

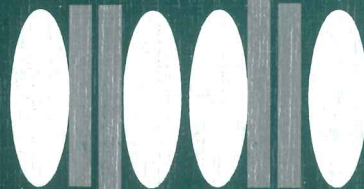
INVESTIGAÇÃO OPERACIONAL

Junho 1996

Número 1

Volume 16

Publicação Científica da



Associação Portuguesa para o Desenvolvimento
da Investigação Operacional

INVESTIGAÇÃO OPERACIONAL

Propriedade:

APDIO — Associação Portuguesa para o Desenvolvimento
da Investigação Operacional

ESTATUTO EDITORIAL

«Investigação Operacional», órgão oficial da APDIO cobre uma larga gama de assuntos reflectindo assim a grande diversidade de profissões e interesses dos sócios da Associação, bem como as muitas áreas de aplicação da I. O. O seu objectivo primordial é promover a aplicação do método e técnicas da I. O. aos problemas da Sociedade Portuguesa. A publicação acolhe contribuições nos campos da metodologia, técnicas, e áreas de aplicação e software de I. O. sendo no entanto dada prioridade a bons casos de estudo de carácter eminentemente prático.

Distribuição gratuita aos sócios da APDIO

INVESTIGAÇÃO OPERACIONAL

Volume 16 - nº 1 - Junho 1996

Publicação semestral

Editor Principal: Joaquim J. Júdice
Universidade de Coimbra

Comissão Editorial

M. Teresa Almeida
Inst. Sup. Economia e Gestão

Jaime Barceló
Univ. de Barcelona

Paulo Barcia
Univ. Nova de Lisboa

Isabel Branco
Univ. de Lisboa

António Câmara
Univ. Nova de Lisboa

C. Bana e Costa
Inst. Superior Técnico

M. Eugénia Captivo
Univ. de Lisboa

Jorge O. Cerdeira
Inst. Sup. de Agronomia

João Clímaco
Univ. de Coimbra

J. Dias Coelho
Univ. Nova de Lisboa

J. Rodrigues Dias
Univ. de Évora

Laureano Escudero
IBM, Espanha

J. Soeiro Ferreira
Univ. do Porto

J. Fernando Gonçalves
Univ. do Porto

Clóvis Gonzaga
Univ. Fed., Rio Janeiro

Luís Gouveia
Univ. de Lisboa

Rui C. Guimarães
Univ. do Porto

J. Assis Lopes
Inst. Superior Técnico

N. Maculan
Univ. Fed., Rio Janeiro

Ernesto Q. Martins
Univ. de Coimbra

Vladimiro Miranda
Univ. do Porto

J. Pinto Paixão
Univ. de Lisboa

M. Vaz Pato
Inst. Sup. Economia e Gestão

Celso Ribeiro
Univ. Católica, Rio Janeiro

A. Guimarães Rodrigues
Univ. do Minho

Mário S. Rosa
Univ. de Coimbra

J. Pinho de Sousa
Univ. do Porto

Reinaldo Sousa
Univ. Católica, Rio Janeiro

L. Valadares Tavares
Inst. Superior Técnico

Isabel H. Themido
Inst. Superior Técnico

B. Calafate Vasconcelos
Univ. do Porto

José M. Viegas
Inst. Superior Técnico

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

Esta Revista é distribuída gratuitamente aos sócios da APDIO. As informações sobre inscrições na Associação, assim como a correspondência para a Revista devem ser enviadas para a sede da APDIO - Associação Portuguesa para o Desenvolvimento da Investigação Operacional - CESUR, Instituto Superior Técnico, Av. Rovisco Pais, 1000 Lisboa.

Este Volume foi subsidiado por :

Junta Nacional de Investigação Científica e Tecnológica (JNICT)

Fundação Calouste Gulbenkian

Para efeitos de dactilografia e composição, foram utilizados equipamentos gentilmente postos à disposição pelo CEAUL (DEIO - Faculdade de Ciências de Lisboa).

Assinatura: 5.000\$00

EXPERIÊNCIA COMPARATIVA DE UTILIZAÇÃO DE ALGORITMOS DE "ASSIGNMENT" EM PROBLEMAS DE ESTIMAÇÃO DE MATRIZES DE ORIGEM/DESTINO

Maria Sameiro Carvalho e A. Guimarães Rodrigues

Universidade do Minho
Braga, Portugal

D.V. Vliet e Paul Timms

Institute for Transport Studies
University of Leeds
Leeds, Inglaterra

Abstract

An important problem in traffic planning occurs when trying to devise efficient methods to estimate Origin/Destination (OD) trip matrices from traffic counts in congested networks. The main objective of this research is to compare the performance of different traffic assignment algorithms in a model that estimates OD matrices. In particular, this study examines the relative efficiency of the traditional Frank-Wolfe algorithm and a new technique based on "Social Pressure". These two algorithms are tested in a small network with two origins and two destinations. Results show that the Social Pressure based algorithm can perform better than the traditional approach leading to an improved efficiency of the matrix estimation model.

Resumo

No tratamento do problema da determinação de matrizes de origem/destino (OD) em redes de tráfego congestionadas, colocam-se questões associadas à eficiência das técnicas de solução dos modelos adoptados, bem como da interacção entre os modelos de "afecção" de tráfego ("assignment") e de "estimação de matrizes OD".

Neste artigo, identificam-se alguns aspectos relacionados com a eficiência das técnicas de solução, e relata-se a experiência adquirida com a utilização de duas técnicas alternativas de solução para o problema da afectação das viagens entre as diversas origens e os diversos destinos.

Uma das técnicas utilizadas consiste no método tradicional de Frank-Wolfe (1956). Nesta abordagem, em cada iteração, um novo "caminho mais curto" (entre cada par OD) é gerado com base nas condições correntes de tráfego saídas da anterior iteração. Assim, em cada iteração, um igual número de viagens (proporcional ao fluxo total) é deslocado de cada "caminho" gerado em soluções anteriores, para o novo "caminho mais curto".

A outra técnica utilizada, que se designa por "Social Pressure Algorithm" procura incorporar no procedimento de "assignment" o conceito de "Pressão Social" a fim de representar a maior ou menor pressão sentida pelos condutores, que os leva a optar por vias de menor "custo". Assim, condutores que viajam em rotas mais caras sentir-se-ão mais fortemente pressionados a transferirem-se para rotas de menor custo em contraposição ao método de Frank-Wolfe que considera um peso igual na transferência de viagens.

Relatam-se alguns resultados obtidos no contexto de um modelo iterativo de estimação de matrizes de origem/destino que comprovam uma maior eficiência do método alternativo baseado em "Pressão Social" numa aplicação a uma rede simples de teste com duas origens e dois destinos.

Keywords

Traffic Assignment, Network Equilibrium, OD Matrix Estimation, Assignment Algorithms.

1. Introdução

Neste artigo pretende-se examinar o desempenho de um algoritmo de afectação baseado em Pressões Sociais (Social Pressure Based Assignment Algorithm-SP) no âmbito de um modelo iterativo de estimação de matrizes de origem/destino (OD).

O algoritmo baseia-se no conceito de Pressão Social que foi aplicado pela primeira vez para implementar algoritmos de afectação dinâmica (Cybis, 1993). A mesma ideia foi posteriormente utilizada no desenvolvimento de algoritmos de afectação (Wang, 1993). Estes algoritmos mostraram ter um desempenho superior ao das técnicas tradicionais (Frank-Wolfe, 1956) constituindo portanto uma alternativa atraente.

Em primeiro lugar serão revistos alguns princípios básicos relativos aos modelos de afectação de tráfego e estimação de matrizes. Em segundo lugar serão referidas as alterações introduzidas no modelo tradicional de afectação necessárias à implementação do modelo "Social Pressure". Testes e resultados são apresentados em terceiro lugar e, na última parte, são apresentadas algumas conclusões.

Uma rede de tráfego urbano é representada por um grafo orientado $G = (X, A)$ onde X é um conjunto de pontos (nodos ou vértices) que representam intersecções de ruas, áreas residências, estações de transportes públicos, etc, ligados por um conjunto de linhas A (arcos ou ramos) que representam as ruas. Cada origem ou destino (correspondendo a uma área) é também representado por um nodo e cada par origem/destino está ligado por um número finito de caminhos (rotas constituídas por uma sequência de arcos). Nas expressões apresentadas adopta-se a seguinte notação:

$\{T\}$ - matriz de viagens cujos elementos T_{ij} representam o número total de viagens entre a origem i e o destino j

O - conjunto origens

D - conjunto de destinos

N - número total de origens

M - número total de destinos

V_a - fluxo de tráfego no ramo a

\bar{V}_a - fluxo de tráfego observado no ramo a

$C_a(.)$ - função custo de viagem no ramo a

p_{ij} - um caminho entre i e j

K_{ij} - conjunto de todos os caminhos entre i e j

P_{ij}^a - proporção de viagens entre i e j que usam o ramo a

$T_{p_{ij}}$ - total de viagens na rota p_{ij}

$C_{p_{ij}}$ - custo total de viajar na rota p_{ij}

$P_{p_{ij}}$ - factor de "pressão social" associado à rota p_{ij}

$$\delta_{pij}^a = \begin{cases} 1 & \text{se a rota } pij \text{ usa o arco } a \\ 0 & \text{caso contrário} \end{cases}$$

2. Modelo de Afectação de Tráfego

2.1 Introdução

A estimação do número total de viagens que é efectuado, num dado período de tempo, entre as diversas zonas de uma rede urbana (normalmente representado por uma matriz de origem-destino OD - $\{T\}$) é uma parte fundamental na análise da gestão de tráfego nomeadamente no desenho e avaliação de esquemas de gestão de tráfego. A grande complexidade e dimensão dos problemas que se colocam neste âmbito requerem o apoio de técnicas eficientes.

Os modelos de estimação de matrizes OD com base em contagens de tráfego têm como objectivo estimar a matriz $\{T\}$ que, quando afectada à rede, gera um conjunto de fluxos de tráfego na rede consistente com as contagens efectuadas num conjunto seleccionado dos seus arcos. A identificação do modo como o total de viagens entre a origem i e o destino j se distribui pelas diferentes rotas que ligam as duas zonas, em particular o cálculo da proporção do número de viagens entre i e j que usam o ramo a (P_{ij}^a) é fundamental num processo de estimação de matrizes OD. A afectação de um conjunto de viagens pelas diferentes rotas de uma rede (traffic assignment) é baseada em princípios que serão discutidos na secção seguinte.

A estimação de uma matriz OD a partir de um conjunto de observações de fluxo de tráfego num subconjunto B de arcos ($B \in A$) pode ser definido como o problema de encontrar os valores de T_{ij} que satisfaçam o seguinte conjunto de equações:

$$\bar{V}_a = \sum_i \sum_j P_{ij}^a \times T_{ij} \quad \forall a \in B \quad (1)$$

A dificuldade associada à resolução deste sistema indeterminado de equações (o número de incógnitas é, normalmente, muito superior ao número de equações disponíveis) é, na maior parte dos modelos existentes, ultrapassada definindo uma função objectivo que visa seleccionar a matriz que satisfaz determinado objectivo (por exemplo, que minimiza uma função que constitui uma medida das diferenças entre a matriz e uma estimativa anterior).

Em redes de tráfego congestionadas os valores de P_{ij}^a são função dos valores de T_{ij} e não são, portanto, conhecidos previamente. A técnica de estimação de matrizes OD, em redes congestionadas de tráfego, com base em contagens de fluxos de tráfego, proposto por Willunsen (1981), consiste num procedimento iterativo onde um modelo de afectação de tráfego e um modelo de estimação de matrizes são executados alternadamente. Assumindo que existe uma estimativa inicial para a matriz $\{T\}$, o modelo de afectação de tráfego estima as proporções de tráfego que usam determinado arco a (P_{ij}^a) e com base nesses valores o modelo de estimação de matrizes deriva uma nova estimativa para $\{T\}$. Este procedimento tenta ultrapassar o problema de combinar num modelo único um procedimento capaz de gerar uma matrix de OD

que, quando afectada a uma rede, reproduza fluxos nos arcos próximos dos valores observados.

A eficiência deste procedimento depende, naturalmente, do uso de técnicas eficientes para a resolução dos dois sub-problemas. Iremos, em particular, analisar o desempenho de duas técnicas alternativas de resolução do problema de afectação: o tradicional algoritmo de Frank-Wolfe e um algoritmo baseado em "pressões sociais" ('Social Pressure') e avaliar o impacto que o desempenho delas tem no desempenho do modelo global de estimação de matrizes.

2.2 Afectação de Tráfego

Dada uma rede de tráfego urbana dividida em E zonas e assumindo que o número total de viagens entre as diferentes zonas, por unidade tempo, é conhecido, o modelo de afectação estima o número de condutores que utiliza determinada rota. A previsão do comportamento dos condutores relativo à decisão de selecção de rotas feitas pelos mesmos permite obter informação relativa aos níveis de fluxos em cada uma das ruas assim como, avaliar o desempenho da rede de tráfego (níveis de congestionamento, identificação de zonas-problema, etc).

Wardrop estabelece em 1952 os princípios básicos que regem essa distribuição de tráfego:

"when there is more than one route from origin to destination and all the routes do not have any capacity restrictions then the journey times by all the used routes would be equal to one another and will not be more than the time that will be taken on an unused route"

ou seja, os tempos (ou, mais genericamente, os custos) de viagem em todas as rotas efectivamente usadas é menor ou igual que os custos que um qualquer condutor pode experimentar em qualquer rota não utilizada, e

"... assigns the traffic in such a way that the total time spent in the network by all the vehicles is minimised"

que implica que, o tempo (custo) médio de viagem em todas as rotas usadas é o menor possível.

Estes dois princípios levaram ao desenvolvimento de duas abordagens distintas: "User Equilibrium" e "System Equilibrium". Enquanto que na primeira (considerada mais realista para representar situações reais de tráfego) cada utilizador pretende minimizar o seu tempo, ou mais genericamente, os seus custos de transporte, na segunda entende-se haver uma atitude concertada entre os diferentes utilizadores da rede de modo a minimizar os tempos (custos) totais de transporte no sistema global.

A abordagem "User Equilibrium" é caracterizada por uma situação estável (de equilíbrio de tráfego) onde cada condutor encontrou o "melhor caminho" e, portanto, nenhum condutor pode melhorar o seu tempo (custo) de viagem trocando de rota.

Assumindo que os custos associados ao percurso de um ramo da rede são função somente do volume de tráfego nesse ramo, e que cada condutor visa minimizar os seus tempos (custos) de transporte, as condições que representam uma situação de "User Equilibrium" podem ser

descritas do modo seguinte:

$$T_{pij} > 0 \Rightarrow C_{pij} = C_{\min} \quad \forall pij \in K_{ij} \quad (2)$$

$$T_{pij} = 0 \Rightarrow C_{pij} \geq C_{\min}$$

onde C_{\min} é o tempo mínimo de viagem entre i e j .

i.e. se um caminho pij tem fluxo positivo então essa rota tem necessariamente que ter um custo mínimo. Se o fluxo num caminho é nulo então o custo de viajar nessa rota é superior ou igual ao custo mínimo.

Beckman (1956) demonstrou que resolver um problema de afectação de tráfego é equivalente a resolver um problema de minimização com restrições. Pode-se provar (Van Vliet, 1979) que o conjunto de fluxos de tráfego (V_a) obtido como solução do problema de minimização que se apresenta de seguida, satisfaz as condições de equilíbrio atrás referidas:

$$\min Z(v) = \sum_{a \in A} \int_0^{V_a} C_a(v) dv \quad (3)$$

sujeito às seguintes condições:

- a soma dos fluxos em todos os caminhos entre i e j é igual ao fluxo total entre origem i e destino j (restrições de conservação de fluxos):

$$\sum_{pij \in K_{ij}} T_{pij} = T_{ij} \quad \forall i \in O, \forall j \in D \quad (4)$$

- os fluxos de tráfego em qualquer rota deverão ser não negativos:

$$T_{pij} \geq 0 \quad \forall i \in O, \forall j \in D, \forall pij \in K_{ij} \quad (5)$$

e pelas relações que exprimem os fluxos nos arcos em função dos fluxos nas rotas:

$$V_a = \sum_{ij} \sum_{pij \in K_{ij}} T_{pij} \delta_{pij}^a \quad \forall a \in A \quad (6)$$

onde δ_{pij}^a é dado por:

$$\delta_{pij}^a = \begin{cases} 1 & \text{se a rota } pij \text{ usa o arco } a \\ 0 & \text{caso contrário} \end{cases}$$

A Função Objectivo não pretende ter qualquer interpretação económica ou de comportamento devendo ser vista, apenas, como uma ferramenta matemática utilizada na solução de problemas de afectação de tráfego em situações de equilíbrio, tal como é referido por Sheffi (1985).

As condições a que deve obedecer a função de custos nos arcos são:

- a função deverá ser positiva e crescente

$$\frac{\partial C_a(V_a)}{\partial V_a} > 0 \quad \forall a \in A \quad (7)$$

- os custos de viajar num arco a deverão ser apenas função do fluxo nesse arco:

$$\frac{\partial C_a(V_a)}{\partial V_b} = 0 \quad \forall a \neq b \quad (8)$$

De facto, é geralmente bem aceite que a evolução dos custos associados ao percurso de um arco é função do fluxo nesse arco e é caracterizado por uma função do tipo da que a seguir se ilustra:

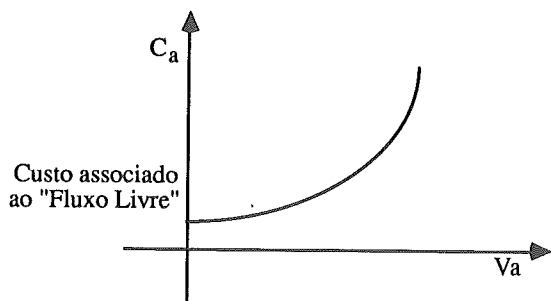


Figura 1 - Custo de viagem num arco em função do fluxo nesse arco

A grande vantagem deste tipo de formulação resulta do facto da solução que satisfaz as condições de equilíbrio de Wardrop poder ser encontrada resolvendo um problema de optimização de redes convexo. O método de Combinação Convexa, conhecido como método de Frank-Wolfe, é utilizado para a resolução deste problema, e consiste numa aplicação sequencial de aproximações lineares à função de custos, resultando, em cada iteração, num problema de programação linear que é resolvido a fim de determinar a direcção de descida, e para o qual é garantida a convergência.

2.2.1 Algoritmo de Frank-Wolfe (FW)

No algoritmo de Frank-Wolfe, a escolha da direcção a tomar com vista a uma diminuição do valor da Função Objectivo é feita do seguinte modo: em cada iteração, assumindo que a solução na iteração n é dada pelo conjunto de fluxos $V^{(n)}$, o algoritmo procura, no espaço de soluções admissíveis, um Ponto Auxiliar $F^{(n)}$ sobre a fronteira do espaço de soluções e numa direcção que garanta um máximo decréscimo do valor da função objectivo.

O algoritmo, quando aplicado à resolução de um problema de afectação de tráfego em condições de equilíbrio (User Equilibrium), pode ser descrito pelo seguinte conjunto de etapas:

PASSO 0 : Inicialização

Assumir condições de fluxo livre nos arcos ($V_a = 0$) e calcular os respectivos custos. Determinar os caminhos mais curtos entre as diversas origens e destinos e afectar o total das viagens aos caminhos de menor custo a fim de produzir $V_a^{(1)}$.

Pôr $n = 1$.

PASSO 1: Actualização dos Custos nos Arcos

$$C_a^{(n)} = C_a(V_a^{(n)})$$

PASSO 2: Determinação do Ponto Auxiliar

O ponto auxiliar $F_a^{(n)}$ é obtido através da solução de um problema de Programação Linear que considera uma aproximação linear da função objectivo segundo a direcção que garante o máximo decréscimo do valor dessa função.

PASSO 3: Determinação do λ ("step-size")

Calcular $\lambda = \bar{\lambda}$ que resolve

$$\bar{\lambda} = \min_{0 \leq \lambda \leq 1} \sum_a \int_0^{V_a^{(n)} + \lambda(F_a^{(n)} - V_a^{(n)})} C_a(v) dv$$

No domínio definido entre a solução actual $V^{(n)}$ e o ponto auxiliar $F^{(n)}$, a minimização desta função encontra o valor de λ necessário à obtenção de uma solução a que corresponde um valor mínimo para a função objectivo original.

PASSO 4: Cálculo de uma Nova Solução

Construir uma nova solução através de:

$$V_a^{(n+1)} = V_a^{(n)} + \bar{\lambda} \times (F_a^{(n)} - V_a^{(n)})$$

PASSO 5: Verificação da Convergência

Se a condição de convergência for verificada parar. Senão incrementar n de 1 e voltar ao passo 1.

A condição de convergência utilizada consiste numa medida adimensional baseada nas alterações dos fluxos nos arcos em iterações sucessivas:

$$\Delta = \frac{\sqrt{\sum_a (V_a^{(n+1)} - V_a^{(n)})^2}}{\sum_a V_a^{(n)}} \leq \epsilon \quad (9)$$

Ou seja, o processo iterativo deverá terminar quando em duas iterações sucessivas, a variação produzida nos fluxos de todos os arcos for inferior a um valor de erro pré-estabelecido (ϵ).

Embora a descrição do algoritmo de Frank-Wolfe tenha sido feita em termos de fluxos nos arcos (V_a), ela também pode ser feita em termos de fluxos nas rotas:

$$T_{pij}^{(n+1)} = (1 - \bar{\lambda}) \times T_{pij}^{(n)} \quad (10)$$

para cada caminho que não o caminho de custo mínimo entre i e j , e

$$T_{pij}^{(n+1)} = (1 - \bar{\lambda}) \times T_{pij}^{(n)} + T_{ij} \times \bar{\lambda} \quad (11)$$

para o caminho de custo mínimo entre i e j .

A expressão (10) mostra que, para uma rota genérica, o fluxo é reduzido de um factor ($\bar{\lambda}$). As rotas de custo mínimo absorvem os fluxos transferidos de outras rotas conforme ilustrado no segundo termo da expressão (11).

2.2.2 Algoritmo "Social Pressure" (SP)

O algoritmo baseado em "pressões sociais" pretende incluir no modelo de afectação atrás referido a ideia de que condutores viajando em caminhos relativamente mais caros se sentirão mais fortemente pressionados a mudar de caminho do que aqueles que viajam em rotas com custo mais próximo do caminho de custo mínimo. Sendo assim, a taxa a que se processa a transferência para os caminhos de menor custo não será somente função do volume de tráfego que esse caminho acomoda mas, também proporcional às diferenças de custo em relação ao custo mínimo - "Pressão-Social" (Wang, 1993).

Em cada iteração, este algoritmo de afectação calcula para cada caminho (p_{ij}) que não o caminho mais curto (q_{ij}) um factor de "pressão social" que é função da diferença de custos:

$$P_{p_{ij}} = f(C_{p_{ij}} - C_{q_{ij}})$$

onde

$C_{p_{ij}}$ - custo na rota p_{ij}

$C_{q_{ij}}$ - custo da rota de menor custo

Os novos fluxos nas rotas, serão agora dados por:

$$T_{p_{ij}}^{(n+1)} = (1 - \bar{\lambda} P_{p_{ij}}) T_{p_{ij}}^{(n)} \quad (12)$$

$$T_{q_{ij}}^{(n+1)} = T_{q_{ij}}^{(n)} + \sum_{p \neq q} T_{p_{ij}}^{(n)} \times \lambda \times P_{p_{ij}} \quad (13)$$

O algoritmo proposto por Wang (1993), que inclui o conceito de "Pressão Social", consiste em:

PASSO 0: Inicialização

Assumir condições de fluxo livre nos arcos ($V_a^{(0)} = 0$) e calcular os respectivos custos. Determinar os caminhos mais curtos entre as diversas origens e destinos e afectar o total das viagens aos caminhos de menor custo a fim de produzir $V_a^{(1)}$.

Pôr $n = 1$.

PASSO 1: Actualização dos Custos nos Arcos

$$C_a^{(n)} = C_a(V_a^{(n)})$$

PASSO 2: Determinação do Ponto Auxiliar

1. Construir árvores de custo mínimo e gerar um novo conjunto ($n^{\text{ésimo}}$) de caminhos de menor custo (q_{ij}). Calcular factores de "pressão social" ($P_{p_{ij}}$) para todos os outros caminhos e normalizar esses factores (dividindo-os pelo maior factor de "pressão social").

$$P'_{p_{ij}} = C_{p_{ij}} - C_{q_{ij}} \quad \forall i \in O, \forall j \in D, \forall p_{ij} \in K_{ij}$$

$$P_{p_{ij}} = \frac{P'_{p_{ij}}}{\max_{ij} \{P'_{p_{ij}}\}}$$

2.. Obter um ponto auxiliar $F_a^{(n)}$:

A expressão para a solução auxiliar, $F_a^{(n)}$, obtém-se definindo a nova solução em termos de fluxos nos arcos ($V_a^{(n+1)}$):

$$V_a^{(n+1)} = \sum_{p \neq q} T_{pij}^{(n+1)} \delta_{pij}^a + T_{qij}^{(n+1)} \delta_{qij}^a$$

Usando as expressões (12) e (13) e fazendo:

$$A_{qij} = \sum_{p \neq q} T_{pij}^{(n)} \times P_{pij}$$

$$A_{pij} = T_{pij}^{(n)} \times P_{pij}$$

obtem-se a expressão para o Ponto Auxiliar:

$$F_a^{(n)} = V_a^{(n)} + (A_{qij} \times \delta_{qij}^a - \sum_{p \neq q} A_{pij} \delta_{pij}^a)$$

PASSO 3: Determinação do $\bar{\lambda}$ ("step size")

Calcular $\bar{\lambda}$ que minimiza a seguinte função:

$$\bar{\lambda} = \min_{0 \leq \lambda \leq 1} \sum_a \int_0^{V_a^{(n)} + \lambda(F_a^{(n)} - V_a^{(n)})} C_a(v) dv$$

PASSO 4: Cálculo de uma Nova Solução

Calcular o novo conjunto de fluxos:

$$T_{pij}^{(n+1)} = T_{pij}^{(n)} - \bar{\lambda} \times A_{pij}$$

para uma rota genérica pij , e:

$$T_{qij}^{(n+1)} = T_{qij}^{(n)} + \bar{\lambda} \times A_{qij}$$

para a rota de custo mínimo qij .

PASSO 5: Verificação de convergência

Parar se a condição de convergência dada em (9) se verificar. Caso contrário, incrementar n de 1 e voltar ao passo 1.

Basicamente, a grande diferença entre os dois algoritmos consiste no modo como a direcção de descida é seleccionada.

3. Testes e Resultados

3.1 Introdução

Com o objectivo de avaliar e comparar o desempenho dos dois algoritmos de afectação de tráfego (Frank-Wolfe (FW) e " Social Pressure" (SP)) foi realizado um conjunto de testes.

A rotina de afectação é um dos modulos básicos do modelo iterativo de estimação de matrizes OD. Alterações no desempenho desta rotina vão certamente reflectir-se no desempenho do procedimento de estimação das matrizes.

Foi desenvolvido um programa de computador (Fortran 77) para implementação do modelo de estimação de matrizes. Uma matriz simples associada à rede da figura seguinte, com 2 origens e 2 destinos, foi a utilizada nos testes.

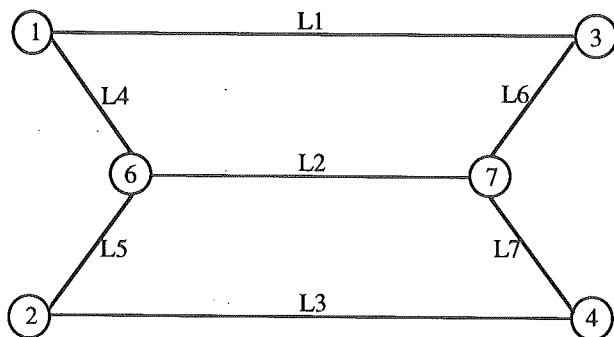


Figura 2 - Rede de teste

Diferentes condições de tráfego foram ensaiadas no modelo.

Foram considerados dois aspectos distintos na análise comparativa dos dois algoritmos. O primeiro pretende avaliar o impacto da utilização dos dois modelos diferentes no contexto do modelo de estimação de matrizes. O segundo, directamente relacionado com o primeiro, pretende explicar possíveis diferenças de desempenho com base na análise dos algoritmos de afectação.

Sendo assim, a primeira análise é baseada no parâmetro "tempo de CPU" - tempo necessário à obtenção da solução.

Para o segundo tipo de análise foram considerados os seguintes parâmetros:

- n - número de iterações para obtenção da solução;
- Δ - medida de variação dos fluxos nos arcos da rede em iterações sucessivas dada por (9);
- Z - valor da Função Objectivo nas iterações sucessivas;
- tempo de CPU para obtenção da solução.

3.2 Método iterativo de estimação de matrizes OD com algoritmo de afectação Frank-Wolfe (FW) versus Social Pressure (SP)

A mesma solução inicial (estimativa da matriz de OD) foi utilizada para a execução do modelo de estimação de matrizes nas duas situações diferentes: algoritmo de afectação Frank-Wolfe (FW) e Social Pressure (SP).

O primeiro teste diz respeito a uma situação de nível moderado de congestionamento da rede de tráfego. Em particular, considera-se que um aumento do fluxo foi observado nos arcos L2 e L5, respectivamente de 4% e 6% em relação ao equilíbrio de tráfego existente.

Os resultados obtidos são:

- 1 - tempos de execução do modelo de estimação de matrizes ao fim de 5 iterações (5 ciclos de execução alternada dos modelos de afectação e estimação de matrizes):

Algoritmo	Tempo CPU	Tempo médio/iter
Modelo iterativo com FW	1.82	0.364
Modelo iterativo com SP.	1.44	0.288

Foi observada uma diferença significativa em termos do tempo necessário para obtenção da solução (nova matriz). Foi obtida uma redução de cerca de 20% de tempo de execução com o modelo que utiliza o algoritmo de afectação "Social Pressure". Sabendo que este algoritmo requer cálculos extra na definição da direcção descendente (cálculos associados à definição dos factores de "pressão social"), uma análise de alguns dos parâmetros do modelo de afectação permitirá explicar tal facto.

- 2 - Análise relativa ao modelo de afectação com base em dados obtidos durante a primeira iteração (afectação da estimativa da matriz de OD disponível):

iter nº	Frank-Wolfe		Social Pressure	
	Δ	Z(F.O.)	Δ	Z(F.O.)
1	0.250332	133097.50	0.287216	133097.50
2	0.136685	92107.83	0.087994	83209.17
3	0.009181	79908.94	0.021944	80033.35
4	0.005869	79867.30	0.000550	79836.46
5	0.003863	79851.36	0.000121	79836.32
6	0.002687	79843.58	0.000074	79836.28
7	0.001838	79840.23		

e os tempos de execução dessa iteração são dados por:

Algoritmo	Tempo CPU	Tempo médio/iter
FW	0.11035	0.00649
SP	0.10100	0.01430

Como pode ser observado, o algoritmo "Social Pressure" apresenta uma melhor taxa de convergência: tanto o valor da Função Objectivo como o parâmetro Δ diminuem mais rapidamente (embora haja uma pequena inversão desta tendência na 3ª iteração). O número de iterações necessário para obter convergência foi, em média, 4 a 5 vezes superior no caso do algoritmo de FW. Para um mesmo número de iterações o valor da F.O. é, com uma única excepção na fase inicial do processo, sempre inferior para o algoritmo SP.

Em relação aos tempos de execução, pode observar-se que o tempo total de execução da rotina de afectação é inferior no caso do algoritmo SP (8%) apesar do respectivo tempo médio por iteração ser superior (120%). O algoritmo "Social Pressure" requer mais tempo de execução por iteração para cálculos relacionados com a geração dos factores de "Pressão Social".

Os mesmos testes foram repetidos para um maior nível de congestionamento. Os resultados estão incluídos nas seguintes tabelas:

1 - tempos de execução do modelo de estimação de matrizes ao fim de 6 iterações (6 ciclos de execução alternada de afectação e distribuição):

Algoritmo	Tempo CPU	Tempo médio/iter
Modelo iterativo com FW	4.39	0.732
Modelo iterativo com SP	2.38	0.397

A diferença dos tempos de execução, já anteriormente verificada, é agora acentuada. Com efeito, para uma rede com nível médio de congestionamento, a redução de tempo obtida com o SP é, agora, de cerca de 46%.

2 - Análise relativa ao modelo de afectação com base em dados gerados na 2ª iteração (aumento do nível da procura da ordem dos 50%)

iter nº	Frank-Wolfe		Social Pressure	
	Δ	Z(F.O.)	Δ	Z(F.O.)
1	0.233829	172714.70	0.299541	172738.30
2	0.134838	128760.70	0.007246	108391.40
3	0.047281	112189.00	0.002103	108358.40
4	0.028943	110696.70	0.001130	108355.30
5	0.023436	110113.80	0.000496	108354.70
6	0.018328	109743.20	0.000386	108354.40
7	0.015979	109515.40	0.000206	108354.30
8	0.013502	109342.90	0.000191	108354.30
9	0.121820	109221.00	0.000112	108354.30
10	0.010711	109120.60	0.000115	108354.20
11	0.009861	109044.70	0.000074	108354.20
12	0.008884	108978.80		108354.20

e, para tempos de execução, obteve-se:

Algoritmo	Tempo CPU	Tempo médio/iter
FW	0.72266	0.00723
SP	0.10156	0.00846

As tendências verificadas nos testes anteriores ficaram aqui confirmadas: convergência mais acelerada para o algoritmo SP quer em termos do parâmetro Δ quer em termos do valor da Função Objectivo. Os valores do parâmetro Δ e da Função Objectivo para o algoritmo SP apresentam sempre valores inferiores ao longo das iterações.

A diferença dos tempos de execução é, agora, ainda maior. O algoritmo de FW requer, para um nível médio de congestionamento, cerca de sete vezes mais tempo para convergir para a solução óptima. Os tempos médios de execução são superiores (+17%) para o algoritmo "Social Pressure".

Foram ainda efectuados testes para uma situação com grande nível de congestionamento onde ainda se acentuaram as diferenças entre o desempenho das duas técnicas.

4. Conclusões

O objectivo deste trabalho consiste na análise comparativa de dois algoritmos de afectação no contexto de um modelo de estimação de matrizes em redes congestionadas de tráfego: o algoritmo de Frank-Wolfe e o algoritmo "Social Pressure". Foi utilizada uma rede simples de 2 origens e dois destinos e foram ensaiados diferentes níveis de congestionamento na rede.

Em primeiro lugar, pela análise dos tempos de execução necessários à obtenção da solução do problema de estimação de matrizes, podemos concluir que o desempenho do procedimento que utiliza o algoritmo "Social Pressure" é superior. Comparando as duas situações descritas (congestionamento de nível moderado e médio), verifica-se que a diferença é ainda mais acentuada para valores crescentes de congestionamento. Isso decorre do facto de maiores níveis de procura requererem um maior número de iterações para a obtenção da solução de equilíbrio.

Em relação à análise feita aos parâmetros gerados pelo modelo de afectação pode-se concluir que embora o algoritmo "Social Pressure" requeira mais tempo de CPU, em cada iteração, para cálculos extra relacionados com a obtenção da direcção descendente (cálculo dos factores de "Pressão Social"), é mais rápido na obtenção da solução final.

Em conclusão, podemos afirmar que, o algoritmo "Social Pressure" é mais eficiente que o algoritmo de Frank-Wolfe. Obtém-se uma aceleração da convergência quando se utiliza a direcção descendente sugerida pelo modelo "Social Pressure".

Admite-se que o ensaio de novas condições, nomeadamente em redes de maior dimensão e complexidade, permitirá generalizar estes resultados.

5. Bibliografia

- [1] Beckmann M.J., McGuire, C.B. and Winsten, C.B., *Studies in the Economics of Transportation*, Yale University Press, New Haven, Conn (1956).
- [2] Cybis, H.B.B., *A Dynamic User Equilibrium Assignment Model*, PhD Thesis, Institute for Transport Studies, University of Leeds (1993).
- [3] Frank, M., Wolfe, P., *An Algorithm for Quadratic Programming*, *Naval Research Logistics Quarterly* 3 (1956) 95-110.
- [4] Luenberger, D.G., *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Publishing Company (1973).
- [5] Sheffi, Y., *Urban Transportation Network: Equilibrium Analysis with Mathematical Programming Methods*, Prentice Hall, Englewood Cliffs, New Jersey (1985).
- [6] Van Vliet, D., *Road Assignment I, II, III*, *Transportation Research* 10 (1976) 137-143.
- [7] Van Vliet, D. and Dow, P.D.C., *Capacity restrained road assignment*, *TEC - Traffic Engineering and Control* 20 (1979) 296-305.
- [8] Van Vliet, D., *The common man's guide to the theory of equilibrium assignment*, *Lecturers notes*, University of Leeds (1979).
- [9] Van Vliet, D., *Traffic assignment models - recent advances*, *Lecturers notes*, University of Leeds (1979).
- [10] Willumsen, L.G., *Estimation of an O-D matrix from traffic counts - a review*, *Working Paper 99*, Institute for Transport Studies, University of Leeds (1978).
- [11] Willumsen, L.G., *Estimation of trip matrices from volume counts; validation of a model under congested conditions*, *Proceedings 10th PTRC Summer Annual Meeting*, University of Warwick, July 1982, England (1982).
- [12] Willumsen, L.G., *An entropy maximizing model for estimating trip matrices from traffic counts*, PhD Thesis, Institute for Transport Studies, University of Leeds. University of Leeds (1981).
- [13] Wang, J., *Accelerated Convergence in Traffic Assignment by "Perturbation" Approach*. MSc Thesis in Transport Planning and Engineering, Institute for Transport Studies, University of Leeds (1993).
- [14] Wardrop, J.G., *Some Theoretical Aspects of Road Traffic Research*, *Proceedings, Institute of Civil Engineers II* (1) (1952) 325-378.

OPTIMAL CONTROL OF DISCRETE SYSTEMS • RECENT RESULTS AND CASE STUDIES •

José A. Soeiro Ferreira

Faculdade de Engenharia do Porto - DEEC
4099 Porto Codex

INESC - Instituto de Engenharia de Sistemas e Computadores
Apartado 4433
4007 Porto Codex

Abstract

The paper has two main purposes: to give an introduction to some fundamental results of discrete-time optimal control and to show that they are useful to solve various types of problems. Maximum principles, which represent fundamental optimization principles within optimal control theory, such as the weak, the strong, the generalized and the global maximum principles are presented. Four case studies will illustrate the application of these maximum principles: an assortment problem, the optimal design of a wall, a pump-pipe system design and a consumer brand choice model.

Resumo

O artigo tem dois objectivos principais: por um lado apresentar, sucintamente, resultados fundamentais em controlo óptimo em tempo discreto e, por outro, mostrar que podem ser úteis para a resolução de diferentes tipos de problemas. Princípios do máximo, que corporizam princípios fundamentais de optimização da teoria do controlo óptimo, são incluídos, nomeadamente os princípios do máximo fraco, forte, generalizado e global. Apresentam-se ainda quatro casos com vista a ilustrar os referidos princípios e a sua aplicação: um problema de sortido, o desenho óptimo dum muro, um sistema "pump-pipe" e um modelo de "marketing".

Keywords

Optimal control, maximum principles, principle of optimality, consumer choice, assortment, wall design, pump-pipe system.

1. Introduction

Optimal control deals with the problem of governing the states of a system, by using controls, in order to maximize/minimize a particular objective or performance of the system.

The paper, which deals only with discrete-time systems, has two main purposes:

- to give an introduction to fundamental, and recent, results of the discrete-time optimal control theory, a family of maximum principles; and
- to show that they are useful to solve real problems in a variety of situations, by briefly presenting some case studies.

Many other specific applications of discrete-time optimal control theory could be mentioned within, for instance, management of energy and other resources, development of economic systems, study of ecological and physical systems, determination of design parameters for technical equipment and control of production processes.

The reason for this is that a lot of problems are naturally formulated or interpreted as discrete-time optimal control problems. This is true when considering systems that evolve in time, but also other situations can be formulated as such. On the other hand it is well-known that for those who deal directly with computers, the most important part of mathematics involved tends not to be the calculus, but the areas of discrete mathematics. Thus, although some problems are more naturally formulated as continuous-time optimal control, their solution by a digital computer needs some form of discretizations - one of which being the conversion to a discrete-time optimal control problem.

Maximum principles constitute fundamental optimization principles in the theory of discrete-time optimal control. One of their main advantages is the decomposition of the problem allowing qualitative type of analysis that is so valuable for better understanding the problem to be solved. The paper includes the weak maximum principle and the strong maximum principle. Papers with serious mistakes are still being published in well-known journals, for instance [Lin and Yang 1989] in IEEE Transactions on Automatic Control, where the differences between the weak and the strong maximum principles are not considered.

As recent results the paper presents the generalized maximum principle and the global maximum principle and reference is made to corresponding algorithms.

Four case studies illustrate the application of the different maximum principles introduced in the paper, namely an assortment problem, the optimal segmentation of a wall, a consumer brand choice model and the design of a pump-pipe system. The suitability of each principle is highly related to the actual problem in question. There is not a "best way" to solve all the problems. Attention has also to be made to the type of algorithms being developed to use each of the maximum principles. Other applications of discrete-time optimal control are referred in [Ravn and Vidal 1990].

The general optimization problem, which will be considered in the paper, denominated the Discrete-time Optimal Control (DOC) problem (also known as multistage decision problem) is defined as follows:

$$J = \max_{\{u(i)\}} \sum_{i=0}^{N-1} r_i(x(i), u(i)) \quad (1.1)$$

sub. to:

$$x(i+1) = f_i(x(i), u(i)), \quad i = 0, 1, \dots, N-1 \quad (1.2)$$

$$u(i) \in U(i) \quad i = 0, 1, \dots, N-1 \quad (1.3)$$

and:

$$x(0) = \bar{x}(0) \quad (1.4)$$

$$x(N) \text{ free} \quad (1.5)$$

where $x(i) \in \mathbb{R}^m$ is denominated the state vector, $u(i) \in \mathbb{R}^n$ is the so-called control vector, $r_i(x(i), u(i)) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}$ is the partial return cost function and $f_i(x(i), u(i)) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^m$ is the transformation function, so that (1.2) defined the dynamics of the systems. J is the total return function and $U(i) \subset \mathbb{R}^n$ denotes the set of feasible values for the control vector $u(i)$.

It is for DOC problems formulated as (1.1)-(1.5) that the main results will be presented, although they can easily be extended for other formulations, like those with final state $x(N)$ fixed or including restrictions in the states $x(i)$.

2. The Weak Maximum Principle: necessary conditions

The so-called Hamiltonian or Hamilton function plays a central role in control theory and is defined for problem (1.1)-(1.5) as follows,

$$H_i(x(i), u(i), p(i+1)) = r_i(x(i), u(i)) + p(i+1) f_i(x(i), u(i)) \tag{2.1}$$

for $i = 0, 1, \dots, N - 1$, where $p(i) \in \mathbb{R}^m$, $i = 1, 2, \dots, N$, are denominated Lagrange multipliers or adjoint variables.

For the weak maximum principle we will have:

- Assumptions 1:*
- a) $r_i(\cdot, \cdot)$ is a known differentiable function;
 - b) $U(i) = \{u(i) \mid h_i(u(i)) \leq 0, h_i(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^{s_i}\}$, for $i = 0, 1, \dots, N-1$, and $h_i(\cdot)$ are known differentiable functions;
 - c) a constraint-qualification for (1.1)-(1.5) is satisfied.

Then, applying directly Karush-Kuhn-Tucker conditions the following result can be easily shown.

Theorem 1

Assume that there exist $\{x^*(i+1)\}$, $\{u^*(i)\}$, $i = 0, 1, \dots, N - 1$, satisfying (1.2)-(1.5), then there exist $\{p^*(i+1)\}$, $\{\lambda^*(i)\}$ satisfying the following necessary conditions for optimality:

- (a) adjoint equations

$$\frac{\partial H_i(\cdot, \cdot, \cdot)}{\partial x^*(i)} = p^*(i), \quad i = 1, 2, \dots, N-1 \tag{2.2}$$

- (b) stationarity equations

$$\frac{\partial H_i(\cdot, \cdot, \cdot)}{\partial u^*(i)} - \frac{\partial h_i(\cdot)}{\partial u^*(i)} \lambda^*(i) = 0, \quad i = 0, 1, \dots, N-1 \tag{2.3}$$

$$\lambda^*(i) h_i(\cdot) = 0, \lambda^*(i) \geq 0, \quad i = 0, 1, \dots, N-1 \tag{2.4}$$

- (c) boundary conditions

$$x(0) = \bar{x}(0) \text{ and } p(N) = 0 \tag{2.5}$$

The stipulations of this theorem for more general cases can be seen in [Vidal 1981] or [Lewis 1986]. For a DOC problem with the fixed final condition $x(N) = \bar{x}(N)$, then the final adjoint variable is free, i.e., $p(N)$ is free in (2.5).

From an application viewpoint the complementary slackness conditions (2.4) are difficult to handle, but for cases where the constraints $U(i)$ are rather simple, for instance lower and upper bounds, a relaxation approach is usually recommended. That is, (1.3) is disregarded and then (2.3) will become

$$\frac{\partial H_i(\dots)}{\partial u^*(i)} = 0, \quad i = 0, 1, \dots, N-1 \quad (2.6)$$

Then from the generated stationary solutions only those where (1.3) is satisfied are selected.

This approach has been used by [Pinho and Ferreira 50/90] while solving the problem of the optimal choice of sizes, which is described in the next section.

3. Assortment Problem

When a product is produced in large quantities, some decisions must be taken regarding the choice of characteristics of the units to be produced or stocked. Only a limited number of types can be produced and these units must meet some client requirements while minimizing waste or giving profit. This problem of determining the "best" set of units to offer or stock is generally known as the assortment problem [Borges and Ferreira 1994] (see figure 1).

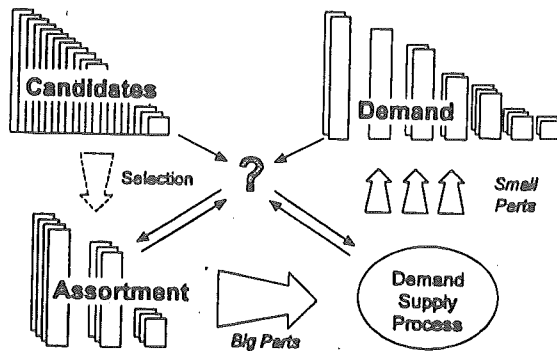


Figure 1 - Schematic representation of assortment problems

Consider, for example, the problem often faced by manufactures of determining the measurement of m sizes of a garment so that expected sales are maximized in a target population of N persons. If a person wants an item with measurement x such that $x_{i-1} \leq x \leq x_i$, where x_i , $i = 1, \dots, m$, are the measurements of the sizes offered, he or she will buy the size x_i , if $|x_i - x|$ is small, for $i \in \{1, \dots, m\}$. So the solution of the maximization of expected sales yields the same solutions as the minimization of the expected discomfort.

In order to formulate the problem consider of the following variables:

- x random variable relative to the size of an article wished by a person, following the normal distribution $h^*(x)$ with mean μ and standard deviation σ
- m numbers of sizes offered by the manufacturer.
- x_i measurement of sizes ($x_{i-1} < x_i$), $x_1 \geq x_e$ and $x_m \leq x_d$
- N number of buyers in the target population

It is assumed the probability that the person, with measurement $x = x_i$, buys an article is $c \in [0,1]$ and let $g^*(x_i, x)$ be the probability that a person with measurement x , in the range $[x_{i-1}, x_i]$ will buy a garment. Assume that each possible buyer will buy only one item so that the number of buyers is equal to the number of items sold. Consider $g^*(x_i, x) = c \exp\left(-\frac{x_i - x}{\sigma}\right)$,

where σ is the standard deviations of h^* . If $x > x_d$, then $g^*(x_m, x) = 0$.

The total expected number of buyers with measurements between x_{i-1} and x_i will be:

$$N \sum_{i=1}^m \int_{x_{i-1}}^{x_i} h^*(x) g^*(x_i, x) dx$$

If f is the unit normal distribution, then the change of variable $u = (x-u)/\sigma$ allows to write

$$u_e = (x_e - \mu)/\sigma, \quad u_d = (x_d - \mu)/\sigma, \quad h(u) = (1/\sigma)f(u), \quad g(u) = \exp(u) h(u)$$

and

$$G(u_i) = \int_{u_0}^{u_i} g(u) du.$$

Introducing the variables v_i 's (controls) where $u_{i+1} = u_i + v_i$, for $i = 0, \dots, m-1$, and defining the functions

$$r_i(u_i, v_i) = \exp(-(u_i + v_i)) [G(u_i + v_i) - G(u_i)]$$

the following DOC problem is formulated:

$$\text{Max } F/N = c\sigma \sum_{i=0}^{m-1} r_i(u_i, v_i) \tag{3.1}$$

sub. to:

$$u_{i+1} = u_i + v_i, \quad i = 0, \dots, m-1 \tag{3.2}$$

$$u_0 = u_e \tag{3.3}$$

$$u_i \in \mathbb{R}, \quad i = 1, \dots, m \tag{3.4}$$

$$v_i \geq 0, \quad i = 0, \dots, m-1 \tag{3.5}$$

The initial state u_i is fixed and the final state u_m is free. According to the model (3.1)-(3.5) the variables v_i may be zero. That will not cause any trouble because an optimal solution with positive v_i will be found and, therefore, conditions $u_{i-1} < u_i$ and then $x_{i-1} < x_i$ will be satisfied.

The Hamiltonian function is defined by

$$H_i(u_i, v_i, p_{i+1}) = r_i(u_i, v_i) + p_{i+1}(u_i + v_i), \quad i = 0, 1, \dots, m-1$$

Now applying the Weak Maximum Principle, it is easy to get the following steps.

The stationary (of the Hamiltonian) condition (2.3) may be simplified to (see also 2.6):

$$\frac{\partial H_i}{\partial v_i} = 0, \quad i = 0, 1, \dots, m-1 \tag{3.6}$$

Equations (2.4) are always verified by $g_i = 0$;

The adjoint equations are:

$$-\exp(-(u_i + v_i)) [G(u_{i+1}) - G(u_i)] + \exp(-(u_i + v_i)) g(u_{i+1}) - g(u_i) + p_{i+1} = p_i \tag{3.7}$$

From (3.6) we get

$$-\exp(-(u_i+v_i)) [G(u_{i+1})-G(u_i)] + \exp(-(u_i+v_i)) g(u_{i+1}) - g(u_i) + p_{i+1} = 0 \quad (3.8)$$

Now subtracting (3.7) to (3.8) comes:

$$p_i = -\exp(-(u_i+v_i)) g(u_i) \quad (3.9)$$

Using this expression for p_i in (3.7) we get,

$$\exp(-(u_{i+1})) [G(u_{i+1})-G(u_i) - g(u_{i+1})] = p_{i+1} \quad (3.10)$$

When $i = m-1$, and using (2.5),

$$\exp(-(u_{m-1}+v_{m-1})) [G(u_m)-G(u_{m-1}) - g(u_m)] = 0 \quad (3.11)$$

Expressions (3.2) and (3.9)-(3.11) are the optimality conditions for the sizing problem (3.1)-(3.5) and they are written in a closed form that allows to compute the variables u_i , u_m , v_i and p_{i+1} , ($i = 1, \dots, m$), after knowing the values of u_0 , p_1 and v_0 .

As a final comment let us say that optimality of the result obtained from the above mentioned expressions cannot be easily proved. The optimality conditions used are only necessary conditions.

In [Pinho and Ferreira 50/90] more details of this case are included, together with computational results.

4. The Strong Maximum Principle: necessary conditions

Let us now add to assumptions 1 of section 2. the following:

- Assumptions 2:* a) $r_i(\cdot, \cdot)$ is concave in $u(i)$, $i = 0, 1, \dots, N-1$, and
b) $U(i)$ is a convex set, $i = 0, 1, \dots, N-1$

Then we have the following result:

Theorem 2

Idem to the theorem 1 with exception of (b) that is replaced by

$$(b') \quad \max_{u(i) \in U(i)} H_i(x^*(i), u(i), p^*(i+1)), \quad i = 0, 1, \dots, N-1 \quad (4.1)$$

This theorem can easily be proved from theorem 1 and the new assumptions. Theorem 2 can be also proved under less restrictive assumptions (directional convexity), see for instance [Canon et al 1970]. These results are not very interesting from a practical viewpoint because such assumptions are difficult to verify a priori, that is before the problem is solved. The interested reader may also see [Boltyansky 1978]. Moreover, if the functions $r_i(\cdot, \cdot)$ are concave in $(x(i), u(i))$ and the $f_i(\cdot, \cdot)$ are linear, then the conditions of theorem 2 are not only necessary but also sufficient for optimality. For more information about sufficiency of the linear (Hamiltonian) maximum principle see [Vidal 1987].

Theorem 2 has been used by [Parreira and Ferreira 1989] to optimize the segmentation of a wall, problem presented in the next section. Another interesting application of theorem 2, concerning production planning is reported by [Kleindorfer et al 1975].

5. Segmentation of a Wall

Walls on nonlinear slopes are often built of horizontal blocks each requiring the digging of a trench in which the depth increases with the length of the blocks, while the height of the wall above ground decreases (see figure 2).

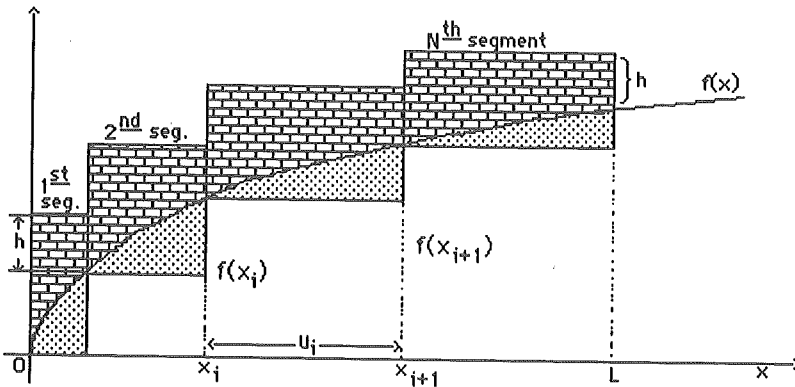


Figure 2 - Wall composed of segments built along a nonlinear surface

It is supposed that the wall to be designed is of defined length and minimum above-ground height (i.e., effective height), composed of adjoining blocks (segments) and it will be built on a surface offering a nonlinear slope contour. The length of each of the segments and the number are not known in advance, rather these will represent the decision variables to be computed, so that total costs involved in the construction are minimized. Three types of costs are considered: a fixed cost per segment, a unit cost of digging a trench for the below ground portion of a segment and a unit cost of building a segment of required effective height. The exact formulation is as follows (for more details see [Pareira and Ferreira 1989]):

$$\text{Min}_{u_i > 0} \sum_{i=0}^{N-1} \left\{ c_1 u_i [f(x_{i+1}) - f(x_i) + h] + c_2 + c_3 \left[\int_{x_i}^{x_{i+1}} f(x) dx - u_i f(x_i) \right] \right\} \quad (5.1)$$

sub. to:

$$x_{i+1} = x_i + u_i, \quad i = 0, 1, \dots, N-1 \quad (5.2)$$

$$u_i \geq 0 \quad (5.3)$$

$$u_0 = 0 \text{ and } x_N = L \quad (5.4)$$

where the x_i 's and the u_i 's are the state and control variables, respectively, and the final state x_N is fixed.

The following notation is used:

- N Number of wall segments (not necessarily known in advance)
- C_1 Building cost per unit area of wall segment
- C_2 Fixed cost for a wall segment independent of its size
- C_3 Per unit excavation cost (side view area)
- h Minimum above-ground height of a segment

- u_i Horizontal length of the $(i+1)$ th wall segment, $i = 0, 1, \dots, N-1$
 x Horizontal distance of wall from its start
 $f(x)$ Elevation of surface above a specified horizontal plane at distance x
 L Overall length of wall measured along a horizontal plane
 x_i Length of wall after segment i , ($i = 1, \dots, N$) is included. $x_0 = 0$

(The costs are relative to a wall of given width).

The optimality conditions given by the classical maximum principle, in a strong form, may be written as [Parreira and Ferreira 1989]:

$$\max H(x_i^*, u_i, p_{i+1}^*), \quad i = 0, 1, \dots, N-1 \quad (5.5)$$

$$\partial H / \partial x_i^* = p_i^*, \quad (\text{adjoint equations}) \quad (5.6)$$

$$\partial H / \partial p_{i+1}^* = x_{i+1}^*, \quad (\text{state equations}) \quad (5.7)$$

and the boundary conditions:

$$x_0 = 0 \quad \text{and} \quad x_N = L \quad (5.8)$$

Furthermore the p_i 's are the adjoint parameters and the Hamiltonian function is defined, at stage i , by:

$$H(x_i, u_i, p_{i+1}) = -r_i(x_i, u_i) + p_{i+1}(x_i + u_i)$$

where r_i is the total cost at stage i , which is included in the objective function (5.1) of problem (5.1)-(5.4).

For the type of surface contours used, and in particular those mentioned in [Parreira and Ferreira 1989], the corresponding $r_i(x_i, u_i)$ functions are strictly convex in u_i , for $u_i \geq 0$ which, along with the linearity of the state equations, implies that the directional convexity property is fulfilled. Moreover, it is easy to show that the Hamiltonian is concave in u_i ($u_i \geq 0$) for given values of x_i and p_{i+1} , thus having a unique stationary point - the global maximum in the region of interest. Therefore, the (necessary) conditions (5.5)-(5.8) also give sufficient optimality conditions. The maximization of the Hamiltonian may be substituted by the stationary condition $\partial H / \partial u_i = 0$, $u_i \geq 0$. A very simple algorithm may be developed based on those conditions. This has been done in the above mentioned paper, where an application to a particular case is also included. Similar applications could have been devised, the main idea being the approximation of a function, or collection of data, by step functions.

6. The Generalized Maximum Principle

First, we introduce the definition of the generalized Hamiltonian as:

$$H_i(x(i), u(i), \pi(i+1)) = r_i(x(i), u(i)) + \pi_{i+1}(f_i(x(i) + u(i))), \quad \text{for } i = 0, 1, \dots, N-1$$

where $\pi(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R}$, is a nonlinear function.

- Assumptions 3:*
- $r_i(x(i), u(i))$, $i = 0, 1, \dots, N-1$, is differentiable with respect to $x(i)$ at the point $x^*(i) \in \text{Int } Y(i)$ (the interior of the sets of reachable states), and
 - the upper boundary functions and the supports $\pi_i^*(x(i))$, $i = 1, 2, \dots, N-1$, are continuously differentiable at $x^*(i)$.

Then we have a nonlinear (Hamiltonian) maximum principle:

Theorem 3

If problem (1.1)-(1.5) has a solution $\{u^*(i)\}, \{x^*(i)\}$, there exists supports π_{i+1}^* to the upper boundary functions, $i = 0, 1, \dots, N-1$, so that:

(a) the generalized adjoint equations

$$\frac{\partial H_i(x^*(i), u^*(i), \pi_{i+1}^*)}{\partial x(i)} = \frac{\partial \pi_i(x^*(i))}{\partial x(i)}, \quad i = 1, 2, \dots, N-1, \quad (6.1)$$

(b) $u^*(i)$ maximizes $H_i(x^*(i), u(i), \pi_{i+1}^*)$ over $U(i)$, $i = 0, 1, \dots, N-1$ (6.2)

(c) boundary conditions

$$x^*(0) = \bar{x}(0), \quad \frac{\partial \pi_N(x^*(N))}{\partial x_N} = 0 \quad (6.3)$$

For a discussion of this matter see [Ferreira 1986] and [Nahorski et al 1983]. This reference also contains the proof of the theorem, under more general conditions. In general if the problem has a solution, (b) is always valid so the question is just to select appropriate forms of $\pi_i(\cdot)$, such that (b) is both necessary and sufficient. The adjoint equations (6.1) are not always valid, and assumptions 3 are needed. Further theoretical developments can be seen in [Nahorski and Ravn 1988].

In practice it is impossible to verify assumptions 3 apriori, usually the validity of theorem 3 is postulated and good experiences have been obtained using quadratic support functions as documented by [Pinho and Ferreira 49/90], while solving the same problem of the optimal choice of sizes presented in sec. 3, and by [Ferreira 1990], while solving a consumer brand-choice problem, which will be analysed in the next section.

Experiences developing algorithms based on the generalized maximum principle can be found in [Ravn and Vidal 1989].

7. A Model of Consumer Brand Choice

Consider a market with two brands, A and B, and assume that the consumers in this market have identical choice behaviour, i.e., their brand choice is described by the same marketing model. Time is partitioned into N periods and in each period all the consumers are assumed to buy one unit of one of the brands. The brand choice is stochastic and the probability of choosing brand A, for example, is assumed to depend on the brand chosen last time and on the price of that brand.

The literature contains a number of examples of this model structure. References may be found in [Ferreira et al 1984]. However, most of those models are based on stationary Markov chains, i.e., the probabilities of brand shifting are assumed constant for a number of periods. This means that the models may only be applied under stable marketing conditions where price, advertising, etc., are kept constant. In particular, the models cannot be used for decision purposes since no decision variables are incorporated. The literature gives a few examples of

models that take into account this problem and incorporate decision variables (often advertising) in a Markov chain model structure.

The optimization model presented in the following belongs to the latter kind. This type of model and its resolution are detailed described in [Ferreira 1990]. The dynamics of the process is constructed within the framework of first-order Markov chain, and the consumer's reaction on price changes is expressed by making the transition probabilities a function of the price of brand A. Moreover, logistic functions are used to ensure that the model becomes logically consistent, i.e., the probabilities remain in the interval $[0,1]$ regardless of the values of the decision variables. The price of brand A is the only decision variables considered because it is assumed that the price of brand B and other decision variables (quality, advertising, etc.) are kept constant. However, the model structure is general enough to allow these variables to be incorporated.

The following concepts (parameters and variables) are used:

- M number of consumers (identical behaviour)
- u price of brand A if it is kept constant for all periods
- u_i price of brand A in period i
- c variable production cost per unit of brand A (independent of i)
- C fixed cost of brand A per time period (independe of i)
- x_i market share of brand A in period i
- $x_L(u)$ equilibrium market share of brand A
- I interest rate of company A
- $p=p(u_i)$ conditional probability that a consumer buys brand B in period i, given that brand A was chosen in the previous period; it is a function of the price u_i
- $q=q(u_i)$ conditional probability that a consumer buys brand A in period i, given that brand B was chosen in the previous period; it is a function of the price u_i
- N number of periods until the time horizon is reached.

Assume that the following functional relationships exist between the price in period i and the conditional probabilities of brand choice:

$$p(u_i) = 1/[1+\exp(\alpha_0+\alpha_1 u_i)] \quad (7.1)$$

$$q(u_i) = 1/[1+\exp(\beta_0+\beta_1 u_i)] \quad (7.2)$$

where α_0 , β_0 , α_1 and β_1 are given constants (parameters).

Introducing the Markov chain assumption, the following dynamic systems is established:

$$[x_{i+1} \quad 1 - x_{i+1}] = [x_i \quad 1 - x_i] \begin{bmatrix} 1 - p(u_i) & p(u_i) \\ q(u_i) & 1 - q(u_i) \end{bmatrix}$$

or since we only consider brand A,

$$x_{i+1} = x_i[1 - p(u_i) - q(u_i)] + q(u_i), \quad i = 0, 1, \dots, N-1 \quad (7.3)$$

Using the above mentioned definitions and assumptions, it is possible to formulate the

following profit function:

$$G = M \sum_{i=0}^{N-1} (u_i - c)x_{i-1}(1 + I)^{-(i+1)} - C \sum_{i=0}^{N-1} (1 + I)^{-(i+1)} \quad (7.4)$$

The function G is the profit function of the company wishing to maximize the net present value of profits over N without posing any restrictions on the choice of u_i . The optimization problem is to determine a price policy, a sequence of prices u_0, u_1, \dots, u_{N-1} , that maximizes (7.4) taking into account that u_i influences profit possibilities in periods $i+1, i+2, N-1$. The problem thus formulated can be characterised as the nonlinear DOC problem:

$$\max G' = \sum_{i=0}^{N-1} (u_i - C)x_{i+1} \quad (7.5)$$

sub.to:

$$x_{i+1} = x_i[1 - p(u_i) - q(u_i)] + q(u_i), \quad i = 0, 1, \dots, N-1 \quad (7.6)$$

$$u_i \geq 0, \quad i = 0, 1, \dots, N-1 \quad (7.7)$$

$$x_0 = \bar{x}_0, \text{ where } \bar{x}_0, \quad 0 < \bar{x}_0 < 1, \text{ is given} \quad (7.8)$$

G' results from (7.4) if $I = 0$, which simplifies the presentation. As it is explained in [Ferreira 1990], this simplification do not compromise the algorithms used for its solution - other values for I mau be taken. Function $p(\cdot)$ and $g(\cdot)$ are defined as in (7.1) and (7.2). State equations (7.6) allow x_i to take values only in the interval $(0,1)$.

Model (7.5)-(7.8) has been solved by an algorithm based on Theorem 3 - The Generalized Maximum Principle (see [Ferreira 1990]). The choice of this approach was due to the fact that the property of directional convexity of the sets of all possibles values of states and criteria is not fulfilled - a convexity condition is required for the derivations of the Strong Maximum Principle.

For problem (7.5)-(7.8) the generalized Hamiltonian function takes the form

$$H_i(x_i, u_i, \pi_{i+1}) = (u_i - c) x_{i-1} + \pi_{i+1}(x_{i-1}), \quad i = 0, 1, \dots, N-1$$

where

$$x_{i+1} = x_i[1 - p(u_i) - q(u_i)] + q(u_i)$$

The functions π_{i+1} are considered as quadratic functions because these are the cases which have been mostly studied and for which some simple examples have been considered [Nahorski et al 1983]. The define

$$\pi_{i+1}(x_{i+1}) = a_{i+1}(x_{i+1})^2 + b_{i+1}x_{i+1}, \quad i = 0, 1, \dots, N-1$$

Details of the algorithms used and some computational results may be found in [Ferreira 1990]. The algorithms is, naturally, more complicated than the typical ones used for maximum principles with linear Hamiltonians.

8. The Global Maximum Principle: sufficient conditions

The global maximum principle decomposes the original control problem in two steps, as stated in the next result.

Theorem 4

If a state vector $\{x^*(i)\}$, and a control vector, $\{u^*(i)\}$, satisfy the following two principles:

- (a) The maximum principle, i.e. for $i = 0, 1, \dots, N-1$, $u^*(i)$ is the solution to

$$H_i(x^*(i), u^*(i+1)) = \max_{\{u(i)\}} r_i(x^*(i), u(i)) \quad (8.1)$$

sub. to:

$$x^*(i+1) = f_i(x^*(i), u(i)) \quad (8.2)$$

$$u(i) \in U(i),$$

and

- (b) The principle of optimality, i.e., for $i = 0, 1, \dots, N-1$, $x^*(i+1)$ is the solution to

$$F_i(x^*(i)) = \max_{\{x_{i+1}\}} [H_i(x^*(i), x(i+1)) + F_{i+1}(x(i+1))] \quad (8.3)$$

and

$$F_N(\bar{x}(N)) = 0,$$

then $\{x^*(i)\}$ and $\{u^*(i)\}$ are optimal.

[Vidal 1986] has shown that for more general problems than (1.1)-(1.5), and under very mild assumptions, these two principles are sufficient for optimality.

The main advantage of this principle is its generality, the $r_i(\dots)$ need not to be continuous and the control and state variables might take only integer values.

Next section describes an application of the Global Maximum Principle to the optimal design of a pump-pipe system. [Vidal 1986] has also applied this principle to solve a production planning problem.

9. Optimization of a Pump-Pipe System

A pump-pipe system is to be designed. This problem is formulated as follows:

$$R = \max_{\{\Delta p(i); d(i)\}} \left\{ - \sum_{i=0}^{N-1} b(\Delta p(i)) + \Psi_i(d(i))(p(i) + \Delta p(i)) + w_i(d(i)) \right\} \quad (9.1)$$

sub.to:

$$p(i+1) = p(i) + \Delta p(i) - \varphi_i(d(i)) \quad i = 0, \dots, N-1 \quad (9.2)$$

$$p(0) = P_0, p(N) = P_N \quad (\text{initial and final conditions}) \quad (9.3)$$

$$p(i) \geq P_i \geq 0 \quad i = 1, 2, \dots, N-1 \quad (9.4)$$

$$\Delta P_i^m \leq \Delta p(i) \leq \Delta P_i^M \quad i = 0, 1, \dots, N-1 \quad (9.5)$$

$$d(i) \in \mathcal{D} \quad i = 0, 1, \dots, N-1 \quad (9.6)$$

where the state variables $p(i)$ denote the inlet pressure at the i^{th} stage, and the control vectors have two elements, $\Delta p(i)$, the pump size, and $d(i)$, the diameter of the pipe, at i^{th} stage.

The total variable costs are to be minimized. At each stage we have the annual pump costs, $b(\Delta p(i))$, and the annual cost of pipe, $\Psi_i(d(i))(p(i) + \Delta p(i)) + w_i(d(i))$. Finally, $\varphi_i(d(i))$ denotes the pressure loss in a pipe of diameter $d(i)$ at the i^{th} stage (see figure 3). The

optimization is subject to lower bounds on the inlet pressure, lower and upper bounds on the pump size, and a set of permissible diameters \mathbb{D} . For more details see [Ferreira and Vidal 1984].

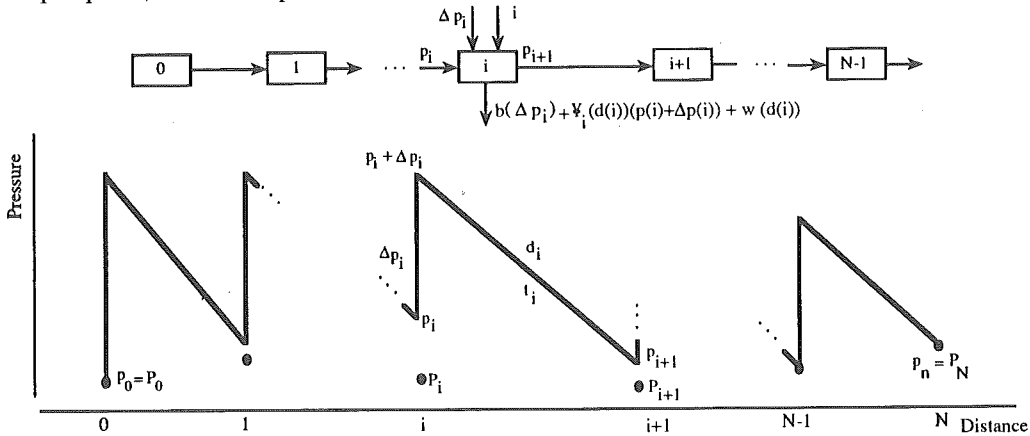


Figure 3 - Multistage process and corresponding pressure-distance profile

Problem (9.1)-(9.6) is somehow a generalization of problem (1.1)-(1.5), since there we have constraints in the state variables, but as it may be seen below these allow for an effective use of the Global Maximum Principle.

Let us start the resolution by studying the following sub-problem

$$H_i(p(i), p(i+1)) = \max_{\{\Delta p(i); d(i)\}} \left\{ -b(\Delta p(i)) - \Psi_i(d(i))(p(i) + \Delta p(i)) - w_i(d(i)) \right\}$$

sub.to:

$$\begin{aligned} \Delta p(i) - \varphi_i(d(i)) &= p(i+1) - p(i) \\ \Delta P_i^m \leq \Delta p(i) &\leq \Delta P_i^M \\ d(i) &\in \mathbb{D} \end{aligned}$$

This optimization problem in two variables can be reduced to an optimization problem in one variable, $d(i)$, by using the relation $\Delta p(i) = \varphi_i(d(i)) + p(i+1) - p(i)$ then, the subproblem becomes

$$H_i(p(i), p(i+1)) = \max_{\{d(i)\}} \left[-b(\varphi_i(d(i)) + p(i+1) - p(i)) - \Psi_i(d(i))(p(i+1)) + \varphi_i(d(i)) - w_i(d(i)) \right]$$

sub.to:

$$\begin{aligned} \Delta P_i^m - p(i+1) - p(i) &\leq \varphi_i(d(i)) \leq \Delta P_i^M - p(i+1) + p(i) \\ d(i) &\in \mathbb{D} \end{aligned}$$

Since usually \mathbb{D} is a finite set with only a few elements, this optimization problem can be easily solved by simple enumeration. Now, we have to find the optimal state vector. It can be shown that if $b(\Delta p(i))$ is convex, decreasing and differentiable then, at the optimal solution, the variables $p(i)$ have to be as small as possible. That is $p^*(i) = P_i$, $i = 1, 2, \dots, N-1$, will be the optimal solution as far as all the sub-problems have at least one feasible solution. Using the recurrence equations of the global maximum principle more cases can be solved [Vidal 1986]. For further discussions and other approaches see [Ferreira and Vidal 1984].

10. Conclusions

The paper includes several approaches, within the optimal control of discrete-time systems theory, which may be useful to solve real problems. These are the weak and the strong maximum principles, the generalized maximum principle and the global maximum principle, which are introduced as theorems.

Four case studies are presented, which have been solved by applying those maximum principles. The variety of situations illustrates, in general, the potentiality of the optimal control theory and shows that the suitability of the maximum principles is highly related to the actual problem in question, in particular it depends on the properties of the objective function. Other types of approaches could have been tried to solve these optimal control problems, such as mathematical programming or dynamic programming, but the advantage of using maximum principles resides in the fact that the n -dimensional problem is decomposed in a series of one-dimensional problems.

It is known that the optimality conditions of the maximum principles give not only quantitative but also qualitative insight to the properties of the optimal solution. Moreover, the straightforward application of those optimality conditions allows the construction of simple algorithms with attractive computational complexity.

Mathematical programming may be used in discrete systems. In fact the DOC problem may be interpreted as a mathematical programming problem with a particular structure and there exist a theoretical foundation involving many fruitful concepts and results and computational methods for solving such problems. However, it looks that stronger theoretical results and better numerical methods may be obtained by directly exploiting the discrete control structure. But for that, it should be of interest to integrate in the optimal control theory the great advances of mathematical programming which took place over the last 20 years. In particular the concepts of duality and sensitivity analysis could have a substantial importance, together with the practical experiences with algorithms and practical problem solving. In this connection we refer to [Ferreira and Vidal 1986], [Nahorski et al 1983] and [Nahorski and Ravn 1988].

On the other hand, it is interesting to observe that the linkage between maximum principles and dynamic programming was never seriously considered. In that line, this paper includes the global maximum principle (and an application), which combines both a maximum principle and the optimality principle, the basis of dynamic programming. Some other investigations may also be found in the papers mentioned above.

References

- [1] Boltyansky, V., *Optimal Control of Discrete Systems*, J. Wiley (1978).
- [2] Borges, P. and Ferreira, J.S., *Overview of Assortment Problems*, Proceedings of the 3rd Meeting of the Nordic Section of the Mathematical Programming Society, Linköping, Sweden (1994).
- [3] Canon, M.D. et al, *Theory of Optimal Control and Mathematical Programming*, Mc-Graw Hill (1970).
- [4] Ferreira, J.S. and Vidal, R.V., *Optimization of a Pump-Pipe System by Dynamic Programming*, Engineering Optimization 7-3 (1984) 241-251.
- [5] Ferreira, J.S., Nilsson, O. and Ravn, H., *Optimal Price Policy in a Dynamic Model of Consumer Brand Choice*, IMSOR 16/84 (1984).
- [6] Ferreira, J.S., *Controlo Ótimo em Tempo Discreto - Uma Introdução*, Ciclo de Conferências em Análise Numérica e Optimização, Dept. Matemática da Universidade de Coimbra (1986).
- [7] Ferreira, J.S. and Vidal, R.V., *On the Connections between Mathematical Programming and Discrete Optimal Control*, Proceedings of the 12th IFIP Conference on System Modelling and Optimization, A. Prékopa (et al) eds, Springer-Verlag (1986) 234-243.
- [8] Ferreira, J.S., *Maximum Principles Applied to a Model of Consumer Brand Choice*, Optimal Control Applications & Methods 11 (1990) 21-37.
- [9] Kleindorfer, P.R. et al, *Discrete Optimal Control of Production Plans*, Management Sciences 22 (1975) 261-272.
- [10] Lewis, F., *Optimal Control*, J. Wiley (1986).
- [11] Lin, T.Y. and Yang, Z.H., *A Discrete Optimal Control Problem for Descriptor Systems*, IEEE Transactions on Automatic Control 34 (1989) 177-181.
- [12] Nahorski, Z., Ravn, H. and Vidal, R.V.V., *Optimization of Discrete Time Systems - The Upper Boundary Approach*, Springer-Verlag (1983).
- [13] Nahorski, Z. and Ravn, H., *The Upper Boundary Approach to Constrained Discrete-time Optimal Control*, Control and Cybernetics 17 (1988) 145-172.
- [14] Parreira, C.M. and Ferreira, J.S., *An Application of the Discrete Maximum Principle to the Segmentation of a Wall for Non-linear Slopes*, Engineering Optimization 14 (1989) 259-266.
- [15] Pinho, M.R. and Ferreira, J.S., *A Sizing Problem Solved by the Generalized Maximum Principle*, INESC AR-49/90 (1990).
- [16] Pinho M.R. and Ferreira, J.S., *Determinação Ótima de Medidas Estandarizadas*, INESC AR-50/90 (1990).
- [17] Ravn, H. and Vidal, R.V., *A Discrete-time Optimal Control Algorithm Based on the Generalized Maximum Principle*, IMSOR, Technical University of Denmark (1989).
- [18] Ravn, H. and Vidal, R.V., *Engineering Applications of Discrete Time Optimal Control*, EJOR 45 (1990) 241-250.
- [19] Vidal, R.V.V., *Notes on Static and Dynamic Optimization*, IMSOR, Technical University of Denmark (1981).
- [20] Vidal, R.V.V., *A Global Maximum Principle for Discrete-time Control Problems*, Engineering Optimization 10 (1986) 77-84.
- [21] Vidal, R.V.V., *On the Sufficiency of the Linear Maximum Principle for Discrete-time Control Problems*, J. of Opt. Theory and Applications 54 (1987) 583-589.

FORECASTING HORTICULTURAL PRICES: A COMPARISON OF METHODS

Diogo Furtado Luzes(*)

Instituto Superior de Agronomia
Universidade Técnica de Lisboa

Isabel Hall Themido

CESUR - Instituto Superior Técnico
Universidade Técnica de Lisboa
Av. Rovisco Pais
1000 Lisboa

Abstract

Forecasts of horticultural prices provide useful information for farmers and retailers. These prices have large fluctuations as a result of seasonal changes in production and the lagged response relationship between a decision to produce and the availability of the crop. Inaccurate price forecasts may lead to a decision not to harvest the crop, with a significant loss to the farmer.

This paper compares different forecasting approaches applied to Portuguese horticultural data prices. The following steps are proposed: 1) spectral analysis for the detection of the dominating and hidden frequencies, 2) modeling the price series using decomposition methods, SARIMA models and causal models followed by an univariate procedure and, finally, 3) diagnostic checking using the last year of available data.

Results show that for vegetables produced mainly under protected conditions (greenhouses) simple univariate models fit well because the planted area has not changed over the last years. When the production comes only from open fields, the use of causal variables representing the planted area, is necessary. Cyclical behaviour is also reported (Cobweb model) in the case of melons, production being dictated by previous season's price level.

Resumo

A previsão dos preços de produtos hortofrutícolas constitui informação de interesse para agricultores e retalhistas. Estes preços têm grandes flutuações resultantes de variações sazonais da produção e do intervalo de tempo que medeia entre a decisão de produzir e a disponibilidade da colheita. Previsões erradas podem conduzir à decisão de não efectuar a colheita causando eventualmente graves prejuízos ao agricultor.

Apresentam-se vários métodos alternativos de previsão de sucessões de preços de produtos hortofrutícolas Portugueses. Seguem-se os seguintes passos: 1) análise espectral para detecção de frequências dominantes e escondidas, 2) modelação das sucessões de preços recorrendo a modelos de decomposição, modelos SARIMA e modelos causais seguidos de modelação univariável e 3) avaliação da qualidade das previsões para o último ano de dados disponíveis.

Os resultados sugerem que para hortaliças produzidas sobretudo em estufa os modelos univariável simples se ajustam bem porque a área plantada não sofreu alteração nos últimos anos. Quando a produção provém de campos descobertos, o uso de variáveis explicativas que representem a área plantada é indispensável. Detectou-se ainda um comportamento cíclico no caso do melão (modelo Cobweb) em que a produção de um ano é determinada pelo preço do ano anterior.

Keywords

Forecasting, horticultural prices, Cobweb model

(*) published after the death of the author

1. Introduction

Producers, processors and distributors of agricultural commodities make decisions in a risky environment. Uncertain and seasonal production, lagged response relationships between a decision to produce and the availability of the crop and relatively low elasticities of demand provide the setting for large fluctuations in prices.

Forecasts of horticultural prices provide useful information for farmers and retailers. Farmers use it for production planning purposes: deciding which crops to introduce each year and when to produce, so that harvest time occurs during the highest price period. Inaccurate price forecasts may lead to a decision not to harvest the crop, with a significant loss to the farmer. Retailers use it for establishing supply contracts (buying when the price is low and then storing products if possible) and for forecasting sales (sales are usually a function of price).

The studies on horticultural prices forecasting are rare. Brandt and Bessler (1983) found SARIMA (Seasonal Autoregressive Integrated Moving Average) models with one seasonal difference, to be the most adequate for prediction of agricultural prices. They used hog prices, which have a different behaviour from horticultural prices. Sapsford and Varoufakis (1987) came to the same conclusion, the SARIMA formulation was appropriate for predicting tea prices. Martin and Albisu (1990) consider that horticultural prices are mainly determined by last week's price (autoregressive process) and by the supply in that week. The problem of predicting supply with accuracy is overcome in the study reported in this paper, introducing a dummy variable which corresponds to a rudimentary prediction of supply. Albisu and Astorquiza (1986) compare econometric models (or causal models) with SARIMA models and choose the econometric model for prediction of lettuce prices. Janecke (1989) is also an apologist of the causal procedure using explanatory variables such as production, imports and consumption, all of them difficult for long-term prediction. He reports that horticultural prices have cycles, for example lettuce prices show a two monthly cycle, equivalent to the growing period of that crop.

The purpose of this paper is to compare the accuracy of several forecasting models for weekly horticultural prices. The paper is organized in four sections: data description, methodology, a summary of results including the statistical performance of alternative forecasting methods and conclusions.

2. Data description

Horticultural prices, calculated as the average of the main Portuguese markets, are published weekly by S.I.M.A. (Information Service about Agricultural Markets).

In our analysis four crops were selected (green bean, lettuce, melon and strawberry). All available prices up to October 1993 were used. This meant a four year series for lettuce and six year series for all other vegetables (Figure 1). Price data was converted to 1993 prices.

Green bean and lettuce prices are available for all 52 weeks of the year. The difference between these two crops is that green bean is mainly cultivated in greenhouses while lettuce is an open field crop, except during the winter period. Melon prices were collected for 13 weeks, from 8 July to 6 October and there is no greenhouse production of this crop. Strawberry prices are for 28 weeks, from 15 January to 28 July, and this crop is of mixed greenhouse and open field production.

All series display a strong seasonal pattern with peaks in the winter period, not so evident in the case of melons. Trend is notorious in lettuce prices; the other series do not exhibit a decrease or increase over time.

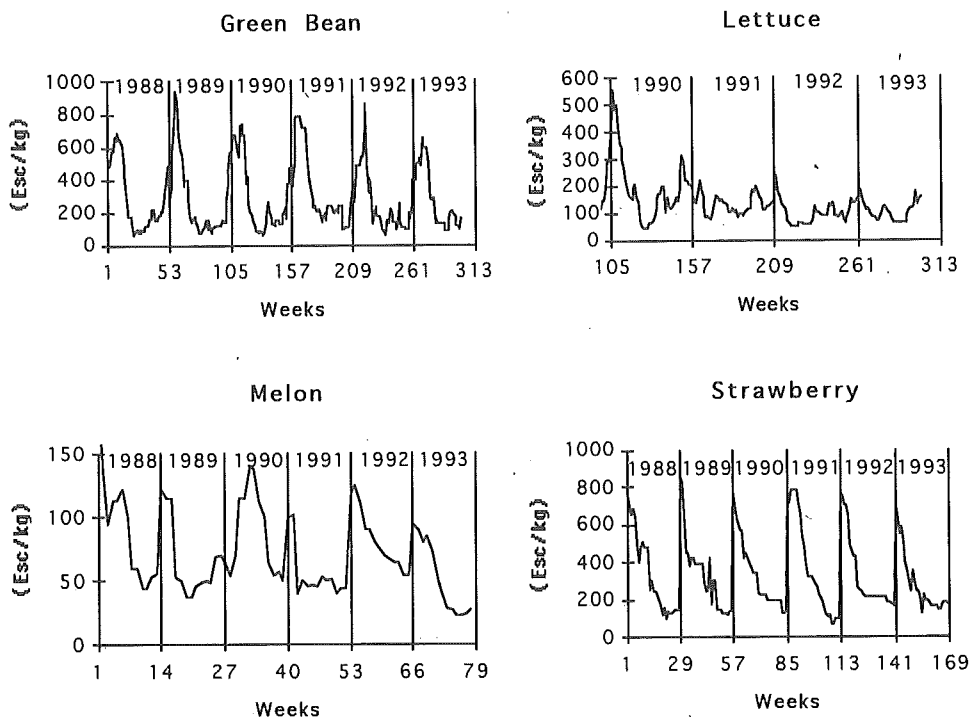


Figure.1 - Weekly prices (Escudos/kg) for four Portuguese horticultural crops

3. Methodology

For forecasting horticultural prices the following steps are proposed: 1) spectral analysis for detection of dominating and hidden frequencies, 2) modeling the price series using appropriate models and 3) diagnostic checking using the last year of available data.

Spectral analysis is the name given to the methods used to estimate the spectral density function or spectrum. It permits the variance to be divided into contributions attributable to different frequencies; a peak in the spectrum indicates an important contribution to variance at

that frequency. With the spectral density function seasonality can be more precisely determined and hidden periodicities identified, after removing trend and obvious seasonality. For a complete presentation of spectral analysis refer to Chatfield (1989).

For forecasting we used decomposition methods, SARIMA models and causal models. Decomposition methods identify and separate components of a time series: trend, seasonal and cyclical factors. As our series are short, we did not consider the cycle. The decomposition methods can be additive or multiplicative. The last form is more appropriate, as in price series seasonality is proportional to trend (Makridakis and Wheelright, 1989; Jarret, 1989);

$$Z_t = T_t \cdot S_t \cdot R_t; \text{ when } T_t = 0, Z_t = L_t \cdot S_t \cdot R_t$$

(at period t : Z_t = time series value; T_t = trend component including level; S_t = seasonal component; L_t = series level; R_t = random component).

Seasonal factors were obtained through moving averages of length determined by the spectral analysis and trend factors were obtained assuming a simple linear regression.

Using the SARIMA models, Box and Jenkins methodology (Box and Jenkins, 1976), we first removed the trend and seasonality obtaining stationary series. After examining the autocorrelation (ACF) and partial autocorrelation (PACF) functions, the model was chosen and parameters were estimated. The ACF and PACF of residuals were computed, to check if all pattern had been removed.

Considering the horticultural crops, the supply available in any one year depends on current production and, to a lesser extent, on imports. As in agriculture a main factor in price variability are the changes in supply, causal models were constructed using the planted area as an input variable, residuals were treated by a decomposition method.

Forecast evaluation should be based on out-of-sample forecasting experience (Brandt and Bessler, 1983). That is, the period over which the evaluation is constructed should not be the same as that over which the parameters of the model were fitted. The twelve months of 1993 will be our evaluation period. Among the commonly used forecasting evaluation methods is the mean squared error (MSE). Because the error is squared, large errors detract substantially more from the performance of the forecasting method than do small errors. The mean absolute percentage error (MAPE) is a valid alternative to MSE. It incorporates a linear loss criteria as opposed to the quadratic loss of the MSE ($MSE = \left[\sum (Z_{tp} - Z_t)^2 \right] / n$; $MAPE = \left[\sum |Z_{tp} - Z_t| / Z_t \right] / n$; Z_{tp} = prediction).

4. Results

4.1. Spectral Analysis

As expected the main peaks in the spectrum occurred at a frequency of one cycle per year: green bean and lettuce with cycles of 52 weeks, melon with a cycle of 13 weeks and strawberry with a cycle of 28 weeks, shown in Figure 2. Spectral analysis would not have been necessary in this case as the seasonal effect is obvious anyway.

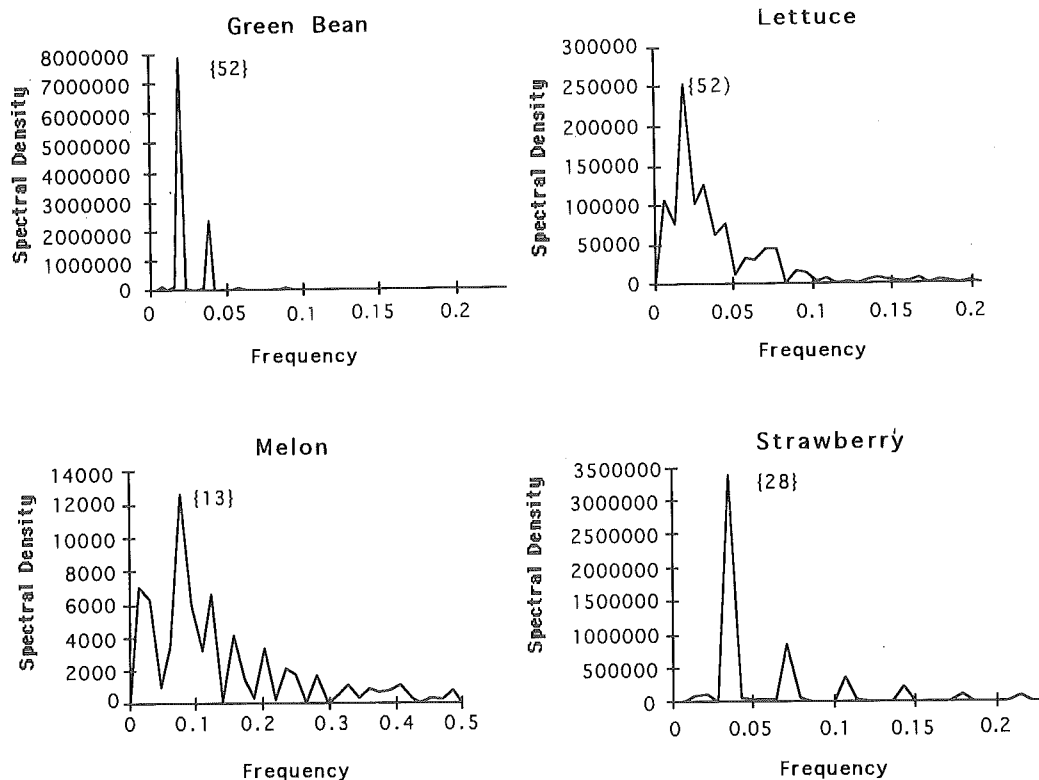


Figure 2 - Spectral analysis: determination of the spectrum of the four series, still nonstationary. Values in { } are the inverse frequencies expressed in weeks

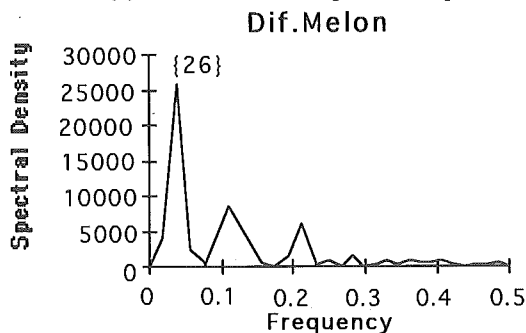


Figure 3 - Spectrum of the melon differenced series. Values in { } are the inverse frequencies expressed in weeks

After differencing the series in order to remove from the data obvious seasonality and trend (lettuce case), we carried out a new spectral analysis. Melon data was the only one with concentrated variance at 0.0385 frequency (36 weeks) (Figure 3).

The other three cases presented a flat spectrum after the removal of the main seasonality. This means that the melon data exhibits two important cycles, one annual and the other biannual. The other series only have an annual cycle, without any hidden periodicities.

4.2 Forecasting methods

4.2.1 Decomposition methods

Seasonal factors were calculated using moving averages of the length determined by spectral analysis. Unfortunately, there was not enough data to calculate the biannual factor for the melon prices. Only lettuce prices exhibit trend. Models are presented below leaving out the random component for the sake of simplicity ($t = \text{week}$):

$$\text{Green bean prices: } Z_t = L_t \cdot S_t = 296.338 S_t$$

$$\text{Melon prices: } Z_t = L_t \cdot S_t = 73.05 S_t$$

$$\text{Lettuce prices: } Z_t = T_t \cdot S_t = (258.115 - 0.686 \text{ Week number}_t) S_t$$

$$\text{Strawberry prices: } Z_t = L_t \cdot S_t = 339.309 S_t$$

4.2.2 SARIMA models

To achieve stationarity the main seasonal patterns were removed. For lettuce, prices trend was also removed. The models are:

Green Bean prices = $(1,0,0) (1,1,0)_{52} = (1-0.702B) (1+0.643B^{52}) (1-B^{52})Z_t = a_t$; residual standard deviation = 69.7

$$Z_t = 0.702.Z_{t-1} + 0.357.Z_{t-52} - 0.252.Z_{t-53} + 0.643.Z_{t-104} - 0.454.Z_{t-105}$$

Lettuce prices = $(0,1,0) (1,1,0)_{52} = (1-B) (1+0.330B^{52}) (1-B^{52})Z_t = a_t$; residual standard deviation = 39.6

$$Z_t = Z_{t-1} + 0.670.Z_{t-52} - 0.670.Z_{t-53} + 0.330.Z_{t-104} - 0.330.Z_{t-105}$$

Melon prices = $(1,0,0) (1,1,0)_{13} = (1-0.608B) (1+0.347B^{13}) (1-B^{13})Z_t = a_t$; residual standard deviation = 31.1

$$Z_t = 0.608.Z_{t-1} + 0.665.Z_{t-13} - 0.397.Z_{t-14} + 0.347.Z_{t-26} - 0.211.Z_{t-27}$$

Strawberry prices = $(1,0,0) (1,1,0)_{28} = (1-0.731B) (1+0.448B^{28}) (1-B^{28})Z_t = a_t$; residual standard deviation = 68.7

$$Z_t = 0.731.Z_{t-1} + 0.551.Z_{t-28} - 0.403.Z_{t-29} + 0.448.Z_{t-56} - 0.327.Z_{t-57}$$

Results show similarity between the different price models; they all have one seasonal difference, and correspond to an autoregressive process, seasonal and nonseasonal (the lettuce model, due to the single differencing, does not have an autoregressive parameter).

The models have small residual variance and seem to be quite adequate. No remaining pattern was evident from the analysis of residuals.

4.2.3 Causal models

Planted area is needed for building the causal models. For green bean and strawberry crops this variable was constant for the years studied, so causal models were not fitted. Significant variation in planted area for the lettuce and melon crops are reported, but information about actual planted area is not available.

The melon case is a typical example of the Cobweb model; production is dictated by previous season's price level (Figure 4). When price level is high, the next year's planted area will be increased resulting in higher production and an accentuated decrease in price. Quantities and prices have a cyclical behaviour and are linked recursively in a causal chain over time. This reaction is valid because a time lag exists, between the decision to produce and harvesting, and farmers base production plans on last season's prices. The Cobweb model was first presented by Ezekiel in 1938; Tomek and Robinson (1981) provide a more recent summary on this subject.

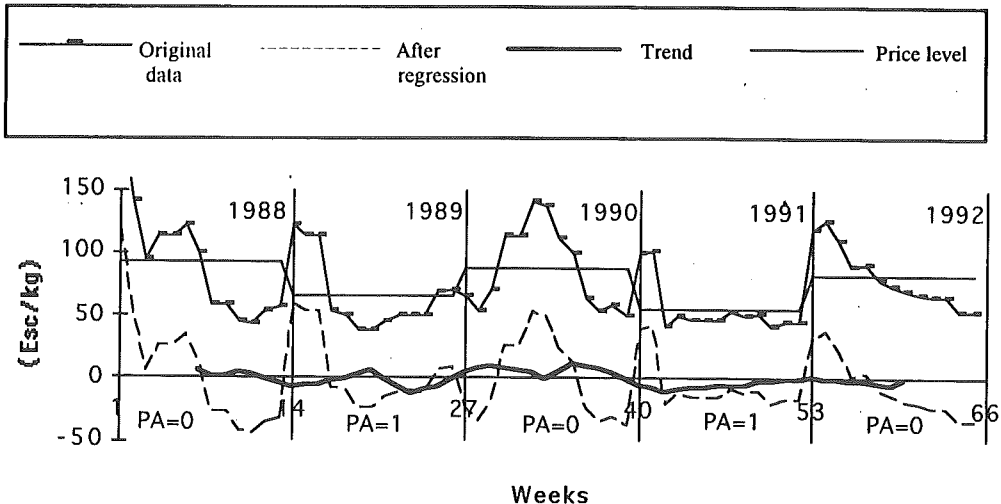


Figure 4. Cyclical behaviour of melon weekly prices (PA = Planted area)

Planted area (PA), a dummy variable, will assume zero value in 1988, 1990 and 1992, meaning that supply was small and it will assume a value equal to 1 in 1989, 1991 and 1993, meaning that supply was large:

$$\text{Melon prices: } Z_t = (87.333 - 26.526 \text{ PA}_{\text{year}}) S_t$$

Planted area for lettuce has been increasing since 1990 and this is the real reason behind the negative price trend. Unfortunately, the information about planted area was not available making it impossible to build a causal model for this crop.

4.3. The evaluation of forecasts

Using the MAPE or the MSE as performance measures, it appears that the multiplicative decomposition models performed better for the green bean and strawberry price series and the causal models for melon price series (Table 1). SARIMA models also performed well, with a small difference from the other univariate procedures and for the lettuce price series performed better than multiplicative decomposition models. SARIMA strawberry forecasts have the lowest MAPE but they show a considerable difference from real data at the beginning of the season, so MSE is very high.

CROP	FORECASTING MODEL	MSE	MAPE
Green bean	Multiplicative Decomposition Method	5037	0.2196
	SARIMA	7324	0.3265
Lettuce	Multiplicative Decomposition Method	2963	0.4293
	SARIMA	834	0.2379
Strawberry	Multiplicative Decomposition Method	6874	0.2482
	SARIMA	11786	0.2444
Melon	Multiplicative Decomposition Method	726	0.7317
	SARIMA	658	0.6870
	Causal followed by a Mult. Decomp.Method	437	0.5850

Table.1 - Statistical performance evaluation; comparison between real (1993) and predicted prices.

The performance of alternative forecasting methods is also shown in Figure 5. For the green bean, lettuce and strawberry crops predicted values were very close to real prices. Melon prices show an important decrease in August, not predicted by any forecasting method. An excessive supply was expected, but not to such an extent.

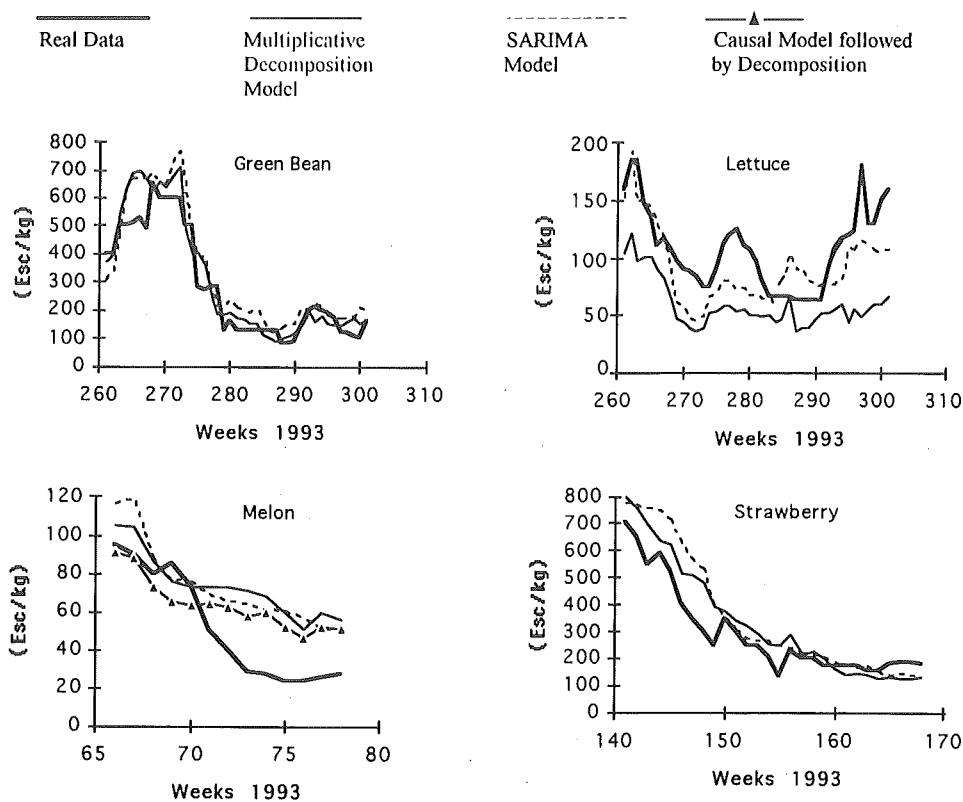


Figure 5 - Predicted and real values for 1993 of green bean, lettuce, melon and strawberry prices.

5. Conclusions

The set of forecasting procedures discussed in this analysis by no means exhausts the list of forecasting alternatives. The objective here was to develop and evaluate a selected group of methods which are relatively simple to use and update.

The autoregressive process in price formation, referred to by several authors, was confirmed; causal models were also appropriate for predicting melon prices. Three conclusions may be drawn from the evaluation of the proposed forecasting models:

- 1) For forecasting prices of greenhouse products simple univariate models, such as the Multiplicative Decomposition or the SARIMA models, fit well. Over the last 4 years these products had the same area under cultivation, because investments are high and subsidies are non-existent.
- 2) To forecast prices of open field products it is necessary to use explanatory variables, such as the planted area.
- 3) Two cycles were identified for the melon price series by spectral analysis; seasonal production is responsible for the annual cycle, last season's price level is responsible for the biannual cycle. This pattern leads to a typical example of a Cobweb model.

References

- [1] Albisu, L. and L. Astorquiza, El comportamiento de los precios de lechuga en Mercabilbao, *Inv.Agrar.Econo.*1 (1986) 61-69.
- [2] Brandt, J. and Bessler, Price forecasting and evaluation: an application in agriculture, *Journal of Forecasting* 2(1983) 237-248.
- [3] Chatfield, C., *The Analysis of Time Series, An introduction*, Chapman & Hall, 4th edit. (1989).
- [4] Ezekiel, J., The Cobweb Theorem, *Quarterly Journal of Economics* LII (1938) 255-280.
- [5] Janecke, T., Market models for analyzing and forecasting the prices of horticultural crops, *Acta Horticulturae* 248 (1989) 151-152.
- [6] Jarret, J., Forecasting Monthly Earnings per share, time series models, *OMEGA* 17 (1989) 37-44.
- [7] Makridakis, S. and S. Wheelwright, *Forecasting Methods for Management*, John Wiley & Sons, 5th edit. (1981).
- [8] Martin, D. and L. Albisu, Determinacion a corto plazo del precio de las hortalizas, *I Congr. Cien. Horticolas ACTA* 5 (1990) 48-53.
- [9] Sapsford, D. and Y. Varoufakis, An ARIMA analysis of Tea prices, *Journal of Agricultural Economics* 38 (1987) 329-334.
- [10] Tomek, W. and K. Robinson, *Agricultural Product Prices*, Cornell University Press, 2th edit. (1981).



DISTRIBUIÇÃO DO PERÍODO DE OCUPAÇÃO DA FILA DE ESPERA $MID|\infty$

Manuel Alberto Martins Ferreira

Dept. de Ciências de Gestão
I.S.C.T.E.
Av. das Forças Armadas
1600 Lisboa

Abstract

The busy period of the $MID|\infty$ system is intensively studied as a particular case of the $MIG|n$ system busy period, to be a reference. We show that it is possible to compute exactly any parameter of its distribution and to interpret its probability density function. The only problem is how to compute its distribution function for which we give an algorithm.

Resumo

O período de ocupação do sistema $MID|\infty$ é muito estudado no âmbito do período de ocupação do sistema $MIG|\infty$, mais geral, por servir como referência. Mostramos que se consegue calcular exactamente qualquer parâmetro da sua distribuição e interpretar a sua função de densidade de probabilidade. O único problema reside no cálculo da sua função de distribuição para o qual apresentamos um algoritmo.

Keywords

Busy period, Probability density function, Distribution function.

1. Introdução

O período de ocupação de uma fila de espera é um período que se inicia quando um cliente chega ao sistema estando ele vazio, termina quando um cliente abandona o sistema deixando-o vazio e em que há, sempre, pelo menos um cliente no sistema.

Pretende-se estudar a distribuição da duração do período de ocupação da fila de espera $MID|\infty$.

Numa fila de espera $MIG|\infty$

- λ é a taxa do processo de Poisson de chegadas,
- $G(\cdot)$ é a f.d. do serviço,
- $\alpha = \int_0^{\infty} [1 - G(t)]dt$ é o tempo médio do serviço,
- $\rho = \lambda\alpha$ é a intensidade de tráfego,
- cada cliente encontra sempre, ao chegar, um servidor disponível.

Designaremos por B a duração do período de ocupação e por $b(t)$, $B(t)$ e $\bar{B}(s)$ as suas f.d.p., f.d. e transformada de Laplace-Stieltjes, respectivamente.

$\bar{B}(s)$ pode ser descrita em termos de $G(\cdot)$ e λ do modo seguinte:

$$\bar{B}(s) = 1 + \lambda^{-1} \left(s - \frac{1}{\int_0^{\infty} e^{-st-\lambda \int_0^t [1-G(\vartheta)]d\vartheta} dt} \right) \quad (1.1)$$

(para detalhes ver, por exemplo, Takács (1962)).

A inversão desta transformada não é fácil, do ponto de vista analítico, com excepção de certos casos particulares (ver Ferreira (1991) e Ferreira (1995)).

No caso dos serviços terem duração constante de valor α (sistema $MID|\infty$) a expressão obtida por inversão de (1.1), embora complexa, é de interpretação fácil devido ao facto de, neste sistema, a ordem de saída dos clientes ser idêntica à de chegada.

Mostraremos que, para este sistema, se podem obter expressões exactas para os momentos de B .

Já para $B(t)$ assim não acontece.

Na Secção 2 apresentamos um estudo de $b(t)$ para o sistema $MID|\infty$.

Na Secção 3 estudamos os momentos de B para o sistema $MID|\infty$.

Na Secção 4 apresentamos os resultados analíticos possíveis para $B(t)$ para este sistema.

Finalmente, na Secção 5, apresentamos um algoritmo que permite o cálculo aproximado de $B(t)$.

2. Função de Densidade de Probabilidade de B para o Sistema $MID|\infty$

Para o sistema $MID|\infty$, (1.1) toma a forma

$$\bar{B}(s) = 1 + \frac{1}{\lambda} \left(s - \frac{(s+\lambda)s}{\lambda e^{-(s+\lambda)\alpha} + s} \right) \quad (2.1)$$

obtendo-se, por inversão da transformada de Laplace-Stieltjes (ver, ainda, Ferreira (1995))

$$b(t) = \sum_{n=0}^{\infty} \left(\frac{d}{dt} \frac{c(t)}{e^{-\rho t}} \right) * \left(\frac{d}{dt} \frac{1-d(t)}{1-e^{-\rho t}} \right)^{*n} e^{-\rho t} (1-e^{-\rho t})^n \quad (2.2)$$

em que

– * é o operador convolução,

$$\frac{c(t)}{e^{-\rho t}} = \begin{cases} 0, & t < \alpha \\ 1, & t \geq \alpha \end{cases} = G(t),$$

$$\frac{1-d(t)}{1-e^{-\rho t}} = \begin{cases} \frac{1-e^{-\lambda t}}{1-e^{-\rho t}}, & t < \alpha \\ 1, & t \geq \alpha \end{cases}$$

Note-se que a saída do sistema M|D|∞, em equilíbrio, é um processo de Poisson de parâmetro λ (Daley (1976)).

Assim, podemos dar a (2.2) a forma

$$b(t) = \sum_{n=0}^{\infty} g(t) * \left(\frac{d}{dt} P(t_0 \leq t | t_0 \leq \alpha) \right)^{*n} e^{-\rho} (1-e^{-\rho})^n \quad (2.3)$$

em que g(·) é a f.d.p. do serviço e t₀ é o tempo que decorre entre duas saídas consecutivas da fila de espera.

Portanto,

- g(t) é a densidade correspondente à contribuição para o período de ocupação do primeiro cliente,
- $\frac{d}{dt} P(t_0 \leq t | t_0 \leq \alpha)$ é a densidade correspondente à contribuição para o período de ocupação do n-ésimo cliente, n ≥ 2,
- $(1-e^{-\rho})^n e^{-\rho}$ é a probabilidade de o período de ocupação acabar na saída n+1, n = 0, 1, 2,...

e, portanto,

- $g(t) * \left(\frac{d}{dt} P(t_0 \leq t | t_0 \leq \alpha) \right)^{*n}$ é a densidade de um período de ocupação com n+1 clientes, n = 0, 1, 2,.... . Esse período terá uma duração inferior ou igual a (n+1)α, n = 0, 1, 2,....

3. Momentos de B para o Sistema M|D|∞

A expressão (1.1) é equivalente a $(\lambda O(B, \bar{)}(s) - 1) (C(s) - 1) = \lambda^{-1} s C(s)$ em que $C(s) = \int_0^{\infty} e^{-st - \lambda \int_0^t [1-G(\vartheta)] d\vartheta} \lambda (1-G(t)) dt$. Derivando n vezes e ordem a s, tendo em conta a fórmula de Leibnitz e fazendo s = 0 obtém-se

$$E[B^n] = (-1)^{n+1} \left\{ \frac{e^{\rho}}{\lambda} n C^{(n-1)}(0) - e^{\rho} \sum_{p=1}^{n-1} (-1)^{n-p} \binom{n}{p} E[B^{n-p}] C^{(p)}(0) \right\}, n = 1, 2, \dots \quad (3.1)$$

em que

$$C^{(n)}(0) = \int_0^{\infty} (-t)^n e^{-\lambda \int_0^t [1-G(\vartheta)] d\vartheta} \lambda (1-G(t)) dt, n = 0, 1, 2, \dots \quad (3.2).$$

Para o sistema M|D|∞

$$\begin{aligned} C^{(n)}(0) &= 1 - e^{-\rho} \\ C^{(n)}(0) &= -e^{-\rho} (-\alpha)^n - \frac{n}{\lambda} C^{(n-1)}(0), n = 1, 2, \dots \end{aligned} \quad (3.3)$$

e podemos calcular valores exactos para os momentos de qualquer ordem em torno da origem, não sendo necessário recorrer a métodos numéricos, como acontece para a generalidade dos sistemas M|G|∞ (ver Ferreira e Ramalhoto (1994)).

Assim, tem-se que

$$- n = 1$$

$$E[B] = \frac{e^{\rho}-1}{\lambda} \quad (3.4)$$

como, aliás, para qualquer sistema $M|G|∞$,

$$- n = 2$$

$$E[B^2] = \frac{2e^{\rho}(e^{\rho}-1-\rho)}{\lambda^2} \quad (3.5)$$

De (3.5) e (3.4) deduz-se que

$$\text{VAR}[B] = \frac{2e^{\rho}(e^{\rho}-1-\rho)-(e^{\rho}-1)^2}{\lambda^2} \quad (3.6)$$

Por sua vez, de (3.6) e (3.4) conclui-se que

$$\gamma_B^2 = \frac{2e^{\rho}(e^{\rho}-1-\rho)}{(e^{\rho}-1)^2} - 1 \quad (3.7)$$

em que γ_B é o coeficiente de variação de B.

De (3.7) conclui-se que

- γ_B só depende de ρ ,

- $\gamma_B < 1$,

- $\lim_{\rho \rightarrow \infty} \gamma_B = 1$ (valor do coeficiente de variação para a distribuição exponencial).

Calculando alguns valores de γ_B obtém-se:

ρ	γ_B
.5	.40655883
1	.56798436
10	.99959129
20	.99999999
50	.99999999
100	.99999999

com 8 algarismos significativos. Note-se que a partir de $\rho = 10$ γ_B , na prática, não se distingue de 1.

Sendo β_1 o coeficiente de simetria de B (ver Kendall e Stuart (1979) e Murteira (1979), por exemplo) obtém-se

$$\beta_1 = \frac{(2e^{3\rho}-6(1-\rho)e^{2\rho}-6(1+\rho)e^{\rho}-12\rho+10)^2}{(e^{2\rho}-2\rho e^{\rho}-1)^3}, \quad c \neq 0 \quad (3.8)$$

Como é evidente, $\lim_{\rho \rightarrow \infty} \beta_1 = 4$ (valor do coeficiente de simetria para a distribuição

exponencial). Calculando alguns valores de β_1 obtém-se

ρ	β_1
.5	6.0360869
1	4.5899937
10	4.0000001
20	4.0000000
50	4.0000000
100	4.0000000

e, assim, a partir de $\rho = 10$ β_1 praticamente não se distingue de 4. Por outro lado tem-se sempre $\beta_1 > 0$ o que indica uma distribuição assimétrica à direita.

Sendo β_2 o coeficiente de achatamento (ou curtose) de B (ver ainda Kendall e Stuart (1979) e Murteira (1979)) obtém-se

$$\beta_2 = \frac{(9e^{4\rho} - 24e^{3\rho} + 96e^{2\rho} + 36e^{\rho} - 117 + 72\rho e^{\rho} - 60\rho e^{3\rho} - 96\rho e^{2\rho} - 72e^{\rho}\rho^2 + 24e^{2\rho}\rho^2 - 4\rho^3 - 36\rho^2 - 160\rho) / (e^{2\rho} - 2\rho e^{\rho} - 1)^2}{p \neq 0} \quad (3.9)$$

Como é evidente, $\lim_{\rho \rightarrow \infty} \beta_2 = 9$ (valor do coeficiente de achatamento para a distribuição exponencial). Calculando alguns valores de β_2 obtém-se

ρ	β_2
.5	11.142336
1	9.6137084
10	9.0000000
20	9.0000000
50	9.0000000
100	9.0000000

e, assim, a partir de $\rho = 10$ β_2 praticamente não se distingue de 9. Tem-se sempre $\beta_2 > 3$ o que indica distribuições, para B, leptocúrticas (com um pico bastante acentuado).

4. Função de Distribuição de B para o Sistema M|D|∞

A partir de (2.1) pode obter-se

$$B(t) = 1 - \lambda^{-1} \sum_{n=1}^{\infty} \left(\frac{d}{dt} \frac{1-d(t)}{1-e^{-\rho}} \right)^* n (1-e^{-\rho})^n \quad (4.1)$$

(para detalhes ver Ferreira (1995)). Mas, para α e ρ suficientemente grandes

$$\frac{d}{dt} \frac{1-d(t)}{1-e^{-\rho}} \cong \lambda e^{-\lambda t}, \quad t \geq 0 \quad (4.2).$$

E, nestas condições, podemos concluir que

$$B(t) \cong 1 - e^{-\lambda e^{-\rho t}}, t \geq 0 \quad (4.3)$$

ou seja: que B é aproximadamente exponencial, em coerência com o que vimos em 3. Para γ_B , β_1 e β_2 .

Por outro lado mostra-se também que (ver ainda Ferreira (1995))

$$e^{-\rho G(t)} \leq B(t) \leq G(t) \quad (4.4).$$

Assim, para valores de ρ próximos de 0, $B(t) \cong G(t)$. E sendo t_p o quantil correspondente à probabilidade de p , para B, tem-se, a partir de (4.4), para o sistema M|D|∞

$$t_p = \alpha, p \leq e^{-\rho} \quad (4.5).$$

$$t_p \geq \alpha$$

Finalmente, recorrendo à desigualdade de Tchebychev mostra-se que (ver ainda Ferreira (1995))

$$B(t) \geq 1 - \frac{e^{2\rho} - 2\rho e^{\rho} - 1}{(1 + \lambda t - e^{\rho})^2} \quad (4.6)$$

desde que $t > \lambda^{-1} \left[e^{\rho} - 1 + \max \{ e^{\rho} - 1; \sqrt{e^{2\rho} - 2\rho e^{\rho} - 1} \} \right]$.

5. Algoritmo para o Cálculo aproximado de B(t) para o Sistema M|D|∞.

Não havendo uma expressão prática para o cálculo de distribuição do período de ocupação do sistema M|D|∞, excepto para valores de ρ muito baixos ou muito altos como vimos em 4, vamos apresentar um método de cálculo para B(t) baseado num algoritmo de Platzman, Ammons e Bartholdi III (1988).

Seja X uma variável aleatória e A um número dado. Platzman, Ammons e Bartholdi III (1988) apresentam um algoritmo para calcular um valor aproximado de $P[X > A]$ que exige o conhecimento da Transformada de Laplace-Stieltjes de X, L(s), em forma simples. Assim, em princípio, este algoritmo poder-se-ia aplicar ao cálculo de B(t), para qualquer sistema M|G|∞, dado que conhecemos $\bar{B}(s)$ (expressão (1.1)). Mas, de facto, não acontece assim porque apenas para os casos particulares referidos em 1. (ver Ferreira (1991) e Ferreira (1995)), em que o cálculo de B(t) não apresenta qualquer problema, e serviços constantes (expressão (2.1.)) $\bar{B}(s)$ tem uma forma simples.

Diz-se, geralmente, que:

- um algoritmo é exacto ("accurate") se procura resolver um problema "próximo" do que é suposto estar a resolver,
- um algoritmo é preciso ("precise") se obtém uma solução "próxima" da verdadeira solução do problema que está a tentar resolver.

Mais concretamente, sendo ΔA ($\Delta A > 0$) a exactidão e Δp ($0 < \Delta p < \frac{1}{2}$) a precisão requeridas a aproximação τ de $P[X > A]$ deve satisfazer a condição

$$P[X \geq A + \Delta A] - \Delta p \leq \tau \leq P[X > A - \Delta A] + \Delta p \quad (5.1).$$

Platzman, Ammons e Bartholdi III (1988) sugerem que se faça

$$\tau = \frac{U - A + \Delta A}{U - L + 2\Delta A} + \sum_{n=1}^N \frac{\alpha^{n^2}}{\pi n} \operatorname{im} \left\{ (\beta^n - \gamma^n) \cdot L(j\omega n) \right\} \quad (5.2)$$

em que

- $K = \ln \left(\frac{2}{\Delta p} \right)$,
- $D = \frac{\Delta A}{\sqrt{2K}}$,
- $\omega = \frac{2\pi}{U - L + 2\Delta A}$,
- $N = \left\lceil \frac{2K}{\omega \Delta A} \right\rceil$, em que $[x]$ é a característica de x ,
- $\alpha = e^{-D^2 \omega^2 / 2}$,
- $\beta = e^{j(U + \Delta A)\omega}$,
- $\gamma = e^{jA\omega}$,
- U e L são números tais que $1 - P[L \leq X \leq U] \ll \Delta p$
- $j = \sqrt{-1}$ e
- $\operatorname{im}(\cdot)$ designa a parte imaginária de um número complexo

e provam que a aproximação assim definida satisfaz (5.1).

Em termos gerais este método pode ser descrito do seguinte modo: para garantir uma execução rápida, calculam-se apenas N valores da transformada. Esses valores são cuidadosamente selecionados de modo a fornecerem tanta informação quanto possível. Calcula-se então o valor exacto da cauda correspondente à função de distribuição mais suave cuja transformada passa por esses N pontos.

Espera-se que um método deste tipo se comporte pelo menos tão bem como qualquer outro método que calcule N valores da transformada, e mais rápido que qualquer algoritmo que tenha que a calcular mais vezes.

Platzman, Ammons e Bartholdi III (1988) mostram, também, que calcular uma cauda a partir de uma transformada é um problema de grau de dificuldade $\#P$ -hard. Isto é indicativo do esforço computacional exigido porque resolver um problema $\#P$ -hard, mesmo apenas com uma certa garantia de aproximação, requer um acréscimo de cálculo que cresce exponencialmente com a dimensão da descrição do problema. Note-se, aliás, que o algoritmo dá uma solução, não do problema original, mas de uma sua aproximação definida por ΔA e Δp .

Repare-se que na definição do erro usada, ΔA refere-se a uma perturbação do parâmetro A enquanto que a definição mais comum de erro se refere apenas a uma perturbação Δp do resultado.

Podemos aplicar este algoritmo no cálculo da função de distribuição do período de ocupação da fila de espera $M|D|\infty$ ($B(t)$) porque, neste caso, a transformada de Laplace-Stieltjes tem uma forma simples (expressão (2.1)).

Então para aplicar o algoritmo ao cálculo da função de distribuição do período de ocupação do sistema $M|D|\infty$ teremos

- $L = a$ (em vez de α por razões óbvias)

$$- U = \lambda^{-1} \left(e^{\rho} - 1 + \sqrt{\frac{e^{2\rho} - 2\rho e^{\rho} - 1}{\Delta p}} \cdot 10^l \right), l = 1, 2, \dots$$

visto que $1 - P[L \leq X \leq U] = 1 - P[a \leq X \leq U] = 1 - P[0 \leq X \leq U]$ tendo que ser $B(U) > 1 - 10^{-l} \Delta p$ e isso acontece desde que $\frac{e^{2\rho} - 2\rho e^{\rho} - 1}{(1 + \lambda U - e^{\rho})^2} = 10^{-l} \Delta p \Leftrightarrow$

$$(1 + \lambda U - e^{\rho})^2 = \frac{e^{2\rho} - 2\rho e^{\rho} - 1}{\Delta p} 10^l \text{ o que conduz ao valor indicado (recordar$$

(4.6)),

- $A = t$ (tempo),

- os valores obtidos $B_c(t)$ são dados por $1 - \tau$.

Tomando $l = 3$ construímos o programa de computador em linguagem FORTRAN para implementar o algoritmo (é necessário indicar os valores de a , $A = t$, $\Delta A = \Delta t$ e Δp) que se segue em Anexo.

Indicamos a seguir resultados da aplicação desse algoritmo ao cálculo da função de distribuição do período de ocupação do sistema $M|D|\infty$, nos seguintes casos:

1. $\alpha = .1$ e $\lambda = 1$,
2. $\alpha = 1$ e $\lambda = 1$,
3. $\alpha = 3$ e $\lambda = 3$.

Confrontamos os valores de $B_c(t)$ obtidos com os extremos inferiores $B_1(t) = 1 - \frac{e^{2\rho} - 2\rho e^{\rho} - 1}{(1 + \lambda t - e^{\rho})^2}$ (ver expressão (4.6)) e $B_2(t) = \begin{cases} 0, & t < a \\ e^{-\rho}, & t \geq a \end{cases}$ (ver expressão (4.4)).

Os valores de $B_c(t)$ satisfazem sempre aos de $B_1(t)$ (que por vezes são triviais) e também aos de $B_2(t)$.

Note-se que o período de ocupação deste sistema tem uma concentração de probabilidade em $t = a$ de valor $e^{-\rho}$ (ver 2). Assim, para testar a validade dos valores obtidos, calculamos a média e a variância a partir dos valores de $B_c(t)$ obtidos, mas considerando $B_c(\alpha) = e^{-\rho}$, e confrontamo-las com os verdadeiros valores.

Os valores obtidos para média estão bastante próximos dos verdadeiros. Já os obtidos para a variância apresentam erros maiores. Tal é natural atendendo a que o cálculo da variância acumula os erros de cálculo dos 1º e 2º momentos centrados na origem. Em suma, em face dos erros observados pode concluir-se que os resultados obtidos através $B_c(t)$ são bastante satisfatórios.

Note-se ainda que, em princípio, os valores obtidos podem ser melhorados diminuindo Δt (exactidão) e Δp (precisão) (ver Anexo). E dizemos em princípio porque a corrida do programa é bastante morosa e essa morosidade aumenta com a diminuição de Δt e Δp .

Cálculo da Função de Distribuição do Período de Ocupação do Sistema $M|D|\infty$

1 - $\alpha = .1$; $\lambda = 1$; $\rho = .1$

$\Delta t = .001$; $\Delta p = .001$

t	$B_1(t)$	$B_2(t)$	$B_c(t)$
.11	-14.805062	.904837	.94131
.15	.816597	.904837	.950782
.2	.959013	.904837	.996209
.25	.982428	.904837	.999575

	Exactos	Calculados a partir de $B_c(t)$ com $B_c(.1) = .904837$	Erro
E[B]	.105170918	.1049714128	0,2%
VAR[B]	.0003685744	.00031661238	14%

2 - $\alpha = 1$; $\lambda = 1$; $\rho = 1$

$\Delta t = .1$; $\Delta p = .001$

t	$B_1(t)$	$B_2(t)$	$B_c(t)$
2	-11.001397	.367879	.741497
3	.420202	.367879	.907228
4	.817048	.367879	.969885
5	.911558	.367879	.992784

	Exactos	Calculados a partir de $B_c(t)$ com $B_c(1) = .367879$	Erro
E[B]	1.718281828	1.6649785	3%
VAR[B]	.9524924414	.70343785	26%

3 - $\alpha = 3$; $\lambda = 1$; $\rho = 3$ $\Delta t = .5$; $\Delta p = .01$

t	$B_1(t)$	$B_2(t)$	$B_c(t)$
4	-.238790	.0497871	.099527
5	-.420929	.0497871	.148885
6	-.646402	.0497871	.198405
7	-.930133	.0497871	.244893
8	-1.294064	.0497871	.288204
9	-1.771539	.0497871	.329391
10	-2.415214	.0497871	.368208
15	-15.889655	.0497871	.530699
20	-336.121704	.0497871	.65134
25	-7.0691347	.0497871	.740937
30	-1.366543	.0497871	.807469
35	-.113102	.0497871	.856896
40	.355496	.0497871	.893608
45	.580208	.0497871	.920880
50	.705018	.0497871	.941125
55	.781435	.0497871	.956144
60	.831591	.0497871	.967298
70	.891248	.0497871	.981726
75	.909828	.0497871	.986298
80	.924024	.0497871	.989706
85	.935113	.0497871	.992233

	Exactos	Calculados a partir de $B_c(t)$ com $B_c(3) = .0497871$	Erro
$E[B]$	19.08553692	18.60845683	2%
$VAR[B]$	281.9155718	250.9048589	11%

6. Conclusões

O período de ocupação do sistema $M|D|∞$ é muito estudado no âmbito do período de ocupação do sistema $M|G|∞$, mais geral, por servir como referência.

Como vimos, conseguimos calcular, sem qualquer dificuldade, qualquer parâmetro relacionado com a sua distribuição, e interpretar perfeitamente a sua função de densidade de probabilidade.

O único problema reside no cálculo da função de distribuição. Precisamente, apresentamos um algoritmo para o seu cálculo, que funciona bem à parte problemas de tempo.

Vimos ainda que para grandes intensidade de tráfego ($\rho \geq 10$) a distribuição do período de ocupação é aproximadamente exponencial, simplificando-se assim bastante a sua aplicabilidade a situações práticas.

7. Bibliografia

- [1] Daley, D.J., *Queueing output processes*, J.A.P. 8 (1976) 395-415.
- [2] Ferreira, M.A.M., *Um Sistema $M|G|∞$ com período de ocupação exponencial*, Actas da XV Jornadas Luso-Espanholas de Matemática, Vol. IV, Universidade de Évora, Évora (1991).
- [3] Ferreira, M.A.M., *Comportamento Transeunte e período de ocupação de sistemas de fila de espera sem espera*, Tese de Doutoramento, I.S.C.T.E. (1995).
- [4] Ferreira, M.A.M. e Ramalhoto, M.F., *Estudo de parâmetros básicos do período de ocupação da fila de espera $M|G|∞$* , A Estatística e o Futuro e o Futuro da Estatística, Actas do I Congresso Anual da Sociedade Portuguesa de Estatística, Edições Salamandra, Lisboa (1994).
- [5] Kendall e Stuart, *The advanced theory of statistics. Distribution Theory*, Charles Griffin and Co., Ltd., London (1979).
- [6] Murteira, B.J.F., *Probabilidades e Estatística*, Vol. I, Editora McGraw-Hill de Portugal, Lisboa (1979).
- [7] Platzman, L.K., Ammons, J.C. e Bartholdi III, J.J., *A simple and efficient algorithm to compute probabilities from transforms*, O.R. 36 (1988) 137-144.
- [8] Takács, L., *An introduction to queueing theory*, Oxford University Press, New York (1962).

ANEXO

PROGRAM TPROG

```

REAL      A, DELTA, DELTP, APEQ, LAMBDA, RO
REAL      KAPA, D, OMEGA, ALFA, U, PI, TAU, X, Y, SOMA, XX, XXX
COMPLEX   BETA, GAMA, CC, CLAMB, CAPEQ, CL
INTEGER   N, I

```

```
DATA      PI/3.14157/
```

```

PRINT *, ' A          '
READ *, ' A          '
PRINT *, ' APEQ       '
READ *, APEQ
PRINT *, ' LAMBDA     '
READ *, LAMBDA
PRINT *, ' DELTA      '
READ *, DELTA
PRINT *, ' DELTP      '
READ *, DELTP

```

```

RO = LAMBDA * APEQ
U = EXP(2*RO) - 2*RO*EXP(RO) - 1
U = (U / DELTP) * 1000
U = SQRT(U)
U = EXP(RO) - 1 + U
U = U / LAMBDA

```

```

PRINT *, ' U = ', U
PAUSE

```

```

KAPA = LOG(2/DELTP)
D = DELTA/SQRT(2*KAPA)
OMEGA = 2*PI / (U - APQ + 2*DELTA)

```

```
N = NINT (2*KAPA/(DELTA*OMEGA))
```

```

PRINT *, ' N = ', N
PAUSE

```

```
ALFA = EXP (-D*OMEGA)**2/2)
```

```

X = COS( (U+DELTA)*OMEGA)
Y = SIN( (U+DELTA)*OMEGA)
BETA = CMPLX(X, Y)

```

```

X = COS(A*OMEGA)
Y = SIN(A*OMEGA)
GAMA = CMPLX(X, Y)

```

```
SOMA = 0
```

```

DO 100 I=1, N
  X = OMEGA * REAL (I)
  CC = CMPLX(0.0, X)
  CLAMB = CMPLX(LAMBDA, 0.0)

```

```
CAPEQ = CMPLX(APEQ, 0.0)
CL     = CC * (CC + CLAMBD)
CL     = CL / (CLAMB * CEXP(-(CC+CLAMBD)*CAPEQ) + CC)
CL     = (CC - CL) - CLAMBD + CMPLX(1, 0)
```

```
CL     = (BETA**I - GAMA**I) * CL
```

```
X      = AIMAG(CL)
```

```
XXX    = PI * REAL(I)
```

```
Y      = ((ALFA**REAL(I)**REAL(I)) / (XXX))
```

```
SOMA   = SOMA + Y*X
```

```
100 CONTINUE
```

```
TAU = 1 - (U-A+DELTA)/(U-APEQ + 2*DELTA) - SOMA
```

```
PRINT *, 'TAU = ', TAU
```

```
STOP
END
```



MODELAÇÃO DE SISTEMAS: UMA EXPERIÊNCIA DE ENSINO NA FCT-UNL

Ruy Araújo da Costa

Dept. Matemática da FCT-UNL
Quinta da Torre
2825 Monte de Caparica

Abstract

Systems Modelling is a very important area of Operational Research and Systems Engineering. Its peculiar nature makes it a difficult subject to teach and to learn.

In this paper we describe the introduction of this subject in the Mathematics Course (5-year course) of the Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT-UNL) in 1993/94. Results of the first two years of teaching that subject already show that the introduction of Systems Modelling in this course was quite important, although we can expect some improvements to be made in the future.

This paper presents the goals to attain, the topics to be taught, the references, the class planning, the assessment of results and the analysis of results.

Resumo

A Modelação de Sistemas tem uma grande importância dentro da Investigação Operacional e Engenharia de Sistemas. A natureza peculiar deste domínio, faz com que o seu ensino, a nível de Licenciatura, se revista de algumas dificuldades tanto para docentes como para discentes.

Neste artigo pretende-se apresentar a experiência vivida na Licenciatura em Matemática da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT-UNL) com a introdução da disciplina de Modelação de Sistemas no respectivo plano de estudos em 1993/94. A experiência obtida com o funcionamento da disciplina em dois anos lectivos permite uma perspetivação da sua importância na Licenciatura em Matemática, embora se reconheça que no futuro serão certamente feitas melhorias.

Relativamente a esta disciplina apresentar-se-ão: os objectivos a atingir, o programa adoptado, a bibliografia recomendada, o planeamento das aulas, a avaliação de conhecimentos e os resultados obtidos.

Keywords

Teaching O.R., Systems modelling.

1. Introdução

A Licenciatura em Matemática da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT-UNL) existe desde 1983/84, permitindo aos alunos a especialização em um dos quatro Ramos seguintes:

- Formação Educacional
- Ciências Estatísticas
- Matemáticas Actuarias
- Investigação Operacional

Os dois primeiros anos curriculares são comuns a todos os alunos da Licenciatura, sendo a especialização feita nos Ramos nos três anos seguintes. O último ano envolve sempre a realização de um Estágio.

No Quadro 1 seguinte são apresentadas as disciplinas da área de Investigação Operacional, o respectivo semestre curricular, carga horária semanal e Unidades de Crédito:

Disciplina	Carga Horária Semanal	U.C.	Ramo da Licenciatura em Matemática			
			Formação Educacional	Ciências Estatísticas	Matemáticas Actuais	Investigação Operacional
Grafos e Aplicações	3T+3P	4,0	4º	4º	4º	4º
Programação Matemática	3T+3P	4,0	6º	6º	6º	6º
Modelação de Sistemas	3T+3P	4,0	-	7º(*)	-	7º
Elementos de I.O.	3T+3P	4,0	6º(*)	8º(*)	8º(*)	-
Investigação Operacional (A)	3T+4TP	5,5	-	-	-	8º
Seminário em I.O.	-	5,0	-	-	-	9º
Estágio Profissionalizante (IO)	600h ⁽¹⁾	20,0	-	-	-	9º/10º

Notas: (*) - As disciplinas assinaladas têm carácter opcional.

(1) - Carga horária mínima total do Estágio Profissionalizante (I.O.)

Quadro 1 - Disciplinas da área de I.O. na Licenciatura em Matemática da FCT-UNL

Para um melhor enquadramento, dever-se-á referir ainda que fazem parte da formação curricular comum dos alunos da Licenciatura em Matemática as disciplinas de "Aspectos Fundamentais da Matemática" (1º semestre curricular), "Introdução aos Computadores e Programação" (1º semestre curricular), bem como "Probabilidades e Estatística (A)" (3º semestre curricular). A primeira das disciplinas referidas é leccionada pelo Departamento de Matemática, possibilitando aos alunos uma panorâmica de diferentes áreas da Matemática, incluindo um primeiro (e breve) contacto com a Investigação Operacional. A segunda disciplina referida é leccionada pelo Departamento de Informática, permitindo aos alunos obter alguma familiarização com os meios informáticos. A terceira disciplina referida é leccionada pelo Departamento de Matemática, permitindo aos alunos a obtenção de uma sólida formação base na área das Probabilidades e Estatística (essencialmente em Estatística Paramétrica).

A propósito do Quadro 1 poder-se-á referir alguns aspectos:

- Todos os alunos da Licenciatura em Matemática da FCT-UNL têm, pelo menos, duas disciplinas obrigatórias da área de I.O. (no 4º e 6º semestres curriculares).
- A disciplina de "Elementos de Investigação Operacional" (E.I.O.) é ainda oferecida como opção aos alunos dos Ramos de Formação Educacional, Ciências Estatísticas e Matemáticas Actuais. Esta disciplina aborda cinco áreas distintas: Gestão de Stocks, Previsão, Redes e Grafos/Gestão de Projectos, Teoria da Decisão e Simulação, estando "emparelhada" com a disciplina de "Investigação Operacional

(A)" (I.O.(A)) destinada exclusivamente aos alunos do Ramo de Investigação Operacional.

A carga horária semanal destas duas disciplinas coincide no que diz respeito às aulas teóricas (3 horas), pelo que as referidas aulas teóricas são dadas em simultâneo. Relativamente à disciplina de E.I.O., a carga horária semanal compreende ainda 3 horas de aulas práticas. Por outro lado, a carga horária semanal da disciplina de I.O.(A) compreende ainda 4 horas de aulas teórico-práticas. Assim, embora se verifique a coincidência dos programas das duas disciplinas e das respectivas aulas teóricas, a sua diferenciação é feita no aprofundamento dos temas nas aulas "não teóricas", reservando-se uma hora semanal em I.O.(A) para aplicações informáticas ligadas aos temas em análise.

- As disciplinas de "Investigação Operacional (A)", "Seminário em I.O." e "Estágio Profissionalizante (I.O.)" são obrigatórias para os alunos do Ramo de Investigação Operacional, respectivamente nos 8º, 9º e 9º/10º semestres curriculares.
- A disciplina de "Modelação de Sistemas" é obrigatória para os alunos do Ramo de Investigação Operacional, sendo ainda uma possível opção dos alunos do Ramo de Ciências Estatísticas. Assim, para todos os possíveis destinatários, esta disciplina (do 7º semestre curricular) antecede a sua disciplina principal da área de I.O. (E.I.O. ou, se os alunos forem do Ramo de Investigação Operacional, I.O.(A)).

A experiência ligada à docência de "Investigação Operacional (A)" permitiu constatar que muitos alunos apresentavam dificuldades na articulação de conceitos de Probabilidades e Estatística, com algumas áreas da Investigação Operacional. Pôde ainda constatar-se um nível médio muito baixo na capacidade de programação de computadores, o que inviabilizava um processo de aprendizagem mais rico que envolvesse a programação de computadores (ainda que a nível básico) para permitir a implementação de alguns modelos (por exemplo, de simulação).

A experiência global ligada à docência de "Investigação Operacional (A)", do "Seminário em I.O." e do "Estágio Profissionalizante em I.O." permitiu constatar que muitos alunos, para além das referidas dificuldades na articulação de conceitos de Probabilidades e Estatística e na programação de computadores, exibiam uma quase total incapacidade de abordar problemas "menos estruturados" (poder-se-ia esperar alguma perplexidade de um aluno que penetra no mundo do trabalho e inicia o seu "Estágio Profissionalizante" de, pelo menos, 600 horas, a decorrer em empresas, mas, em muitos casos, assistiu-se a uma completa falta de iniciativa em abordar, ainda que de modo simples, os problemas propostos).

Por todos os motivos referidos pareceu ser importante reforçar a preparação prévia dos alunos, nomeadamente a sua capacidade de "integração" de diferentes conhecimentos, antes da abordagem das disciplinas de E.I.O./I.O.(A) e, especialmente, antes do Seminário em I.O. e do estágio Profissionalizante em I.O. Eis, resumidamente, a justificação para a génese da disciplina de "Modelação de Sistemas" da Licenciatura em Matemática da F.C.T.-U.N.L.

2. A Disciplina de "Modelação de Sistemas" da Licenciatura em Matemática da FCT-UNL

Como já se referiu, esta disciplina é obrigatória para os alunos do Ramo de Investigação Operacional, sendo ainda uma possível opção dos alunos do Ramo de Ciências Estatísticas, surgindo no 7º semestre curricular da Licenciatura em Matemática da FCT-UNL.

A carga horária semanal é de três horas de aulas teóricas e de três horas de aulas práticas, a que corresponde um total de 4 Unidades de Crédito.

2.1. Objectivos

O grande objectivo desta disciplina é "reforçar a integração de conhecimentos de áreas distintas, preparando assim os alunos para, com maior sucesso, poderem tirar partido da interdisciplinaridade que caracteriza muitas aplicações da Investigação Operacional e habilitá-los a abordar a modelação de diferentes sistemas".

Para se atingir este objectivo, decidiu-se apontar três "sub-objectivos":

- habilitar os alunos a programar computadores;
- habilitar os alunos a utilizar técnicas estatísticas e
- habilitar os alunos a integrar diferentes conhecimentos, nomeadamente de programação de computadores e de técnicas estatísticas com vista a modelar diferentes sistemas.

Relativamente à programação de computadores optou-se por uma abordagem virada para o desenvolvimento de aplicações. Foi escolhida a linguagem Pascal, que por ser estruturada nos pareceu mais facilmente dominável pelos alunos mais renitentes em programar. Com efeito, pôde observar-se que, vencida a relutância inicial de alguns alunos com mais dificuldades, a possibilidade de estruturar um problema em diferentes módulos e associar-lhe um programa principal muito reduzido (e facilmente assimilável) em Pascal e um conjunto de subrotinas associadas aos diferentes módulos veio a mostrar que a opção por esta linguagem foi muito útil.

Tendo em conta que os alunos a frequentar a disciplina já tinham uma formação base relativamente às Probabilidades e Estatística (essencialmente Paramétrica), embora nem sempre com grande capacidade para utilizar os conhecimentos adquiridos, resolveu dar-se maior ênfase às técnicas de Estatística Não-Paramétrica, com vista à análise de dados. Para além de técnicas de Estatística Descritiva básica, aborda-se o ajustamento de uma distribuição a uma amostra (Teste de Qui-Quadrado e Teste de Kolmogorov-Smirnov), a comparação de duas amostras independentes (Teste de Qui-Quadrado e Teste de Kolmogorov-Smirnov para duas amostras) e a comparação de mais de duas amostras independentes (Teste de Qui-Quadrado e Teste de Kruskal-Wallis).

Relativamente à integração dos conhecimentos de programação de computadores e de técnicas estatísticas, com vista à modelação de diferentes sistemas, julgou-se adequado proporcionar aos alunos a participação na análise de Problemas Formativos (análise de pequenos "Casos" a efectuar nas aulas, que, por uma questão de pudor, se insistiu em designar

por Problemas Formativos, já que, se por um lado, não têm a dimensão e complexidade clássicas dos tradicionais *Case Studies*, por outro não são meros problemas de aplicação de técnicas ou conceitos, mas antes problemas menos estruturados, que se crê, terem importante papel formativo).

2.2. Programa

1 - Elementos de Programação de Computadores

- 1.1 - Fluxogramas
- 1.2 - Linguagem Pascal
- 1.3 - Aplicações

2 - Elementos de Estatística Aplicada

- 2.1 - Estatística Descritiva
 - 2.1.1 - Histogramas; Representações gráficas
 - 2.1.2 - Valor médio, desvio padrão, coeficiente de assimetria e coeficiente de Kurtosis - estimação de valores amostrais e formulação de hipóteses plausíveis de distribuição com base nessas estimativas
- 2.2 - Estatística Não-Paramétrica
 - 2.2.1 - Verificação da aleatoriedade de uma amostra: Teste do Sinal
 - 2.2.2 - Ajustamento de uma distribuição univariada: Testes de Qui-Quadrado e de Kolmogorov-Smirnov
 - 2.2.3 - Comparação de duas amostras independentes: Testes de Qui-Quadrado e de Kolmogorov-Smirnov
 - 2.2.4 - Comparação de mais de duas amostras independentes: Testes de Qui-Quadrado e de Kruskal-Wallis

3 - Modelação de Sistemas/Aplicações

- 3.1 - Introdução. Sistema. Modelo
- 3.2 - Análise de Dados *versus* Modelação de Sistemas
- 3.3 - Aplicações

2.3 - Avaliação de Conhecimentos

Um aluno só poderá obter aprovação na disciplina se tiver participado em dois trabalhos de grupo, bem como na análise de problemas formativos e em, pelo menos, um exame escrito individual.

I - Obtenção de Frequência

Um aluno obtém "Frequência" na disciplina se e só se verificar cumulativamente as seguintes condições:

- a) Participação no 1º Trabalho (T1) - trata-se de um trabalho de grupo (de três alunos) em que os alunos devem proceder à análise de dois ficheiros de dados e posterior

ajustamento de distribuições. O enunciado e os dados são facultados aos alunos após a abordagem do ponto 2.2.2 do programa, devendo os grupos entregar os seus relatórios no prazo de duas semanas.

A constituição dos grupos é da responsabilidade dos alunos. Embora não seja feita uma discussão do relatório com cada grupo (o que permitiria que o docente pudesse tomar a iniciativa de eventualmente classificar diferentemente os diferentes membros do grupo, mas que se traduziria numa grande sobrecarga para o docente) é dada a opção a cada grupo de, antes de saberem a classificação do trabalho, solicitarem classificações diferenciadas entre os membros do grupo. Caso optem por essa via, os três membros do grupo deverão chegar a acordo entre si e indicarem ao docente quais as diferenças relativas de classificações que consideram justas.

- b) Participação no 2º Trabalho (T2) - trata-se de um trabalho de grupo (de três alunos) em que os alunos devem proceder à análise de três ficheiros de dados. Os grupos deverão caracterizar as variáveis a que correspondem os três ficheiros, bem como as suas eventuais interdependências. O enunciado e os dados são facultados aos alunos após a abordagem do ponto 2.2.4 do programa, devendo os grupos entregar os seus relatórios no prazo de quatro semanas.

As referências feitas (a propósito do 1º Trabalho) à constituição dos grupos e à classificação dos seus membros são também aplicáveis ao 2º Trabalho.

- c) Participação nos Problemas Formativos (PF) - trata-se da análise de pequenos "casos" cujo enunciado é apresentado no início de cada aula. Segue-se um período que varia entre os 60 e os 90 minutos em que os diferentes grupos (de cinco alunos) analisam os referidos problemas e preparam um "mini-relatório" para apresentarem as suas conclusões. Em seguida cada grupo nomeia um porta-voz que apresenta as referidas conclusões aos colegas. Segue-se a discussão do problema e a comparação das diferentes abordagens sugeridas. Com base nos relatórios apresentados e na discussão referida cada grupo é classificado.

Os PF começam a ser apresentados após a abordagem do ponto 3.1 do programa, prosseguindo até final do semestre. Refira-se que a constituição dos grupos para a abordagem dos PF é feita aleatoriamente no início de cada aula, visando assegurar a maior *rotação* dos alunos entre si e evitar alguns males resultantes de grupos fixos. De referir ainda que se teve o cuidado de incluir entre os PF vários tópicos que serão desenvolvidos na disciplina subsequente de Investigação Operacional(A)/Elementos de Investigação Operacional.

Para efeitos de classificação atribuída à parcela "Problemas Formativos", será calculada a média das n melhores classificações obtidas nos diferentes PF em que cada aluno participou; sendo n um número inteiro que se situa entre $2/3$ e $3/4$ do número total de PF apresentados na disciplina.

- d) Nenhuma das três classificações anteriores (no 1º Trabalho, 2º Trabalho e Problemas Formativos) deverá ser inferior a 9,5 valores.

II - Classificação de Frequência

Sejam CT1, CT2 e CPF as classificações obtidas por um aluno, respectivamente, no 1º Trabalho, 2º Trabalho e Problemas Formativos. Se as referidas classificações forem todas não inferiores a 9,5 valores, calcula-se a Classificação de Frequência (CF) do modo seguinte:

$$CF = (27,5 \cdot CT1 + 27,5 \cdot CT2 + 5,0 \cdot CPF)/60$$

III - Admissão a Exame

É admitido a exame (sem consulta e com três horas de duração) todo o aluno com Frequência. Cada aluno só se poderá apresentar em Época Normal de avaliações a uma das duas chamadas de exame. Caso não tenha obtido aprovação em Época Normal, um aluno com Frequência poderá apresentar-se à única chamada de exame em Época de Recurso.

Seja CEX a classificação obtida por um aluno em exame.

Se CEX for inferior a 9,5 valores, o aluno estará Reprovado; caso contrário, estará Aprovado, sendo a sua classificação final (CFin) o valor inteiro mais próximo de $(0,6 \cdot CF + 0,4 \cdot CEX)$.

2.4 - Planeamento do Semestre Lectivo

Considera-se, para efeitos de planeamento, um semestre lectivo com catorze semanas efectivas de actividades lectivas (para atender a feriados e alguns imponderáveis).

• Aulas:

Ao primeiro tópico do programa - "Elementos de Programação de Computadores" - são reservadas três semanas. Tal como se referiu anteriormente, a abordagem deste tópico é feita desde logo voltada para as aplicações. Assim é dado relevo ao desenvolvimento de rotinas de ordenação, cálculo estatístico básico, bem como à manipulação de ficheiros de dados.

Ao segundo tópico do programa - "Elementos de Estatística Aplicada" são reservadas cinco semanas. Sempre que possível refere-se a existência de software de apoio, aproveitando-se também para o desenvolvimento de alguns programas de aplicação, com a eventual integração de rotinas já desenvolvidas anteriormente.

Ao terceiro tópico do programa - "Modelação de Sistemas/Aplicações" - são dedicadas seis semanas.

• Trabalhos de Grupo:

No 1º Trabalho de grupo os alunos devem proceder à análise de dois ficheiros de dados (um relativo a uma amostra proveniente de uma população discreta e o outro relativo a uma amostra proveniente de uma população contínua), formular hipóteses de ajustamento de distribuições e, finalmente, aos testes estatísticos adequados. Assim, o enunciado e os dados são facultados aos alunos após a abordagem do ponto 2.2.2 do programa - aproximadamente após cinco semanas de aulas, devendo os grupos entregar os seus relatórios no prazo de duas semanas.

No 2º Trabalho de grupo os alunos devem proceder à análise de três ficheiros de dados, caracterizando as correspondentes variáveis, bem como as suas eventuais interdependências. O enunciado e os dados são facultados aos alunos no final da abordagem do ponto 2.2.4 do programa - aproximadamente após oito semanas de aulas, devendo os grupos entregar os seus relatórios no prazo de quatro semanas.

Os Problemas Formativos decorrem nas últimas seis semanas do semestre.

• Exame:

Trata-se de uma prova escrita sem consulta com a duração de três horas, que decorrerá após o período lectivo (um aluno com Frequência poderá optar por uma das duas chamadas da Época Normal, ou pela chamada única da Época de Recurso). Cada Exame consiste em três grupos - um primeiro ligado às técnicas estatísticas (a que correspondem 11,5 valores), um segundo dedicado à programação em Pascal (5,0 valores) e um último grupo dedicado à modelação (3,5 valores), com uma questão "menos estruturada", devendo o aluno propor a sua abordagem, referir que dados gostaria de poder analisar e que análises pretendia levar a cabo - trata-se assim de um grupo mais ligado à modelação na sua vertente criativa, do que à análise de dados ou à utilização directa de técnicas mais objectivas.

2.5 - Bibliografia

Relativamente ao primeiro tópico do programa - "Elementos de Programação de Computadores", não se aconselha qualquer referência em particular, dada a existência de um grande número de títulos no mercado (alguns dos quais em Português, o que é certamente preferencial para a maioria dos alunos). Refira-se, no entanto, e a título de exemplo, "Programming in Pascal" de P. Grogono da Editora Addison-Wesley.

Relativamente ao segundo tópico do programa - "Elementos de Estatística Aplicada", recomenda-se a consulta de "Nonparametric Statistics for the Behavioral Sciences" de S. Siegel e N. Castellan Jr da Editora Mc Graw Hill Internacional.

Relativamente ao último tópico do programa - "Modelação de sistemas/Aplicações", recomenda-se o livro "How to Model It - Problem Solving in the Computer Age" de A. Starfield, K. Smith e A. Bleloch da Editora Mc Graw Hill Internacional. Trata-se de um livro extremamente interessante, com exemplos aparentemente muito simples, cuja análise é feita de forma gradual, sendo deixadas ao leitor muitas tarefas e sugestões, bem como diversas indicações de leituras adicionais.

Ainda relativamente a este último tópico do programa, refira-se que, por amabilidade dos autores, tivemos acesso à versão final (anterior à publicação) do livro "Strategies for Creative Problem Solving" de H. S. Fogler e S. E. LeBlanc, entretanto editado pela Editora Prentice Hall. Trata-se de um livro muito interessante, com vários exemplos de problemas reais (essencialmente da área das Engenharias), a partir dos quais se discute a sua "resolução".

3. Funcionamento da Disciplina de "Modelação de Sistemas" da Licenciatura em Matemática da FCT-UNL em 1993/94 e 1994/95

Como se esperaria, o funcionamento desta disciplina em 1993/94 ainda não foi "em pleno" e, obviamente, admite-se ainda que, apesar das melhorias introduzidas em 1994/95, ainda haja lugar a mais algumas beneficiações.

Em 1993/94 a disciplina funcionou pela primeira vez. Assim, notou-se um menor ritmo nas aulas, comparativamente ao ano seguinte - a experiência prévia de um docente relativamente a uma disciplina (ainda mesmo que não por via directa) revela-se claramente importante.

Em 1993/94 a população a que se destinava a disciplina não foi "homogénea", isto é, para além dos alunos que pela primeira vez frequentavam o 7º semestre curricular e que se inscreveram nesta disciplina (25 alunos), houve um grupo relativamente numeroso (14 alunos) que sendo já finalistas decidiram frequentar a título extraordinário esta disciplina para se sentirem mais apoiados nos respectivos Estágios Profissionalizantes, que então iniciavam. Estes alunos vieram a revelar-se muito interessados, não se distinguindo *aparentemente* dos seus demais colegas. No entanto, a sua presença foi um dos factores mais importantes que inviabilizou o recurso sistemático aos Problemas Formativos nesse ano lectivo (especialmente com o seu papel de apresentação de tópicos que depois são abordados no 8º semestre curricular nas disciplinas de I.O.(A)/E.I.O.). Refira-se, no entanto, uma grande vantagem na presença destes alunos "mais adiantados" - dado que esses alunos frequentaram os seus Estágios Profissionalizantes ainda em 1993/94, foi desde logo possível, por comparação com o desempenho dos seus colegas de anos anteriores, notar uma maior facilidade de abordagem dos seus problemas de Estágio, o que obviamente era uma das expectativas criadas com a introdução desta disciplina.

A população de alunos desta disciplina em 1994/95 já não apresentou tais problemas - os 26 alunos que a constituíam frequentavam o 7º semestre curricular, não tendo sido alunos das disciplinas subsequentes de I.O.(A)/E.I.O., pelo que a introdução dos Problemas Formativos se veio a revelar muito interessante sob diversos pontos de vista: não só se introduziram alguns dos tópicos que vieram a ser subsequentemente tratados nas disciplinas do 8º semestre curricular, como ainda, se aumentou consideravelmente o nível de participação dos alunos na discussão dos Casos apresentados, já que 5 % da sua classificação final dependeria da classificação individual obtida nos Problemas Formativos.

Em 1993/94 não foi possível utilizar equipamento informático portátil nas aulas, que ainda não estava disponível. Esta utilização (ainda que esporádica) de equipamento informático portátil durante o ano lectivo de 1994/95 traduziu-se num maior interesse dos alunos e numa maior dinamização das aulas. À medida que se vulgariza a utilização desses meios, espera-se um maior nível de participação e de aproveitamento dos alunos.

O programa referido em 2.2 foi cumprido nos dois anos lectivos. A metodologia de avaliação de conhecimentos referida em 2.3 e o planeamento do semestre referido em 2.4 dizem

respeito ao ano de 1994/95, já que contemplam a existência formal dos Problemas Formativos, correspondendo a um estado de desenvolvimento da disciplina superior, em termos qualitativos, ao do ano precedente.

O Quadro 2 seguinte sintetiza algumas informações relativas aos resultados obtidos na disciplina de "Modelação de Sistemas" nos anos lectivos de 1993/94 e 1994/95:

Ano Lectivo →	1993/94			1994/95			
Resultados observados	1º Trab.	2º Trab.	Resultado Final	1º Trab.	2º Trab.	Problemas Formativos	Resultado Final
Mínimo	13,1	10,0	12	16,0	11,5	12,4	10
Máximo	18,2	18,5	17	19,0	16,0	15,7	17
Valor Médio	15,63	12,73	14,32	17,31	13,53	14,37	14,71
Desvio Padrão	2,96	3,73	2,80	1,19	1,43	0,88	1,52
Nº alunos ⁽¹⁾	38			26			
Nº aluno c/freq ^a	37			26			
%o aluno c/freq ^a	97,40%			100,00%			
Nº aluno aprov. ⁽²⁾	37			24			
%o aluno aprov.	100,00%			92,31%			

Notas: (1) - Nº total alunos que frequentaram a disciplina

(2) - Após as duas épocas de avaliação

Quadro 2 - Resultados obtidos em "Modelação de Sistemas" em 1993/94 e 1994/95.

A propósito do Quadro 2 poder-se-á referir alguns aspectos:

- A taxa final de aprovação (após as duas Épocas de Avaliação) desta disciplina é muito elevada, o que parece justificável, por um lado, pelo interesse que este tipo de disciplina desperta nos alunos (até aí muitas vezes com disciplinas "menos aplicadas") e, por outro lado, pelo próprio nível de desenvolvimento dos alunos (no 7º semestre curricular).
- Em geral, os resultados obtidos no 1º Trabalho de avaliação são claramente superiores aos resultados obtidos no 2º Trabalho de avaliação. Tal deve-se obviamente à maior complexidade do 2º Trabalho face ao 1º Trabalho. Crê-se, no entanto, ser fundamental iniciar o processo de avaliação de conhecimentos por um 1º trabalho "mais suave" para, de algum modo, não amedrontar os alunos com mais dificuldades (nomeadamente, na utilização de meios informáticos). Além disso, o 1º trabalho apresentado permite o desenvolvimento de software que depois poderá, com as devidas adaptações, ser utilizado no 2º Trabalho. Sentiu-se claramente que os alunos teriam tido muito mais dificuldade em executar um único Trabalho de avaliação do tipo do 2º trabalho, se não tivessem passado previamente pelo Trabalho anterior.

- Os resultados obtidos nos Problemas Formativos (em 1994/95) são de algum modo intermédios entre os resultados dos dois Trabalhos de avaliação, apresentando uma maior homogeneidade. (Será fruto do efeito de rotação dos alunos nos grupos constituídos aleatoriamente ?)
- Para o cômputo do Resultado Final dos alunos a classificação obtida no Exame individual tem 40 % de peso. Como se esperaria, este Resultado Final apresenta-se "intermédio" relativamente às classificações mais extremas dos dois Trabalhos. É interessante notar não ter havido uma diferença muito acentuada entre os valores médios do Resultado Final nos dois anos lectivos estudados (14,32 e 14,71). No entanto, parece poder-se referir que a variabilidade das várias parcelas classificativas é menor em 1994/95. (Será tal devido à presença em 1993/94 dos "alunos mais avançados" que já então se encontravam a frequentar o 9º semestre curricular ?)
- É importante referir que, relativamente ao Resultado Final, não há uma clara relação entre as contribuições correspondentes às diferentes parcelas (1º Trabalho, 2º Trabalho, Problemas Formativos e Exame) para todos os alunos. Tal parece suportar a importância dessas diferentes parcelas no processo de avaliação, que para além de admitir uma classificação proveniente de trabalhos realizados em grupo, exige sempre uma parcela de carácter individual (sendo esta última determinante para diferenciar os alunos).

4. Conclusão

O interesse demonstrado pela generalidade dos alunos desta disciplina traduziu-se, não só em bons resultados na própria disciplina, mas também num efeito positivo detectável em disciplinas subsequentes da área: "Investigação Operacional (A)"/"Elementos de Investigação Operacional", "Seminário em I.O." e, muito especialmente, no "Estágio Profissionalizante em I.O." (foi particularmente gratificante, que grande número dos alunos finalistas em 1993/94, tivessem optado por, *a título extraordinário*, frequentar a disciplina, tendo vindo quase todos esses alunos a manifestar a "grande ajuda" que sentiram nos seus estágios resultante da frequência dessa disciplina; idênticos comentários foram já feitos por alunos em estágio no ano lectivo de 1994/95, que frequentaram a disciplina de "Modelação de Sistemas" em 1993/94).

Comparativamente com o passado, os docentes ligados às várias disciplinas "avançadas" da área de Investigação Operacional constataram já uma maior capacidade dos alunos em integrar áreas que anteriormente lhes pareciam "estanques". Foi ainda observada uma maior iniciativa dos alunos numa primeira abordagem dos problemas e, em particular, do "Estágio Profissionalizante".

A utilização (ainda que esporádica) de equipamento informático portátil durante o ano lectivo de 1994/95, leva a antecipar um maior interesse dos alunos e uma maior dinamização das aulas à medida que se vulgariza a utilização desses meios.

Considera-se que o processo de avaliação da disciplina é "relativamente pesado", quer para alunos, quer para um único docente que leccione a disciplina. Crê-se que, no futuro, se o número de docentes da disciplina aumentar, se poderá eventualmente conceber um único Trabalho de Grupo cujo desenvolvimento inicial poderia ser acompanhado em directo pelos docentes durante as aulas. Seguir-se-ia uma discussão oral do Trabalho de Grupo com cada um dos Grupos, o que viria a proporcionar uma maior diferenciação dos alunos, podendo eventualmente vir a tornar desnecessária a realização de um Exame escrito individual. Uma equipa docente mais alargada poderia ainda permitir que alguns Problemas Formativos evoluíssem naturalmente, passando a verdadeiros *Case Studies* cujos enunciados poderiam ser distribuídos com alguma antecedência relativamente à sua apresentação pelos alunos e eventual discussão.

A experiência de ensino vivida com a disciplina de "Modelação de Sistemas" da Licenciatura em Matemática da FCT-UNL é, ainda, relativamente breve. No entanto poder-se-á desde já realçar que para além do óbvio trabalho associado ao lançamento de uma nova disciplina (especialmente, quando obriga não só à preparação de aulas, mas também à geração de ficheiros de dados, preparação de enunciados de Trabalhos, Problemas Formativos e Exames, no caso vertente agravado por se tratar apenas de um único docente) se pode referir a profunda satisfação com que se constata a maior facilidade dos alunos desta Licenciatura (muito especialmente os do Ramo de Investigação Operacional) na abordagem de problemas "menos estruturados", o que de algum modo os deixa com uma melhor preparação para enfrentar, a curto prazo, o mundo do trabalho.

Referências

- [1] Fogler, H.S., LeBlanc, S.E., *Strategies for Creative Problem Solving*, Prentice Hall (1994).
- [2] Grogono, P., *Programming in Pascal*, Addison-Wesley (1984).
- [3] Siegel, S. Castellan Jr., N., *Nonparametric Statistics for the Behavioral Sciences*, Mc Graw Hill Int. (1988).
- [4] Starfield, A., Smith, K. e Bleloch, A., *How to Model It - Problem Solving in the Computer Age*, Mc Graw Hill Int. (1990).

MODELO DE SEGURANÇA DO TRABALHO

Hamilton Bezerra Fraga da Silva

Universidade Federal Fluminense UFF
Niterói, Brasil

Abstract

This article has as its main goal, to show the possibility of using an Operational Research Model known as Data Envelopment Analysis - DEA (see Charnes et al. [2]) relating to safety at work studies. The DEA model will indicate the position of the participant companies related to the accidents and, now, as we know the "ranking", one can accomplish a benchmarking, among them, in order to find out which management actions (procedures) allowed to achieve a good performance.

Resumo

Este artigo tem como objectivo mostrar a possibilidade de utilização de um modelo de Investigação Operacional conhecido como Data Envelopment Analysis - DEA (Charnes et al. [2]) em estudos relativos à segurança do trabalho. O modelo DEA indicará a posição das empresas participantes no tocante aos acidentes e, agora, ao se conhecer o "ranking", pode realizar-se um "benchmarking" entre elas, no intuito de descobrir que acções gerenciais permitiram lograr o bom desempenho.

Keywords

Operational Research, Benchmarking, Accidents.

1. Introdução

O Brasil figura no cenário mundial como um dos países recordistas em acidentes do trabalho. Os motivos apontados são diversos: falta de consciencialização de todos os envolvidos, baixos índices educacionais, de saúde e alimentares, o que torna o trabalhador brasileiro um excelente hospedeiro para o acidente.

Os modelos de Pesquisa Operacional estão consagrados e são amplamente empregados em maximização do lucro, minimização de custos, problemas de transporte, localização de postos de serviço em grafos [8]. O artigo se propõe a mostrar como o modelo DEA poderá vir a ser útil como uma ferramenta quantitativa de avaliação e tomada de decisão com relação aos acidentes.

2. Data Envelopment Analysis

O Data Envelopment Analysis é uma técnica de programação matemática que foi inicialmente utilizada na avaliação de organizações governamentais sem fins lucrativos e ao longo dos anos foi aplicada nos mais diversos tipos de segmento. Dentre os trabalhos de cunho

social podem ser destacados os referentes a escolas (Bessent et al. [1]), justiça (Lewin et al. [6]), distritos florestais (Kao C. [5]) e outros.

Hoje, o "DEA" se expandiu para uma variada gama de aplicação desde a comparação entre produtos de informática (Doyle JR. e Green RH. [3]) até à comparação entre serviços bancários conforme documentado nos artigos de (Giokas D. [4]) e (Oral Muhitin [9]).

A técnica pode ser usada sempre que houver o interesse em medir a eficiência relativa de unidades comparáveis que se utilizem de entradas (insumos) para produzir saídas múltiplas.

2.1 O Modelo DEA

O modelo proposto para estudo comparativo envolvendo acidentes de trabalho se baseia em programação linear fracionária, onde o numerador com as parcelas T_r assume a forma do número de trabalhadores segurados ou homem-hora, ou alguma outra saída que se deseja maximizar. Poderão existir "R" saídas.

O denominador com as parcelas " A_i " representa, neste modelo, os acidentes ocorridos em um período de tempo (acidente típico, acidente de trajeto, doença ocupacional, etc). Poderão existir "I" entradas.

Um modelo DEA conhecido como C.C.R. e que privilegia a eficiência é então mostrado a seguir:

$$\text{MAX } H_B = \left\{ \sum_{r=1}^R U_{rB} T_{rB} \right\} / \left\{ \sum_{i=1}^I V_{iB} A_{iB} \right\} \quad (1)$$

sujeito a

$$\left\{ \sum_{r=1}^R U_{rB} T_{rJ} \right\} / \left\{ \sum_{i=1}^I V_{iB} A_{iJ} \right\} \leq 1, \quad J = 1, 2, \dots, n \quad (2)$$

$$U_{rB}, V_{iB} \geq \epsilon > 0 \quad \forall r, i,$$

Onde:

H_B = representa um índice de uma unidade B que está sendo avaliada razão entre saídas e entradas comparadas. O resultado de H_B será medido por um número escalar compreendido entre [0,1],

B = representa a unidade (base) que está sendo avaliada,

T_{rJ} = quantidade observada de saída r produzida por J,

A_{iJ} = quantidade observada de entrada i usada por J,

U_{rB} = peso dado à saída r relativo à base B,

V_{iB} = peso dado à entrada i relativo à base B,

ϵ = número pequeno positivo (10^6)

O modelo de programação fracional do "DEA" é linearizado e mostrado a seguir:

$$\text{MAX } H_B = \sum_{r=1}^R \mu_{rB} T_{rB} \quad (4)$$

sujeito a

$$\sum_{i=1}^R v_{iB} A_{iB} = 1 \quad (5)$$

$$\sum_{r=1}^R \mu_{rB} T_{rJ} - \sum_{i=1}^R v_{iB} A_{iJ} \leq 0, J = 1, 2, \dots, n \quad (6)$$

$$\mu_{rB}, v_{iB} \geq \epsilon > 0 \quad \forall r, i, \quad (7)$$

onde os indicadores (μ_{rB}, v_{iB}) representam os multiplicadores virtuais.

2.2 Aplicação Teórica do Modelo DEA em Segurança do Trabalho

O modelo proposto constitui-se de duas fases: a primeira fase é quantitativa e se utiliza de programação matemática conforme consta dos procedimentos 1 a 5; a segunda fase é qualitativa.

2.2.1 Primeira Fase

- 1) Proceda uma prospecção no mercado e escolha um grupo de empresas com um mesmo ramo de actividades, possuidoras, portanto, de um mesmo grau de risco conforme Quadro 1 do código de actividades incluídas na NR4 - Ministério do Trabalho.
- 2) Elabore um quadro resumo. Neste quadro serão listados as empresas "J" e os seus dados agrupados em dois setores: dados que se desejam maximizar (numerador: homem-hora, número de segurados, etc.); dados que se desejam minimizar (denominador: acidentes típicos, acidentes de trajecto, doenças ocupacionais, etc.), ou seja, a formação de dois conjuntos. O primeiro varia de (1 a "r") e o segundo de (1 a "I").
- 3) Determinação da eficiência, razão entre o numerador e o denominador, levando-se em conta todas as empresas envolvidas na análise. Para esse cálculo pode-se usar o programa "LINDO" [7], de pesquisa operacional, amplamente utilizado pelas instituições de ensino.
- 4) A eficiência relativa de cada empresa é calculada através do modelo de equações (4-7), que ainda informa o conjunto referência para cada empresa analisada e as folgas se houver. Neste momento, faz-se um quadro resumo listando por coluna as unidades que estão sendo avaliadas, a eficiência relativa em termos de segurança, o conjunto referência de cada empresa, folgas de entrada e saída, assim como os multiplicadores virtuais.
- 5) Proceda a uma análise com base na eficiência relativa e suas folgas com o objectivo de determinar os excessos no denominador ou escassez no numerador, fonte geradoras de ineficiências.

A metodologia descrita tem a pretensão de fornecer subsídios para a gerência. Após a função ter sido otimizada é possível saber onde se pode "mexer" para que a eficiência seja um.

Nos dados de entrada, diminuindo as quantidades de entrada, ou nos dados de saída, aumentando as quantidades de saída.

Serão analisados os excessos das quantidades de entrada nas empresas que foram consideradas ineficientes do ponto de vista da segurança do trabalho. O excesso de cada entrada é obtido multiplicando cada entrada pelo factor $(1 - H_B) + F_{iB}$ onde F_{iB} representa as folgas de entrada.

A deficiência nos dados de saída em comparação ao conjunto referência foi obtida multiplicando-se cada valor de saída pela expressão $(1/H_B - 1 + F_{rB})$, onde F_{rB} representa as folgas de saída.

Escolha as unidades que apresentaram eficiência menor do que "1" e compare com o seu conjunto de referência (unidades que obtiveram eficiência "1") para descobrir a fonte de suas ineficiências.

Para alcançar o conjunto referência serão analisados os excessos nos quantitativos do denominador nas empresas que foram consideradas ineficientes assim como a escassez nos quantitativos do numerador.

Com este procedimento, a empresa que não obteve o desempenho "1" consegue se situar, e agora é sabedora da sua posição no "ranking".

2.2.2 Segunda Fase

A segunda fase do modelo é qualitativa. A empresa que não tenha obtido conceito "1" deve estudar o seu conjunto referência para ver como as unidades participantes atuam. Organiza-se um "Benchmarking", que objective descobrir as acções gerenciais que evitem ou minimizem os acidentes de trabalho; por exemplo, o que as outras empresas estão fazendo com os seus funcionários no que diz respeito aos recursos humanos, quais os processos de que estão se utilizando, soluções ergonômicas, treinamento e capacitação de pessoal, soluções anti-stress, estudos visando à diminuição da monotonia, etc. Realiza-se, para isso, um "brainstorming" entre os seus dirigentes com o objectivo da busca e da implementação de melhores processos que evitem os acidentes.

Dando sustentação à análise retro-citada desenvolve-se uma avaliação relativa do peso de cada dado individual de excesso de entrada no escopo total de entrada, fazendo-se o mesmo para cada ineficiência de saída. Estes pesos serão utilizados na decisão da relevância das medidas gerenciais a serem implementadas já que se constituirão de pesos ponderados.

Determine o peso relativo de cada parcela referente aos excessos de entrada e/ou referente às ineficiências de saída k (%) de cada " X_i " ou " Y_r " pertencentes à unidade "J" escolhida para estudo.

O peso de cada parcela K (%) representa a importância relativa do "custo" do excesso de cada entrada ou a importância relativa do "custo" da escassez de cada saída, na empresa que está sendo avaliada, quando comparada com o seu conjunto referência. O termo K (%) será utilizado como peso no cálculo do diagnóstico dos níveis de importância das acções gerenciais.

O modelo de programação linear, apesar de informar as quantidades de entrada ou saída que devem ser mudadas para se atingir eficiência "1", não identifica que acções gerenciais devem ser empregadas para atingir esta meta.

O passo seguinte é a identificação de forma qualitativa e com a ajuda de técnicos da área, das possíveis acções gerenciais que deverão ser tomadas no intuito de diminuir as entradas e/ou aumentar as saídas conforme estabelecido pelo modelo de Programação Linear. Observe as acções gerenciais que foram utilizadas no seu conjunto referência ou nas empresas líderes com o objectivo de copiá-las.

Estabeleça o quadro conforme a seguir. As colunas são as entradas "X_i" ou saídas "Y_r" com o seu respectivo peso relativo K (%).

Estabeleça uma relação de peso forte, médio, fraco, como por exemplo (9,3,1), respectivamente, entre todas as acções gerenciais e entre todos os dados de entrada ou saída, onde cada célula a_{ij} terá a sua relação com a acção gerencial i-ésima e com o j-ésimo dado de entrada. Em outras palavras, o valor 9, colocado na diagonal superior da célula a₁₁, indica que a primeira acção gerencial tem forte relação com a primeira entrada. O estabelecimento desta relação é feita de forma qualitativa pelos tomadores de decisão.

Multiplique cada peso da relação (a_{ij}) estipulada, como por exemplo (9,3,1), pelo peso da coluna correspondente, peso relativo K (%), perfazendo todos a_{ij}, fazendo variar "j". Faça o mesmo procedimento para todas as linhas conforme exemplo retirado do quadro 1, mostrado a seguir.

Quadro 1 - Peso das acções gerenciais considerando uma unidade "J" com H_{B<1}

Acção Gerencial	k(%)	k(%)	...%	k(%)
AG ₁	9	a ₁₁	3 a _{1n}
AG ₂				
.				
.				
AG _m	3	a _{m1}	1 a _{mn}

Determine o peso relativo, em (%), de cada linha (I), de cada acção gerencial (AG_I), através da expressão:

$$AG_I = \left(\sum_{j=1}^n a_{Ij} K_j \right) / \left(\sum_{i=1}^m \sum_{j=1}^n a_{ij} K_j \right) \quad I = 1, 2, 3, \dots, m$$

O maior peso ponderado, por linha, em ordem decrescente, indicará a ordem de prioridade das acções gerenciais, ou seja, a maior percentagem encontrada corresponderá à acção gerencial mais importante que primeiro sofrerá um processo de "benchmarking".

Finalmente, tendo utilizado um valor ponderado para os dados de entrada ou saída e uma relação de intensidade entre as acções gerenciais e estes dados de entrada ou saída, chega-se à

importância da contribuição de cada medida gerencial para que a eficiência no tocante a minimização dos acidentes se torne "1" numa próxima verificação.

3. Conclusão

Este trabalho propõe o uso de um procedimento sistêmico que se adoptado pelos responsáveis pela segurança e saúde do trabalhador poderá vir a ser uma das ferramentas de controle no combate à prevenção de acidentes. O modelo fornece a posição relativa das unidades no que diz respeito a segurança de seus funcionários, assim como permite uma melhor visualização das empresas que lograram melhor êxito no aspecto segurança. A realização de uma "benchmarking" entre as unidades que participam do estudo e a técnica do "brainstorming" permitem a descoberta da importância das acções gerenciais pertinentes levando a extinção ou minimização dos acidentes.

O conhecimento da realidade dos outros, nos moldes como é apresentado pelo modelo, acaba por conduzir a uma nova modelação empresarial com a definição clara do novo rumo a ser seguido. O motivo é que, agora, se conhecem as diferenças existentes e necessárias para atingir o alvo de eficiência 100% e a partir desta constatação, o estabelecimento das directrizes fundamentadas nestes desvios.

Por último, a utilização do modelo DEA poderá servir como um modelo base comparativo para a Previdência Social quanto ao aspecto de como cobrar e multar as empresas reincidentes no que diz respeito a segurança do trabalho, já que agora se conhece a posição das empresas no "ranking".

Bibliografia

- [1] Bessent, A. e Bessent, W., *Determining the comparative efficiency of schools through Data Envelopment Analysis*, Educations Administration Quarterly 16 (1980) 57-75.
- [2] Charnes, A., Cooper, W. e Rhodes, E., *Measuring the efficiency of decision-making units*, European Journal of Operational Research 2 (1978) 429-444.
- [3] Doyle, JR. and Green, RH., *Comparing products using data envelopment analysis*, Omega 19 (1991) 631-638.
- [4] Giokas, DI., *Bank branch operating efficiency: A comparative application of DEA and the loglinear model*, Omega 19 (1991) 549-557.
- [5] Kao, C., *Efficiency improvement in data envelopment analysis*, European Journal of Operational Research 73 (1994) 487-494.
- [6] Lewin, A.Y., Morey, R.C. e Cook, T.J., *Evaluation the administrative efficiency of courts*, Omega 10 (1982) 401-411.
- [7] *User's manual for linear interger and quadratic programming with lindo*, Scientific Press (1987).
- [8] Pizzoloto, N.D., *Programação Linear*, Livros Técnicos Científicos Ltd. (1987).
- [9] Oral, M., Ossama, K., Reha Uolalan, *An empirical study on analyzing the productivity of bank branches*, IIE Transactions 24 (1992).

UM MODELO DE PREVISÃO DE VENDAS PARA O RETALHO ALIMENTAR

Ana Paula Barroso

Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
Quinta da Torre
2825 Monte Caparica - Portugal

Isabel Themido

IST/CESUR
Universidade Técnica de Lisboa
Av. Rovisco Pais
1000 Lisboa - Portugal

Diogo Furtado Luzes

Instituto Superior de Agronomia
Universidade Técnica de Lisboa
Tapada da Ajuda
1300 Lisboa - Portugal

Abstract

This paper presents a forecasting model for the daily sales of a food retailing chain developed on the basis of real data from hypermarkets. Sales variability is high as a result of marked seasonality and intensive marketing activity. The model is applicable to daily sales of individual items, only parameters vary with the type of article and shop. To forecast the daily sales the model takes into consideration the position of that day in the week, and year and the proximity of the day to an holiday. A classification of the ABC type provides the basis for an appropriate forecasting and replenishment method for each one of the items in the range. The flexible structure adopted will allow the incorporation of the price variation effect on sales for class A items and class B items, when promoted.

Resumo

Neste trabalho, apresenta-se um modelo de previsão de vendas para o retalho alimentar desenvolvido com base em dados reais, de uma cadeia de hipermercados, que apresentam grande variabilidade resultante de sazonalidade e actividade promocional. O modelo é aplicável a todas as lojas da cadeia logística variando de loja para loja apenas o valor dos parâmetros. Na previsão de vendas de um artigo para um determinado dia, o modelo toma em consideração o dia da semana, o dia do mês, a proximidade de um feriado e a época do ano onde se incluem as quadras festivas. É proposta uma classificação de todos os artigos da gama em três classes, A, B e C, sendo recomendado um modelo de previsão adequado a cada situação. A estrutura flexível do modelo permitirá incorporar o efeito da alteração dos preços nos artigos da classe A e nos artigos da classe B quando em promoção.

Keywords

Forecasting, Retailing, Logistics, Distribution, Modelling.

1. Introdução

No ambiente fortemente competitivo que se vive no sector do retalho alimentar, principalmente nas grandes superfícies, é de capital importância a gestão racional de recursos escassos como o espaço e os meios financeiros envolvidos (Mercer, 1993). Saber antecipadamente quanto é que se vai vender em cada dia, de cada artigo, pode permitir uma considerável redução nos custos de operação, redução esta frequentemente acompanhada por uma melhoria do nível de serviço em resultado de uma melhor previsão.

O estudo apresentado teve como objectivo desenvolver modelos de previsão de vendas que permitam apoiar a gestão de aprovisionamentos de toda a cadeia logística, lojas e entreposto.

Uma das preocupações que esteve sempre presente foi a de desenvolver modelos de previsão de vendas de pouca complexidade, adequados à frequência de utilização. Tipicamente, uma grande superfície vende uns largos milhares de artigos em relação aos quais toma decisões de aprovisionamento todos os dias.

A evolução do retalho alimentar em Portugal foi idêntica à verificada noutros países (ver Lopes (1991)) tendo-se assistido recentemente à reestruturação e consolidação dos grandes grupos económicos detentores desta área do mercado e ao aumento do número de lojas de grande superfície.

A utilização generalizada de códigos de barras, introduzidos com o intuito de obter economias de mão-de-obra e diminuição do tempo de espera junto às caixas registadoras, as quais dispõem actualmente de leitores ópticos, veio permitir, tal como sugerido por Themido (1988), a introdução de sistemas logísticos mais sofisticados e eficientes que procuram tirar partido de informação, até há pouco inexistente. Este estudo insere-se na tentativa de aumento de eficiência e de diminuição de custos logísticos a que actualmente se assiste.

2. Metodologia de Abordagem

A informação histórica disponível e o objectivo de prever a curto prazo (tipicamente poucos dias), conduziu à utilização de métodos de previsão quantitativos. Como se pretendia fazer previsões das vendas diárias de milhares de artigos o modelo devia ser pouco complexo para que os resultados pudessem ser obtidos em tempo aceitável.

Uma loja está organizada em secções que por sua vez estão divididas em famílias de produtos que agregam artigos cobrindo idênticas necessidades dos clientes.

A análise das vendas de algumas dessas famílias de produtos, indicia que a série de vendas apresenta sazonalidade semanal, mensal e anual, que há perturbações nas vendas causadas pelos feriados e que a série apresenta um nível e uma tendência estáveis. As características, de facto, são idênticas às da série para a qual Nóvoa e Guimarães (1993) desenvolveram um modelo Holt-Winters generalizado. Esse modelo permite prever as receitas globais de uma loja, da mesma cadeia, para fins de planeamento financeiro. Contudo, o modelo desenvolvido por aqueles autores não é adequado para aplicação a milhares de artigos porque exige a optimização

de seis taxas de alisamento, para cada artigo, o que implicaria custos computacionais elevados e muito tempo dispendido no desenvolvimento dos modelos de previsão. Para além desta limitação, de ordem técnica, o modelo de Holt-Winters generalizado não é recomendável para prever vendas de artigos individuais que, na actual situação do retalho alimentar, em Portugal apresentam grandes flutuações. Estas flutuações resultam no essencial de três factores:

- **efeitos sazonais** - veja-se, por exemplo, a série de vendas diárias de um artigo (Figura 9), que variam entre um máximo de 30 mil unidades num sábado, do fim do mês anterior à Páscoa, e um mínimo de 4 mil unidades no domingo de Páscoa, o dia com vendas mais baixas em todo o ano;
- **roturas de stock** - deficiências de previsão associadas a dificuldades de fornecimento, ainda vulgares em Portugal, levam frequentemente a roturas de stock, totais ou parciais, que influenciam as vendas do artigo em rotura e as de outros artigos substitutos que vêm as suas vendas aumentar;
- **actividade promocional** - associando frequentemente redução de preço a exposição massificada em topos de corredor as promoções afectam positivamente as vendas do artigo promovido e negativamente as dos restantes artigos de família, seus substitutos directos.

Quando o objectivo é prever as vendas agregadas de uma loja, ou mesmo de uma secção, as perturbações nas vendas, causadas por roturas de stock ou actividade promocional, anulam-se, ficando, no essencial, apenas as perturbações resultantes dos efeitos sazonais o que torna possível a utilização do modelo de Holt-Winters. O mesmo não sendo verdadeiro para artigos individuais, cujas vendas apresentam, como vimos, flutuações resultantes de roturas de stock e actividade promocional dos próprios artigos e dos seus concorrentes directos, a utilização de um modelo de amortecimento exponencial para o cálculo dos factores sazonais resultaria sempre no arrastamento, por períodos sucessivos, de perturbações que nada têm a ver com a sazonalidade mas seriam forçosamente com ela confundidos. Nesta situação, optou-se por um modelo de decomposição clássica, sendo a sazonalidade modelada ao nível da família de produtos e os factores sazonais assim determinados utilizados nos modelos de previsão das vendas individuais de cada artigo da família, como veremos adiante. Esta decisão de modelar a sazonalidade ao nível da família, resultou da forte convicção dos gestores de loja de que não faz sentido considerar diferentes sazonalidades para artigos directamente substituíveis e portanto com características idênticas. Esta convicção foi confirmada, pela análise de séries de vendas de vários artigos, tendo constituído uma decisão fundamental em termos da abordagem seguida.

Tone (1993) desenvolveu um modelo causal de previsão do número de clientes que visitam uma loja Japonesa, que, embora suportado numa análise de regressão múltipla e num contexto provavelmente diverso do verificado em Portugal, considera factores sazonais diários idênticos aos utilizados no modelo estatístico de previsão que seguidamente se descreve.

3. Tratamento da Sazonalidade

Por análise da série de vendas diárias da secção de mercearia, Figura 1, verifica-se que esta apresenta uma variabilidade marcada. Uma análise mais detalhada, apresentada na Figura 2, indicia que o volume de vendas é afectado pelos seguintes factores:

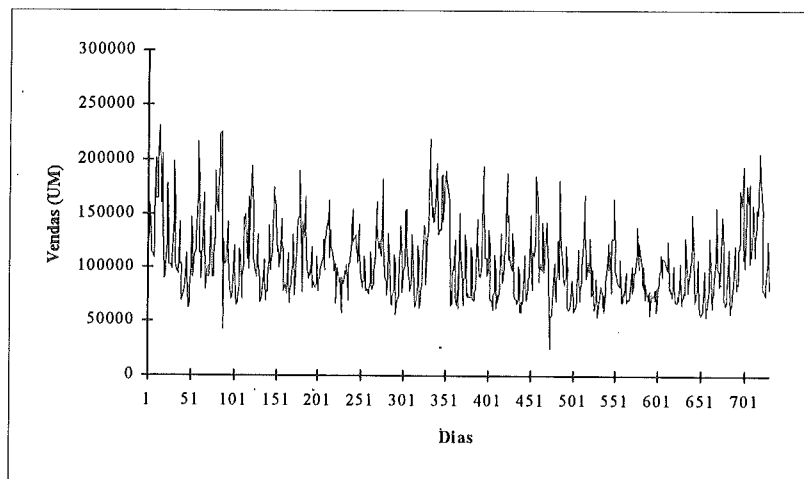


Figura 1 - Volume de vendas diário da secção de mercearia ao longo de 2 anos

- Dia da semana - diferentes dias da semana implicam valores de vendas distintos: no sábado o valor das vendas é superior a qualquer um dos outros dias da semana.
- Proximidade de um feriado - um dia da semana pode passar a ter vendas superiores ou inferiores no caso de se situar próximo ou de coincidir com um feriado. O feriado de Carnaval, por exemplo, cuja localização na semana é sempre à terça-feira, induz um aumento substancial no volume de vendas no dia anterior, segunda-feira, um dia em que o volume de vendas é usualmente baixo. Os feriados de sexta-feira Santa e da Páscoa têm efeitos opostos: na sexta-feira verifica-se um aumento das vendas enquanto que no domingo de Páscoa se verifica uma diminuição significativa das mesmas relativamente aos outros domingos.
- Dia do mês - o volume de vendas é maior no fim e no início do mês. Depois do dia 25 de cada mês, dia a partir do qual a maioria das pessoas começa a receber o salário, as vendas apresentam um valor superior à média, mantendo-se elevadas sensivelmente até ao dia 8 do mês seguinte.
- Época do ano - existem famílias de produtos que registam aumentos das vendas nos períodos que antecedem quadras festivas, como o Natal e a Páscoa, enquanto que outras são influenciadas pela época do ano. Por exemplo, verifica-se um aumento do volume de vendas:
 - do arroz no Natal.
 - dos chocolates na Páscoa e Natal.
 - das conservas de peixe nos meses de Verão.

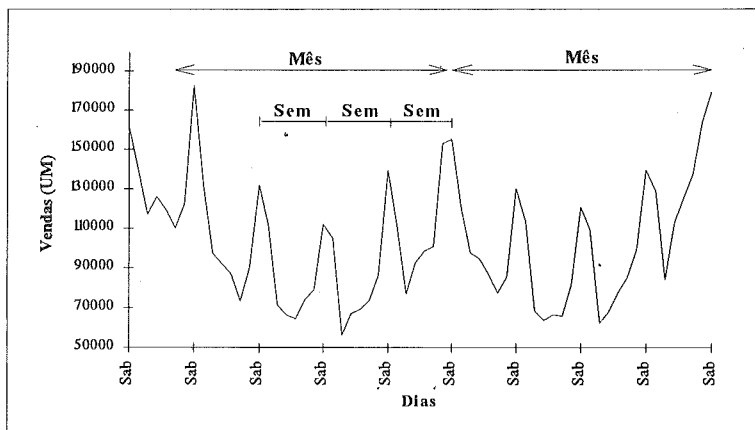


Figura 2 - Volume de vendas diário da secção de mercearia ao longo de 2 meses

Reflectindo esta realidade, o factor sazonal para um determinado dia terá que ser função do dia da semana (FS_s), do dia do mês (FS_m), da época do ano (FS_a) e da proximidade de um feriado (FS_f). Uma vez que o objectivo é prever vendas diárias, o factor sazonal será diário, sendo dado pelo produto das quatro sazonalidades, como se representa na expressão seguinte:

$$FS = FS_s \times FS_m \times FS_a \times FS_f \quad (1)$$

A análise das vendas de uma selecção de artigos, permitiu concluir que o comportamento sazonal é semelhante para todos os artigos pertencentes à mesma família de produtos. As excepções verificam-se apenas quando existe rotura de *stock* de uma marca ou quando algum artigo está em promoção, principalmente se esse artigo for um *leader*. Estas ocorrências introduzem perturbações tanto nas vendas do artigo em causa como nas vendas dos restantes artigos da família, por haver substituição entre artigos.

A constatação de que o comportamento sazonal de todos os artigos de uma família é idêntico, permitiu fazer o cálculo das sazonalidades (semanal, feriados, mensal e anual) ao nível da família de produtos reduzindo assim, drasticamente, o tempo de preparação do modelo a aplicar aos artigos individuais.

Os factores sazonais foram calculados, aplicando o método da decomposição clássica aos valores históricos¹, pela ordem: semanal, feriado, mensal e anual correspondendo o seu produto ao factor sazonal desse dia. A metodologia é apresentada em pormenor em Barroso (1994).

Tal como se procura sugerir na Figura 3, verifica-se que a amplitude da variação das vendas diárias é menor nos meses de Julho, Agosto e Setembro do que nos outros meses. Por esse

¹ Os valores das vendas foram ajustados relativamente à inflação utilizando o Índice de Preços do INE para bens alimentares e bebidas. Esta correcção não é indispensável para os modelos apresentados neste artigo mas torna-se necessária quando se pretende modelar o efeito do preço sobre as vendas.

motivo, determinam-se factores de sazonalidade semanal e mensal distintos para os meses de Verão (Julho, Agosto e Setembro) e para os outros meses.

4. Análise Comparativa de Algumas Famílias de Produtos

Uma análise comparativa, dos factores sazonais de várias famílias de produtos, permitiu concluir que as diferenças entre os factores sazonais semanais, dessas famílias e os das vendas agregadas da secção de mercearia, são desprezáveis. A título de exemplo representam-se na Figura 4 os factores sazonais semanais para o arroz, óleo e conservas de peixe, apresentando todas as outras famílias testadas comportamento idêntico.

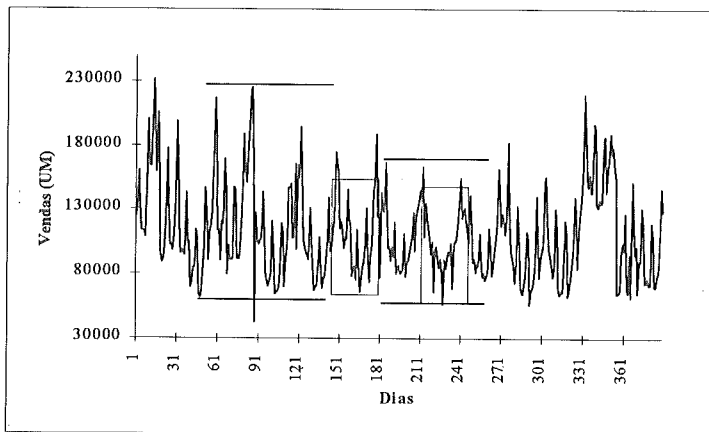


Figura 3 - Amplitude dos valores diários semanais e mensais ao longo de 1 ano

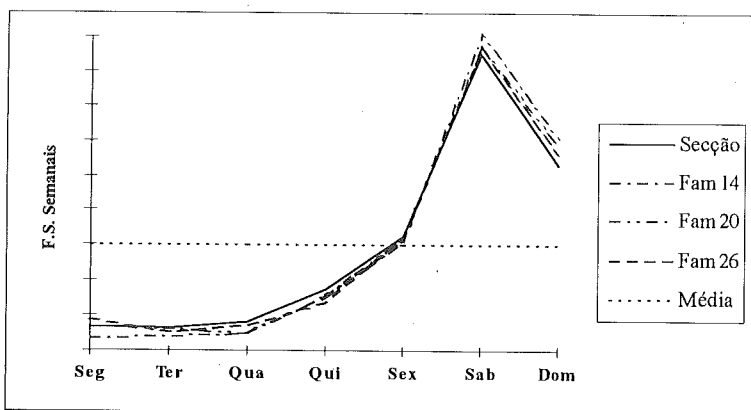


Figura 4 - Factores sazonais semanais para algumas famílias de produtos

Já os factores sazonais mensais, desse mesmo grupo de famílias de produtos, embora com andamento idêntico ao da secção de mercearia, apresentam algumas variações (ver Figura 5), sendo a diferença máxima, aproximadamente, 18%. Este valor, já significativo, reflecte um comportamento diverso das vendas, de cada família de produtos, ao longo do mês.

Em termos de sazonalidade anual, existem comportamentos muito distintos para as várias famílias de produtos, o que implica factores sazonais anuais distintos (ver Figura 6). Note-se, por exemplo, as diferenças entre as vendas da família "conservas de peixe" (Família 26) e da família "açúcar" (Família 23) durante a época do Verão e do Natal. Consomem-se muitas conservas no Verão, chás e infusões (Família 6) no Inverno e açúcar ao longo de todo o ano, com um pico na época do Natal. Estes comportamentos reflectem diferentes apetências para os produtos, ao longo do ano, geralmente associadas a diferenças climatéricas ou festas tradicionais.

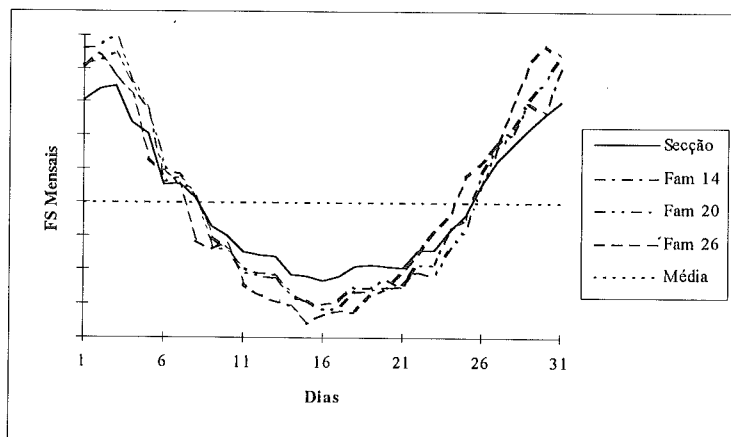


Figura 5 - Factores sazonais mensais para algumas famílias de produtos

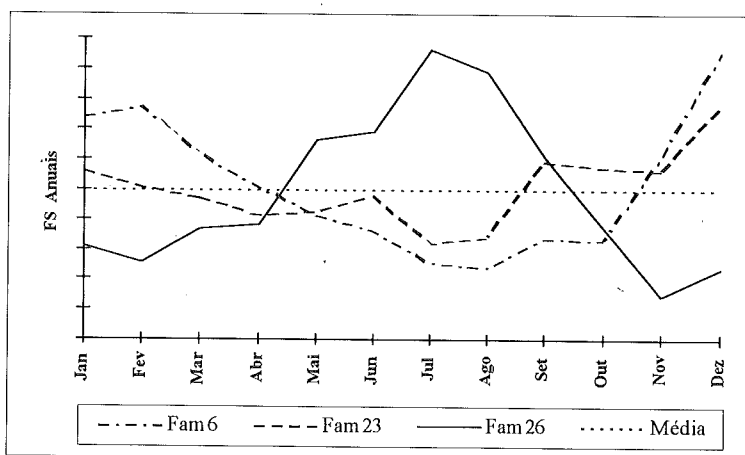


Figura 6 - Factores sazonais anuais para algumas famílias de produtos

5. Análise Comparativa das Lojas

As lojas que fazem parte da cadeia estão situadas em áreas distintas: periferia, interior, litoral, etc. Verifica-se que há lojas que apresentam mais variabilidade nas vendas do que outras, provavelmente porque os padrões de visita dos clientes são diferentes. Por exemplo, ao longo de uma semana, verifica-se que há lojas que vendem relativamente pouco durante os dias de semana e muitíssimo no fim-de-semana enquanto outras, não vendem tanto ao fim de semana mas nos dias de semana apresentam vendas mais elevadas do que as primeiras.

Por esta razão, as lojas foram classificadas em duas categorias distintas:

- Loja Conv - Lojas que se aproximam das lojas de conveniência, pois estão estrategicamente colocadas no percurso emprego-casa e, portanto, apresentam valores mais uniformes para os factores sazonais semanais e mensais.
- Loja Wend - Lojas de fim-de-semana que, tal como o nome sugere, têm uma grande afluência de clientes ao fim-de-semana em detrimento do resto da semana. Assim, há uma grande amplitude entre os factores sazonais semanais dos dias de semana e os factores sazonais semanais de sábado e domingo. O mesmo acontece com os factores sazonais mensais relativamente aos dias do fim/início do mês e meio do mês.

Nas Figuras 7 e 8 ilustram-se as diferenças entre os factores sazonais das lojas dos tipos Loja Conv e Loja Wend. Embora com um padrão idêntico, os factores sazonais apresentam maior amplitude nas lojas com maior afluência de clientes ao fim de semana.

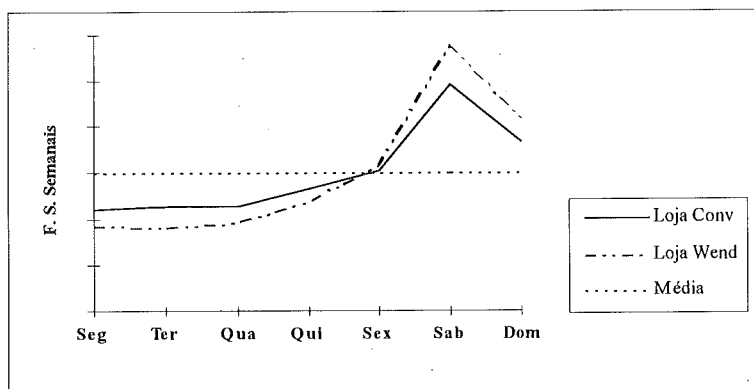


Figura 7 - Factores sazonais semanais para os dois tipos de lojas

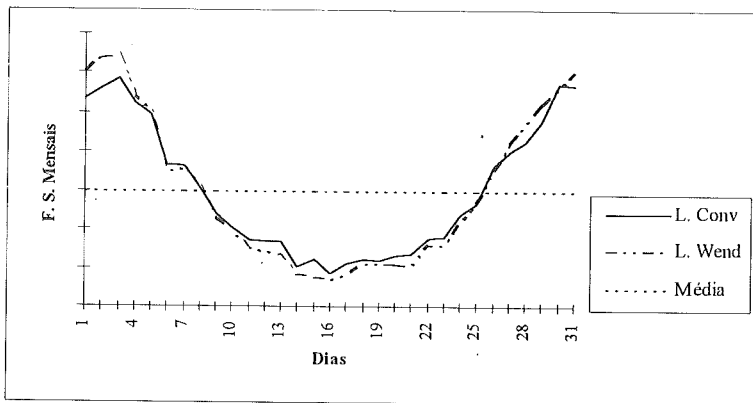


Figura 8 - Factores sazonais mensais para os dois tipos de lojas

6. Modelo de Previsão

A análise de vendas de alguns artigos, permitiu verificar que o nível e a tendência da série de vendas não apresentam um comportamento estável. Para efectuar previsões a curtíssimo prazo, optou-se pela adopção de um modelo de amortecimento exponencial sem tendência. Os erros resultantes da não inclusão da tendência não são relevantes, uma vez que esta vai sendo incorporada no nível à medida que este valor é actualizado.

A previsão das vendas de um artigo para o dia $t+k$, efectuado com base na informação disponível no dia t , é definida de acordo com a expressão seguinte:

$$P_t(k) = N_t \cdot FS_{t+k} \quad (2)$$

onde

$P_t(k)$ - previsão de vendas para o dia $t+k$, efectuada com base na informação disponível no dia t .

N_t - estimativa do nível de vendas obtida com base na informação disponível no dia t .

FS_{t+k} - factor sazonal corresponde ao dia $t+k$ para a família de produtos a que o artigo pertence.

A estimativa do nível, N_t , é obtida recursivamente de acordo com as expressões

$$N_t = N_{t-1} + \alpha e_t \quad (3)$$

e

$$e_t = \frac{V_t}{FS_t} - N_{t-1} \quad (4)$$

onde

N_{t-1} - estimativa do nível obtido com base na informação disponível do dia $t-1$.

α - constante de alisamento, que pode tomar valores entre 0 e 1.

e_t - desvio entre o nível observado e previsto para t (dado pela diferença entre V_t/FS_t e a estimativa anterior do nível, N_{t-1}).

V_t - vendas reais relativas ao período t .

O parâmetro do modelo, α , é calculado para cada artigo por forma a minimizar o somatório dos desvios quadráticos, e_t^2 , observados ao longo da série cronológica.

O período para o qual se pretende prever poderá ser igual a um ou mais dias, dependendo da loja e do esquema de abastecimento do artigo. A previsão de vendas de um artigo para um período de n dias pode obter-se somando as previsões efectuadas para cada um dos n dias.

7. Qualidade do Modelo de Previsão

Aplicando o modelo de previsão de vendas a um artigo típico, pertencente à secção de mercearia, para um período igual a um dia, e considerando uma constante de alisamento igual a 0.5 (ver Figura 9), verifica-se que a sazonalidade se ajusta perfeitamente à série real das vendas. De referir, em particular, o bom ajustamento das previsões realizadas para a Páscoa (11 de Abril de 1993). Deve notar-se que o cálculo dos factores sazonais e a determinação da constante de alisamento não foram efectuados com o conjunto de valores que serviram para testar o ajustamento do modelo.

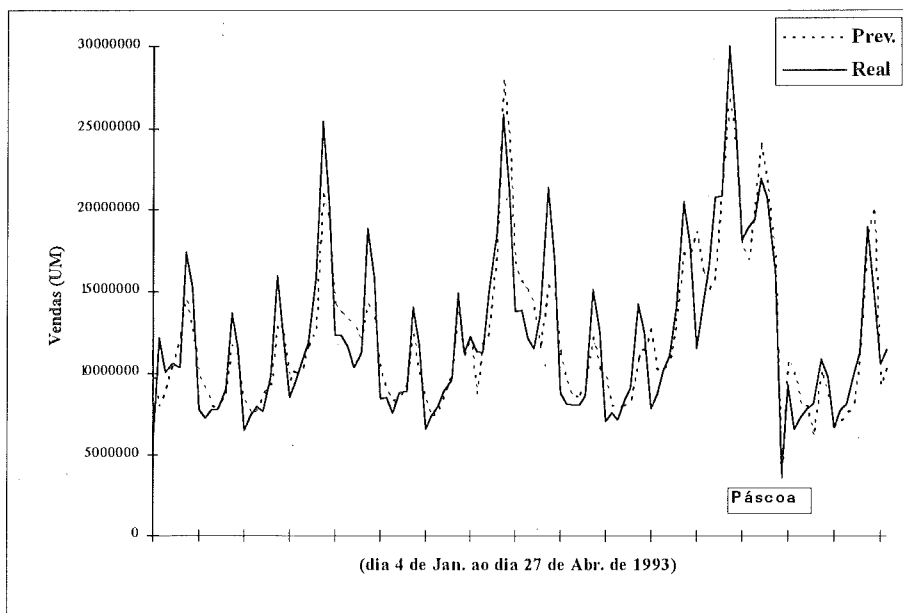


Figura 9 - Previsão de vendas de um artigo da secção de mercearia para o primeiro quadrimestre de 1993

8. Aplicação do Modelo: Implicações da Análise ABC aos Artigos

A aplicação de uma análise ABC aos cerca de 1900 artigos da secção de mercearia permitiu verificar que uma pequena percentagem dos artigos são responsáveis por um elevado volume de vendas. O conjunto total de artigos foi subdividido em três classes:

- Classe A - Aproximadamente 10% dos artigos que correspondem a 65% do volume total de vendas
- Classe B - Cerca de 30% dos artigos que correspondem aproximadamente a 25% do volume total de vendas
- Classe C - Os restantes artigos, cerca de 60%.

Para a classe A, à qual pertence um número reduzido de artigos com elevado volume de vendas bem como os artigos em promoção da classe B, o modelo de previsão desenvolvido é susceptível de ser melhorado, designadamente, através da incorporação de elasticidades directas e cruzadas das vendas, em relação ao preço, que permitem ajustar mais rapidamente o nível das vendas de cada artigo, face a variações do preço desse artigo e dos seus concorrentes (ver Themido (1988)). Para a classe B, com um número de artigos considerável, o modelo desenvolvido afigura-se adequado. Finalmente, não parece haver necessidade de fazer previsões para os artigos da classe C pois, sendo o seu consumo muito baixo e a frequência de abastecimento elevada, as recomendações de encomenda, do modelo de previsão, seriam muito pequenas, quando comparadas com a quantidade mínima de encomenda. Outras técnicas podem ser utilizadas para a gestão de *stocks* destes artigos como por exemplo encomendar a quantidade mínima quando há espaço suficiente no expositor.

Na Tabela 1 identificam-se as diferentes alternativas de aplicação de modelos às três classes de artigos. Desta aplicação selectiva de métodos de previsão, poderão obter-se as seguintes vantagens:

- i) Economias informáticas significativas, já que para os artigos do tipo C não são feitas previsões (notar que estes artigos representam cerca de 60% dos artigos de mercearia, sendo possível que a percentagem possa ser ainda maior noutras secções).
- ii) Concentração da atenção dos gestores nos artigos de tipo A e de tipo B, por serem aqueles em que a redução dos custos de aprovisionamento pode conduzir à obtenção de melhores resultados e, em relação aos quais, é mais importante oferecer ao cliente um bom nível de serviço.

MODELO DE PREVISÃO A UTILIZAR	
A	Prever vendas diárias. O efeito preço influencia o nível. Modelo de previsão com NÍVEL (função do preço) e sazonalidade: SEMANAL, FERIADOS, MENSAL e ANUAL à família de produtos.
B	Prever vendas diárias utilizando o modelo de previsão com NÍVEL e sazonalidade: SEMANAL, FERIADOS, MENSAL e ANUAL à família de produtos.
C	Não é realizada qualquer previsão. Encomenda-se a quantidade mínima quando há espaço no expositor.

Tabela 1 - Aplicação do modelo de previsão mais adequado em função do tipo de artigo

9. Conclusões

Foi possível, no âmbito deste estudo, modelar a sazonalidade das vendas diárias, ao nível das famílias de produtos. Na Tabela 2 identificam-se os factores que afectam o comportamento sazonal das vendas. Uma vez modelados os factores sazonais, estes foram incorporados num modelo de amortecimento exponencial, sem tendência, desenvolvido com o objectivo de efectuar previsões a curto prazo.

	VERÃO/OUTROS	FAMÍLIAS	LOJAS
SAZ. SEMANAL	Varia	Varia Pouco (Máx.5,8%)	Varia
SAZ. FERIADOS		Varia Pouco	Varia
SAZ. MENSAL	Varia	Variação Intermédia (Máx.18%)	Varia
SAZ. ANUAL		Varia Muito	Varia

Tabela 2 - Resumo da variação dos factores sazonais

Considera-se que, para uma vasta gama de artigos, o modelo satisfaz totalmente, carecendo apenas de actualização do cálculo dos parâmetros, à medida que informação mais recente se torne disponível.

Efectivamente, todas as experiências feitas com o modelo de previsão demonstram que a sazonalidade se ajusta perfeitamente à série real das vendas do ano de 1993. O abandono prematuro da aplicação informática anterior responsável pelas previsões, e respectivos factores sazonais, não permitiu a comparação dos resultados do modelo anterior com os deste modelo de previsão. No entanto, a implementação imediata dos factores sazonais determinados e os bons resultados obtidos nas lojas, com factores sazonais ainda relativamente grosseiros, permite concluir que os resultados do modelo são bons. Em particular, o ajustamento das previsões feitas para a Páscoa de 1993 sugere que a modelação proposta é robusta e flexível.

Para os artigos da classe A e, eventualmente, para os da classe B que sejam objecto de promoções será desejável incluir, no modelo, o efeito de variáveis exógenas, tais como o preço dos artigos.

Bibliografia

- [1] Lopes, Pedro Durão, *A Gestão de Stocks no Retalho Alimentar*, Tese de Mestrado em Investigação Operacional e Engenharia de Sistemas, Instituto Superior Técnico, Lisboa (1991).
- [2] Themido, Isabel Hall, *Modelos de Investigação Operacional para Apoiar a Tomada de Decisão em Supermercados*, Investigação Operacional 8 (1988) 3-16.
- [3] Nóvoa, Henriqueta e Guimarães, Rui, *Forecasting of Daily Sales of Hypermarkets Using a Modified Holt-Winters Procedure*, Comunicação apresentada na Conferência IFORS93, Lisboa (1993).
- [4] Barroso, Ana Paula, *Um Modelo de Previsão de Vendas para o Retalho Alimentar*, Tese de Mestrado em Investigação Operacional e Engenharia de Sistemas, Instituto Superior Técnico, Lisboa (1994).
- [5] Tone, Kaoru, *Methods for Utilizing POS (Point-of-Sale) Data*, Comunicação apresentada na Conferência IFORS93, Lisboa (1993).
- [6] Mercer, Alan, *Developments in Implementable Retailing Research*, European Journal of Operational Research 68 (1993) 1-8.

INSTRUÇÕES AOS AUTORES

Os autores que desejem submeter um artigo à Investigação Operacional devem enviar três cópias desse trabalho para:

Prof. Joaquim J. Júdice
Departamento de Matemática
Universidade de Coimbra
3000 Coimbra, Portugal

Os artigos devem ser escritos em Português ou Inglês. A primeira página deve conter a seguinte informação:

- Título do artigo
- Autor(es) e instituição(ões) a que pertence(em)
- Abstract (em inglês)
- Resumo
- Keywords (em inglês)
- Título abreviado

As figuras devem aparecer em separado de modo a poderem ser reduzidas e fotocopiadas. As referências devem ser numeradas consecutivamente e aparecer por ordem alfabética de acordo com os seguintes formatos:

Artigos: autor(es), título, título e número da revista (livro com indicação dos editores), ano, páginas.

Livros: autor(es), título, editorial, local de edição, ano.

**Fotografia, Montagem
Impressão e Acabamentos**
Tip. Nocamil
COIMBRA

ÍNDICE

M. S. Carvalho, A. Guimarães Rodrigues, D. V. Vliet e P. Timms, Experiência comparativa de utilização de algoritmos de "assignment" em problemas de estimação de matrizes de origem/destino	3
J. A. Soeiro Ferreira, Optimal Control of discrete systems: recent results and case studies	17
D. F. Luzes, I. H. Themido, Forecasting horticultural prices: a comparison of methods	33
M. A. M. Ferreira, Distribuição do período de ocupação da fila de espera $MIDI_{\infty}$	43
R. A. Costa, Modelação de sistemas: uma experiência de ensino na FCT-UNL.....	57
H. B. Fraga da Silva, Modelo de segurança do trabalho.....	69
A. P. Barroso, I. H. Themido e D. F. Luzes, Um modelo de previsão de vendas para o retalho alimentar	75



Associação Portuguesa para o Desenvolvimento
da Investigação Operacional

CÉSUR - Instituto Superior Técnico - Avenida Rovisco Pais
1000 Lisboa - Telef. 80 74 55