

IMPLEMENTAÇÃO DO ALGORITMO DE BOX E HILL PARA A DISCRIMINAÇÃO ENTRE MODELOS COMPETITIVOS.

Roberto Francisco Marques Mendes

Aluno da Universidade Braz Cubas - Bolsa PIBIC/CNPQ

Orientador: Dr. Ralf Gielow, Pesquisador, Divisão de Ciências Meteorológicas

O algoritmo de Box e Hill, com base no conceito de entropia da informação e no teorema Bayes, é um procedimento sequencial para a discriminação entre m modelos que competem para representar um determinado fenômeno ou sistema n -dimensional, assim como indica uma maneira para melhorar esta discriminação, considerando dados observacionais e respectivos erros de medida.

Assim, para cada modelo atribui-se inicialmente sua probabilidade (igual ou não às dos demais) e, dispondo de M pontos experimentais ou observacionais (M maior que o número de parâmetros em qualquer modelo), ajusta-se cada modelo a estes pontos - por regressão não linear - obtendo-se os parâmetros correspondentes e a variância de cada ajuste. Então, utilizando-se o teorema de Bayes, atualizam-se as probabilidades dos modelos. A seguir, através da maximização de uma função discriminação, determina-se o ponto para a realização de nova observação, cujo resultado - após reajuste dos parâmetros de cada modelo e de sua probabilidade -, é utilizado para melhorar a discriminação; prossegue-se assim sucessivamente, até clara discriminação em favor de um dos modelos.

A implementação do algoritmo, feita de forma amigável para o usuário, divide-se em dois módulos: o primeiro, utilizando um analisador sintático simplificado, gera um arquivo de dados referentes aos modelos, enquanto o segundo, a partir deste arquivo e dos dados observados, realiza os ajustes e determina a discriminação entre modelos, indicando também como realizar observações adicionais para melhorar a discriminação. Para os ajustes não-lineares de parâmetros, utiliza-se o algoritmo de Marquardt, modificado por Nash. Os códigos estão escritos em linguagem C e Mathematica, mas seu uso não exige conhecimento destas linguagens.

Finalmente, apresenta-se a aplicação do algoritmo à discriminação entre quatro modelos para representar a cinética química de uma reação $A \rightarrow B$, em que a concentração de A depende da temperatura e do tempo de reação, chegando-se a uma perfeita discriminação, mesmo quando se parte de probabilidades a priori completamente erradas, deste modo demonstrando a robustez do método.

ALGORITMO DE BOX E HILL PARA DISCRIMINAÇÃO ENTRE MODELOS COMPETITIVOS

ORIENTADOR: Ralf Gielow

BOLSISTA (PIBIC): Roberto Francisco Marques Mendes

PERÍODO: Agosto de 1995 A Julho de 1996

RELATÓRIO FINAL

PARECER DO ORIENTADOR

Analisando o relatório final, verificamos que ele retrata bem o progresso do Bolsista, o qual foi bem satisfatório, conforme pudemos verificar acompanhando o seu trabalho. O algoritmo, de grande complexidade teórica, foi bem entendido pelo Bolsista, que o implantou de forma amigável para o usuário, utilizando várias ferramentas computacionais. Esta tarefa não foi simples, envolvendo um analisador sintático, derivação analítica automática e ajuste não-linear de parâmetros de funções, os quais foram programados de modo invisível para o usuário, que deve preocupar-se apenas com os modelos, os dados e a interpretação dos resultados. O funcionamento do programa resultante foi testado com sucesso utilizando dados sintéticos e modelos concernentes a uma situação de cinética química, demonstrando que a ferramenta resultante é de fácil operação e um meio objetivo para discriminar modelos. Como continuação, pretende-se tornar a interface mais amigável ainda, ao mesmo tempo procedendo-se à determinação do melhor modelo para a estimativa de componentes do balanço de energia na interface solo-planta-atmosfera, utilizando dados observacionais obtidos na Região Amazônica, preparando o Bolsista para atuar na área de Computação Científica. O Bolsista, em si, é muito responsável e tem boa iniciativa, especialmente na área computacional, além de grande interesse científico, o que o recomenda para uma carreira de pesquisa



Ralf Gielow
Orientador



PARA:	NOSSA REF.	SUA REF.
DE: ROBERTO FRANCISCO MARQUES MENDES	05 107196	RAMAL 25-6644

ASSUNTO

Encaminhamento do relatório final

TEXTO

Estou encaminhando cópia do relatório final, referente ao período de agosto de 1995 a julho de 1996.

Agradeço pelas providências

Atenciosamente.

IMPLEMENTAÇÃO DO ALGORITMO DE BOX E HILL PARA DISCRIMINAÇÃO
ENTRE MODELOS COMPETITIVOS.

RELATÓRIO FINAL

ORIENTADOR: Ralf Gielow

BOLSISTA (PIBIC): Roberto Francisco Marques Mendes

PERÍODO: agosto de 1995 a julho 1996

Algoritmo de Box e Hill para discriminação entre modelos competitivos

1. Introdução

Não raro, fenômenos naturais ou processos artificiais como evapotranspiração, reações químicas, acessibilidade urbana ou rendimento agrícola, podem ser representados matematicamente por mais que um modelo, de acordo com o mecanismo de funcionamento do sistema ou ajuste estatístico-matemático que o observador ou experimentador considerar. Surge assim o problema de determinar qual o modelo que melhor representa o sistema em estudo.

O algoritmo de Box e Hill (1967) constitui um procedimento para a discriminação entre modelos que competem para representar um determinado fenômeno ou processo, em que uma variável dependente é função de várias variáveis independentes e parâmetros, e do qual se tem dados observados ou medidas experimentais sob diversas condições, cobrindo todo o domínio das variáveis, e se conhece o erro observacional ou experimental. Em adição, no caso de se poder realizar observações ou experimentos adicionais após a discriminação inicial entre os modelos, o algoritmo indica quais os valores das variáveis independentes a serem examinados subsequentemente para melhorar a discriminação.

Neste trabalho, o objetivo é implementar o algoritmo de Box e Hill na forma de um programa de computador utilitário de interface ser-humano-máquina amigável, que permita resolver problemas de discriminação do tipo acima descrito. Apresenta, também, sua aplicação em situação efetiva.

2. Análise do material bibliográfico

Os modelos objeto deste estudo têm a forma de uma variável Y que depende de variáveis independentes $X = (X_1, X_2, \dots, X_k)$ e parâmetros $TETA = (TETA_1, TETA_2, \dots, TETA_p)$, como seja, $Y = f(TETA, X)$.

Box e Hill (1967) desenvolveram um algoritmo, baseado no conceito de entropia da informação e no teorema de Bayes para discriminar entre M modelos que competem entre si para representar um sistema, conhecendo-se observações ou medidas experimentais, e respectiva variância, cobrindo todo o domínio das variáveis independentes. O número de variáveis independentes e parâmetros pode ser diferente em cada modelo. Inicialmente impõe-se uma probabilidade para cada modelo, não necessariamente $1/M$, a qual é corrigida após consideração dos dados iniciais. Não havendo uma clara discriminação em favor de um dos modelos, e sendo possível observações ou medidas adicionais, o algoritmo, comparando os valores Y observados ou medidos inicialmente com os previstos pelos diversos modelos, determina sequencialmente sob que condições (valores das variáveis independentes) proceder para melhorar a discriminação, repetindo-se

sucessivamente o processo até se ter uma clara discriminação.

Wadsworth (1990) cita tão somente o algoritmo de Box e Hill para este tipo de discriminação, dito bayesiano. Os métodos bayesianos (Feigelson e Babu, 1992) consideram a probabilidade como uma medida da plausibilidade de uma hipótese (modelo), em contraposição à visão frequencial, que identifica a probabilidade com a frequência relativa de ocorrência de um resultado de uma infinidade de repetições "idênticas" de um experimento ou observação. A inferência bayesiana enfoca hipóteses alternativas, enquanto a estatística frequencial enfoca conjuntos de dados. Para avaliar uma hipótese H, o enfoque bayesiano compara a probabilidade de H com as probabilidades de outras hipóteses; já os métodos frequenciais supõem H verdadeira e comparam a probabilidade dos dados observados ou medidos com as probabilidades de outros conjuntos de dados preditos por H. Pragmaticamente, há fortes evidências da superioridade dos métodos bayesianos em aplicações reais. Não obstante, em qualquer caso, a maior dificuldade matemática usualmente está na estimativa dos valores dos parâmetros que conectam não-linearmente as variáveis independentes, levando a procedimentos iterativos, ditos regressões não-lineares, que necessitam de estimativas iniciais e podem levar a resultados que dependem fortemente destas. Há muitos métodos de regressão não-linear, como os baseados na eliminação de Gauss e suas variantes (Wadsworth, 1990), sendo cada um mais indicado para certos tipos de funções.

O algoritmo de Box e Hill foi aplicado por Adeodato de Souza (1970) em problema de cinética química (equilíbrio oxigênio-hemoglobina - 8 modelos competitivos) e por Silva Filho (1976) em problema urbano (função acessibilidade entre células urbanas - 5 modelos). Por sinal, a dissertação de Silva Filho (1976), q.v, detalha muito bem o algoritmo de Box e Hill e suas fundamentações. Entretanto, a implementação computacional de ambos os trabalhos é em linguagem Fortran, com rotinas que devem ser reprogramadas para cada caso. Na presente implementação, pretende-se minimizar os esforços de programação computacional, tornando os programas mais amigáveis. Para tanto, faz-se necessário um analisador sintático (Schildt, 1991), para identificar cada modelo em estudo.

3. Procedimentos metodológicos

A implementação do algoritmo de Box e Hill divide-se em dois módulos (programas). O primeiro tem a finalidade de gerar um arquivo com dados referentes aos modelos contando internamente com um analisador sintático simplificado. O segundo programa deverá acessar um arquivo gerado pelo primeiro módulo e, a partir dos dados lidos, determinar a discriminação entre os modelos. Utiliza-se a linguagem C para implementação do primeiro módulo, enquanto o segundo

foi desenvolvido em Mathematica, com possível comunicação com programas em outras linguagens de diversos níveis.

3.1 Primeiro módulo

O programa inicialmente questiona o usuário com relação ao número de modelos a serem testados. Com base na informação fornecida, o programa solicita a entrada dos M modelos especificados pelo usuário, sendo necessária uma observação com relação as regras de entrada. Cada modelo será passado ao computador na forma mais amigável possível, como uma expressão algébrica, onde as variáveis e os parâmetros são definidos na sua forma literal, admitindo-se também constantes numéricas e funções aceitáveis pelo analisador sintático (sen, cos, exp, ...). Os parâmetros são representados por qualquer sequência de caracteres, alfabéticos e maiúsculos (TETA, BETA, ALGO), enquanto as variáveis o são por letras minúsculas (x, var, t, z).

3.1.1 Analisador sintático

A utilização de expressões algébricas necessita um analisador sintático, cuja função é verificar se a expressão que representa um modelo está dentro das regras estabelecidas. O analisador executa uma avaliação geral de cada modelo, identificando erros e divulgando sua natureza. Caso um erro seja identificado, o modelo não será aceito, sendo exigida sua redigitação. Ver resultados no Apêndice A.

3.1.1.1 Regras de sintaxe

As regras de sintaxe para o formato dos modelos são apenas cinco, especificadas da seguinte maneira:

1ª) Toda expressão deve possuir algum elemento, não sendo aceitas expressões vazias.

2ª) Não são admitidos espaços em branco entre os componentes da expressão.

3ª) Quando algum parêntese for aberto, deve-se fechá-lo.

4ª) A expressão deve ser matematicamente válida.

5ª) Um modelo pode conter apenas caracteres alfabéticos, números e símbolos de operações aritméticas. Qualquer símbolo fora deste domínio será considerado um erro.

3.2 Segundo módulo

A função do segundo módulo, inicialmente, é ler o arquivo de informações que foi previamente gerado pelo primeiro, o qual identifica os modelos, seus componentes, e suas respectivas probabilidades a priori. As derivadas

parciais são obtidas automaticamente, de forma algébrica. Finalmente executa-se o algoritmo de Box e Hill, retornando os resultados de interesse do usuário.

3.2.1 O software Mathematica

Para a implementação do segundo módulo utiliza-se o software Mathematica. A implementação do algoritmo de Box e Hill pode ser feita de vários modos, utilizando-se qualquer linguagem, sendo necessário porém que esta manipule matrizes e, de certa forma, equações algébricas. Dadas essas necessidades, uma busca foi realizada para a identificação da ferramenta mais adequada à tarefa, tendo sido escolhido o software Mathematica. A implementação do código nesta linguagem foi decidida em virtude da sua facilidade de uso; e por incluir funções de cálculo vetorial e, também, uma gama de funções para a manipulação algébrica de expressões. Sua escolha resultou em uma maior facilidade de implementação e de operação do algoritmo, corroborando a afirmação de Nordemann (1994), de que o Mathematica representa um ambiente poderoso para a execução de cálculos e operações matemáticas. sendo muito valioso para cálculos complexos e sofisticados, que podem incluir operações sobre símbolos literais e fórmulas, como derivação e integração.

3.2.2 Implementação do algoritmo de Box e Hill

Esta implementação de Box e Hill é composta de três partes principais ou submódulos de programação. Cada submódulo pode, inclusive, funcionar de forma autônoma: No primeiro implementa-se o algoritmo de Marquardt melhorado por Nash(1990), para o ajuste não-linear de parâmetros; no segundo realiza-se a discriminação utilizando a estatística Bayesiana e no terceiro maximiza-se a função discriminação, procedimento que indica, se forem possíveis medidas adicionais, quais os valores das variáveis independentes permitirão uma melhora na discriminação entre os modelos.

O programa foi escrito em Mathematica, necessitando pouca intervenção do usuário, quando da adaptação do código ao seu problema específico. Exige apenas pequenas configurações em sua estrutura. Ver apêndice B em especial diagrama de fluxo.

3.2.2.1 Implementação do algoritmo para o ajuste de parâmetros

Internamente o software Mathematica, apresenta uma função já implementada (**NonlinearFit**), a qual não exige estimativa inicial de parâmetros a ajustar, mas que não foi muito eficiente para as funções que representam os modelos estudados no caso teste. Uma solução foi encontrada com a implementação do algoritmo de Marquardt melhorado por Nash (que exige estimativa inicial dos parâmetros). Para a situação estudada ele mostrou-se eficiente, ajustando bem os parâmetros dos modelos testados; não obstante, o ajuste de

parâmetros pode ser feito internamente pelo algoritmo, ou inserido de forma já ajustada pelo usuário.

4. Resultados

Os dois módulos que compõem o sistema completo estão implementados e operacionais. O primeiro módulo foi totalmente desenvolvido em linguagem C possuindo um código fonte de aproximadamente 510 linhas, atingindo o objetivo de particionar cada modelo em seus componentes, e gravar o resultado em um arquivo de informações.

Após a leitura do arquivo de dados gerado pelo primeiro módulo, todos os modelos e seus componentes são identificados pelo segundo módulo, exigindo apenas que sejam introduzidas algumas informações com respeito a parâmetros e variáveis.

No que concerne a aplicação prática novel do algoritmo de Box e Hill na área de Micrometeorologia, a grande dificuldade é a não disponibilidade de conjuntos completos de dados observacionais; entretanto, para o teste de funcionamento do programa, foram utilizados dados sintéticos e modelos concernentes a uma situação de cinética química, tendo o programa identificado o modelo que melhor representa a situação.

4.1 Aplicação a uma situação de cinética química

O mecanismo de uma certa reação química do tipo $A \rightarrow B$, pode ser representado por 4 modelos independentes

$$\text{Modelo 1 - } E(y) = \text{Exp}[-\text{TIMI} \cdot x_i \cdot \text{Exp}[-\text{TIIMI}/x_{ii}]]$$

$$\text{Modelo 2 - } E(y) = 1 / (1 + \text{TIMII} \cdot x_i \cdot \text{Exp}[-\text{TIIMII}/x_{ii}])$$

$$\text{Modelo 3 - } E(y) = 1 / (1 + 2 \cdot \text{TIMIII} \cdot x_i \cdot \text{Exp}[-\text{TIIMIII}/x_{ii}])^{1/2}$$

$$\text{Modelo 4 - } E(y) = 1 / (1 + 3 \cdot \text{TIMIV} \cdot x_i \cdot \text{Exp}[-\text{TIIMIV}/x_{ii}])^{1/2}$$

sendo y a concentração de A ; x_i o tempo de reação e x_{ii} a temperatura. Testaram-se três situações, como sejam, (i) equiprobabilidade, a probabilidade igual para todos os modelos, (ii) aumento da probabilidade a priori do modelo mais provável, determinado por (i) e (iii) alocação de menor probabilidade ao modelo mais provável. Os modelos e os dados, sintetizados por perturbação gaussiana sobre um modelo de cinética química de 2ª ordem, com variância de 0,05 são os utilizados por Box e Hill(1967).

5. Discussão dos resultados e conclusões

A Tabela 1 mostra os dados utilizados, retirados de Box e Hill(1967)- inicialmente quatro medidas, e as probabilidades a posteriori, Π_i , resultantes.

Como mostra a Tabela 1, após três rodadas o algoritmo demonstra claramente que o modelo 2 é o que melhor se ajusta aos valores medidos, sendo os demais pouco relevantes.

Uma interpretação visual do resultado, consta do apêndice B, através de gráficos referentes aos valores sintéticos e aos previstos pelos modelos em teste.

Tabela 1 - Discriminação entre quatro modelos
 $\sigma^2=0,05 \quad \Pi_1=\Pi_2=\Pi_3=\Pi_4=0,25$

n	xi	xii	Y	Π_1	Π_2	Π_3	Π_4
1	25	575	0,3961				
2	25	475	0,7232				
3	125	475	0,4215				
4	125	575	0,1297	0,0455	0,8804	0,0576	0,0164
5	138	558	0,1399	0,0391	0,9609	0,0000	0,0000
6	120	546	0,1812	0,0272	0,9728	0,0000	0,0000

Se atribuída probabilidade a priori maior para o melhor modelo, o processo tende a se acelerar, como pode ser visto na Tabela 2.

Tabela 2 - Discriminação entre quatro modelos
 $\sigma^2=0,05 \quad \Pi_1=0,2 \quad \Pi_2=0,4 \quad \Pi_3=0,2 \quad \Pi_4=0,2$

n	xi	xii	Y	Π_1	Π_2	Π_3	Π_4
1	25	575	0,3961				
2	25	475	0,7232				
3	125	475	0,4215				
4	125	575	0,1297	0,0242	0,9364	0,0306	0,0087
5	138	558	0,1399	0,0199	0,9801	0,0000	0,0000
6	120	546	0,1812	0,0138	0,9862	0,0000	0,0000

Finalmente, mesmo atribuindo aos modelos probabilidades a priori desfavorecendo o melhor modelo, com um número adicional de experimentos chega-se ao resultado correto, conforma se vê na Tabela 3, deste modo demonstrando a robustez do método.

Tabela 3 - Discriminação entre quatro modelos
 $\sigma^2=0,05 \quad \Pi_1=0,4 \quad \Pi_2=0,1 \quad \Pi_3=0,2 \quad \Pi_4=0,3$

n	xi	xii	Y	Π_1	Π_2	Π_3	Π_4
1	25	575	0,3961				
2	25	475	0,7232				
3	125	475	0,4215				
4	125	575	0,1297	0,1485	0,7175	0,0938	0,0402
5	138	558	0,1399	0,1398	0,8601	0,0000	0,0000
6	120	546	0,1812	0,1006	0,8993	0,0000	0,0000

No Apêndice B mostra-se o material utilizado nas apresentações públicas deste trabalho.

Concluindo, o algoritmo de Box e Hill mostra-se como uma poderosa ferramenta para a discriminação entre modelos competitivos. Outrossim, na forma amigável ora implementada, espera-se que aumente sua utilização constituindo-se num meio objetivo para discriminar modelos. Não obstante, recomenda-se aperfeiçoamentos adicionais em sua forma, a fim de torná-lo mais amigável ainda.

Referências bibliográficas

- Abell, M. L.; Braselton J. P. **The Mathematica Handbook**. London, Academic Press, 1992. 789 p.
- Abell, M. L.; Braselton J. P. **Mathematica by Example**. London, Academic Press, 1994. 523p.
- Adeodato de Souza Neto, J. **Discrimination among mechanistic models for oxigen-hemoglobin equilibrium**. Doctoral Dissertation, Gainesville, University Of Florida, 1970.
- Box, G.E.; Hill, W.J. Discrimination among mechanistic models. **Technometrics**, 9(1):57-71, Feb. 1967.
- Feigelson, E.F.; Babu G.J. **Statistical Challenges in Modern Astronomy**. New York, Springer, 1992.
- Nash, J.C. **Compact Numerical Methods for Computers**. New York, Adam Hilger, 1990. 262p.
- Nordemann, D.J.R. **Introdução ao Mathematica for Windows**. São José dos Campos, Transtec, 1994. 171 p.
- Schildt, H. **C Completo e Total**. São Paulo, Makron Books, 1991. 890 p.
- Silva Filho, J. F. da. **Discriminação entre modelos competitivos**. Dissertação de mestrado em Análise de Sistemas e Aplicações - Instituto de Pesquisas Espaciais, São José dos Campos, junho de 1977. 109 p. (INPE-1027-TPT/051).
- Wadsworth, H.M.J. **Handbook of Statistical Methods for Engineers and Scientists**. New York, McGraw-Hill, 1990. 525p.

Apêndice A

Resultados do analisador sintático

```
4
4
2
2
0.05
Exp[-TIMI*xi*Exp[-TIIMI/xii]]
0.25
2
TIMI
TIIMI
2
xi
xii
1/(1+TIMII*xi*Exp[-TIIMII/xii])
0.25
2
TIMII
TIIMII
2
xi
xii
1/(1+2*TIMIII*xi*Exp[-TIIMIII/xii])^1/2
0.25
2
TIMIII
TIIMIII
2
xi
xii
1/(1+3*TIMIV*xi*Exp[-TIIMIV/xii])^1/2
0.25
2
TIMIV
TIIMIV
2
xi
xi
```

Apêndice B

Trabalhos resultantes

Os resultados deste trabalho, sob o título "Implementação do algoritmo de Box e Hill para a discriminação entre modelos competitivos", foram apresentados no 2º Seminário de Iniciação Científica, realizado no INPE, de 27 a 28 de maio de 1996, tendo também sido aceitos para apresentação na 3ª JORNADA NACIONAL DE INICIAÇÃO CIENTÍFICA (Área C.5 - Matemática, Computação e Estatística), a realizar-se de 08 a 12 julho de 1996, na Pontifícia Universidade Católica de São Paulo.

A seguir, apresentam-se as folhas que constituirão o painel que ficará exposto das 13:00 h do dia 10 às 10:00 h do dia 11 de julho de 1996, com presença do autor das 13:00 às 15:00 h do primeiro dia, no recinto da 48ª Reunião Anual da SBPC; forma preliminar deste trabalho foi apresentada, em transparências, no seminário acima citado.

Introdução

O algoritmo de Box e Hill, com base no conceito de entropia da informação e no teorema de Bayes, é um procedimento seqüencial para a discriminação entre m modelos que competem para representar um determinado fenômeno ou sistema n -dimensional, assim como indica uma maneira para melhorar esta discriminação, considerando dados observacionais e respectivos erros de medida.

Assim, para cada modelo atribui-se inicialmente sua probabilidade (igual ou não às dos demais) e, dispondo de M pontos experimentais ou observacionais (M maior que o número de parâmetros em qualquer modelo), ajusta-se cada modelo a estes pontos - por regressão não linear - obtendo-se os parâmetros correspondentes e a variância de cada ajuste. Então, utilizando-se o teorema de Bayes, atualizam-se as probabilidades dos modelos. A seguir, através da maximização de uma função discriminação, determina-se o ponto para a realização de nova observação, cujo resultado - após reajuste de parâmetro de cada modelo e de sua probabilidade -, é utilizado para melhorar a discriminação; prossegue-se assim sucessivamente, até clara discriminação em favor de um dos modelos.

Equações do algoritmo de Box e Hill

Variância do ajuste de cada modelo i aos dados experimentais:

$$\sigma_i^2 = g_n^{(i)} \left(G_i' G_i \right)^{-1} g_n^{(i)'} \sigma^2$$

Função densidade de probabilidade para cada modelo:

$$p_i = \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma_i^2)}} \exp \left\{ -\frac{1}{2(\sigma^2 + \sigma_i^2)} (y_n - \tilde{y}^{(i)})^2 \right\}$$

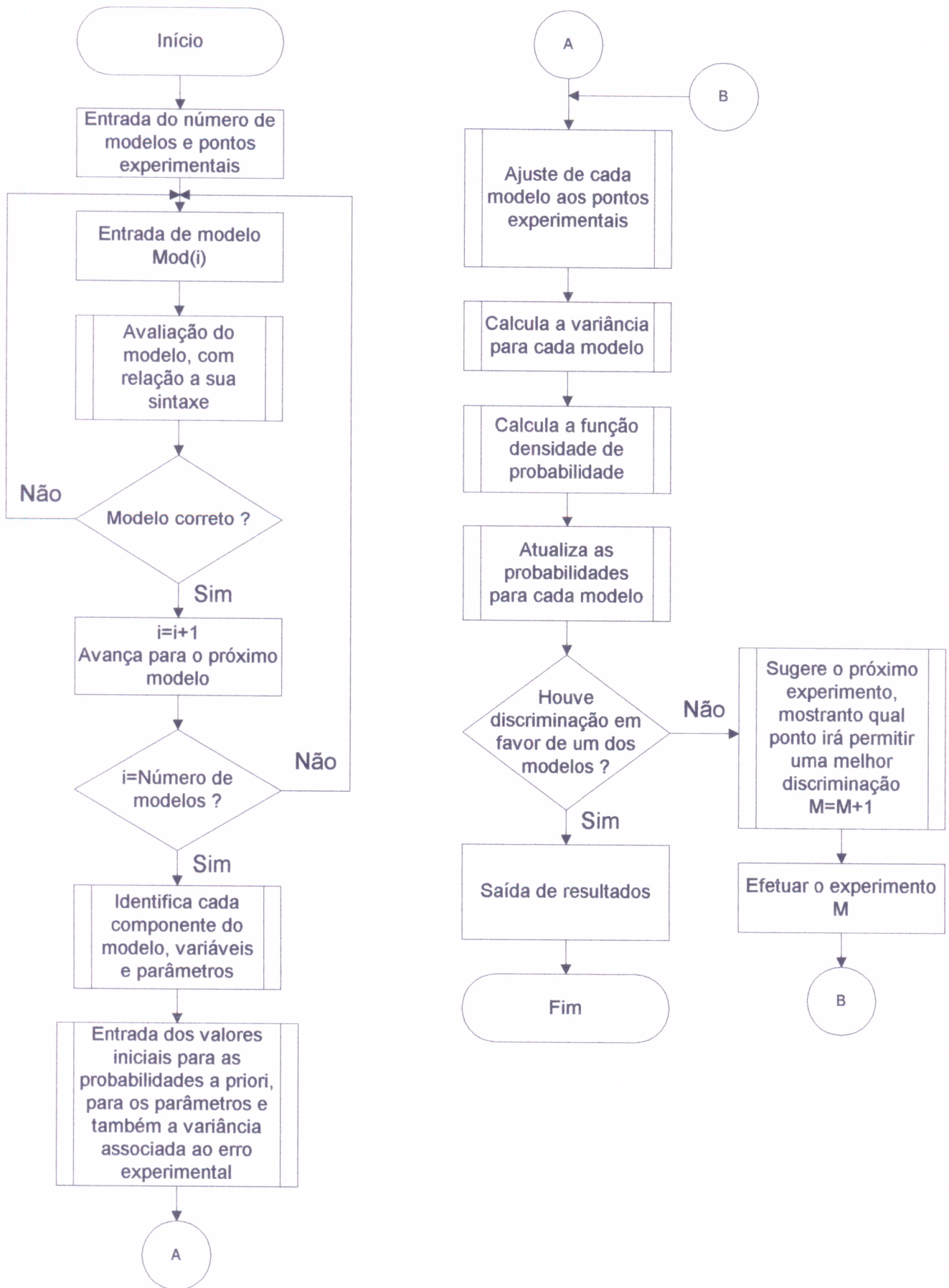
Atualização da probabilidade de cada modelo (Bayes):

$$\Pi_{in} = \frac{\Pi_{in-1} p_i}{\sum_{i=1}^m \Pi_{in-1} p_i}$$

Função discriminação:

$$D = \frac{1}{2} \sum_{i=1}^m \sum_{j=i+1}^m \Pi_{in-1} \Pi_{jn-1} \left\{ \frac{(\sigma_i^2 - \sigma_j^2)^2}{(\sigma^2 + \sigma_i^2)(\sigma^2 + \sigma_j^2)} + (\tilde{y}_n^{(i)} - \tilde{y}_n^{(j)})^2 \left(\frac{1}{\sigma^2 + \sigma_i^2} + \frac{1}{\sigma^2 + \sigma_j^2} \right) \right\}$$

Diagrama de fluxo da implementação do algoritmo



Caso teste

Para uma reação do tipo $A \rightarrow B$, sendo y a concentração de A; x_i o tempo de reação; x_{ii} a temperatura; T_{IMI} , T_{IIMI} , T_{IMII} , T_{IIMII} , T_{IMIII} , T_{IIMIII} , T_{IMIV} , T_{IIMIV} são os parâmetros dos modelos, considerando-se 4 modelos:

Modelo 1

$$y = \text{Exp}[-T_{IMI} * x_i * \text{Exp}[-T_{IIMI} / x_{ii}]]$$

Modelo 2

$$y = 1 / (1 + T_{IMII} * x_i * \text{Exp}[-T_{IIMII} / x_{ii}])$$

Modelo 3

$$y = 1 / (1 + 2 * T_{IMIII} * x_i * \text{Exp}[-T_{IIMIII} / x_{ii}])^{\frac{1}{2}}$$

Modelo 4

$$y = 1 / (1 + 3 * T_{IMIV} * x_i * \text{Exp}[-T_{IIMIV} / x_{ii}])^{\frac{1}{2}}$$

Resultados do analisador sintático

4	Número de modelos
4	Número de pontos experimentais
2	Número máximo de variáveis
2	Número máximo de parâmetros
0.05	Variância associada ao erro experimental
$\text{Exp}[-\text{TIMI} \cdot \text{xi} \cdot \text{Exp}[-\text{TII MI} / \text{xii}]]$	Modelo 1
0.250000	Probabilidade a priori para o modelo 1
2	Número de Parâmetros no modelo 1
TIMI	Parâmetro
TII MI	Parâmetro
2	Número de variáveis no modelo 1
xi	Variável
xii	Variável
$1/(1+\text{TIMII} \cdot \text{xi} \cdot \text{Exp}[-\text{TII MII} / \text{xii}])$	Modelo 2
0.250000	Probabilidade a priori para o modelo 2
2	Número de parâmetros no modelo 2
TIMII	Parâmetro
TII MII	Parâmetro
2	Número de variáveis no modelo 2
xi	Variável
xii	Variável
$1/(1+2 \cdot \text{TII MIII} \cdot \text{xi} \cdot \text{Exp}[-\text{TII MIII} / \text{xii}])^{0.5}$	Modelo 3
0.250000	Probabilidade a priori para o modelo 3
2	Número de parâmetros
TII MIII	Parâmetro
TII MIII	Parâmetro
2	Número de variáveis
xi	Variável
xii	Variável
$1/(1+3 \cdot \text{TII MIV} \cdot \text{xi} \cdot \text{Exp}[-\text{TII MIV} / \text{xii}])^{0.5}$	Modelo 4
0.250000	Probabilidade a priori para o modelo 4
2	Número de parâmetros
TII MIV	Parâmetro
TII MIV	Parâmetro
2	Número de variáveis
xi	Variável
xii	Variável

Resultados obtidos

Discriminação entre quatro modelos

Tabela 1 - $\sigma^2=0,05$ $\Pi_1 = \Pi_2 = \Pi_3 = \Pi_4 = 0,25$

n	xi	xii	Y	Π_1	Π_2	Π_3	Π_4
1	25	575	0,3961				
2	25	475	0,7232				
3	125	475	0,4215				
4	125	575	0,1297	0,0455	0,8804	0,0576	0,0164
5	138	558	0,1399	0,0391	0,9609	0,0000	0,0000
6	120	546	0,1812	0,0272	0,9728	0,0000	0,0000

Tabela 2 - $\sigma^2=0,05$ $\Pi_1=0,2$ $\Pi_2=0,4$ $\Pi_3=0,2$ $\Pi_4=0,2$

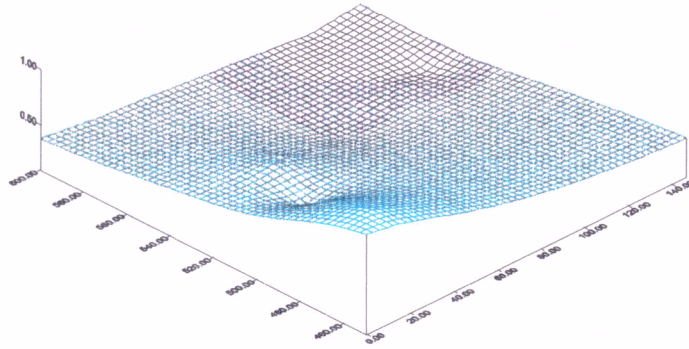
n	xi	xii	Y	Π_1	Π_2	Π_3	Π_4
1	25	575	0,3961				
2	25	475	0,7232				
3	125	475	0,4215				
4	125	575	0,1297	0,0242	0,9364	0,0306	0,0087
5	138	558	0,1399	0,0199	0,9801	0,0000	0,0000
6	120	546	0,1812	0,0138	0,9862	0,0000	0,0000

Tabela 3 - $\sigma^2=0,05$ $\Pi_1=0,4$ $\Pi_2=0,1$ $\Pi_3=0,2$ $\Pi_4=0,3$

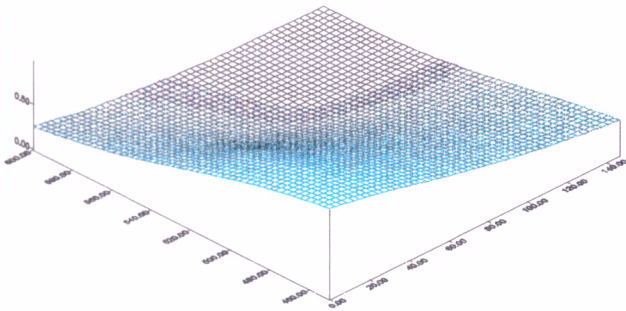
n	xi	xii	Y	Π_1	Π_2	Π_3	Π_4
1	25	575	0,3961				
2	25	475	0,7232				
3	125	475	0,4215				
4	125	575	0,1297	0,1485	0,7175	0,0938	0,0402
5	138	558	0,1399	0,1398	0,8601	0,0000	0,0000
6	120	546	0,1812	0,1007	0,8993	0,0000	0,0000

Gráficos dos modelos

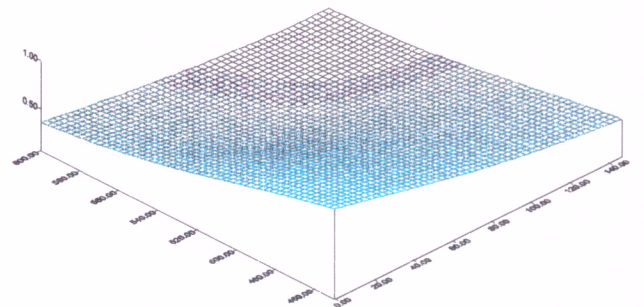
Dados



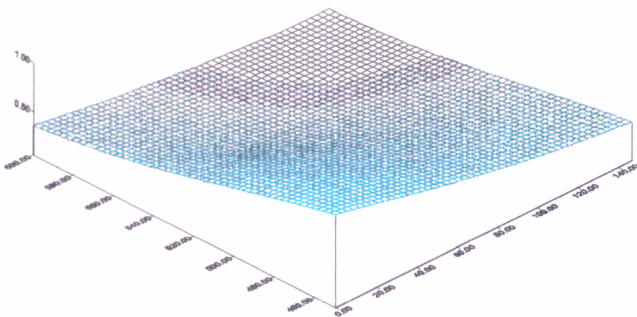
Modelo 1



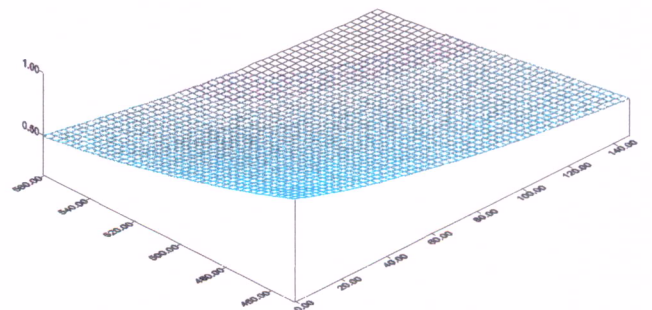
Modelo 2



Modelo 3



Modelo 4



Conclusão

O algoritmo de Box e Hill mostra-se como uma poderosa ferramenta para a discriminação entre modelos competitivos. Na forma ora implementada, demonstrou grande robustez ao identificar o melhor modelo mesmo atribuindo-se menor probabilidade para este. E identificando-o com mais rapidez quando favorecido com uma probabilidade maior que a dos demais.

O objetivo da implementação desta ferramenta foi alcançado, destacando-se que a facilidade de operação deve à implementação de um analisador sintático e de ferramentas que a tornam mais fácil de ser utilizada, com conseqüente aumento de sua utilização. Não obstante pretende-se torná-la mais amigável ainda.

Referências bibliográficas

- Abell, M. L.; Braselton J. P. **The Mathematica Handbook**. London, Academic Press, 1992. 789 p.
- Abell, M. L.; Braselton J. P. **Mathematica by Example**. London, Academic Press, 1994. 523p.
- Adeodato de Souza Neto, J. **Discrimination among mechanistic models for oxigen-hemoglobin equilibrium**. Doctoral Dissertation, Gainesville, University Of Florida, 1970.
- Box, G.E.; Hill, W.J. Discrimination among mechanistic models. **Technometrics**, 9(1):57-71, Feb. 1967.
- Feigelson, E.F.; Babu G.J. **Statistical Challenges in Modern Astronomy**. New York, Springer, 1992.
- Nash, J.C. **Compact Numerical Methods for Computers**. New York, Adam Hilger, 1990. 262p.
- Nordemann, D.J.R. **Introdução ao Mathematica for Windows**. São José dos Campos, Transtec, 1994. 171 p.
- Schildt, H. **C Completo e Total**. São Paulo, Makron Books, 1991. 890 p.
- Silva Filho, J. F. da. **Discriminação entre modelos competitivos**. Dissertação de mestrado em Análise de Sistemas e Aplicações - Instituto de Pesquisas Espaciais, São José dos Campos, junho de 1977. 109 p. (INPE-1027-TPT/051).
- Wadsworth, H.M.J. **Handbook of Statistical Methods for Engineers and Scientists**. New York, McGraw-Hill, 1990. 525p.