecimento, biomassa

1 - Introdução

Um grande número de países estão empenhados em aumentar o consumo de energias renováveis e amigas do ambiente, para reduzir a dependência de energia fóssil e as emissões de gases com efeito de estufa [European Union, 2006; United Nations, 1998]. Em 2006, a Comissão Europeia definiu objectivos energéticos ambiciosos para 2020: "reduzir as emissões de gases com efeito de estufa em 20%, aumentar a quota das energias renováveis para 20% e melhorar 20% na eficiência energética".

Embora existam vários tipos de fontes de energia renováveis, neste trabalho vamos concentrar-nos especificamente na biomassa. A biomassa é um termo geral que incorpora material biológico como a madeira (resíduos da colheita de árvores, resíduos

da industria de madeira, serrações, resíduos da construção, etc) ou os resíduos biodegradáveis [EUBIONET III and Alakangas, 2010]. A biomassa de madeira pode ser incluída em várias actividades económicas, tais como: produção de produtos derivados de madeira, aquecimento, geração de electricidade ou no sector dos transportes.

O plano de acção para a biomassa [European Union, 2005] encoraja a implementação de medidas na promoção da indústria de biomassa. A biomassa florestal é considerada uma energia limpa e renovável com um ciclo de carbono neutro: as árvores capturam o CO_2 da atmosfera; o combustível florestal é queimado e o CO_2 é libertado para a atmosfera. Esse ciclo repete-se sucessivamente e por esta razão, a queima de biomassa pode não causar mais emissões de gases com efeito estufa. Em [Johnson, 2009] é realçada a necessidade de estabelecer um equilíbrio entre actividades de desflorestação e reflorestação para manter o equilíbrio entre as emissões e armazenamento de CO_2.

A energia da biomassa tem sido considerada uma alternativa aos combustíveis fósseis, mas a sua competitividade é altamente dependente da eficiência da cadeia de abastecimento, visto que os custos de transporte representam aproximadamente 50% dos custos totais, em alguns casos podem chegar a 65%.

A cadeia da biomassa tem características especiais que a tornam mais complexa que a generalidade de cadeias de abastecimento. A biomassa tem a oferta e a procura sazonal, está dispersa por vastas áreas, é sensível ao clima, tem valores de densidade baixa e pode apresentar teores de humidade variável [Sokhansanj et al., 2006].

Uma cadeia de abastecimento de biomassa eficiente deve optimizar toda a cadeia, equilibrar os custos e o nível de serviço, de forma a alcançar altos níveis de desempenho, pelo que o recurso a sistemas de apoio à decisão é essencial no processo de análise e optimização de um sistema tão complexo.

Neste trabalho, é desenvolvido um modelo de programação inteira mista capaz de incorporar decisões tácticas e operacionais da cadeia de abastecimento de biomassa. Enquanto minimizamos os custos totais do sistema, esta abordagem tenta incorporar algumas das especificidades e complexidades destes sistemas logísticos.

Na secção 2, é feita uma revisão da literatura relacionada com a abordagem aqui proposta e descrito o problema abordado. Na secção 3, é apresentado em detalhe o modelo de programação inteira mista proposto. Na secção são apresentados resultados

computacionais. Por fim, na secção 5, são discutidas as principais conclusões deste trabalho.

2 - Descrição do problema

Em termos de cadeias de abastecimento de biomassa, o número de publicações científicas tem vindo a aumentar, devido a uma maior preocupação sobre os recursos energéticos renováveis. [Iakovou et al., 2010] apresenta as características especiais das cadeias de abastecimento de resíduos de biomassa e salienta que a logística e a gestão de cadeias são áreas de importância crítica para uma utilização bem sucedida da biomassa. Além disso, neste trabalho, o autor salienta a necessidade de abordagens abrangentes e integradoras e que a maioria dos trabalhos publicados foca apenas um aspecto particular do processo de decisão havendo ainda muito trabalho a desenvolver para integrar toda a complexidade dos sistemas de biomassa no mesmo modelo de decisão.

Um modelo de programação inteira mista de cadeia de abastecimento de biomassa foi proposto em [Gunnarsson et al., 2004] que integra decisões estratégicas e tácticas. O problema abordado é multi-período (base mensal e com um horizonte de planeamento de um ano), multi-produto, com instalações com limites de capacidade, definido sobre uma estrutura de dois níveis: os fornecedores (áreas de recolha, serrações, fornecedores externos), terminais (para armazenamento) e clientes (centrais termoeléctricas) com a procura conhecida de combustível florestal. As decisões relativas à escolha dos fornecedores e dos terminais, o momento e a localização de onde estilhar, as quantidades de armazenamento e os fluxos entre as instalações são abordadas. O modelo matemático é resolvido através de uma abordagem heurística.

O modelo proposto neste artigo é uma extensão do modelo de [Gunnarsson et al., 2004]. A principal diferença diz respeito à inclusão do nível de decisão operacional e uma escala de tempo de base diária é usada para incorporar as operações de transporte (por exemplo, quantos veículos são necessários) e operações de processamento (alocação de recursos para recolha e operações de estilha). Além disso, como os custos de transporte constituem uma das fracções mais importante dos custos totais, uma abordagem mais realista é adoptada: os custos de transporte não são uma função linear de unidades transportadas, mas uma função em degrau onde cada nível corresponde a um veículo, permitindo assim uma maior aproximação aos custos reais.

Na construção do modelo considerou-se uma empresa que faz a recolha de recursos florestais e os entrega a clientes, após a sua transformação em estilha ("sub-peças em forma rectangular com um tamanho de partícula definido, produzido por tratamento mecânico, geralmente, de facas" [EUBIONET III and Alakangas, 2010]). O objectivo da empresa é entregar a quantidade de produtos desejada no momento pretendido aos clientes, minimizando o custo total da cadeia. A estilha pode ser transportada directamente das florestas para os clientes, mas também é possível usar armazéns intermédios para minorar os efeitos da sazonalidade. Os armazéns intermédios também podem receber produtos por estilhar, sendo essa operação executada lá posteriormente.

Um aspecto importante do problema é a gestão optimizada dos equipamentos e recursos humanos envolvidos nas operações de recolher, estilhar, transportar e armazenar a biomassa. Assumiu-se, tal como referido anteriormente, que a duração de todas as operações é um dia.

Para um horizonte de planeamento definido, em períodos de tempo discretizados, o modelo permite as seguintes decisões: selecção de áreas de recolha; localização de armazéns (se necessários); onde estilhar os produtos (uso ou não de estilhadores na floresta e em armazéns, localização de estilhadores fixos e móveis); decisões de armazenamento (quantidades a serem armazenados em todos os períodos de tempo, em diferentes armazéns) e fluxos de transporte entre diferentes elos da cadeia.

3 - Modelo matemático

Nesta secção descreve-se em pormenor o modelo de programação inteira mista desenvolvido que permite a optimização do problema descrito na secção anterior. Na concepção do modelo definiram-se os seguintes conjuntos: I – conjunto das áreas florestais, indexadas por i; W - conjunto dos potenciais armazéns, indexados por w; J – conjunto dos clientes, indexados por j; P – conjunto dos produtos, indexados por p; T – conjunto dos períodos de tempo, indexados por t.

3.1 - Parâmetros

Usou-se uma notação para os parâmetros baseada em três letras Agh. A letra A representa o tipo de parâmetro e pode ter os 5 significados seguintes: N – número; V – volume; S – custo de armazenagem; C – custo excepto os de armazenagem; D – distância. As letras g e h representam entidades associadas ao tipo de parâmetro A,

podendo ter 7 significados: *m* – estilhador móvel; *f* – estilhador fixo; c – produto estilhado; *n* – produto não estilhado; *q* – custo fixo; *v* – custo variável; *k* – camião.

Para o tipo de parâmetro *N* definiu-se: *Nm* - número de estilhadores móveis disponíveis; *Nf* – número máximo de estilhadores fixos e *Ni* – número máximo de equipas de recolhas.

Para o tipo de parâmetro *V* definiu-se: V_i^p- volume da oferta do produto *p* na área *i*; V_i^{pt}- volume de procura do cliente *j* do produto *p* no período *t*; *Vc*- capacidade de cada armazém de produto estilhado; *Vn* - capacidade de cada armazém de produto não estilhado; *Vm* - capacidade de cada estilhador móvel; *Vf* - capacidade de cada estilhador fixo; *Vi* - capacidade de recolha de cada equipa; *Vk* - capacidade de cada camião.

Para o tipo de parâmetro *C* definiu-se: *Ciq* - custo fixo de recolha de qualquer produto em qualquer área num período; *Civ* - custo variável de recolha de qualquer produto em qualquer área num período; *Cmq* - custo fixo de usar um estilhador móvel num período; *Cmv* - custo variável de usar um estilhador móvel num período; *Cfq* - custo fixo de instalar um estilhador fixo num armazém; *Cfv* - Custo variável de usar um estilhador fixo num armazém (proporcional à quantidade estilhada); *Ck* - custo por quilómetro de transporte.

Para o tipo de parâmetro *S* definiu-se: *Siv* - Custo variável de armazenar produto não estilhado na floresta por período e quantidade armazenada; *Snv* - Custo variável de armazenar produto não estilhado em armazém por período e quantidade armazenada; *Scv* - Custo variável de armazenar produto estilhado em armazém por período e quantidade armazenada; *Swq* - custo fixo por usar um armazém.

Para o tipo de parâmetro *D* definiu-se: d_{ij} - distância entre área *i* e cliente *j*; d_{iw} - distância entre área *i* e o armazém *w*; d_{wj} - distância entre armazém *w* e cliente *j*.

3.2 – Variáveis de decisão

As variáveis de decisão podem ser divididas em três tipos: *X* - variáveis contínuas; *Y* - variáveis binárias; *Z* - variáveis inteiras.

As variáveis contínuas indicam o volume de produtos recolhidos, estilhados, armazenados e transportados de cada produto *p* em cada período *t* e podem ser divididas nesses quatro grupos. Para o volume transportado foram definidas quatro conjuntos de variáveis: Xn_{iw}^{pt} - produto não estilhado entre uma floresta e um armazém; Xc_{iw}^{pt} - produto estilhado entre uma floresta e um armazém; Xc_{wj}^{pt} - produto estilhado entre um armazém

e um cliente; Xc_{ij}^{pt} - produto estilhado entre uma floresta e um cliente. Para o volume estilhado foram definidos três conjuntos de variáveis: Xm_i^{pt} - na floresta com um estilhador móvel; Xm_w^{pt} - no armazém com estilhador móvel; Xf_w^{pt} - no armazém com estilhador fixo. Para o volume de existências em armazém definiram-se três conjuntos de variáveis: Xn_i^{pt} - produto não estilhado na floresta; Xn_w^{pt} - produto não estilhado em armazém; Xc_w^{pt} - produto estilhado em armazém. Para o volume de produto recolhido em cada área foi definido X_i^{pt}.

As variáveis binárias permitem a tomada de decisões sobre o uso de certos recursos, neste caso definem-se: $Y_i^t = 1$ se na área florestal *i* é realizada recolha no período *t*, $Y_i^t = 0$, caso contrário; $Y_w = 1$, se o armazém intermediário *w* é usado, $Y_w = 0$, caso contrário; $Yf_w = 1,$ se é instalado um estilhador fixo no armazém *w*, $Yf_w = 0,$ caso contrário.

As variáveis inteiras são: Zm_i^t e Zm_w^t - número de estilhadores móveis usados em cada período em cada área e em cada armazém, respectivamente. As seguintes variáveis indicam o número de viagens necessárias para o transporte em cada período e de cada produto: Zn_{iw}^{pt} - os produtos não estilhados entre as áreas florestais e os clientes; Zc_{iw}^{pt} - os produtos estilhados entre as áreas florestais e os armazéns; Zc_{wj}^{pt} - os produtos estilhados entre armazéns e clientes; Zc_{ij}^{pt} - os produtos estilhados entre áreas florestais e clientes.

Para o cálculo dos custos de transporte usou-se uma estrutura de custo em degrau: a capacidade do veículo define o patamar do degrau, permitindo determinar o número de viagens necessárias para transportar a totalidade dos produtos. O custo é determinado pela multiplicação do número de viagens pela distância entre a origem e o destino.

3.3 - Função objectivo

A função objectivo é constituída pela soma dos custos fixos e variáveis das operações de recolher, estilhar, armazenar e transportar, visando a minimização do custo total.

$$Min\ Z = Ch + Cc + Cs + Ct \tag{1}$$

$$Ch = \sum_{i\in I}\sum_{t\in T} Ciq\ Y_i^t + \sum_{i\in I}\sum_{p\in P}\sum_{t\in T} Civ\ X_i^{pt} \tag{2}$$

$$Cc = \sum_{i\in I}\sum_{t\in T} Cmq\ Zm_i^t + \sum_{w\in W}\sum_{t\in T} Cmq\ Zm_w^t + \sum_{i\in I}\sum_{p\in P}\sum_{t\in T} Cmv\ Xm_i^{pt} + \sum_{w\in W}\sum_{p\in P}\sum_{t\in T} Cmv\ Xm_w^{pt} + \sum_{w\in W} Cfq\ Yf_w + \sum_{w\in W}\sum_{p\in P}\sum_{t\in T} Cfv\ Xf_w^{pt} \tag{3}$$

$$Cs = \sum_{i\in I}\sum_{p\in P}\sum_{t\in T} Siv\ Xf_i^{pt} + \sum_{w\in W}\sum_{p\in P}\sum_{t\in T} Scv\ Xc_w^{pt} + \sum_{w\in W}\sum_{p\in P}\sum_{t\in T} Snv\ Xn_w^{pt} + \sum_{w\in W} Swq\ Y_w \quad (4)$$

$$Ct = \sum_{i\in I}\sum_{w\in W}\sum_{p\in P}\sum Ck\ d_{iw}\ Zn_{iw}^{pt} + \sum_{i\in I}\sum_{w\in W}\sum_{p\in P}\sum_{t\in T} Ck\ d_{iw}\ Zc_{iw}^{pt} + \sum_{w\in W}\sum_{j\in J}\sum_{p\in P}\sum_{t\in T} Ck\ d_{wj}\ Zc_{wj}^{pt} + \sum_{i\in I}\sum_{j\in J}\sum_{p\in P}\sum_{t\in T} Ck\ d_{ij}\ Zc_{ij}^{pt} \quad (5)$$

A função objectivo (1) tem quatro parcelas (2-5) associadas aos diferentes tipos de custos.

Os custos de recolher (2) estão divididos em dois componentes. O primeiro corresponde aos custos fixos e o segundo aos variáveis.

O custo de estilhar (3) está dividido em seis componentes. O primeiro e o segundo referentes aos custos fixos de utilizar cada um dos estilhadores móveis; o terceiro e quarto referentes ao custo de estilhar cada unidade de volume de produto através dos estilhadores móveis; o quinto refere-se ao custo de instalar estilhadores fixos em armazém e o último refere-se ao custo variável da utilização dos estilhadores fixos.

O custo de armazenar (4) tem quatro componentes. O último refere-se ao custo fixo de abertura de um armazém intermédio. Os primeiros três referem-se aos custos variáveis na floresta, nos armazéns de produto estilhado e não estilhado respectivamente.

O custo de transporte (5) tem quatro componentes. O primeiro e segundo referem-se ao custo do transporte entre as florestas e os armazéns intermédios, no primeiro caso, produto não estilhado e no segundo, produto estilhado. O terceiro refere-se ao transporte entre armazéns intermédios e clientes e o último entre as florestas e os clientes.

3.4 - Restrições

O problema está sujeito a uma série de restrições, que são detalhadas abaixo:

As restrições (6) e (7) garantem que o material recolhido não excede a oferta existente em cada área de cada produto.

$$X_i^{pt} \le V_i^p\ Y_i^t\ , \forall i \in I, \forall p \in P, \forall t \in T \quad (6)$$

$$\sum_{t\in T} X_i^{pt} \le V_i^p, \forall i \in I, \forall p \in P \quad (7)$$

$$\sum_{i\in I} Y_i^t \le Ni\ \ , \forall t \in T \quad (8)$$

$$\sum_{p\in P} X_i^{pt} \le Vi\ \ Y_i^t, \forall i \in I, \forall t \in T \quad (9)$$

Como existe um número limitado de equipas de recolha e em cada área florestal só é permitido o trabalho de uma equipa, é necessário restringir o número de equipas usadas simultaneamente. Este aspecto está salvaguardado na restrição (8). A restrição (9) limita a capacidade de recolha.

$$Xn_i^{pt} + X_i^{pt} = Xn_i^{p,t+1} + Xm_i^{p,t+1} + \sum_{w \in W} Xn_{iw}^{p,t+1}, \forall i \in I, \forall p \in P, \forall t = 1, \dots, T-1 \quad (10)$$

$$Xm_i^{pt} = \sum_{w \in W} Xc_{iw}^{p,t+1} + \sum_{j \in J} Xc_{ij}^{p,t+1}, \forall i \in I, \forall p \in P, \forall t = 1, \dots, T-1 \quad (11)$$

$$\sum_{i \in I} Zm_i^t + \sum_{w \in W} Zm_w^t \leq Nm, \forall t \in T \quad (12)$$

$$\sum_{p \in P} Xm_i^{pt} \leq Vm\, Zm_i^t, \forall i \in I, \forall t \in T \quad (13)$$

$$\sum_{p \in P} Xm_w^{pt} \leq Vm\, Zm_w^t, \forall w \in W, \forall t \in T \quad (14)$$

$$\sum_{p \in P} Xf_w^{pt} \leq Vf\, Yf_w, \forall w \in W, \forall t \in T \quad (15)$$

$$\sum_{w \in W} Yf_w \leq Nf \quad (16)$$

$$Yf_w \leq Y_w, \forall w \in W \quad (17)$$

$$Zm_w^t \leq Nm\ \ Y_w, \forall w \in W, \forall t \in T \quad (18)$$

$$Xn_w^{pt} + \sum_{i \in I} Xn_{iw}^{pt} = Xn_w^{p,t+1} + Xm_w^{p,t+1} + Xf_w^{p,t+1}, \forall w \in W, \forall p \in P, \forall t = 1, \dots, T-1 \quad (19)$$

$$Xc_w^{pt} + \sum_{i \in I} Xc_{iw}^{pt} + Xm_w^{pt} + Xf_w^{pt} = Xc_w^{p,t+1} + \sum_{j \in J} Xc_{wj}^{p,t+1}, \forall w \in W, \forall p \in P, \forall t = 1, \dots, T-1 \quad (20)$$

As restrições (10), (19) e (20) permitem a conservação de fluxo nos armazéns. Nestas equações, o lado esquerdo representa o que está no armazém no início do período adicionado do fluxo de entrada durante o período; o lado direito representa o que está em armazém no final do período adicionado ao fluxo de saída durante o período.

A restrição (10) refere-se ao armazém de produto não estilhado nas florestas, onde o fluxo de entrada é o volume de produto recolhido, o fluxo de saída é o volume do produto estilhado na floresta e o volume do produto não estilhado transportado para armazéns intermédios e a (11) garante que o material estilhado na floresta é transportado para os clientes ou para armazéns intermédios.

Considerou-se um número limitado de estilhadores móveis, todos com a mesma capacidade, que podem ser utilizados nas florestas ou armazéns intermédios. A restrição

(12) garante que em cada período o número de estilhadores utilizados não seja ultrapassado. As restrições (13), (14) e (15) limitam a capacidade de estilhar, a (13) para estilhadores móveis na floresta, a (14) para o estilhadores móveis em armazéns e a (15) para estilhadores fixos.

A restrição (16) limita o número de estilhadores fixos em armazéns e as (17) e (18) permitem estilhar nos armazéns, apenas se estes forem utilizados. A (17) para os estilhadores fixos e a (18) para os móveis.

A restrição (19) refere-se ao armazém de produto não estilhado em armazéns intermédios, onde o fluxo de entrada é o volume de produto por estilhar recepcionado das florestas, o fluxo de saída é o volume de produtos estilhados nesse período através de estilhadores fixos e móveis. A restrição (20) refere-se ao armazém de produto estilhado em armazéns intermédios, onde o fluxo de entrada é o volume de produto estilhados recepcionado das florestas mais o volume de produtos estilhados nesse período através de estilhadores fixos e móveis, o fluxo de saída é o volume de produtos transportados para os clientes no período.

$$\sum_{p\in P} Xc_w^{pt} + \sum_{p\in P}\sum_{j\in J} Xc_{wj}^{pt} \leq Vc_w\ Y_w\ , \forall w \in W, \forall t \in T \quad (21)$$

$$\sum_{p\in P} Xn_w^{pt} + \sum_{p\in P} Xm_w^{pt} + \sum_{p\in P} Xf_w^{pt} \leq Vn_w\ Y_w\ , \forall w \in W, \forall t \in T \quad (22)$$

$$\sum_{w\in W} Xc_{wj}^{pt} + \sum_{i\in I} Xc_{ij}^{pt} = V_j^{p,t+1}, \forall j \in J, \forall p \in P, \forall t \in T \quad (23)$$

As restrições (21) e (22) limitam a capacidade dos armazéns. A (21) para os produtos estilhados e a (22) para produtos não estilhados.

A (23) garante que a procura dos clientes é satisfeita.

$$Xn_{iw}^{pt} \leq Vk\ Zn_{iw}^{pt}\ , \forall i \in I, \forall w \in W, \forall p \in P, \forall t \in T \quad (24)$$

$$Xc_{iw}^{pt} \leq Vk\ Zc_{iw}^{pt}\ , \forall i \in I, \forall w \in W, \forall p \in P, \forall t \in T \quad (25)$$

$$Xc_{wj}^{pt} \leq Vk\ Zc_{wj}^{pt}\ , \forall w \in W, \forall j \in J, \forall p \in P, \forall t \in T \quad (26)$$

$$Xc_{ij}^{pt} \leq Vk\ Zc_{ij}^{pt}\ , \forall i \in I, \forall j \in J, \forall p \in P, \forall t \in T \quad (27)$$

As restrições de (24) a (27) limitam a capacidade de transporte.

Existem ainda restrições que garantem as condições iniciais do modelo (por exemplo, a quantidade de produtos estilhados no primeiro período é nula) e o domínio das variáveis de decisão de acordo com a sua definição apresentada previamente.

4 - Resultados

O modelo matemático foi implementado em C + + e utilizou-se o Cplex 12.1 para obter soluções. O Cplex usa *branch-and-cut* para a resolução de modelos de programação inteira mista (MIP). O computador utilizado para realizar os testes foi um processador Intel ® Core ™ 2 Duo de 2.00GHz com 2.00GB de RAM.

Para construir as instâncias usadas nos testes utilizou-se informação contida em [EUBIONET II, 2007] para que os custos, as restrições tecnológicas, oferta e procura, estejam as mais próximas possíveis da realidade. A procura, a oferta e as localizações foram gerados aleatoriamente, com base em informações do mesmo relatório.

Os conjuntos de parâmetros que influenciam o tamanho do modelo foram variados, para verificar o seu comportamento para instâncias pequenas e de dimensão mais próximas da realidade. Desta forma, testou-se o modelo com 5 e 150 áreas de florestais, com 1 e 4 armazéns potenciais, com 3 e 7 clientes, com 1 e 4 produtos, com 7, 30, 90 e 150 períodos de tempo. Na prática, um período de tempo está associado a um dia.

De seguida apresentam-se duas tabelas com os resultados computacionais obtidos. Cada tabela é composta por quatro colunas (esquerda) que indicam quatro dos parâmetros variados do modelo (|I| - número de florestas, |*W*| - número de potenciais armazéns, |*J*| - número de clientes, |P| - número de produtos) e dois conjuntos de seis colunas, referentes a dois períodos de tempo. Para cada período de tempo as primeiras três colunas correspondem ao tamanho do modelo (*Linhas* - número de restrições, *Colunas* - número de variáveis, *Não-zeros* - o número de elementos não negativos da matriz associada às restrições) e as últimas três referentes aos resultados obtidos (*Z* - melhor solução inteira, $GAP = 100\ |Z\text{-}LI| / |Z|\ \%$ em que Z corresponde ao valor da melhor solução encontrada e *LI* corresponde ao limite inferior para o valor da solução óptima dado pelo *Cplex*. *Tempo* - o tempo de execução em segundos). Note-se que, por exemplo, um GAP de 0.04% significa que a solução obtida tem um custo superior à solução óptima em não mais de 0.04%. Definiram-se dois critérios de paragem, um relativamente ao tempo máximo de execução (3600 segundos) e o outro referente ao GAP (10^{-4}).

Para a maioria das instâncias, a aplicação foi interrompida por atingir o limite de tempo de execução. Quando na coluna de Z aparece o símbolo * ou ** significa que não se

obteve uma solução inteira ou existiu um erro devido à falta de memória do Cplex, respectivamente.

Tabela 1 – Resultados para as instâncias com o período de tempo igual a 7 e 30

|I|	|W|	|J|	|P|	|T|											
				7						30					
				Linhas	Colunas	Não-zeros	Z	GA P	Tempo	Linhas	Colunas	Não-zeros	Z	GA P	Tempo
5	1	3	1	513	604	1,445	206,880	0.01%	122	2,008	2,582	6,183	842,725	0.17%	3,600
5	1	3	4	1,689	2,179	5,204	597,071	0.04%	3,600	6,496	9,332	22,293	3,584,108	0.30%	3,600
5	1	7	1	757	940	2,173	406,504	0.06%	3,600	2,896	4,022	9,211	1,938,850	0.25%	3,600
5	1	7	4	2,665	3,523	8,116	1,477,708	0.02%	3,600	10,048	15,092	34,405	6,755,222	0.41%	3,600
5	4	3	1	1,029	1,261	3,050	209,389	0.01%	131	3,904	5,378	13,032	803,498	0.28%	3,602
5	4	3	4	3,429	4,726	11,156	587,944	0.04%	3,600	12,721	20,228	47,772	3,453,325	0.59%	3,601
5	4	7	1	1,381	1,765	4,222	409,068	0.01%	3,600	5,176	7,538	17,884	1,943,465	0.24%	3,608
5	4	7	4	4,837	6,742	15,844	1,489,029	0.01%	3,600	17,809	28,868	67,180	8,066,747	0.18%	3,601
150	1	3	1	11,968	15,829	35,810	198,940	4.48%	3,604	46,813	67,832	153,938	851,028	0.59%	3,603
150	1	3	4	41,419	56,989	130,484	687,353	0.75%	3,602	159,616	244,232	561,113	3,926,461	7.97%	3,607
150	1	7	1	17,432	24,285	53,358	333,517	0.87%	3,605	66,261	104,072	227,146	2,053,567	1.79%	3,604
150	1	7	4	63,275	90,813	200,676	1,546,040	1.60%	3,605	237,408	389,192	853,945	8,952,701	3.28%	3,609
150	4	3	1	20,314	28,666	61,340	195,933	3.48%	3,601	76,549	122,828	264,752	837,781	1.85%	3,604
150	4	3	4	74,479	108,256	232,136	687,189	1.05%	3,604	277,201	463,928	1,002,452	*		
150	4	7	1	25,886	37,290	79,332	333,042	0.82%	3,602	96,381	159,788	339,784	1,967,425	1.32%	3,605
150	4	7	4	96,767	142,752	304,104	1,535,634	1.59%	3,605	356,529	611,768	1,302,580	8,952,701	3.28%	3,600

Para as instâncias com menos períodos (Tabela 1) foi quase sempre possível obter soluções de qualidade no tempo muito aceitável de uma hora. A excepção foi a 14ª instância. Note-se que metade destas instâncias têm 150 áreas florestais.

Tabela 2 - Resultados para as instâncias com o período de tempo igual a 90 e 150

|I|	|W|	|J|	|P|	|T|											
				90						150					
				Linhas	Colunas	Não-zeros	Z	GA P	Tempo	Linhas	Colunas	Não-zeros	Z	GA P	Tempo
5	1	3	1	5,908	7,742	18,543	2,561,156	0.24%	3,600	9,808	12,902	30,903	4,815,603	0.80%	3,601
5	1	3	4	19,036	27,992	66,873	12,048,978	0.23%	3,601	31,576	46,652	111,453	19,092,615	0.32%	3,602
5	1	7	1	8,476	12,062	27,571	6,755,222	0.41%	3,600	14,056	20,102	45,931	11,304,564	0.44%	3,601
5	1	7	4	29,308	45,272	102,985	26,472,796	0.15%	3,601	48,568	75,452	171,565	43,780,685	0.16%	3,603
5	4	3	1	11,404	16,118	39,072	2,600,874	0.30%	3,601	18,904	26,858	65,112	4,315,313	0.72%	3,603
5	4	3	4	36,961	60,668	143,292	10,975,805	0.30%	3,605	61,201	101,108	238,812	17,599,099	0.28%	3,604
5	4	7	1	15,076	22,598	53,524	6,175,014	0.27%	3,601	24,976	37,658	89,164	10,324,530	0.34%	3,600
5	4	7	4	51,649	86,588	201,100	25,656,848	0.32%	3,605	85,489	144,308	335,020	42,573,105	0.32%	3,607
150	1	3	1	137,713	203,492	462,098	2,787,543	1.01%	3,628	228,613	339,152	770,258	4,874,245	3.34%	3,612
150	1	3	4	467,956	732,692	1,684,493	**			776,296	1,221,152	2,807,873	**		
150	1	7	1	193,641	312,212	680,506	*			321,021	520,352	1,133,866	**		
150	1	7	4	691,668	1,167,572	2,558,125	**			1,145,928	1,945,952	4,262,305	**		
150	4	3	1	223,249	368,468	795,392	2,758,312	3.57%	3,612	369,949	614,108	1,326,032	*		
150	4	3	4	806,041	1,391,768	3,011,972	**			1,334,881	2,319,608	5,021,492	**		
150	4	7	1	280,281	479,348	1,019,224	6,487,737	4.75%	3,621	464,181	798,908	1,698,664	**		
150	4	7	4	1,034,169	1,835,288	3,907,300	**			1,711,809	3,058,808	6,512,020	**		

Para as instâncias com mais períodos (Tabela 2), para as instâncias com menos áreas florestais foi sempre possível chegar a soluções de qualidade. Para as instâncias com 150 áreas florestais, com 90 períodos foi ainda possível chegar a algumas soluções de qualidade no tempo disponível. Para as instâncias com 150 períodos tal já não se verificou, devido ao modelo ter uma dimensão inviável para o *Cplex*.

5- Conclusões

Neste artigo foi proposto um modelo de programação inteira mista que apoia decisões tácticas e operacionais, de modo integrado, de toda a cadeia de abastecimento de biomassa florestal. O modelo foi implementado em C++, utilizou-se o Cplex 12.1 para obter soluções e foram realizados testes computacionais para 64 instâncias baseadas em informação encontrada na literatura, com 5 e 150 áreas de florestais, com 1 e 4 armazéns potenciais, com 3 e 7 clientes, com 1 e 4 produtos, com 7, 30, 90 e 150 períodos de tempo.

Para instâncias de pequena e média dimensão conseguiram-se obter soluções, numa hora de tempo computacional, cujo valor é próximo do valor da solução óptima (a maior parte dos casos a diferença é de menos de 1%).

6- Agradecimentos

Este trabalho foi parcialmente financiado pela FCT (Fundação para a Ciência e Tecnologia) através do projecto "SearchCol: Metaheuristic search by column generation" (PTDC/EIA-EIA/100645/2008) e pela bolsa de doutoramento da FCT do primeiro autor (SFRH/BD/44601/2008).

7- Referências

EUBIONET II (2007). Factsheet 17 - Wood chips for a district heating plant - case district heating plant in Central Carinthia, Austria.

EUBIONET III, and Alakangas, E. (2010). EN 14961-1 - Classification of biomass origin in European solid biofuel standard.

European Union, C. (2005). Biomass action plan.

European Union, C. (2006). Green Paper - A European Strategy for Sustainable, Competitive and Secure Energy.

Gunnarsson, H., Ronnqvist, M., and Lundgren, J. T. (2004). Supply chain modelling of forest fuel. *European Journal of Operational Research* 158, 103–123.

Iakovou, E., Karagiannidis, A., Vlachos, D., Toka, A., and Malamakis, A. (2010). Waste biomass-to-energy supply chain management: A critical synthesis. *Waste Management* 30, 1860-1870.

Johnson, E. (2009). Goodbye to carbon neutral: Getting biomass footprints right. *Environmental Impact Assessment Review* 29, 165-168.

Sokhansanj, S., Kumar, A., and Turhollow, A. F. (2006). Development and implementation of integrated biomass supply analysis and logistics model (IBSAL). *Biomass and Bioenergy* 30, 838-847.

United Nations (1998). Kyoto Protocol to the United Nations Framework Convention on Climate Change.

Adaptive forecasting methods for time series with complex trend patterns

Vítor Mendonça[a], *António J.L. Rodrigues*[b*]

[a] *CIO-FCUL, Universidade de Lisboa, Lisboa, Portugal*
vitormendonca@gmail.com

[b] *CIO-FCUL, Universidade de Lisboa, Lisboa, Portugal*
ajrodrigues@fc.ul.pt
[*] *Corresponding Author*

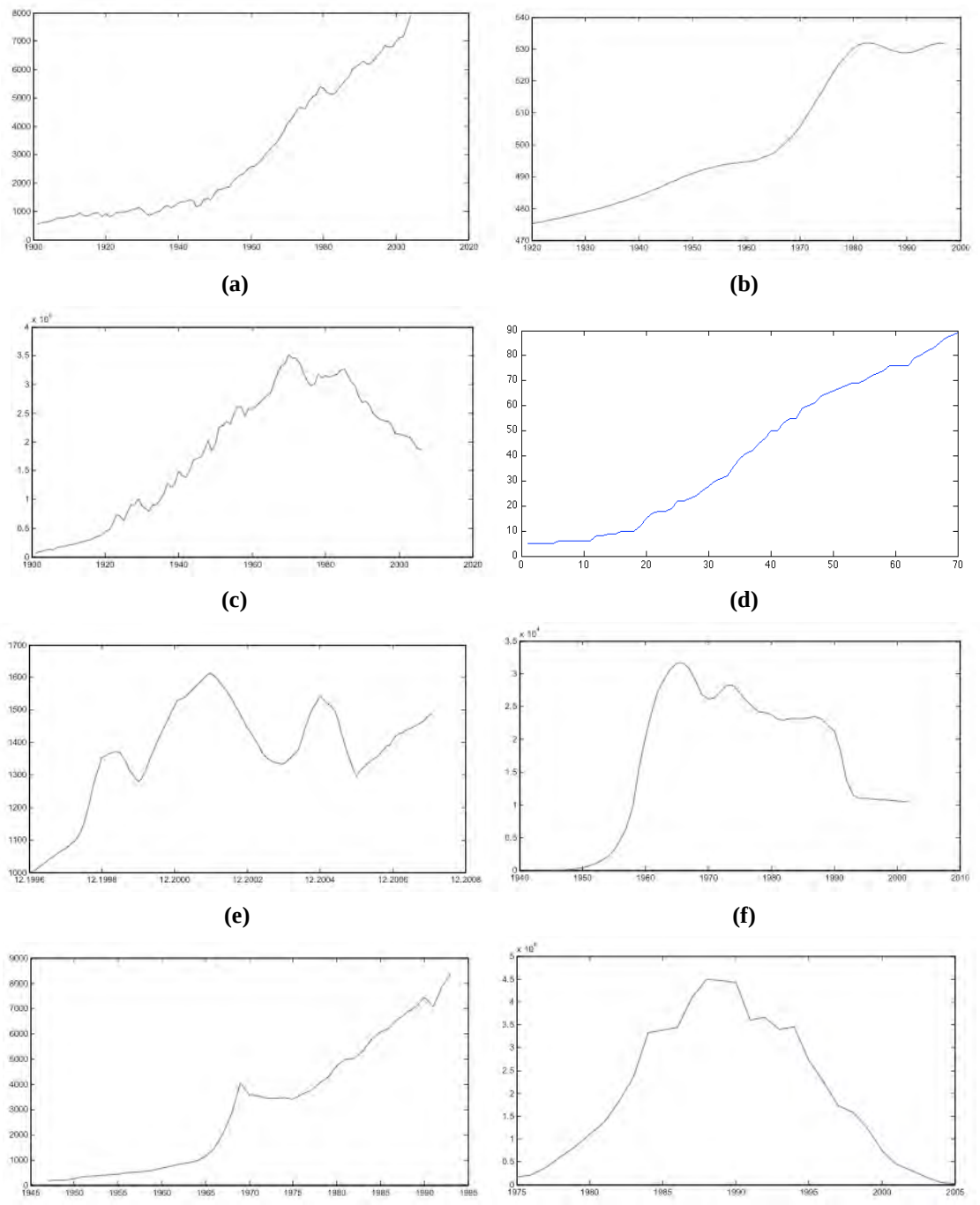
Abstract

Growth or development processes may exhibit trend patterns more complex than usually assumed by traditional models, including pure sigmoidal growth in innovation diffusion. We propose new modelling approaches, and introduce two new recursive estimation methods, designed to better account for modulated but aperiodic changes in curvature and convexity in trends. This is confirmed from comparative experiments of their application to a heterogeneous set of time series from very different fields — marketing, energy, climate, technology, health, etc.

Keywords: time series forecasting; recursive estimation methods; exponential smoothing; growth curves.

1- Introduction

Curvilinear trends in time series — i.e., trends with non-periodic, quasi-periodic, or cyclical variations in convexity — are quite common, and new modelling approaches are needed in order to track and forecast those variations adequately. Time series related to marketing, energy, technology or the environment often display such effects, as examplified in Figure 1. Usually, but not always, those are time series observed along relatively large timespans — say, years —, with no seasonal effects present.



(a) **(b)**

(c) **(d)**

(e) **(f)**

(g) **(h)**

Figure 1 – Sample time series:
(a) World CO2 emissions from fossile fuels, yearly data, 1991-2004;
(b) CH_3Cl concentration at Cape Grim, yearly data, 1920-1997 ;
(c) Crude oil production in the USA, yearly data, 1901-2006;
(d) H1N1 related deaths in Portugal, accumulated daily data, Nov-2009-Jan-2010;
(e) Lisbon hotel rooms occupancy index (VTQD-96), monthly data, Dec-1996-Jan-2008;
(f) Nuclear weapons stockpile in the USA, yearly data, 1945-2002;
(g) Air carriers, mail ton-miles in the USA, yearly data, 1947-1993;
(h) Musical album sales on audio tapes in the USA, yearly data, 1975-2005.

Most time series forecasting methods aim at producing reliable forecasts for relatively short horizons, namely to support decisions at an operational level. To support planning endeavours, reliable forecasts for longer horizons are *also* required. This is relatively easy to achieve when the time series exhibits simple, clearly defined and sufficiently regular low-frequency effects, i.e., trend and/or cycles.

In simple cases, such as most of those concerning the diffusion of innovations (technologies, brands, etc.), the time series under study naturally displays, or is expected to follow a sigmoidal growth pattern (as in Figure 1.d) or other relatively simple trend patterns — linear, quadratic, exponential, etc. It is then customary to use growth curves to fit the past data and, simultaneously, produce forecasts [Meade and Islam, 2006]. However, there are some fundamental or potential problems and limitations with this approach. Most notably, the principles and criteria that underlie the choice and assessment of *explanatory models* are different from those of truly *predictive models*.

Even when curve fitting leads to robust and sufficiently reliable long-term forecasts, the short-term forecasts are usually poor — this is due to the intrinsic inflexibility of models with just a few parameters, all constant in time, and the inability to cope with the local variations in the time series. It is also important to stress that forecasts are usually sought to be made along the whole observation period, and not only when enough data is available to comfortably conclude on the shape of the trend. Moreover, it is not possible to estimate beforehand the accuracy of the extrapolation estimates.

These difficulties can be circumvented in special cases — for instance, Meade [1985] demonstrated how some growth curves can be formulated recursively, thus enabling the estimation of the model parameters — then interpreted as time-varying parameters — through the Extended Kalman Filter. However, in practice, this approach has limited applicability and little appeal.

In this paper, we look at forecasting time series with complex trend patterns from a new angle, aiming at producing reliable forecasts for short-to-medium horizons. For that purpose, we introduce (in Sections 2 and 3) two new forecasting methods, and empirically assess and compare them (in Section 4). Both are extensions of well-known exponential smoothing methods (see Gardner [2006] for an overview):

— the *Adaptive Gardner-McKenzie* (AGMK) method is an adaptive version of what is best known as the *damped trend exponential smoothing* method (proposed by Gardner and McKenzie [1985]; see also Taylor [2003]);

— the *Curvilinear Exponential Smoothing* (CES) method is an extension of Holt's method [see Holt, 2004].

Firstly, we introduce the main notation used in the rest of the paper:

y_k	the scalar observation of a univariate time series at time k;
$\hat{y}_{k+p\|k}$	the p-steps-ahead forecast, with base k and horizon $p \geq 1$;
$e_{k+p\|k} = y_k - \hat{y}_{k+p\|k}$	the p-steps-ahead forecasting error;
$\Delta x_k = x_k - x_{k-1}$	first-difference operator;
$a_k,\ b_k,\ c_k$	time-varying parameters;
$\alpha,\ \beta,\ \gamma,\ \phi,\ \phi_k,\ \delta$	hyperparameters.

2 – The Adaptive Gardner-McKenzie (AGMK) method

The Gardner-McKenzie (GMK) method is defined by the following estimation formulae — the first two are similar to those of Holt's method and define how two time-varying parameters (*level* and *slope*) are recursively updated with the availability of new information; the third one is for the computation of forecasts at any moment in time, and any horizon:

$$\begin{aligned} a_k &= \alpha\, y_k + (1-\alpha)(a_{k-1} + \phi\, b_{k-1}) \\ b_k &= \beta(a_k - a_{k-1}) + (1-\beta)\phi\, b_{k-1} \\ \hat{y}_{k+p|k} &= a_k + \left(\sum_{i=1}^{p} \phi^i\right) b_k = \hat{y}_{k+p-1|k} + \phi^k\, b_k \end{aligned} \qquad (1)$$

There are 3 hyperparameters, which should be optimized with respect to an appropriate predictive performance measure: $0 < \alpha, \beta < 1$ and $\phi > 0$. The sequence of forecasts, for different time horizons, can be linear (when $\phi = 1$, in which case the method *coincides* with Holt's method), damped towards an asymptote ($\phi < 1$), or exponential ($\phi > 1$). Note that we consider also the latter case in the optimization of the method while, in the literature, it is usually assumed that a nonlinear transformation applied to the original data may change a supralinear growth into a linear or infralinear one.

In the literature, it is neither discussed the fact that, for many time series, a constant value of ϕ may not describe adequately the overall trend pattern. This is illustrated in Figure 2, with the infamous "peak oil" series (also shown in Fig. 1.c). In the first part of the series there was evidence of supralinear growth, and one would find $\phi^* > 1$ in the optimization, while, at a later stage, one finds $\phi^* < 1$. Therefore, it is reasonable to adapt the value of ϕ along time, by monitoring the sequence of one-step-ahead prediction errors, produced by the method. Among other possible candidate procedures, we propose a simple heuristic, universally applicable for the adaptation of any similar hyperparameter.

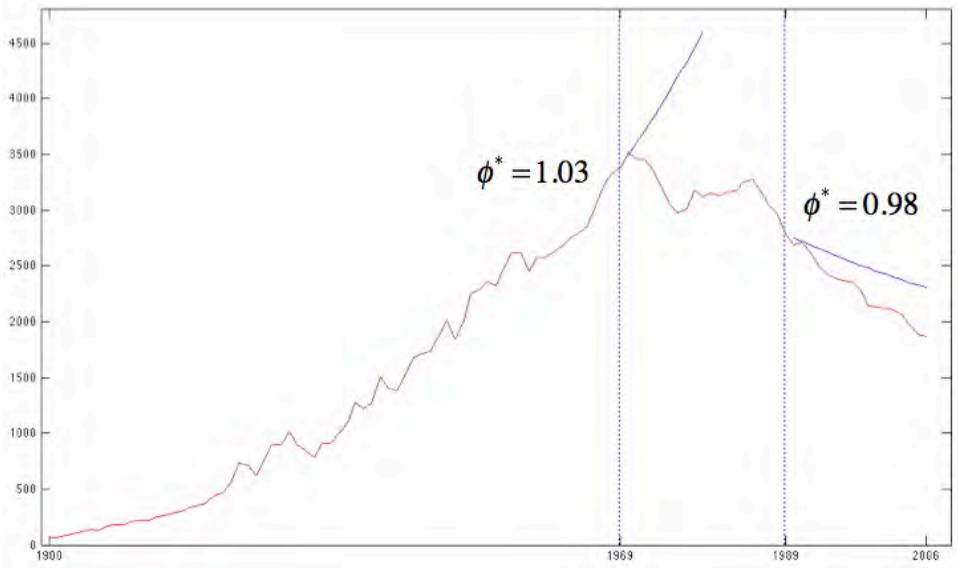


Figure 2 – Finding different optimal ϕ values at different moments.

Specifically, we consider making a small adjustment to the current value, ϕ_k, to be used in the next step of the recursive estimation algorithm, in case we recognize that such adjustment had been beneficial if already introduced in the *previous* step. Specifically, we consider 3 possible courses of action:

$$\phi_{k+1} = \begin{cases} (1-\delta)\,\phi_k \\ \phi_k \\ (1+\delta)\,\phi_k \end{cases}$$

according to which alternative would have produced a smaller value of $\left|e_{k|k-1}\right|$. With a carefully chosen (or, rather, optimized) value for δ, close to zero, more or less gradual changes are then possible. A decrease could also be accomplished by taking $\phi_{k+1} = (1+\delta)^{-1}\phi_k$, but our experiments have shown no significant difference in the results. By monitoring the sequence of ϕ_k values, it is then possible to estimate or detect the occurrence of turning points in the trend.

3- Curvilinear Exponential Smoothing (CES)

The newly proposed Curvilinear Exponential Smoothing method adds a curvature parameter, c_k, to Holt's method formulation, and correspondingly a different procedure for computing forecasts:

$$\begin{aligned} a_k &= \alpha\, y_k + (1-\alpha)(a_{k-1} + b_{k-1} + c_{k-1}) \\ b_k &= \beta\,(a_k - a_{k-1}) + (1-\beta)(b_{k-1} + c_{k-1}) \\ c_k &= \gamma\,(b_k - b_{k-1}) + (1-\gamma)\,c_{k-1} \\ \hat{y}_{k+p|k} &= a_k + pb_k + (p+1)\frac{p}{2}c_k \\ &= \hat{y}_{k+p-1|k} + \left(b_k + pc_k\right) \end{aligned} \qquad (2)$$

Figure 3 illustrates the pattern that the sequence of forecasts may form when the slope is positive, $b_k > 0$, and depending upon the sign of the curvature parameter. For instance, $c_k > 0$ corresponds to convex behaviour, with a quadratic increase in curvature, milder than the exponential increase from GMK. As shown in Figure 4, by monitoring the sequence of c_k values, it is then possible to estimate (i.e., to detect) convexity variations in the trend.

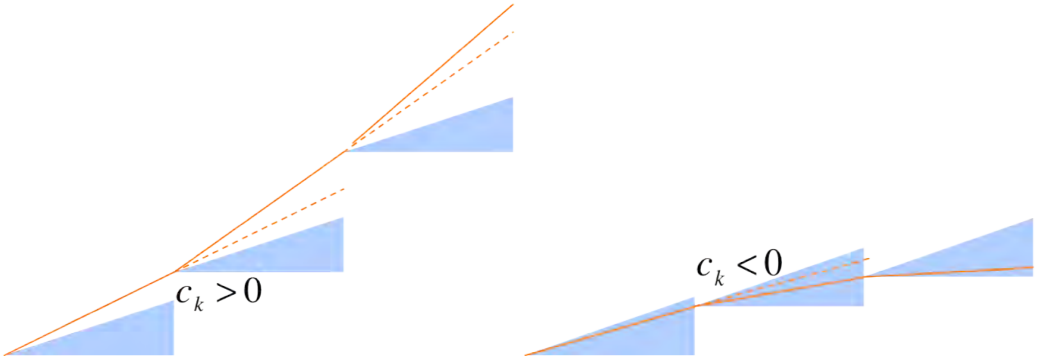


Figure 3 – Patterns of behaviour of CES forecasts.

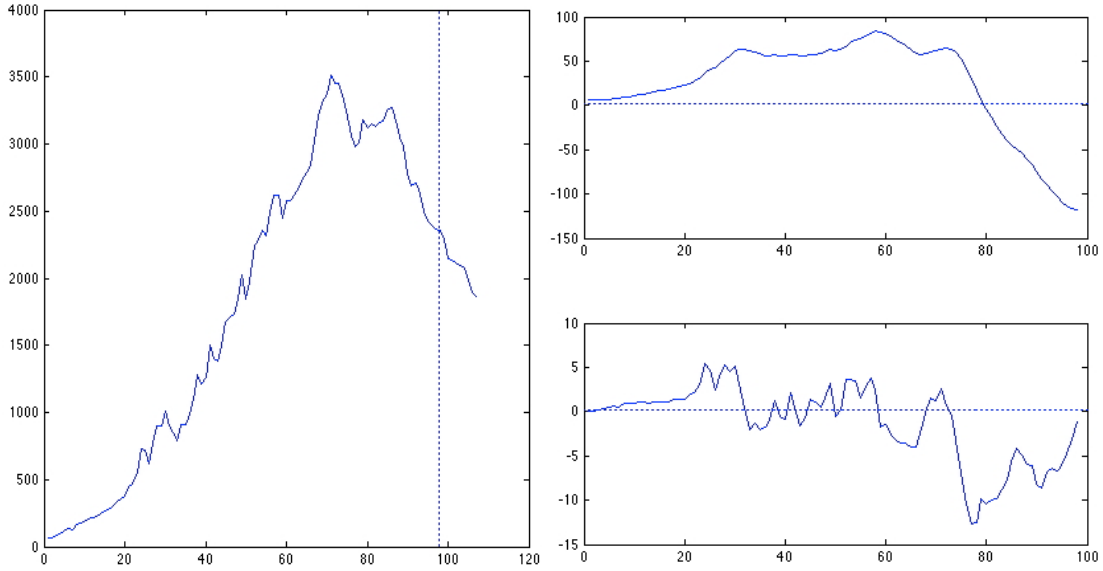


Figure 4 – Evolution of CES parameters b_k (top right) and c_k (bottom right) throughout the time series (left).

From Equation 2, it is easy to deduce how the changes in the 3 time-varying parameters depend on the 3 hyperparameters and on the prediction errors:

$$\begin{aligned}\Delta a_k &= \alpha e_{k|k-1} + b_{k-1} + c_{k-1}\\ \Delta b_k &= \alpha\beta e_{k|k-1} + c_{k-1}\\ \Delta c_k &= \alpha\beta\gamma e_{k|k-1}\end{aligned}$$

Moreover, following Hyndman et al [2008], the method can be formulated as a *single source of error* (SSOE) state space model:

$$\begin{aligned}y_k &= \mathbf{h}_k\mathbf{x}_k + \varepsilon_k\\ &= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} a_{k-1} & b_{k-1} & c_{k-1} \end{bmatrix}^{\mathrm{T}} + \varepsilon_k\end{aligned}$$

with a system equation, $\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \boldsymbol{\omega}\varepsilon_k$, mainly defined by the state transition matrix

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

This state space formulation allows the recursive estimation of the 3 time-varying parameters through the Kalman filter algorithm [Kalman, 1960], with forecasts given, equivalently, by $\hat{y}_{k+p|k} = \mathbf{h}_{k+p}\mathbf{F}^p\hat{\mathbf{x}}_{k|k}$. It also enables further insights on the properties of the method, and the extension to multivariate, or to multiple source of error formulations.

4- An experimental comparison

To test the two newly proposed methods, 26 non-trivial time series were collected, from Portuguese and international sources, including those of Figure 1. For comparison, Holt's method and GMK were also considered in the experiment. Predictive performance was optimized on the basis of two measures, MAPE (*mean average percentage error*) and MASE (*mean average scaled error*) — see Hyndman and Koehler [2006]. Both have the advantage of being expressed in abstract measuring units, allowing a safe computation of averages across time series measured in different units or with different data magnitudes.

It is little acknowledged in the literature that a model or method optimized for one-step-ahead prediction performance is, in general, suboptimal for longer horizons. As some of the time series on hand are not very long, we optimized the methods also on the basis of 5-steps-ahead forecasting accuracy, but not longer. In summary, the measures considered were MAPE(1), MAPE(5), MASE(1) and MASE(5), defined by:

$$\text{MAPE}(p) = \frac{100}{N-n} \sum_{k=n+1}^{N} \frac{\left|e_{k|k-p}\right|}{y_k}$$

$$\text{MASE}(p) = \sum_{k=n+1}^{N} \left|e_{k|k-p}\right| \Big/ \sum_{k=n+1}^{N} \left|\Delta y_k\right|$$

We have ignored the first n=10 errors in the computation of these measures, in order to reduce the possible bias effect of the parameters initialization — a common practice when applying recursive estimation methods.

The 26 series were divided into two groups: 15 smaller series (with 30 to 49 observations) and 11 larger series (with at least 50 observations). The main results of these tests are shown on Tables 1 and 2.

Table 1 – Average results for the shorter time series

	MAPE		**MASE**	
	1 s.a.	**5 s.a.**	**1 s.a.**	**5 s.a.**
Holt	6.2549	18.4670	0.8160	2.1633
GMK	5.4346	17.5386	0.7907	2.0255
AGMK	5.4026	**17.4467**	**0.7872**	**1.9898**
CES	**5.0770**	18.0497	0.7895	2.0659

Table 2 – Average results for the longer time series

	MAPE		**MASE**	
	1 s.a.	**5 s.a.**	**1 s.a.**	**5 s.a.**
Holt	2.5461	11.5413	0.5207	2.2986
GMK	2.5232	10.4176	0.5129	2.1471
AGMK	2.5089	**10.0643**	0.5059	**2.0899**
CES	**2.4819**	10.6394	**0.5035**	2.1835

5- Conclusions and further comments

From the experimental results one can firstly notice the improvement of GMK over Holt, of AGMK over GMK, and of CES over Holt. This was expectable beforehand. In fact, by extending the formulation of a model or method — $M_1 \prec M_2$, with M_1 being a particular case, or a limit case of M_2 —, one extends the feasible set of solutions (i.e., hyperparameters values). Then, one is bound to achieve better performance — or, at least, asymptotically equal performance —, provided the optimization is carried out carefully and thoroughly enough.

One of the main conclusions of our study is that, for time series with complex trends, the improvement achieved by AGMK or CES, while not being impressive, can be very valuable in practice, especially for long forecasting horizons.

AGMK produced, in most cases, better 5-steps-ahead forecasts than CES. However, it is yet to be determined the possible contribution of a nonlinear transformation of the original data — such as the logarithm or the square root — to the relative performance between these two methods. In doubt, we suggest to optimally combine the forecasts of both methods, rather than ‘betting” on only one of them.

Also, one should not discard the potential benefit of combining any of the new methods with the results from curve fitting, at least in some cases. This can be especially useful in setting more feasible paths or workplans for controlling processes in the short-to-medium, or even the long-term future. Figure 5 illustrates this idea with the world CO2 emissions data (shown up to 1997). Given the possibility of intervention in the evolution of a phenomenon, one can set goals for the long-term. Then, one may linearly combine the predictive estimates that assume a no-intervention scenario (say, resulting from using CES) with the path defined by a curve fitted also to the desired future target value. By progressively changing the relative contribution of the two sets of forecasts in forming the combined forecasts, one can determine reasonable intermediate target values, useful for policy planning purposes.

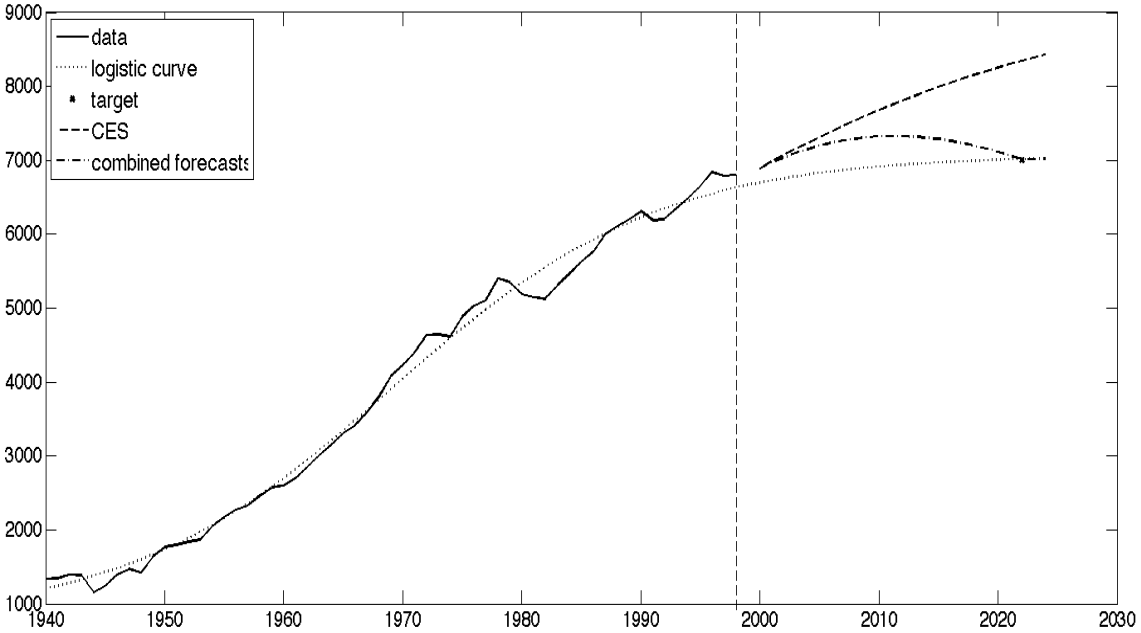


Figure 5 – Adjusting forecasts to accommodate future goals or constraints.

6- References

Gardner Jr., E.S., McKenzie, E. (1985). Forecasting trends in time series. *Management Science*. 31, 1237-1246.

Gardner Jr., E.S. (2006). Exponential smoothing: The state of the art – Part II. *International Journal of Forecasting*. 22, 637-666.

Holt, C.C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*. 20, 5-10 (reprint of the 1957 original report paper).

Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer-Verlag, Berlin.

Hyndman, R.J., Koehler, A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*. 22, 679-688.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problem. *Journal of Basic Engineering*. 82(1), 35–45.

Meade, N. (1985). Forecasting using growth curves – An adaptive approach. *Journal of the Operational Research Society*. 36(12), 1103-1115.

Meade, N., Islam, T. (2006). Modelling and forecasting the diffusion of innovation – A 25-year review. *International Journal of Forecasting*. 22, 519-545.

Taylor, J.W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*. 19, 715-725.

Índice de Autores

A. Miguel Gomes..... 165
Aguinaldo Aparecido Pereira..... 153
Ana Paula Ferreira Dias Barbosa Póvoa 13, 79, 103, 129, 199, 237, 249, 261
Ana Paula Teixeira 1
Ana S. Camanho 37, 65, 115
André Lopes Fradinho..... 13
André Luís Marques Marcato..... 213
Andreia Filipa Antunes..... 25
Andreia Zanella..... 37
António J. L. Rodrigues..... 287
Armando Brito de Sá..... 177
Augusto Queiroz Novais..... 249
Bruno Henriques Dias..... 213
Carlos A. Bana e Costa..... 177
Carlos Alberto Rodrigues..... 53
Clara Vaz..... 65
Cláudia Duarte..... 79
Edméa Cássia Baptista..... 153
Eduarda Pinto Ferreira..... 165
Elsa Silva..... 91
Fernando Luiz Cyrino Oliveira..... 213
Filipe Pereira e Alvelos..... 91, 275
Frederico Gonçalves..... 103
Helena Monteiro..... 1
Isabel M. Horta..... 115
João Carlos Correia Baptista Soares de Mello..... 187
Jorge Moreira da Costa..... 115
José Manuel Valério de Carvalho..... 79, 91
Leão José Fernandes..... 129
Luís Alçada-Almeida..... 225
Luís Dias..... 225
Manuel Rodrigues..... 141

Margarida Moz 25
Maria Cláudia Sávio Masiero 153
Maria Helena Correia 53
Maria Isabel Gomes 13, 199, 261
Maria Manuel Nascimento 1
Maria Sameiro Carvalho 275
Mariana Vieira Rangel Nunes 187
Marisa Oliveira 165
Mónica D. Oliveira 141, 177
Nissia Carvalho Rosa Bergiante 187
Nuno Miguel Santana 199
Reinaldo Castro Souza 213
Sandra Silva 225
Sónia R. Cardoso 103, 237
Susana Relvas 103, 129, 237
Tânia Pinto-Varela 249
Tânia Rodrigues Pereira Ramos 261
Teresa C. Rodrigues 177
Teresa G. Dias 37
Teresa Peña 141
Tiago Costa Gomes 275
Vanusa Alves de Sousa 153
Vítor Mendonça 287