

**ALGORITMO DE BOX E HILL PARA DISCRIMINAÇÃO ENTRE
MODELOS COMPETITIVOS**

RELATÓRIO FINAL

Roberto Francisco Marques Mendes
Agosto de 1996 a fevereiro de 1997

Atila Madureira Bueno
Abril de 1997 a julho de 1997

Bolsistas PIBIC/INPE

Orientador: Ralf Gielow

Algoritmo de Box e Hill para discriminação entre modelos competitivos

Atila Madureira Bueno
Universidade Braz Cubas
Bolsista PIBIC/INPE

Ralf Gielow
Pesquisador Titular - DCM/INPE
Orientador

1. Introdução e material bibliográfico

O algoritmo de Box e Hill (1967), implementado de forma amigável por Mendes (1997), discrimina modelos que competem para representar um determinado fenômeno natural ou processo artificial. Em cada modelo, a variável dependente é função de variáveis independentes e de parâmetros; para a discriminação, deve-se ter dados observados ou medidas experimentais da variável dependente cobrindo todo o domínio das variáveis envolvidas.

A descrição dos procedimentos metodológicos e a aplicação do algoritmo de Box e Hill (BH) a seis casos de competição entre modelos micrometeorológicos e hidrológicos encontrados na literatura, é apresentada por Mendes (1997) - cópia no Anexo A, que descreve as atividades do Bolsista R. F. M. Mendes até o final de sua bolsa em fevereiro de 1997. Em adição, a utilização do algoritmo BH, com exemplificação através de um dos casos mencionados, foi submetida em março de 1997 por R. Gielow, R. F. M. Mendes e G. A. Degrazia, este último da Universidade Federal de Santa Maria (Gielow et al., 1997 - ver Anexo B), tendo sido aceita para apresentação no X Congresso Brasileiro de Agrometeorologia a realizar-se em julho de 1997. Anteriormente, Mendes (1996) reproduziu o caso teste de cinética química, com quatro modelos com parâmetros a ajustar, abordado originalmente por Box e Hill (1967).

Neste trabalho, que cobre as atividades do bolsista que substituiu R. F. M. Mendes a partir de 1º de abril de 1997, após (i) familiarização com as realizações anteriores relatadas por Mendes (1997) e respectiva bibliografia, implementou-se (ii) um conjunto de informações de ajuda "clicáveis" na interface gráfica GUI, para tornar o pacote BH mais amigável ainda, seguido (iii) da aplicação do BH a um caso de discriminação entre modelos aeronômicos encontrado na literatura, isto substituindo, em razão do pouco tempo de atividade do novo bolsista, a aplicação micrometeorológica originalmente planejada.

2. Procedimentos metodológicos

2.1 Informações de ajuda para a utilização do algoritmo de Box e Hill

Para implementação das informações de ajuda "clicáveis" na interface gráfica GUI ("Graphical User Interface"), geraram-se textos em linguagem HTML ("Hyper Text Markup Language"), que permite a criação de hipertextos, ou seja, textos em que o usuário pode "navegar" através das informações, sem a necessidade de menus, o que facilita sobremaneira a interação Homem-Máquina. Para a implementação dos hipertextos foi utilizado o "toolbox uitools" do MATLAB, que executa a maioria das funções das linguagens HTML. O conjunto de informações de ajuda implementado é mostrado no Anexo D.

2.2 Discriminação entre modelos para representar o conteúdo eletrônico da ionosfera

A determinação do conteúdo eletrônico da ionosfera (CEI) é de importância tanto para o estudo do comportamento da atmosfera superior, quanto em aplicações em Geodésia Espacial e na navegação de satélites e espaçonaves.

Para representar o CEI existem dois modelos: **IRI90** - International Reference Ionosphere (Ionosfera de Referência Internacional) e **SUPIM** - Sheffield University Plasmasphere-Ionosphere (Plasmasfera e Ionosfera da Universidade de Sheffield), sendo o primeiro um modelo empírico global, e o segundo um modelo mecanístico (Paula et al., 1996).

Assim, para representar observações de CEI realizadas em Cachoeira Paulista, SP (22°S, 45°W, 14°S latitude dipolar) - região de baixa latitude, confrontam-se através do BH esses dois modelos, utilizando observações realizadas na época equinocial dos anos de 1982, 1984, 1986 e 1989 - com atividade solar de níveis baixo, médio e alto, e resultados obtidos com aplicações de cada um dos modelos, mostradas na Figura 2 do trabalho de Paula et al. (1996). Souza (1997) proveu as matrizes com as médias das observações de CEI, determinadas a cada 5 minutos nas 24 horas do dia, com desvio-padrão igual a 0,5, assim como os resultados gerados pelo modelo IRI90 para as horas cheias, e aqueles gerados pelo modelo SUPIM para cada 30 minutos, estratificados para baixo, médio e alto níveis de atividade solar. Aplicou-se BH (i) para uma base horária e, após interpolação linear dos resultados do IRI90, (ii) para cada 30 minutos de 0 a 24 horas local (HL), seguindo-se (iii) aplicação à estratificação dos dados em três períodos correspondentes às fases da evolução diária do CEI.

3 Resultados e conclusões

A aplicação do algoritmo de Box e Hill à discriminação entre os modelos IRI90 e SUPIM para representar o conteúdo eletrônico da ionosfera sobre Cachoeira Paulista, considerando o período integral de 24 horas, teve os resultados mostrados na Tabela 1, abaixo, e Figuras C.7, C.8 e C.9, do Anexo C.

Tabela 1 - Conteúdo eletrônico da ionosfera
Cachoeira Paulista - 22,5°S, 45°W
Equinócio - 0:00 a 24:00 HL

ATIVIDADE SOLAR	MODELO			
	IRI90		SUPIM	
	Π	DMQ	Π	DMQ
Mínima	0,35 - 0,37	2,08 - 2,08	0,65 - 0,62	1,96 - 1,96
Média	0,21 - 0,21	2,81 - 2,80	0,79 - 0,79	1,63 - 1,63
Máxima	0,08 - 0,14	5,39 - 5,42	0,92 - 0,86	1,91 - 1,94

Π - Probabilidade do modelo DMQ - Desvio médio quadrático
O primeiro valor de cada coluna refere-se a dados horários, e o segundo, a cada 30 minutos.

Portanto, em todos os níveis de atividade solar (mínima, média e máxima) o modelo SUPIM é claramente superior ao IRI90 pelo critério de BH. Estes resultados são coerentes com o valor do desvio médio quadrático entre os resultados dos modelos e os dados observacionais; no entanto, a diferença das probabilidades calculadas com as frequências horária e semi-horária é menor que 7% em todos os casos, exceto um (IRI90/ atividade solar máxima, com baixa probabilidade).

Outrossim, como a evolução temporal diária do CEI tem três fases (Figura 1), como sejam: (i) do nascer do sol até o seu ocaso, com um máximo em torno de 15:00 HL; (ii) do mínimo no ocaso até um máximo relativo aproximadamente às 22:00 HL (o intervalo ocaso - reversão noturna) e (iii) reversão noturna até o nascer do sol - hora de menor CEI, pode-se estratificar a análise conforme estas fases. Deste modo, utilizando dados semi-horários, têm-se os resultados mostrados na Tabela 2.

Tabela 2 - Conteúdo eletrônico da ionosfera
 Cachoeira Paulista - 22,5°S, 45°W
 Equinócio - Fases diárias

ATIVIDADE SOLAR	MODELO			
	IRI90		SUPIM	
	II	DMQ	II	DMQ
<i>Nascimento-Ocaso</i>				
Mínima(6 às 22:3)	0,39	2,21	0,61	2,32
Média	0,14	2,78	0,86	1,93
Máxima	0,50	5,86	0,50	2,085
<i>Ocaso-Reversão Noturna</i>				
Mínima(22:3 às 24)	0,40	1,64	0,60	0,44
Média	0,22	3,89	0,78	1,32
Máxima	0,13	7,30	0,87	2,27
<i>Reversão noturna</i>				
Mínima	0,42	1,74	0,58	0,64
Média	0,46	2,32	0,54	0,91
Máxima	0,24	2,73	0,76	1,30

II - Probabilidade do modelo

DMQ - Desvio médio quadrático

Deste modo, estratificando a análise conforme as fases temporais diárias de comportamento do CEI, confirma-se a superioridade do Modelo SUPIM, porém com situações de maior aproximação com o modelo IRI90. Evidentemente, cabe ao especialista em Aeronomia interpretar o significado final dos resultados.

4. Comentários finais

Neste segundo ano do projeto (i) aperfeiçoou-se a implementação do algoritmo de Box e Hill, tornando-a mais amigável através de uma interface GUI, a qual inclui informações de ajuda 'cliváveis', (ii) incorporou-se a possibilidade de entrada de resultados finais concernentes aos modelos em competição, (iii) testaram-se sete casos de discriminação entre modelos, abrangendo situações micro-meteorológicas, hidrológicas e aeronômica, com dados extraídos da literatura e (iv) houve o aceite da apresentação da técnica, com exemplo de aplicação, no X Congresso Brasileiro de Agrometeorologia, a realizar-se em julho de 1997. o resumo estendido e o material de apresentação para o Seminário de Iniciação Científica, a realizar-se em julho de

1997, abordando os sete casos de discriminação estudados, estão mostrados no Anexo C, ilustrando a versatilidade do BH como ferramenta auxiliar para os especialistas.

Agradecimentos: Aos que proveram os dados observacionais e os modelos testados; ao CNPq, pelas Bolsas de Iniciação Científica PIBIC.

Referências bibliográficas

Box, G. E.; Hill, W. J. Discrimination among mechanistic models. **Technometrics**, 9(1):57-71, Feb. 1967.

Gielow, R.; Mendes, R. F. M.; Degrazia, G. A Utilização do algoritmo de Box e Hill para discriminação entre modelos competitivos. Aceito para apresentação no X Congresso Brasileiro de Agrometeorologia a realizar-se em Piracicaba, SP em julho de 1997 e publicação nos respectivos **Anais**.

Mendes, R. F. M. **Algoritmo de Box e Hill para discriminação entre modelos competitivos**. Relatório de Bolsista PIBIC/INPE referente ao período agosto de 1995 a julho de 1996. São José dos Campos, INPE, jul. 1996.

_____ Relatório de Bolsista PIBIC/INPE referente ao período agosto de 1996 a fevereiro de 1997. São José dos Campos, INPE, fev. 1997.

Souza, J. R. de **Dados e simulações de conteúdo eletrônico da ionosfera em Cachoeira Paulista**. Comunicação pessoal, jun. 1997.

Paula, E. R. de; Souza, J. R. de; Abdu, N. A. ; Bailey, G.J.; Batista, I. S.; Bittencourt, J. A.; Bonelli, E. Ionospheric electron content over Brazilian low latitude and its comparison with the IRI and SUPIM models. **Advanced Space Research**, 18 (6):245-248, 1996.

Anexo A

Relatório Parcial

Agosto de 1996 a fevereiro de 1997

**ALGORITMO DE BOX E HILL PARA DISCRIMINAÇÃO ENTRE MODELOS
COMPETITIVOS**

RELATÓRIO PARCIAL

PERÍODO: Agosto de 1996 a fevereiro de 1997

BOLSISTA (PIBIC/INPE): Roberto Francisco Marques Mendes

ORIENTADOR: Ralf Gielow

Algoritmo de Box e Hill para discriminação entre modelos competitivos

Roberto Francisco Marques Mendes
Universidade Braz Cubas
Bolsista PIBIC/INPE

Ralf Gielow
Pesquisador Titular/INPE
Orientador

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, que cobre o primeiro semestre do segundo ano do projeto, implementou-se uma interface do tipo GUI ("Graphical User Interface" - Mathworks, 1994), tornando mais amigável o uso do programa de computador utilitário que foi implementado durante o primeiro ano para resolver problemas de discriminação do tipo acima descrito. Apresenta-se, também, sua aplicação em situações efetivas de discriminação entre modelos.

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 - 4 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. Assim, com o propósito de tornar mais amigável o uso do algoritmo, no primeiro ano deste projeto ele foi implantado automatizando operações como a obtenção analítica de derivadas e a regressão não-linear para ajuste de parâmetros, devendo o usuário apenas digitar as expressões para os modelos e alguns dados de entrada, enquanto o processamento em MATLAB e C é feito em "background" (Mendes, 1996). Para tornar essa implementação mais amigável ainda, neste primeiro semestre do segundo ano de projeto, (i) implantou-se interface tipo GUI, (ii) estendeu-se o programa para admitir situações em que já se tenha os resultados previstos por modelos competitivos e, (iii) apresentam-se aplicações em situações efetivas, com dados obtidos na literatura.

3. Procedimentos metodológicos

A implementação do algoritmo de Box e Hill foi efetivada utilizando o software MATLAB. O objetivo desta implementação é gerar um programa dotado de todo o formalismo proposto por Box e Hill e de um ambiente gráfico tipo GUI ("Graphical User Interface"). Tal ambiente proporciona ao usuário facilidades quando da utilização do programa.

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 às regras de entrada, verificadas automaticamente por um analisador sintático. Cada modelo é 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 MATLAB (sen, cos, exp, ...). Os parâmetros são representados por qualquer sequência de caracteres, alfabéticos e maiúsculos (T, B, A), enquanto as variáveis o são por letras minúsculas (x, v, t, z). Outra alternativa é permitir a entrada dos resultados já calculados para cada modelo.

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.

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.

3.2.1 O software MATLAB

O MATLAB (Matrix Laboratory) se constitui em um pacote destinado à computação numérica e programação algébrica. Inicialmente era um software que tinha como objetivo criar um ambiente "fácil" para a manipulação matricial, possuindo características revolucionárias, como a não-definição prévia de índices matriciais e a facilidade de se executarem operações matriciais diretamente. À medida que novos recursos foram incorporados, o MATLAB se transformou em uma linguagem computacional, contando com todos os recursos disponíveis em qualquer linguagem de alto nível (laços, decisões, interfaces gráficas, recursividade, etc); contudo, possui uma diferença fundamental em relação a outras linguagens: a possibilidade de ser programado algebricamente.

Com o MATLAB torna-se possível desenvolver aplicações de alto nível, envolvendo desde rotinas simples até interfaces GUI ("Graphical User Interface").

Sendo uma linguagem interpretada, o MATLAB é ideal para o trabalho experimental, pois os programas podem ser alterados e testados rapidamente. Outro fator a se destacar com relação a este software é a sua grande capacidade de evolução. Uma vez que a estrutura básica foi implementada, as adaptações para as mais variadas necessidades geram rotinas específicas. Rotinas utilizadas para processamento de sinais, estatística, redes neurais, matemática simbólica, simulação, processamento de imagens e outros. Essas rotinas

formam pacotes denominados "ToolBox" (Caixa de Ferramentas), permitindo dessa forma que o software evolua de acordo com a necessidade e a aplicação.

3.2.2 Elaboração do ambiente gráfico "GUI"

O MATLAB provê alguns recursos que facilitam a elaboração de interfaces gráficas. Esses recursos combinados tendem a gerar por sua vez novas ferramentas que podem ser aplicadas ao desenvolvimento de aplicações mais elaboradas. Dentre esses recursos, têm-se: janelas, botões de pressão, marcadores, barras de rolagem, janelas de edição e janelas de exibição.

O desenvolvimento de uma interface gráfica, demanda um conhecimento prévio com relação ao perfil do usuário ao qual se destina a aplicação. A interface deve estar adaptada de maneira a facilitar o trabalho de quem a utiliza. Neste caso, tendo em vista que a implementação do algoritmo de Box e Hill será utilizada por experimentadores, esta foi elaborada de forma a permitir que o programa possa ser utilizado por pessoas com pouco ou nenhum conhecimento de computação.

3.2.2.1 Facilidades disponíveis na interface

A interface implementada permite ao usuário inserir os dados na forma de arquivos, dispensando a necessidade destes serem digitados. Com relação aos modelos a serem testados, estes podem ser introduzidos no programa de duas maneiras distintas: (i) na forma simbólica, caso em que o analisador de expressões irá realizar a identificação de cada componente da expressão; (ii) na forma numérica, onde são fornecidos ao sistema os resultados correspondentes a cada modelo e os valores observados.

4. Resultados

O algoritmo de Box e Hill está implementado e operacional com uma interface ser-humano-máquina do tipo GUI bastante amigável, utilizando MATLAB e linguagem C em "background". Os dados podem ser introduzidos de duas maneiras: (i) expressões simbólicas dos modelos ou (ii) resultados numéricos referentes aos modelos.

Como aplicação prática do algoritmo de Box e Hill, em adição à situação de cinética química tratada no primeiro ano, a qual usa todo o potencial do algoritmo, no presente

abordam-se diversos casos utilizando resultados de modelos em competição publicados na literatura.

5 Discussão dos resultados e conclusões

Aplicou-se o algoritmo de Box e Hill a seis casos de competição entre modelos micrometeorológicos e hidrológicos encontrados na literatura. Cópia das referências respectivas, já publicadas, encontram-se em anexo.

Caso #1

Concentração de poluentes na camada limite convectiva utilizando modelo Gaussiano vs soluções da equação de difusão.

A discriminação foi feita entre três modelos para a obtenção de concentrações superficiais integradas lateralmente, analisando os resultados observados e os simulados constantes da Tabela 2 do trabalho de Degrazia et al. (1996), conforme mostrado na Tabela 1.

Tabela 1 - Caso #1

Obs	Mod1	Mod2	Mod3
6,48	6,00	7,10	6,29
2,31	3,90	3,81	3,50
5,38	3,59	4,37	3,96
2,95	2,63	2,99	2,54
8,20	7,28	6,57	6,46
6,22	5,16	3,70	3,53
4,30	4,16	2,67	2,51
6,70	3,91	5,41	5,17
3,25	2,49	3,24	2,92
2,23	2,11	2,57	2,24
4,16	4,10	4,89	4,76
2,02	3,07	3,46	3,19
1,52	2,51	2,52	2,27
4,58	3,48	4,39	4,60
3,11	2,31	2,99	2,96
2,59	1,83	2,22	2,12

$\sigma = 0,1$

Probabilidade

Modelo 1 0,3309

Modelo 2 0,3374

Modelo 3 0,3317

O algoritmo de Box e Hill indica a equivalência entre os três modelos (empate técnico).

Caso #2

Concentração de poluentes emitidos por fonte linha na camada limite convectiva utilizando soluções analíticas e numéricas.

A discriminação foi feita entre os modelos analíticos e uma simulação numérica para a obtenção de concentrações na superfície, constantes na Tabela 1 do trabalho de Carvalho et al. (1996), comparando com os resultados de experimentos de laboratório, conforme mostrado na tabela 2.

Tabela 2 - Caso #2

Obs	Mod1	Mod2	Mod3
0,40	0,67	0,98	0,23
1,25	1,68	2,13	1,42
2,50	2,57	2,98	2,15
3,55	3,35	3,59	2,55
3,65	3,57	2,73	2,54
3,50	3,77	3,84	2,43
3,35	3,87	3,86	2,27
2,70	3,87	3,77	1,95
1,50	3,52	3,40	1,13

$\sigma = 0,1$

Probabilidade

Modelo 1 0,3025

Modelo 2 0,3031

Modelo 3 0,3944

Apesar da ligeira vantagem do modelo numérico 3, o resultado indica um empate técnico com os modelos analíticos 1 e 2.

Caso #3

Concentração de poluentes ao nível da superfície da camada limite convectiva utilizando soluções analíticas e numéricas.

De modo semelhante ao caso #2, a discriminação foi feita utilizando os dados constantes na Tabela 2 do trabalho de Carvalho et al. (1996), conforme mostrado na Tabela 3.

Tabela 3 - Caso #3

Obs	Mod1	Mod2	Mod3
0,90	2,54	3,71	0,66
4,90	4,49	5,70	4,02
5,59	5,29	6,23	5,30
5,59	5,40	6,27	5,37
4,90	5,61	6,12	5,00
4,00	5,45	5,70	4,23
3,20	5,13	5,22	3,47
2,20	4,77	4,75	2,84
1,60	4,07	3,96	1,95
0,90	2,81	2,72	1,18

$\sigma = 0,1$

Probabilidade

Modelo 1 0,2425
Modelo 2 0,2326
Modelo 3 0,5249

O modelo numérico 3 é o mais provável, com probabilidade em torno de 0,5, frente aos modelos analíticos 1 e 2.

Caso #4

Nível do rio Paraguai em quatro meses consecutivos utilizando redes neurais.

A discriminação foi feita entre dois modelos utilizando redes neurais e o modelo de regressão iterativa e comparando com dados observacionais, todos eles constantes do trabalho de Weigang e Nordemann(1996), conforme mostrado na Tabela 4, referente aos meses de março a junho de 1995.

Tabela 4 - Caso #4

Obs	Mod1	Mod2	Mod2
543	446	477	310
650	605	606	332
622	637	641	336
588	583	567	331

$\sigma = 1$

Probabilidade

Modelo 1 0,9916
Modelo 2 0,0084
Modelo 3 0,0000

O modelo 1 (Rede Neural com Atraso no Tempo - TDNN) é o mais provável frente aos modelos 2 (Rede de Retropropagação de Erro - BPNN) e 3 (Regressão Iterativa).

Caso #5

Nível do rio Paraguai em quatro meses consecutivos utilizando quatro modelos de redes neurais.

A discriminação foi feita entre quatro modos de treinamento de redes neurais apresentados na Tabela 3 do trabalho de Weingang et al. (1996).

Os dados observacionais e das quatro simulações foram fornecidos por Weingang (1997), levando aos resultados mostrados na Tabela 5, referentes aos meses de março a junho de 1995.

Tabela 5 - Caso #5

Obs	Mod1	Mod2	Mod3	Mod4
543	453	506	453	456
650	503	540	494	510
622	522	541	509	535
588	503	510	495	526

$\sigma = 1$

Probabilidade

Modelo 1 0,0000
Modelo 2 0,0000
Modelo 3 0,0000
Modelo 4 1,0000

O modelo 4, que utiliza 1142 pontos entre 1900 e 1995 é o mais provável, em contraposição aos demais, que utilizam menos pontos iniciando em anos bem mais recentes.

Caso #6

Nível do rio Paraguai em quatro meses consecutivos utilizando dois modelos de redes neurais.

A discriminação foi feita entre dois modos de treinamento de redes neurais de modo semelhante ao caso #5, mas considerando a segunda e terceira linha da Tabela 2 do trabalho de Weingang et al. (1996), levando aos resultados

constantes da Tabela 6, referentes aos meses de maio a agosto de 1959.

Tabela 6 - Caso #6

Obs Mod1 Mod2

585	614	632
549	592	604
491	530	544
436	453	473

$\sigma = 1$

Probabilidade

Modelo 1 0,9999

Modelo 2 0,0001

A discriminação favorece o modelo 1, que utiliza 344 pontos entre 1900 e 1930, contra o modelo 2, que utiliza 700 pontos entre 1900 e 1959.

6. Comentários finais

Neste primeiro semestre do segundo ano do projeto (i) aperfeiçou-se a implementação do algoritmo de Box e Hill, tornando-a mais amigável através de uma interface GUI, (ii) incorporou-se a possibilidade de entrada de resultados finais concernentes aos modelos em competição e, (iii) testaram-se seis casos de discriminação entre modelos, com dados extraídos na literatura. No segundo semestre, o objetivo é a aplicação do algoritmo a uma situação micrometeorológica utilizando dados observacionais existentes no INPE. Entretanto, como o bolsista foi aceito no curso de mestrado em Ciência da Computação da Universidade de São Paulo, graduando-se em março de 1997 na Universidade Bráz Cubas, sua bolsa se encerra em 28 de fevereiro de 1997, esperando-se que um substituto conclua o projeto.

Referências bibliográficas

Adeodato de Souza Neto, J. **Discrimination among mechanistic models for oxygen-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.

Carvalho, J. C.; Velho, H. F. de C.; Degrazia, G. A. Um estudo numérico da dispersão de poluentes na camada limite

- convectiva. **Anais**, IX Congresso Brasileiro de Meteorologia, Campos do Jordão, SP, nov. 1996. p.4-9.
- Degrazia, G. A.; Moreira, D. M.; Friedrich H. R. Uma comparação entre um modelo Gaussiano e a equação de difusão no cálculo da concentração de poluentes na camada limite convectiva. **Anais**, IX Congresso Brasileiro de Meteorologia, Campos do Jordão, SP, nov. 1996. p.86-89.
- Feigelson, E.F.; Babu G.J. **Statistical challenges in modern astronomy**. New York, Springer, 1992.
- Mathworks. **MATLAB - High-Performance Numeric Computation and Visualization Software**. External Interface Guide. Natick, MA, 1994.
- Mendes, R. F. M. **Algoritmo de Box e Hill para discriminação entre modelos competitivos**. Relatório de Bolsista PIBIC/INPE referente ao período ago. 1995 a jul. 1996. São José dos Campos, INPE, jul. 1996.
- 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.
- Weigang, L. Comunicação pessoal, fev. 1997.
- Weigang, L.; Nordemann, D. J. R. Study and prediction of the Paraguay river level by harmonic analysis and neural networks. **Revista Brasileira de Geofísica**, 14(2):195-202,1996.
- Weigang, L.; Sá, L. D. de A.; Galvão, G. P.; Bevilaqua, R. M. Prediction of the Paraguay river level using neural networks. Aceito para publicação no **Pesquisa Agropecuária Brasileira**, 1996.

Anexo

IX Congresso Brasileiro de Meteorologia

Campos do Jordão, 6 a 13 de novembro de 1996

Os benefícios das modernas técnicas de previsão de tempo e clima para as atividades sócio-econômicas

Volume 1

Promoção

SBMET / Sociedade Brasileira de Meteorologia

Realização

INPE / Instituto Nacional de Pesquisas Espaciais

Um Estudo Numérico da Dispersão de Poluentes na Camada Limite Convectiva

Jonas C. Carvalho
Departamento de Ciências Atmosféricas, Instituto Astronômico e Geofísico, USP. Cep 05508-900, São Paulo, SP

E-mail: carvalho@model.iag.usp.br

Haroldo F. de Campos Velho

Laboratório Associado de Computação e Matemática Aplicada, INPE. Cep 12201-970, S.J. dos Campos, SP.

Gervásio A. Degrazia

Departamento de Física, UFSM. Cep 97119-900, Santa Maria, RS, Brasil

Abstract

The diffusion equation is solved numerically in order to describe the dispersion of pollutants in a convective boundary layer for two situations: the emission from surface instantaneous area source and the emission from elevated continuous line source. A eddy diffusivity expression obtained from statistical diffusion theory and expressions for turbulent velocity spectra is compared with other obtained from data of Large-Eddy Simulation.

1 Introdução

A aproximação clássica para estudar a dinâmica atmosférica na transferência de poluentes é examinar a equação de difusão. Os termos nesta equação são frequentemente fechados pelo emprego da hipótese fluxo-gradiente, a qual relaciona os fluxos turbulentos com as quantidades médias por um coeficiente de difusão K . Desta forma, este esquema fecha o conjunto de equações somente em certo grau, pois o coeficiente de difusão ainda precisa ser determinado. Neste trabalho a equação de difusão é resolvida numericamente para investigar a dispersão de poluentes passivos e conservativos na camada limite convectiva (CLC), em duas situações: a emissão a partir de uma fonte área instantânea na superfície e a emissão a partir de uma fonte linha contínua elevada. Compara-se duas formulações conhecidas para o coeficiente de difusão. Uma obtida pela combinação da teoria de difusão estatística e de expressões para o espectro de energia turbulenta e a outra obtida a partir de dados de *Large-Eddy Simulation*. Estas parametrizações são não-locais e apresentam um caráter universal expresso em termos da escala de velocidade convectiva w_c e da altura da CLC z_c . Os resultados das simulações são comparados entre si, de acordo com a parametrização utilizada. Os valores de concentração resultantes da emissão por uma fonte linha elevada são comparados com valores obtidos em experimentos de laboratório e outras simulações numéricas.

2 Modelos de Difusão

A difusão de poluentes na atmosfera é dada pela equação

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} + v \frac{\partial C}{\partial y} + w \frac{\partial C}{\partial z} = \frac{\partial}{\partial x} \left(K_x \frac{\partial C}{\partial x} \right) + \frac{\partial}{\partial y} \left(K_y \frac{\partial C}{\partial y} \right) + \frac{\partial}{\partial z} \left(K_z \frac{\partial C}{\partial z} \right) + S, \quad (1)$$

onde C é a concentração em qualquer localização (x, y, z) e tempo t , u e v são as componentes horizontais do vento, w é a componente vertical do vento, K_x and K_y são os coeficientes de difusão horizontal, K_z é o coeficiente de difusão vertical e S é o termo de fonte.

2.1 Fonte Área

O modelo utilizado para simular a dispersão vertical de poluentes abandonados por uma fonte área é escrito pela simplificação da equação (1), na forma adimensional, como

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z} \left(K \frac{\partial C}{\partial z} \right) \quad 0 < z < 1, \quad (2)$$

com as seguintes condições iniciais e limites

$$C(z, t) = Q \delta(z - z_s) \quad t = 0 \quad (3)$$

$$K \frac{\partial C}{\partial z} = 0 \quad z = 0, z = z_c \quad (4)$$

onde z_s é a altura da fonte área, $\delta(z - z_s)$ é a função delta de Dirac e Q é a intensidade de emissão.

As variáveis são adimensionalizadas da seguinte forma

$$C = C' z_c / Q \quad t = t' w_c / z_c \quad z = z' / z_c \quad K = K' / (w_c z_c) \quad z_c = z'_c / z_c,$$

onde as linhas representam variáveis dimensionais.

A equação (2) é integrada pelo método de Cranck-Nicholson e as magnitudes seleccionadas para w_c e z_c são 2.0 m/s e 2000 m, respectivamente.

2.2 Fonte Linha

O modelo utilizado para simular a dispersão de poluentes emitidos por uma fonte linha normal a direção do vento é obtido da simplificação da equação (1), na forma adimensional, como

$$u \frac{\partial C_x}{\partial x} = \frac{\partial}{\partial z} \left(K \frac{\partial C_x}{\partial z} \right) \quad 0 < z < 1, \quad (5)$$

com as seguintes condições iniciais e limites

$$C_1(z, x) = Q \delta(z - z_1) \quad x = 0 \quad (6)$$

$$K \frac{\partial C_1}{\partial z} = 0 \quad z = 0, z = z_1, \quad (7)$$

onde C_1 é a concentração integrada normal a direção do vento.

As variáveis tornam-se adimensionais como segue

$$C_1 = (C, z, u)/Q \quad u = u'/u \quad z = z'/z_1 \quad K = K'/(w, z_1) \quad z_1 = z_1'/z_1 \quad x = (x' w_1)/(z_1 u)$$

onde as linhas representam variáveis dimensionais.

A equação (5) é integrada pelo método de Crank-Nicholson e as magnitudes seleccionadas para u , w_1 e z_1 são 4.0 m/s, 2.0 m/s e 2000m, respectivamente.

2.2.1 Cálculo da Concentração ao Nível da Superfície

O perfil lateral de uma pluma lembra uma distribuição Gaussiana. Conhecendo a concentração integrada normal a direção do vento ao nível da superfície, $C_1(x, z = 0)$, é possível calcular a concentração na superfície, $C(x, y, 0)$, em qualquer ponto, utilizando um modelo Gaussiano de concentrações laterais

$$C(x, y, 0) = \frac{C_1(x, 0)}{\sqrt{2\pi} \sigma_y} \exp\left[-\frac{y^2}{2\sigma_y^2}\right], \quad (8)$$

onde y é distância lateral adimensional ($y = y'/z_1$) e σ_y é o parâmetro de dispersão lateral adimensional

$$[\sigma_y = \sigma_y'/z_1 = 0.6 x/(1 + 1.5 x)].$$

3 Coeficientes de Difusão

Degrazia et al (1996) obtiveram uma expressão para o coeficiente de difusão, combinando a teoria de difusão estatística de Taylor com expressões para o espectro de energia turbulenta

$$\frac{K_{zz}}{w, z_1} = 0.22 \left(\frac{z}{z_1}\right)^{1/3} \left(1 - \frac{z}{z_1}\right)^{1/3} \left[1 - \exp\left(-4 \frac{z}{z_1}\right) - 0.0003 \exp\left(8 \frac{z}{z_1}\right)\right]. \quad (9)$$

Uma outra expressão para o coeficiente de difusão foi derivada por Holtslag e Moeng (1991) baseada em dados de *Large-Eddy Simulation*

$$K_c = \frac{(1 - z/z_1 + R_c z/z_1) K_b K_t}{(1 - z/z_1) K_t + R_c (z/z_1) K_b}, \quad (10)$$

onde R_c é a razão entre o fluxo de entranhamento e de superfície na CLC e K_b e K_t são os coeficientes de difusão *bottom-up* e *top-down*

$$\frac{K_b}{w, z_1} = \left(\frac{z}{z_1}\right)^{4/3} \left(1 - \frac{z}{z_1}\right)^2 \quad \frac{K_t}{w, z_1} = 7 \left(\frac{z}{z_1}\right)^2 \left(1 - \frac{z}{z_1}\right)^3.$$

Na figura 1, os perfis normalizados de K_c para diferentes R_c (0.5, 1.0, 1.5) são comparados com K_{zz} . A principal diferença entre as duas parametrizações ocorre em regiões elevadas da CLC. Nas proximidades da capa de inversão K_{zz} , ao contrário de K_c , mantém um nível turbulência elevado.

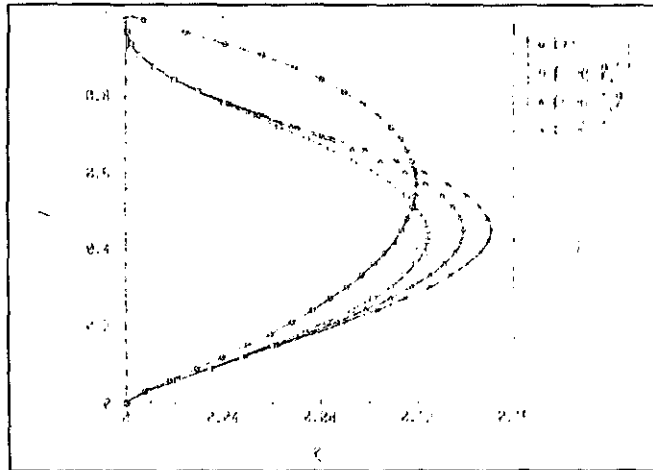


Figura 1 - Perfis verticais dos coeficientes de difusão (2)- K_{zz} e (2)- K_c .

4 Resultados e Conclusões

4.1 Fonte Área

As figuras 2 e 3 apresentam a distribuição vertical de concentração obtida através dos modelos (2)- K_{zz} e (2)- K_c , para uma fonte área localizada em $z_s = 0.025 z_t$. Com o passar do tempo o poluente é transportado rapidamente para as regiões médias e superiores da CLC. A elevada magnitude dos coeficientes de difusão na região central da camada é responsável por esta eficiente mistura. Desta maneira, a forte ação dispersiva da turbulência na CLC é o mecanismo que ocasiona a sensível redução na concentração superficial. Pode-se concluir que os coeficientes de difusão são tão grandes nos níveis médios da camada que as concentrações nestas regiões são efetivamente independentes da altura. Um resultado adicional é que o emprego do coeficiente de difusão K_c não reproduz um perfil vertical de concentração homogêneo próximo ao topo da camada. Tal situação ocorre, pois os valores de K_c aproximam-se de zero antes que o topo seja atingido (fig. 1).

4.2 Fonte Linha

Na tabela 1, apresentam-se os valores de concentração integrada na superfície, $C_s(x,0)$, para uma fonte linha localizada em $z_s = 0.25 z_t$. Estes resultados são comparados com valores obtidos de experimentos em laboratório por Willis e Hukari (1984) e de simulações numéricas de Misra (1982). Nota-se uma concordância razoável entre os resultados obtidos numericamente pela equação advecção-difusão (5) e os observados por Willis e Hukari (W-H). Os valores de concentração de W-H apresentam magnitudes elevadas na vizinhança da fonte, estendendo-se a distâncias da ordem de $x=1.0$. Neste intervalo de distância, os resultados fornecidos pelos modelos (5)- K_{zz} e (5)- K_c reproduzem magnitudes semelhantes. A magnitude máxima de C_s observada por W-H é de 3.65 em $x=0.60$, enquanto as magnitudes calculadas por (5)- K_{zz} e (5)- K_c são da ordem de 3.87 em $x=0.80$. Nota-se que estas magnitudes são maiores do que aquelas obtidas na simulações numéricas de Misra. Tanto para (5)- K_{zz} quanto para (5)- K_c , os valores de concentração não decaem em magnitude tão acentuadamente para maiores distâncias. Todavia, pode-se concluir que o emprego de coeficientes de dispersão K_{zz} e K_c conduz a magnitudes satisfatórias de concentrações superficiais na vizinhança da fonte.

4.2.1 Concentração ao Nível da Superfície

Na tabela 2, os valores de concentração ao nível da superfície, $C(x,0,0)$, são comparados com resultados obtidos de experimentos em laboratório por Willis e Deardorff (1978) e de simulações numéricas de Misra (1982). Os valores de concentração calculados a partir da expressão (8) com o emprego dos modelos (5)- K_{zz} e (5)- K_c apresentam uma magnitude máxima da ordem de 6.00 para x entre 0.4 e 0.5. Este resultado está em boa concordância com os dados experimentais de Willis e Deardorff e os valores numéricos fornecidos por Misra.

Um resultado de grande interesse prático na difusão atmosférica é a previsão das isolinhas de concentração ao nível da superfície, como mostrado nas figuras 4 e 5. Pode-se observar que existe uma acentuada concordância entre os

modelos. Pequenas diferenças são notadas na posição do máximo de concentração devido a assimetria dos coeficientes de difusão (fig. 1).

É necessário salientar que devido às inúmeras idealizações envolvidas, os resultados apresentados representam um limite superior para os valores de concentração. Estes valores estão relacionados a uma situação de forte convecção em um terreno plano, considerando-se a dispersão de um poluente passivo e conservativo, sem que haja deposição na superfície. Contudo, estes modelos podem ser utilizados na determinação da magnitude e posição dos valores máxima concentração superficial para emissões elevadas. Os coeficientes de difusão utilizados na parametrização da equação advecção-difusão permitem a obtenção destes resultados.

Agradecimentos

Os autores agradecem à CAPES, ao CNPq, e à FAPERGS pelo apoio prestado à realização deste trabalho.

Referencial Bibliográfico

Degrazia, G.A., Campos Velho, H.F. and Carvalho, An analytical method to evaluate nonlocal Lagrangian exchange coefficients in the convective boundary layer. Aceito para publicação na Contributions to Atmospheric Physics(1996).

Hollslag, A.A.M. and Moeng, C.H., Eddy diffusivity and countergradient transport in the convective atmospheric boundary layer, J. Atmos. Sci., 48: 1690-1698(1991).

Misra, P.K., Dispersion of non-buoyant particles inside a convective boundary layer, Atmosf. Environment, 16: 239-243(1982).

Willis, G.E. and Deardorff, J.W., A laboratory study of dispersion from an elevated source within a modeled convective planetary boundary layer, Atmosf. Environment, 12: 1305-1311(1978).

Willis, G.E. and Hukari, N., Laboratory modeling of buoyant stack emissions in the convective boundary layer. 4th Joint Conference on Applications of Air Pollution Meteorology. Amer. Meteor. Soc., Boston, 24-26(1984).

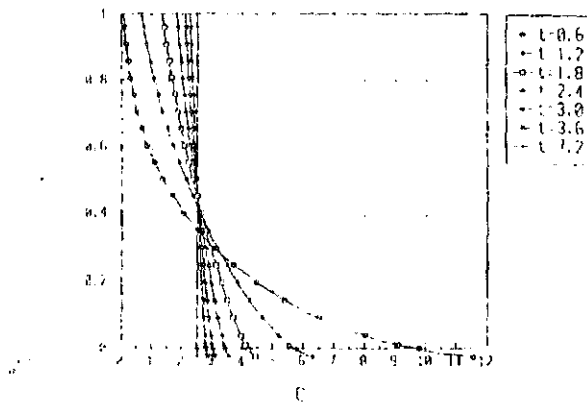


Figura 2 Perfil de concentração para diferentes tempos obtido por (2)- K_v .

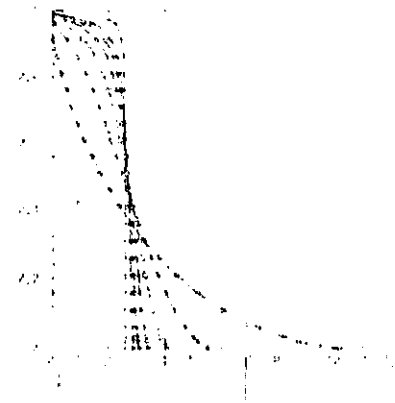


Figura 3 Perfil de concentração para diferentes tempos obtido por (2)- K_c com $R_c=0.5$.

Tabela 1

x	(5)- K_v	(5)- K_c	Willis-Hukari	Misra
0.20	0.67	0.98	0.40	0.23
0.30	1.68	2.13	1.25	1.42
0.40	2.57	2.98	2.50	2.15
0.54	3.35	3.59	3.55	2.55
0.60	3.57	3.73	3.65	2.54
0.70	3.77	3.84	3.50	2.43
0.80	3.87	3.86	3.35	2.27
1.00	3.87	3.77	2.70	1.95
1.50	3.52	3.40	1.50	1.13

Tabela 2

x	(8)-(5)- $K_{//}$	(8)-(5)- K_{\perp}	Willis-Deardorff	Misra
0.20	2.54	3.71	0.90	0.66
0.30	4.49	5.70	4.90	4.02
0.38	5.29	6.23	5.59	5.30
0.40	5.40	6.27	5.59	5.37
0.50	5.61	6.12	4.90	5.00
0.60	5.45	5.70	4.00	4.23
0.70	5.13	5.22	3.20	3.47
0.80	4.77	4.75	2.20	2.84
1.00	4.07	3.96	1.60	1.95
1.50	2.81	2.72	0.90	1.18

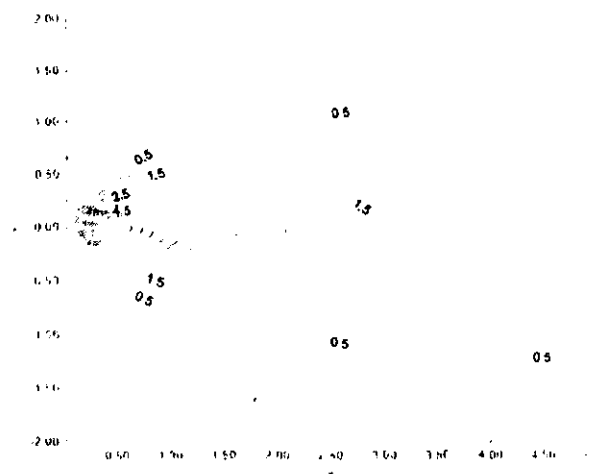
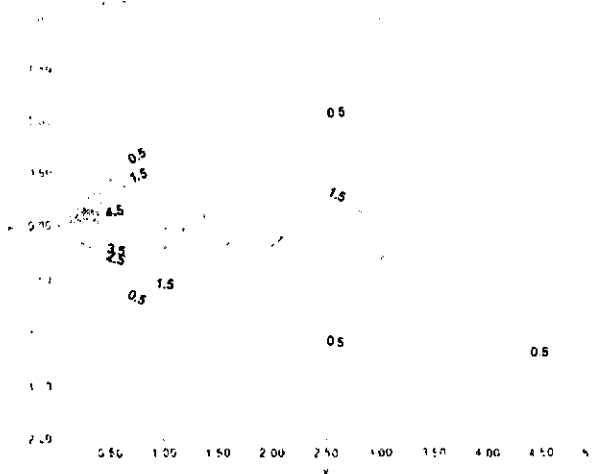


Figura 4 Isolinhas de concentração superficial no plano x-y obtidas pela expressão (8) com (5)- $K_{//}$

Figura 5 Isolinhas de concentração superficial no plano x-y obtidas pela expressão (8) com (5)- K_{\perp}

Uma Comparação entre um Modelo Gaussiano e a Equação de Difusão no Cálculo da Concentração de Poluentes na Camada Limite Convectiva

G.A. Degrazia

D.M. Moreira

Departamento de Física, Universidade Federal de Santa Maria

H.R. Friedrich

Departamento de Estatística, Universidade Federal de Santa Maria 97119-900, Santa Maria, RS

Abstract

In this work two different models are used in order to predict ground level concentrations due to elevated sources in a continuous convective boundary layer (CBL). The first model is the advection-diffusion equation with the application of the Laplace transformation and Gaussian Quadrature method and the second is an update Gaussian model. The advection-diffusion equation is closed using eddy diffusivities proposed by Degrazia et alii (1994) and Holtslag-Moeng (1991).

With the utilization of the observations of ground-level concentrations, collected in Copenhagen (Gryning et alii, 1987), the performance of the advection-diffusion equation model are evaluated against the update Gaussian model. Although the prediction of the Gaussian Model attained to a good concordance with observed ground-level concentrations, statistical indices indicates that better results are obtained with those proposed by advection-diffusion equation.

1 Introdução

Os problemas ambientais causados pelo rápido desenvolvimento industrial e tecnológico têm levado, nas últimas décadas, à investigação da dispersão e do transporte de poluentes na atmosfera. Esta investigação é uma atividade fundamental na proteção da qualidade do ar. A partir da estimativa do campo de concentração de contaminantes próximo à fonte é possível avaliar o impacto ambiental causado e agir no sentido de solucionar o problema da forma mais conveniente.

O estudo da dispersão de poluentes na Camada Limite Convectiva (CLC) é especialmente importante no caso de emissões a partir de fontes elevadas, pois estas geralmente contribuem para a máxima concentração ao nível do solo durante a convecção.

Neste trabalho, resolve-se a equação da difusão e advecção analiticamente e obtém-se a concentração superficial de contaminantes emitidos na CLC. Pelo emprego de observações de concentrações superficiais realizadas na cidade de Copenhagen (Gryning et alii, 1987), a performance do presente modelo é confrontada com a de um modelo Gaussiano atualizado.

2 Método de Solução da Equação de Difusão e Advecção

Integrando-se lateralmente a equação da difusão e advecção e desprezando-se a difusão longitudinal em comparação ao transporte médio obtém-se:

$$U \frac{\partial c^y}{\partial x} = \frac{\partial}{\partial z} \left(K_{zz} \frac{\partial c^y}{\partial z} \right) \quad (2.1)$$

onde c^y é a concentração média integrada lateralmente.

Considera-se uma CLC, de altura Z_i , como sendo uma superfície multicomposta de diversas regiões n . O problema de difusão vertical é resolvido considerando-se as fronteiras da superfície da terra e o topo da CLC. A altura Z_i foi dividida em regiões de 50 m cada, considerando-se um contato perfeito entre elas, com as seguintes condições de contorno:

$$U c_n^y(0, z) = Q \delta(z - z_i) \quad (2.2)$$

$$c_n^y = c_{n+1}^y \quad (2.3)$$

$$K_n \frac{\partial c_n^y}{\partial z} = 0 \quad \text{em } z = 0 \text{ e } z = Z_i \quad (2.4)$$

$$K_n \frac{\partial c_n^y}{\partial z} = K_{n+1} \frac{\partial c_{n+1}^y}{\partial z} \quad (2.5)$$

Aplicando-se a Transformada de Laplace na equação (2.1), na variável x , e as condições de interface, resulta um sistema linear para coeficientes constantes. Neste trabalho utiliza-se o cálculo desta integral de inversão por Quadratura Gaussiana, substituindo-se a variável complexa s pelos valores dos pontos da quadratura.

Assim a solução do problema, conseguida através da fórmula de Quadratura Gaussiana apropriada é:

$$c_n^y(x, z) = \sum_{k=1}^n A_k \cdot \frac{P_k}{x} \left[A_n e^{-\sqrt{\frac{P_k U}{x K_n}}} + B_n e^{\sqrt{\frac{P_k U}{x K_n}}} + \frac{1}{2} \frac{Q}{\sqrt{\frac{P_k K_n U}{x}}} \left(e^{-\sqrt{-\frac{P_k U}{x K_n}}} - e^{\sqrt{-\frac{P_k U}{x K_n}}} \right) \right] \quad (2.6)$$

Pode-se calcular agora a concentração $C_n^y(x, z)$. Os termos A_k e P_k encontram-se tabelados.

No fechamento dos fluxos turbulentos de concentração utilizam-se dois coeficientes de difusão turbulentos: um derivado por Degrazia et alli (1994)(eq. 2.7), que emprega o modelo de difusão estatístico de Taylor em conjunção com a teoria de similaridade convectiva e o outro construído a partir da técnica de Simulação dos Grandes Turbilhões (Holtslag e Moeng, 1991)(eq. 2.8), que considera o transporte contrário ao gradiente e a difusão *bottom-up/top-down*.

$$\frac{K_{zz}}{w \cdot Z_i} = 0.22 \left(\frac{z}{Z_i} \right)^{1/3} \left(1 - \frac{z}{Z_i} \right)^{1/3} \left[1 - \exp\left(-\frac{4z}{Z_i}\right) - 0.0003 \exp\left(\frac{8z}{Z_i}\right) \right] \quad (2.7)$$

$$K_c = \frac{(1 - z/Z_i + R_c \cdot z/Z_i) K_b K_t}{(1 - z/Z_i) K_t + R_c (z/Z_i) K_b} \quad (2.8)$$

3 Um Modelo Gaussiano Atualizado

Começaremos com a representação generalizada, que pode ser escrita na seguinte forma:

$$\sigma_u^2 = \frac{\sigma_i^2 \beta_i^2}{\pi^2} \int_0^\infty S_i(n) \frac{\sin^2(n\pi/\beta_i)}{n^2} dn \quad (3.1)$$

Os espectros de velocidade Eulerianos em condições instáveis podem ser expressos como funções das escalas convectivas da seguinte maneira:

$$\frac{n S_i(n)}{w_i^2} = \frac{0.98 c (f/q) (\psi/q)^{2/3} (z/z_i)^{2/3}}{(f_m)_i^{5/3} \left[1 + \frac{1.5 f}{(f_m)_i q} \right]^{5/3}} \quad (3.2)$$

Integrando-se analiticamente a equação (3.2) sobre o domínio inteiro de frequências resulta uma variância generalizada da forma:

$$\sigma_i^2 = \frac{0.98 c}{(f_m)_i^{2/3}} \left(\frac{\psi}{q} \right)^{2/3} \left(\frac{z}{z_i} \right)^{2/3} w_i^2 \quad (3.3)$$

que é usada para normalizar a função espectral (3.2). Assim, o valor do espectro Euleriano normalizado por σ_i^2 pode ser escrito como:

$$S_i(n) = \frac{1}{(f_m)_i} \frac{(z/U)q}{\left[1 + \frac{1.5 \cdot n z}{(f_m)_i U q} \right]^{5/3}} \quad (3.4)$$

A substituição de (3.3), (3.4) e de $\beta_i = (\pi U^2 / 16 \sigma_i^2)^{1/2}$ na expressão (3.1) fornece um parâmetro de dispersão generalizado para a CLC:

$$\frac{\sigma_u^2}{z_i^2} = \frac{1.5(z/z_i)^2}{16\pi(f_m)_i^2 q^2} \int_0^\infty \frac{\text{sen}^2 \left\{ \frac{4\pi(0.98c)^{1/2} [(f_m)_i^2 q^2 \psi]^{1/3} X n'}{\sqrt{\pi} 1.5(z/z_i)^{2/3}} \right\}}{(1+n')^{5/3} n'^2} dn' \quad (3.5)$$

A concentração integrada lateralmente pode ser calculada em termos do modelo Gaussiano pelo uso da relação:

$$c^y(x, z) = \frac{Q}{\sqrt{2\pi}\sigma_z U} \left\{ \exp\left[-\frac{(z-z_i)^2}{2\sigma_z^2}\right] + \exp\left[-\frac{(z+z_i)^2}{2\sigma_z^2}\right] \right\} \quad (3.6)$$

A relação (3.6) pode ainda ser escrita em uma forma adimensional pelo emprego dos parâmetros de escala convectivos, de modo que

$$C^y(x, z) = \frac{c^y U z_i}{Q} = \frac{1}{\sqrt{2\pi}\sigma'_z} \left\{ \exp\left[-\frac{(Z-Z_i)^2}{2(\sigma'_z)^2}\right] + \exp\left[-\frac{(Z+Z_i)^2}{2(\sigma'_z)^2}\right] \right\} \quad (3.7)$$

A grandeza fundamental no emprego de (3.7) é o parâmetro de dispersão vertical σ'_z que pode ser obtido da eq. (3.5).

Vamos considerar agora condições de convecção forte e abandonos elevados na CLC. Nesta região a escala de comprimento dos grandes turbilhões é proporcional a Z_i de modo que o comprimento de onda associado ao máximo do espectro de velocidade vertical pode ser escrito como $(\lambda_m)_w = Z_i$, implicando em:

$$q = \frac{(f_m)_w}{(f_m)} = \frac{z}{(\lambda_m)_w 0.35} = 2.86 \frac{z}{Z_i} \quad (3.8)$$

onde $(f_m)_w = 0.35$. Introduzindo-se (3.8), $(f_m)_w = 0.35$ e $c = 0.4$ na equação (3.5) obtém-se a seguinte fórmula para o parâmetro de dispersão vertical na CLC:

$$\sigma_z'^2 = \frac{0.094}{\pi} \int_0^\infty \frac{\text{sen}^2 [2.96 \psi^{1/3} X n']}{(1+n')^{5/3} n'^2} dn' \quad (3.9)$$

onde a função de dissipação adimensional é fornecida pela relação:

$$\psi = \left[\left(1 - \frac{z}{z_i}\right)^2 \left(-\frac{z_i}{L} \frac{z}{z_i}\right)^{-2/3} + 0.75 \right]^{3/2} \quad (3.10)$$

Os dados experimentais mostram que a razão de dissipação é elevada na camada limite superficial, assumindo um valor praticamente constante em regiões médias da CLC. O valor médio de $\psi^{1/3}$ calculado a partir da relação (3.10) é de aproximadamente 1,05. A substituição deste valor em (3.9) resulta:

$$\sigma_z'^2 = \frac{0.094}{\pi} \int_0^\infty \frac{\text{sen}^2 [3.11 X n']}{(1+n')^{5/3} n'^2} dn' \quad (3.11)$$

4 Conclusões

A confrontação dos dois modelos com resultados observados de concentração permite concluir que ambas aproximações são capazes de reproduzir adequadamente as observações. Uma análise baseada em índices estatísticos indica que o modelo da solução da equação da difusão e advecção reproduz melhor as concentrações experimentais, porém, devido a praticidade do modelo Gaussiano ele é mais indicado para ser usado. Os resultados dos modelos estão confrontados com os dados observacionais na tabela 2.

Tabela 1. Índices estatísticos dos resultados dos modelos.

Modelo	NMSE	r	FB	FS
Gaussiano	0,08	0,79	-0,12	-0,28
mod. (2.6), (2.7)	0,08	0,76	-0,04	-0,30
mod. (2.6), (2.8)	0,09	0,77	-0,11	-0,34

Tabela 2. Concentrações superficiais integradas lateralmente.

Exp.	Distância (km)	Observada (10^{-4}sm^{-2})	Gaussiano (10^{-4}sm^{-2})	mod. (2.6), (2.7) (10^{-4}sm^{-2})	mod. (2.6), (2.8) (10^{-4}sm^{-2})
1	1,9	6,48	6,00	7,10	6,29
	3,7	2,31	3,90	3,81	3,50
2	2,1	5,38	3,59	4,37	3,96
	4,2	2,95	2,63	2,99	2,54
3	1,9	8,20	7,28	6,57	6,46
	3,7	6,22	5,16	3,70	3,53
	5,4	4,30	4,16	2,67	2,51
7	2,0	6,70	3,91	5,41	5,17
	4,1	3,25	2,49	3,24	2,92
	5,3	2,23	2,11	2,57	2,24
8	1,9	4,16	4,10	4,89	4,76
	3,6	2,02	3,07	3,46	3,19
	5,3	1,52	2,51	2,52	2,27
9	2,1	4,58	3,48	4,39	4,60
	4,2	3,11	2,31	2,99	2,96
	6,0	2,59	1,83	2,22	2,12

5 Referências Bibliográficas

Degrazia, G.A., Moraes, O.L.L. e Oliveira, A.P., An analytical formulation for mixing length scales and eddy diffusivities in a convective PBL, 21st NATO/CCMS International Technical Meeting on Air Pollution Modelling and Its Application, American Meteorological Society, 1994.

Gryning et alli, Applied dispersion modelling based on meteorological scaling parameters Atmos. Environment, 21, 79-89, 1987.

Holtslag, A.A.M., and Moeng, C.H., Eddy diffusivity and counter-gradient transport in the convective atmospheric boundary layer, J. Atmos. Sci., 48, 1690-1698, 1991.

Moreira, D.M., Comparação entre um modelo Gaussiano atualizado e a equação de difusão no cálculo da concentração superficial de poluentes na camada limite convectiva: o caso de fontes pontuais contínuas elevadas. Santa Maria, 53p. Dissertação (Mestrado em Física). Universidade Federal de Santa Maria, 1996.

* Trabalho parcialmente financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo a Pesquisa do Rio Grande do Sul (FAPERGS).

STUDY AND PREDICTION OF THE PARAGUAY RIVER LEVEL BY HARMONIC ANALYSIS AND NEURAL NETWORKS

Li Weigang & Daniel J. R. Nordemann

Neural networks and harmonic analysis by iterative regression are implemented for the prediction of the level of Paraguay River. The selected neural networks include Feed-forward Neural Networks with Backpropagation learning law and Time Delay Neural Network. Using the 1145 monthly mean values, the trained networks predict the levels of the 12 next months with Normalized Mean Squared Error NMSE < 0.06 and Negative Average Log Likelihood NALL < 9.0 . On the other hand, the determination of the most important sine functions embedded in the same series allows to reconstruct the main features of the variations of the river level and to predict yearly mean values for the next two decades (assuming no major change of the environment). The results obtained show that both methods seem to be efficient to process time series related to phenomena influenced by complex climatic and geophysical processes, even not dealing with causal relationships involved in the phenomena studied. They may be used to predict future behavior of such phenomena, at ranges depending specifically on the method used, interval, size and quality of data available.

Key words: Prediction; Neural networks; Harmonic analysis; Iterative regression; Paraguay river; Paraguay river level.

ESTUDO E PREVISÃO DO NÍVEL DO RIO PARAGUAI USANDO ANÁLISE HARMÔNICA E REDES NEURAIS - *Redes neurais e análise harmônica por regressão iterativa foram usadas para a previsão do nível do Rio Paraguai. As redes neurais selecionadas para este trabalho compreendem a rede neural "Feed-forward" com lei de aprendizagem com retropropagação e rede neural com atraso de tempo. A partir de 1145 valores mensais médios, a rede neural treinada prevê o nível dos doze meses seguintes com erro médio normalizado ao quadrado NMSE < 0.06 e probabilidade logarítmica média negativa NALL < 9.0 . Além disto, a determinação das mais importantes funções senoidais embutidas na mesma série permite a reconstrução das principais variações do nível do rio e a previsão dos valores médios anuais para as duas décadas seguintes (supondo nenhuma mudança importante do meio ambiente). Os resultados obtidos mostram que ambos os métodos parecem ser eficientes para processar séries temporais influenciadas por processos climáticos e geofísicos complexos, mesmo sem considerar as relações causais envolvidas nos fenômenos estudados. Eles podem ser usados para prever o comportamento de tais fenômenos em escalas de tempo que dependem especificamente do método usado, do intervalo de tempo e do tamanho e da qualidade dos dados disponíveis.*

Palavras-chave: Previsão; Redes neurais; Análise harmônica; Regressão Iterativa; Rio Paraguai; Níveis do Rio Paraguai

Instituto Nacional de Pesquisas Espaciais -INPE, C.P. 515,
12201-970 São José dos Campos, SP, Brazil, E-mail: wei@met.inpe.br

INTRODUCTION

The Paraguay River (length: 2550 km) rises in the Mato Grosso region of Brazil, at 300 m above sea level, and runs southward between highlands at the west and the Brazilian plateau at the east. Its basin, with an area of approximately 500000 km², consists of a series of huge alluvial plains drained by a complex network of rivers interspersed with marshes, in a region called Pantanal. In this region, many areas suffer a succession of droughts and severe floods with their obvious economic and social consequences. Paraguay River levels are influenced by several different factors from micro to macro scales. Therefore, predicting the level of the Paraguay River with convenient antecedence (and so estimating the area to be flooded) is a relevant scientific goal. In this paper, Feed-forward Neural Network, Time Delay Neural Network and harmonic analysis with iterative regression are implemented for the reconstruction and prediction of the level of the Paraguay River at Ladário near Corumbá (Mato Grosso do Sul State, Brazil). The harmonic analysis with iterative regression was used to investigate the most important periodicities of the time series of the Paraguay River levels. The neural networks method was further used to predict the future behavior of the levels which were measured up to the present time. As shown in the following parts of this work, the neural networks can treat all the involved factors simultaneously, which was an incentive to apply this method to complex time series such as river levels. The daily data of the series studied in this work were collected from 1900 to June of 1995. About 1145 monthly mean data are used to train the neural network, which in return gives the monthly predictions. The initial results show successful predictions within three to four month scale.

PERIODICITIES ANALYZED BY ITERATIVE REGRESSION ANALYSIS

Variations of the Paraguay River (Fig. 1) clearly show a complex behavior with long periods of drought and flood. Periodicities are not evident and it also may be seen that the flow rate is not stationary. For these reasons, these data were analyzed by several mathematical methods including Fourier analysis, periodogram/cyclogram, dynamic spectral analysis and iterative regression and a methodology developed for this purpose (Nordemann, 1995). In spite of the river level

(Fig. 1) not being stationary, (DNOS/UNESCO-PNUD, 1974; Nordemann, 1995), an attempt was done to select the most important periodicities embedded in the signal and to use them in order to predict the near future of the river level.

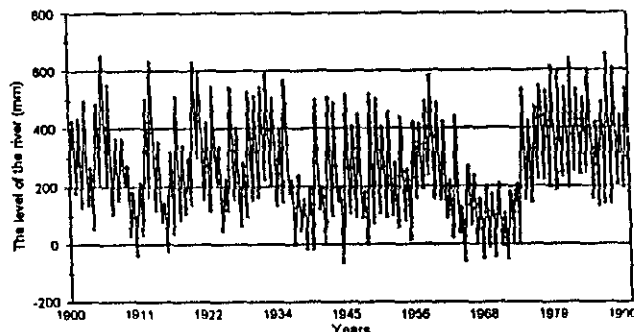


Figure 1 - The observed monthly means of the level of Paraguay River.

Figura 1 - Médias mensais dos níveis do Rio Paraguai.

METHODS

In order to analyze geophysical time series, it is recommended to use methods which are more adequate to the problem and which are chosen within computational tools and software compatible with available hardware and processing time (Dettinger et al., 1995). Several algorithms were developed by us within *Mathematica* for Windows environment (Wolfram, 1991; Nordemann, 1994). All treatments were processed by a 486DX4 100 MHz machine with 16 MB RAM and 500 MB hard disk.

In order to study better River Paraguay behavior, the iterative regression method was chosen among the methods previously used and applied to the 1900-1995 level value series (Nordemann, 1995). Most of classical harmonic analysis are performed on the time series using methods such as Fourier transform. Here, we used a different method which searches one by one the 3-parameter sine functions which fit better with data by a minimum square iterative regression fit (Wolberg, 1967). For a better accuracy on the period values, the method may be applied through the sweep of allowed frequencies or periods. In our case a preliminary search using periodogram restricts the search only to the most important sine function embedded in the signal as detected by the maxima of periodogram. Following this step, the iterative regression method is applied to the regions of

the maxima, in decreasing order of their amplitude. This iterative regression was applied for every sine function to the initial time series stripped of the reconstructed function from the previously detected periodicities.

RESULTS FROM ITERATIVE REGRESSION

These determinations gave convergence for about 50 sine functions with amplitude greater than the corresponding standard deviation. Tab. 1 presents the most important periodicity parameters. The major amplitudes correspond to 1-year period (Earth orbital revolution) followed by the 28-year period and several amplitudes for periods 2 to 4 years (Quasi Biannual Oscillation/El Niño Southern Oscillation QBO/ENSO) and others.

Period (yr.)	Amplitude \pm s.d. (cm)	Observation
1.0000 \pm 0.0002	130.0 \pm 5.2	Orbital revolution
28.4 \pm 0.77	77.0 \pm 13.7	
14.6 \pm 0.34	44.5 \pm 12.0	
8.9 \pm 0.16	33.7 \pm 11.4	
7.8 \pm 0.14	33.6 \pm 12.8	
6.6 \pm 0.12	20.7 \pm 10.2	
3.8 \pm 0.03	32.6 \pm 10.9	QBO/ENSO
4.8 \pm 0.06	23.9 \pm 10.7	
2.8 \pm 0.02	22.7 \pm 10.5	
2.3 \pm 0.02	17.3 \pm 10.1	
...	...	

Table 1 - Most important periodicity parameters (Paraguay river, 1900-1995).

Tabela 1 - Parâmetros das periodicidades mais importantes (Rio Paraguai 1900-1995).

Of course, the most important periodicity found in the studied time series corresponds to one year period, being due to the orbital revolution of the Earth. But it may be seen that the amplitude which corresponds to 1-year period represents only a small fraction of the total amplitude of the river height variations (half difference between extreme heights $[657 - (-61)]/2 = 359$ cm). For this reason, components of the following periods play a very important part. It may also be seen that the sum of the nine major amplitudes is about 400 cm, which means that, in the case of this study, about nine amplitudes

may be enough to account for the observed extreme values. With the method used, about 50 periodicities present an amplitude greater than their respective standard deviation. This means that besides the first nine periodicities quoted, others may be significant and that there may be other natural cycles which may also influence the river height.

To show tendencies embedded in the time series, a dynamic spectral analysis was also performed as "classical" dynamic spectral analysis with constant length samples and also as "wavelet style" dynamic spectral analysis with constant number of periods per sample (Farge, 1995). It showed clearly the evolution of the 2 to 5 years periodicity (ENSO-QBO), stronger from 1900 to 1950 than for the rest of the series as well as other features of the evolution of the river behavior. Among these features is the higher amplitude of the 28.4 year period versus time during the last decades of the series.

The 28.4 year period is within the 28-32 yr interval periodicity shown by Kane & Teixeira (1990) in air temperatures for both hemispheres, but the remaining periodicities at 5-6 yr, 10-11 yr, 15 yr 20 yr and 55-80 yr detected by these authors do not appear or appear only as small signals in the 1900-1995 Paraguay River series.

As a preliminary effort, and in spite of the recognized evolution of the series, the reconstruction, for each hydrological year, was made after the determination by iterative regression analysis of the sine functions embedded in the whole series. The results obtained are presented in Fig. 2 which shows the reconstruction of the Paraguay River height using the nine most significant sine components (Tab. 1). It may be seen that, for this model, some large amplitude floods (before 1920) or droughts (1963 to 1974) are not described with fidelity, which corresponds to their exceptional occurrence, opposed to the hypothesis of being stationary. An attempt was also made using the same functions to predict the behavior of the annual means of the Paraguay River during the next 20 years, assuming no important change in the regional and global environment for this interval. Drastic change in the environment such as land use in the Pantanal and adjacent regions or works to improve the navigability of the Paraguay River should invalidate such medium or long term forecasting. It may be seen that this model predicts a large amplitude drought for the near future, up to about year 2000, followed by alternate periods of normal level and mild flood.

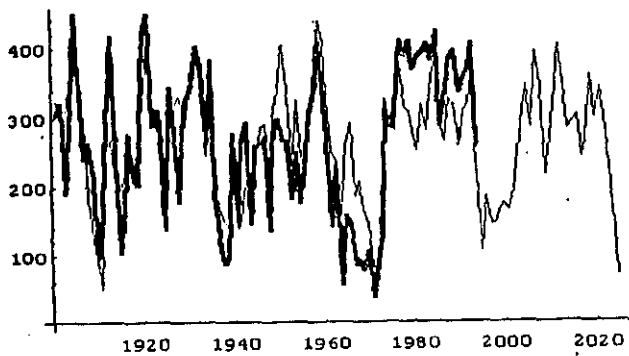


Figure 2 - Yearly means of the level of the Paraguay River at Ladário 1900 to 1994 (thick curve) and prediction based on 9 most important periodicities up to 2025 (thin curve). A rather good fit may be observed between observed and reconstructed values except in the case of several periods of flood or drought after 1950, which characterize their exceptionality. This model predicts a rather long period of lower than mean values from present to about 2005 and variable values higher than mean after 2005.

Figura 2 - Médias anuais dos níveis do Rio Paraguai, em Ladário, de 1900 a 1994 (traço grosso) e previsão baseada nas nove mais importantes periodicidades até 2025 (traço fino). Uma razoável concordância é notada entre os valores observados e os previstos, exceto em casos dos vários períodos de inundação ou seca após 1950, o que caracteriza a sua excepcionalidade. Este modelo prediz um período bem longo de níveis mais baixos do que a média, do presente momento até aproximadamente o ano de 2005, e valores variáveis maiores do que a média após 2005.

APPLICATION OF THE NEURAL NETWORKS

Prediction of time series is an exciting recent application of neural networks. There are a number of prediction methods available for this kind of problem (Casdagli, 1989). Neural networks were found to be useful and competitive with the best recent approximation methods (Lapedes & Farber, 1987; Gallant & White, 1992; Gershenfeld & Weigend, 1993; Li et al., 1995a; 1995b). To predict the future behavior of the Paraguay River level time series, the Feed-forward Neural Network (BPNN) and Time Delay Neural Network (TDNN) were used and both of them were implemented in the neural networks simulator SNNS (Zell et al., 1995). To analyze the prediction quality, we used a simple method-independent technique (Gershenfeld & Weigend, 1993; Nordemann & Li, 1996).

Feed-forward neural network

The most popular network is the Feed Forward Network with backpropagation learning law (Rumelhart et al., 1986). For one hidden layer, the input values of time series $x(t-1), x(t-2), \dots, x(t-d)$ are received through d input units, which simply pass the input forwards to the hidden units $u_j, j = 1, 2, \dots, q$. Each connection performs a linear transformation determined by the connection strength w_{ij} , so the total input for every hidden unit u_j is $\sum_{i=1}^d w_{ij} x(t-i)$. Each unit performs a nonlinear transformation on its total input, producing the output:

$$u_j = \Psi \left(W_{0j} + \sum_{i=1}^d W_{ij} X(t-i) \right).$$

The activation function Ψ is the same for all units. Here, Ψ is a sigmoid function with limiting value 0 and 1 as $u_j \rightarrow -\infty$ and $u_j \rightarrow +\infty$, respectively:

$$\Psi(u_j) = \frac{1}{(1 + e^{-u_j})}.$$

The hidden layer outputs u_j are passed along to the single output unit with connection strength β_j , which performs an affine transformation on its total input. Then, the network's output $x(t)$ can be represented as:

$$x(t) = \beta_0 + \sum_{j=1}^q \beta_j \cdot \Psi \left(w_{0j} + \sum_{i=1}^d w_{ij} \cdot x(t-i) \right),$$

for d inputs and q units in the hidden layer.

Time Delay Neural Network

The Time Delay Neural Network is a layered network in which the outputs of a layer are buffered by several time lags and then fed fully connected to the next layer (Waibel et al., 1989; Wan, 1993). The activation of an unit is normally computed by passing the weighted sum of its inputs to an activation function, usually a threshold or sigmoid function. For TDNN, this behavior is modified through the introduction of delays (Zell et al., 1995). Training for this kind of network is performed by a procedure similar to backpropagation, that takes the special semantics of coupled links into account. To enable the network to achieve the

desired behavior, a sequence of patterns has to be presented to the input layer with the feature of interest shifted within the patterns.

The results from neural networks

The selected prediction methods include Feed-forward Network with Backpropagation (BPNN) and Time Delay Neural Network (TDNN). We used the notation input units: hidden units: hidden units: ... : output units to describe the structure of the network. For BPNN, the selected structure is 12:48:48:1; for TDNN, the selected structure is 36:9:1. In both networks, the training rate used was 0.2. Fig. 3 shows the reconstruction (from 01/90 to 06/95) and the predictions (from 07/95 to 06/96) of the monthly level of Paraguay River. The trained networks gave the next 12 months prediction values. For the period 07/95 to 12/95, the predictions from two networks gave very similar results; for the next 6 months, the differences between predictions were slightly higher. The confidence one may have in such prediction is shown with prediction error bars which characterize the probability of having the result of a future measurement within a given interval (one standard deviation) near the predicted value. Figs. 4 and 5 show the prediction error bars (from 06/95 to 07/96) obtained by using Feed-forward Neural Network with Backpropagation and Time Delay Neural Network.

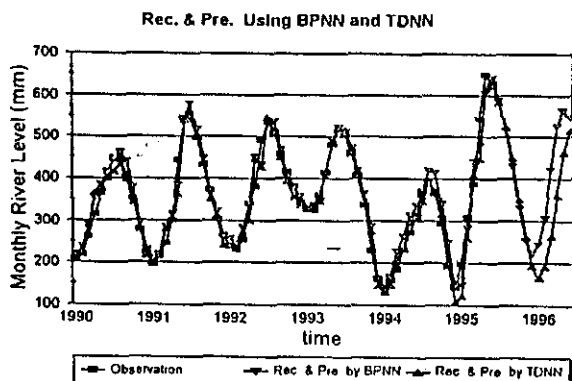


Figure 3 - Reconstruction (01/90-06/95) & prediction (07/95-06/96) of the monthly level of Paraguay River

Figura 3 - Reconstrução (01/90-06/95) e previsão (07/95-06/96) dos níveis mensais do Rio Paraguai.

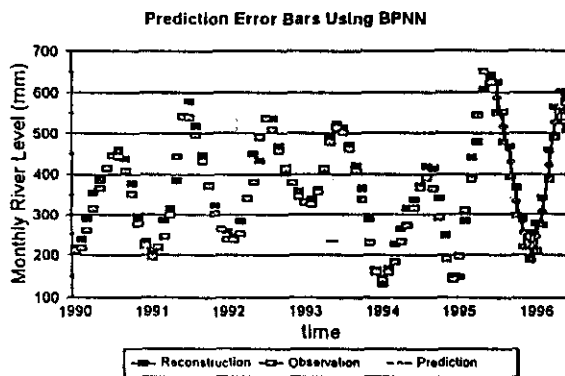


Figure 4 - Prediction error bars (06/95-07/96) using BPNN.

Figura 4 - Barras de erro das previsões (06/95-07/96) utilizando BPNN.

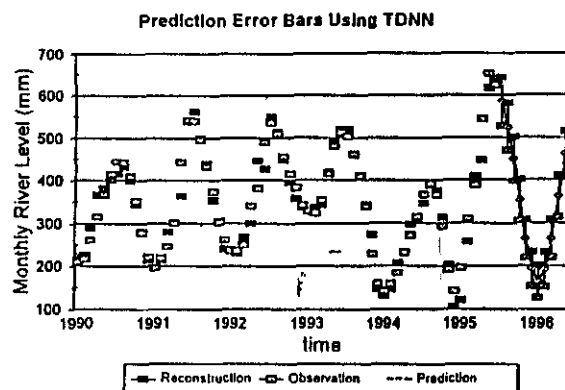


Figure 5 - Prediction error bars (06/95-07/96) using TDNN.

Figura 5 - Barras de erro das previsões (06/95-07/96) utilizando TDNN.

Tab. 2 shows the prediction quality of two neural networks using 1000 monthly mean data with 2000 training cycles. The index *nmse1* and *nall1* denote the quality of the reconstruction and the index *nmse2* and *nall2* denote the quality of the prediction (Nordemann & Li, 1996). The results in Tab. 2 show that Time Delay Neural Network gave the best reconstruction, $nmse1 = 0.0305$ and $nall1 = 8.2288$, and Feed-forward Network with Backpropagation was located in second, $nmse1 = 0.1114$ and $nall1 = 9.5082$. For the future prediction, Time Delay Neural Network also gave the best results, $nmse2 = 0.0588$ and $nall2 = 8.5317$ and and Feed-forward Network with Backpropagation was located in second, $nmse2 = 0.0584$ and $nall2 = 8.5498$.

Net models	Training time	Reconstr. quality		Prediction quality	
		nmse1	na11	nmse2	na12
Backpropagation (BPNN)	2 hours	0.1114	9.5082	0.0584	6.5498
Time Delay Network (TDNN)	20 min.	0.0305	8.2288	0.0588	8.5317

Table 2 - Prediction quality of two neural network models.

Tabela 2 - Qualidade de previsão dos dois modelos de redes neurais.

COMPARISON BETWEEN NEURAL NETWORKS AND ITERATIVE REGRESSION METHOD

For further evaluation of the prediction results, the same prediction has been done by means of the iterative regression method and neural networks (Nordemann et al., 1995).

Comparison of the methods used

Common Points

Neural Networks	Iterative regression
Using the minimization of the mean-square (ms) error as prediction error criterion.	Using the minimization of the mean-square (ms) error as prediction error criterion.
More data, better prediction	More data, better prediction

Differences

Neural Networks	Iterative regression
Parallel processing	Series processing
Using exponent function	Using sine function
Using the learning technique	Using the successful iterations
Taking long time for training neural networks (days)	Short processing time if considering only the most significant periodicities
Good fit of reconstructed data to observed data	Medium quality reconstruction
Good prediction for short range	Prediction supposed to be satisfactory for the middle range (about or less than 20% of the sampled interval, depending on the most significant periodicities)

Comparison of the prediction results

Short range (Months)	March/95	April/95	May/95	June/95
→ Observations(cm)	543	650	622	588
→ TDNN	446	605	637	583
Relative error (%)	7.76	5.34	2.5	0.7
→ BPNN	477	606	641	567
Relative error (%)	12.0	6.0	3.0	0.21
→ Iterative regression	310	332	336	331
Relative error (%)	42.77	48.88	45.95	43.65
Long range (Years)		1996—2001	2002—2023	
Neural Networks		381		
Iterative regression		152		299

Analysis of comparison

From the above analysis, we may draw the following conclusions:

- 1) Both methods are suitable to predict the level of the Paraguay River;
- 2) For short time period, the predicted results from Neural Networks are better than those from iterative regression;
- 3) For medium range period, the iterative regression shows better potential for prediction.

CONCLUSIONS

Feed-forward Neural Network, Time Delay Neural Network and iterative regression methods are suitable tools for mathematical reconstruction and prediction of natural multiple cause complex phenomena such as the level of the Paraguay River. Comparing both methods used, the predictors gave acceptable results for different duration ranges. Although the results presented are preliminary, they are promising. Improving the prediction with the help of predictors carefully chosen and investigating theoretical aspects of the predictability of such geophysical and climatologic phenomena will be the objects of further works.

ACKNOWLEDGMENTS

This research was supported by CNPq, under contracts 300585/94-2 and 300031/94-7, FAPESP under contract 93/2715-1.

REFERENCES

- CASDAGLI, M.-1989- Nonlinear Prediction of Chaotic Time Series. *Physica, D* 35: 335-356.
- DETTINGER, M. D., GHIL, M., STRONG, C. M., WEIBEL, W. & YOU, P.-1995- Software Expedites Singular-Spectrum Analysis of Noisy Time Series. *EOS (AGU)* 76(2):12-21, Jan. 10.
- DNOS/UNESCO-PNUD -1974- Estudos Hidrológicos da Bacia do Alto Paraguai, Relatório Técnico, Vol. 1. Ministério do Interior, Departamento de Obras de Saneamento - DNOS, Brasil and UNESCO, Programa das Nações Unidas Para o Desenvolvimento - PNUD (BRA-66.521).
- FARGE, M.-1995- Personal communication.
- GALLANT, A. R. & WHITE, H. -1992 - On learning the derivatives of an unknown mapping with multilayer Feedforward Networks. *Neural Networks*, Vol.5, pp. 129-138.
- GERSHENFELD, N. A. & WEIGEND, A. S. - 1993 - The future of Time Series: Learning and Understanding. *Time Series Prediction: Forecasting the Future and Understanding the Past*, Eds. A.S. Weigend & N.A. Gershenfeld-1993 - SFI Studies in the Sciences of Complexity, Proc. Vol. XV, Addison-Wesley.
- KANE, R. P. & TEIXEIRA, N. R. -1990- Power Spectrum Analysis of the Time-series of Annual Mean Surface Air Temperatures. *Climatic Change* 17:121-130.
- LAPEDES, A. & FARBER R. -1987 - Nonlinear signal processing using neural networks: prediction and signal modeling. Research Report, Los Alamos.
- LI W. G., SÁ, L. D. A., MANZI, A. O., PRASAD, G. S. S., NOWOSAD, A.G. AND CULF, A. D.-1995a- Neural networks for nonlinear prediction of turbulent signals from data measured above Amazon forest and Pasture. XVIII Congresso Nacional de Matemática Aplicada e Computacional, Curitiba, Brazil.
- LI W. G., SÁ, L. D. A., GALVÃO G. & BEVILAQUA, R. M.-1995B- Prediction of the Paraguay River Level's Time Series Using Neural Networks. Encontro sobre Sensoriamento Remoto Aplicado a Estudos no Pantanal. Corumbá - MS.
- NORDEMANN D. J. R. - 1994 - Introdução ao *Mathematica* for Windows. Transtec Editorial, São José dos Campos, Brazil.
- NORDEMANN, D. J. R. -1995 - Periodicidades e tendências nas médias mensais do nível do Rio Paraguai em Corumbá de 1900 a 1995; Previsão baseada no estudo das médias anuais do nível do rio Paraguai em Corumbá de 1900 a 1995. Encontro sobre Sensoriamento Remoto Aplicado a Estudos no Pantanal, Corumbá, MS.
- NORDEMANN, D. J. R. LI, W. G. & BEVILAQUA, R. M. -1995 - Previsões sobre o nível do rio Paraguai em Corumbá: comparação entre dois métodos (Redes neurais e análise harmônica por regressão iterativa). Encontro sobre Sensoriamento Remoto Aplicado a Estudos no Pantanal. Corumbá, MS.
- NORDEMANN, D. J. R. & LI W. G. - 1996 - Climatic Change Prediction Using Neural Networks and Prediction Quality Analysis: Application to Sea Surface Temperature Time Series. Conference on Environmetrics in Brazil. São Paulo, SP.
- RUMELHART, D. E., MCCLELLAND, J.L. & THE PDP RESEARCH GROUP - 1986- Parallel distributed processing: explorations in the microstructure of cognition, Cambridge, MIT Press, Vol. 1.
- WAN, E. A. -1993 - Time series prediction by using a connectionist network with internal delay lines. *The Future of Time Series: Learning and Understanding. Time Series Prediction: Forecasting the Future and Understanding the Past*, Eds. A.S. Weigend and N.A. Gershenfeld, SFI Studies in the Sciences of Complexity, Proc. Vol. XV, Addison-Wesley.
- WAIBEL, A., T. HANAZAWA, T., HINTON, G., SHIKANO K. & LANG, K. - 1989- Phoneme Recognition using Time Delay Neural Networks. *IEEE Trans. on Acoust. Speech, Signal Proc.* 37, pp.328-339.
- WOLBERG J. R. - 1967 - Prediction Analysis. Van Nostrand Company, Princeton.
- WOLFRAM S. - 1991 - *Mathematica*, A System for Doing Mathematics by Computer. Addison-Wesley Pub. Co.
- ZELL, A. et al. - 1995- Stuttgart Neural Network Simulator -SNNs, User Manual, Version 4.0. University of Stuttgart, Report No. 6/95.

Submetido em: 12/03/96

Revisado pelo(s) autor(es) em: 04/06/96

Aceito em: 10/06/96

ESTUDO E PREVISÃO DO NÍVEL DO RIO PARAGUAI USANDO ANÁLISE HARMÔNICA E REDES NEURAI

Os níveis do Rio Paraguai são influenciados por diversos fatores geofísicos complexos que atuam em várias escalas. Torna-se, portanto, bastante relevante, nos aspectos científicos, sociais e econômicos, a previsão do nível do Rio Paraguai (na cidade de Ladário) com uma certa antecedência (e assim pode-se estimar a área a ser inundada). Neste artigo, redes neurais e análise harmônica por regressão iterativa foram usadas para a previsão do nível do Rio Paraguai. As redes neurais selecionadas para este trabalho compreendem a rede neural "Feed-forward" com lei de aprendizagem com retropropagação e rede neural com atraso de tempo. Ambas foram implementadas no simulador de redes neurais SNNS (Zell et al., 1995). Para analisar a qualidade da previsão nós usamos uma técnica simples independente do método empregado (Gershenfeld & Weigend, 1993; Nordemann & Li, 1996). A partir de 1145 valores mensais médios, a rede neural treinada prediz o nível dos doze meses seguintes com erro médio normalizado ao quadrado NMSE < 0,06 e probabilidade logarítma média negativa NALL < 9,0. Nós comparamos os resultados obtidos pelos dois modelos de redes neurais, de acordo com a análise da qualidade da previsão. O resultado preliminar indica que a TDNN produziu a melhor reconstrução e previsão. Por outro lado, variações no nível do Rio

Paraguai baseadas em determinações diárias em Ladário, de janeiro de 1900 até o presente (médias mensais), foram estudadas para se obter o histórico do clima e a relação com fenômenos geofísicos. Estas variações mostram de maneira bem clara, o comportamento complexo dos longos períodos de seca e de inundações. As periodicidades não são evidentes e também pode ser observado que a taxa de vazão não é estacionária. Por estas razões, estes dados foram analisados por diversos métodos matemáticos incluindo a análise de Fourier, periodogramas/ciclogramas, análise espectral dinâmica e regressão iterativa.

Além disto, a determinação das mais importantes funções senoidais embutidas na mesma série permite a reconstrução das principais variações do nível do rio e a previsão dos valores médios anuais para as duas décadas seguintes (supondo nenhuma mudança importante do meio ambiente). Os resultados obtidos mostram que ambos os métodos parecem ser eficientes para processar séries temporais influenciadas por processos climáticos e geofísicos complexos, mesmo sem considerar as relações causais envolvidas nos fenômenos estudados. Eles podem ser usados para prever o comportamento de tais fenômenos em escalas de tempo que dependem especificamente do método usado, do intervalo de tempo e do tamanho dos dados disponíveis.

NOTA SOBRE OS AUTORES NOTES ABOUT THE AUTHORS

Li Weigang

- B. Tec. (Civil Engineering), He Hai University, China, 1982
- M. Sc. (Operation Research & Urban Planning), China, Nanjing University, 1988
- D. Sc. (Operation Research & Air Transport), Instituto Tecnológico de Aeronáutica, Brazil
- Professional Activities:
1982-1987 - Ass. Professor - Nanjing Engineering Institute, Nanjing, China
1988-1989 - Lecturer - Nanjing Engineering Institute, Nanjing, China
Since 1994 - Research Fellow of CNPq (Recém Doutor) at INPE/DCM, São José dos Campos, Brazil

Daniel Jean Roger Nordemann

- Engineer (Physics), Superior School of Industrial Physics and Chemistry - ESPCI, Paris - France, 1957
- B. Sc. (Mathematical Physics), Paris University, France, 1963
- Ph. D. (Physics), Paris University, France, 1966
- Specialization Areas: Space Geophysics, Geophysical and Climatological Data Processing, Nuclear Geophysics
- Professional Activities:
1957-1970 and 1974-1979 - French Atomic Energy Commission - CEA at Low Radioactivity Center - CFR, France
1970-1973 - UNESCO Expert - Professor of Physics at Bahia Federal University, Bahia - Brazil
Since 1979 - Researcher at INPE, São José dos Campos, Brazil.
Head of Space Geophysics Division - DGE (since 1990)

Anexo B

Trabalho aceito para apresentação

X Congresso Brasileiro de Agrometeorologia

Piracicaba - 13 a 18 de julho de 1997

UTILIZAÇÃO DO ALGORITMO DE BOX E HILL PARA DISCRIMINAÇÃO ENTRE MODELOS COMPETITIVOS

Ralf GIELOW¹, Roberto Francisco Marques MENDES², Gervasio A. DEGRAZIA³

RESUMO

O algoritmo de Box e Hill, com base no conceito de entropia da informação e no teorema de Bayes, é um procedimento para a discriminação entre M modelos que competem para representar um determinado fenômeno ou sistema N -dimensional. Apresenta-se uma versão computacional amigável desse algoritmo, assim como os resultados de sua aplicação na comparação do desempenho, frente a dados experimentais, de três modelos para representar a concentração ao nível da superfície de um gás ou vapor em dispersão na atmosfera.

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.

O algoritmo é baseado no conceito de entropia da informação e no teorema de Bayes; ele efetua a discriminação entre M modelos que competem entre si para representar um sistema, conhecendo-se observações ou medidas experimentais, e respectiva variância, em 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 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

¹ Dr., Pesquisador Titular, Divisão de Ciências Meteorológicas, C. P. 515, 12201-970 São José dos Campos, SP. E-mail: ralf@met.inpe.br

² Pós-graduando em Ciência da Computação, IME/USP. E-mail: mendes@met.inpe.br

³ Dr., Professor, Departamento de Física, UFSM, 97119-900 Santa Maria, RS

“idênticas” de um experimento ou observação. A inferência bayesiana enfoca hipóteses alternativas, enquanto a estatística freqüencial 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 freqüenciais 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 - 4 modelos). Por sinal, a dissertação de Silva Filho (1976), q.v., detalha muito bem o algoritmo de Box e Hill e suas fundamentações.

O presente trabalho mostra uma implementação computacional amigável do algoritmo de Box e Hill, devendo o usuário apenas digitar as expressões para os modelos, ou os resultados da aplicação destes, mais alguns dados de entrada. Apresentam-se, também, os resultados de uma aplicação do algoritmo à dispersão de um gás ou vapor na camada superficial da atmosfera.

MATERIAL E MÉTODOS

Aplicou-se o algoritmo de Box e Hill, na forma acima mencionada, a dados experimentais e três modelos para representá-los, sendo dois analíticos e um numérico, conforme detalhado em Carvalho et al. (1996).

RESULTADOS E CONCLUSÃO

Efetivou-se a discriminação entre os três modelos citados utilizando os dados constantes do trabalho de Carvalho et al. (1996), em que a distância está expressa em metros e as concentrações observadas e as calculadas, em %, conforme mostrado na tabela abaixo, à qual se seguem os resultados obtidos, também reproduzidos na Figura 1:

Distância	Observação	Modelo 1	Modelo 2	Modelo 3
0,20	0,90	2,54	3,71	0,66
0,30	4,90	4,49	5,70	4,02
0,38	5,59	5,29	6,23	5,30
0,40	5,59	5,40	6,27	5,37
0,50	4,9	5,61	6,12	5,00
0,60	4,00	5,45	5,70	4,23
0,70	3,20	5,13	5,22	3,44
0,80	2,20	4,77	4,75	2,84
1,00	1,60	4,07	3,96	1,95
1,50	0,90	2,81	2,72	1,18

Variância do erro experimental = 0,1

Probabilidade

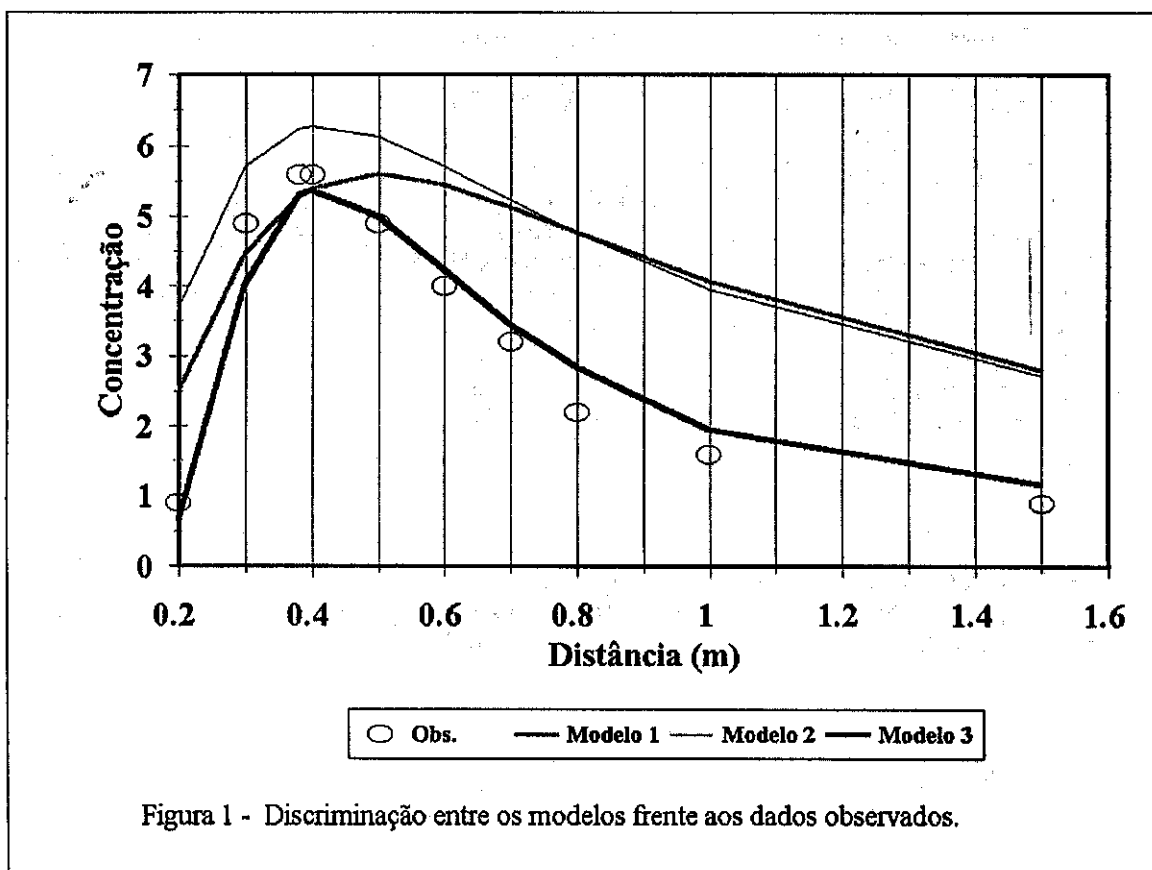
Modelo 1	0,2425
Modelo 2	0,2326
Modelo 3	0,5249

Portanto, o Modelo 3 (numérico) é o mais provável, com probabilidade em torno de 0,5, frente aos modelos analíticos 1 e 2.

AGRADECIMENTO: Ao CNPq, pela concessão de Bolsa de Iniciação Científica PIBIC no INPE a um dos autores, R. F. M. Mendes, de 1995 a 1997.

BIBLIOGRAFIA

- ADEODATO DE SOUZA NETO, J. Discrimination among mechanistic models for oxygen-hemoglobin equilibrium. Gainesville, FL, University Of Florida, 1970. Tese de Doutorado em Engenharia Química.
- BOX, G.E.; HILL, W.J. Discrimination among mechanistic models. *Technometrics*, v. 9, n.1, p.57-71, 1967.
- CARVALHO, J. C.; VELHO, H. F. de C.; DEGRAZIA, G. A. Um estudo numérico da dispersão de poluentes na camada limite convectiva. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 9. Campos do Jordão, SP. *Anais*. Sociedade Brasileira de Meteorologia, p.4-9. 1996.
- FEIGELSON, E.F.; BABU G. J. **Statistical challenges in modern astronomy**. New York, Springer, 1992.
- SILVA FILHO, J. F. da. Discriminação entre modelos competitivos. São José dos Campos, SP: INPE, 1977. 109p. Dissertação de Mestrado em Análise de Sistemas. (INPE - 1027- TPT/051).
- Wadsworth, H. M. J. **Handbook of statistical methods for engineers and scientists**. New York, McGraw-Hill, 1990.



Anexo C

Resumo e resultados do Projeto

Para apresentação no Seminário de Iniciação Científica

INPE - 2 e 3 de julho de 1997

Aplicação do algoritmo de Box e Hill para discriminação entre modelos competitivos

Atila Madureira Bueno

Aluno da Universidade Braz Cubas - Bolsa PIBIC/CNPq

Orientador: Dr Ralf Gielow - Pesquisador Titular DCM/INPE

Fenômenos naturais ou processos artificiais, como evapotranspiração, reações químicas, dispersão de poluentes, acessibilidade urbana ou rendimento agrícola podem ser representados matematicamente por mais de um modelo, de acordo com o mecanismo de funcionamento do sistema ou ajuste estatístico-matemático que o observador ou experimentador considerar.

O algoritmo de Box e Hill (BH), implementado de forma amigável por Mendes (1997), com interface ser-humano-máquina do tipo GUI ("Graphical User Interface" - ao qual se acrescentaram informações de ajuda "clicáveis"), e utilizando o MATLAB e linguagem C em "background", constitui um procedimento para a discriminação entre modelos que competem para representar um determinado fenômeno ou processo. O algoritmo determina a probabilidade a posteriori de cada modelo, utilizando um procedimento bayesiano e considerando dados observados ou medidas experimentais e respectivos erros, cobrindo todo o domínio das variáveis envolvidas.

Neste trabalho, o autor apresenta a aplicação prática de BH a seis casos de competição entre modelos encontrados na literatura, processados por Mendes (1997), mais um caso adicional, mostrando-se a seguir as probabilidades Π_i determinadas para cada modelo i:

Caso #1 - Concentração de poluentes na camada limite convectiva utilizando modelo Gaussiano versus soluções da equação de difusão. Cf. Tab. 2 de Degrazia et al. (1996) - 3 modelos - $\Pi_1 = 0,33$ $\Pi_2 = 0,34$ $\Pi_3 = 0,33$ (empate).

Caso #2 - Concentração de poluentes emitidos por fonte linha na camada limite convectiva utilizando soluções analíticas e numéricas. Cf. Tab. 1 de Carvalho et al. (1996) - 3 modelos - $\Pi_1 = 0,30$ $\Pi_2 = 0,30$ $\Pi_3 = 0,40$

Caso #3 - Concentração de poluentes ao nível da superfície da camada limite convectiva utilizando soluções analíticas e numéricas. Cf. Tab. 3 de Carvalho et al. (1996) - 3 modelos - $\Pi_1 = 0,24$ $\Pi_2 = 0,23$ $\Pi_3 = 0,53$

Caso #4 - Nível do rio Paraguai em quatro meses consecutivos utilizando redes neurais e regressão iterativa. Cf. Weigang e Nordemann (1996) - meses 03 a 06/95 - 3 modelos - $\Pi_1 = 0,99$ $\Pi_2 = 0,01$ $\Pi_3 = 0,00$

Caso #5 - Nível do rio Paraguai em quatro meses consecutivos utilizando quatro modelos de redes neurais. Cf. Weigang et al. (1996) e Weigang (1997) - meses 03 a 06/95 - 4 modelos - $\Pi_1 = \Pi_2 = \Pi_3 = 0,00$ $\Pi_4 = 1,00$

Caso #6 - Nível do rio Paraguai em quatro meses consecutivos utilizando dois modelos de redes neurais. Cf. Weigang et al. (1996) - meses 05 a 08/95 - 2 modelos - $\Pi_1 = 0,9999$ $\Pi_2 = 0,0001$

Caso #7 - Conteúdo ionosférico de elétrons (CIE) sobre Cachoeira Paulista. Conforme Paula et al. (1996) - 2 modelos, SUPIM e IRI90, estratificados cf. o nível de atividade solar no Equinócio (baixo, médio ou alto), referentes a 1982, 1984, 1986 e 1989; resultam para o modelo SUPIM as probabilidades 0,62, 0,79 e 0,86, enquanto para o IRI90 têm-se 0,38, 0,21 e 0,14, respectivamente para a atividade solar em baixo, médio e alto níveis. Outrossim, estratificando-se os dados segundo as três fases temporais diárias do comportamento do CIE (nascimento - ocaso solar, ocaso - reversão noturna e reversão noturna), confirma-se a superioridade do modelo SUPIM, porém com situações de maior aproximação com o IRI90.

Assim, no caso #1 há igualdade entre os modelos, no caso #2 há empate técnico, no caso #3 há discriminação com probabilidade 0,52 para um dos três modelos, enquanto para os casos #4, #5 e #6 a probabilidade para um dos modelos respectivos é maior que 0,99. Finalmente, no caso #7, os resultados são favoráveis para o modelo SUPIM com probabilidades 0,52, 0,60 e 0,67.

Portanto, pode-se concluir que o BH é uma ferramenta de utilidade para a apreciação objetiva do desempenho de modelos que competem para representar realidades para as quais se têm dados observacionais ou experimentais.

Referências Bibliográficas

Carvalho, J. C.; Velho, H. F. de C.; Degrazia, G. A. Um estudo numérico da dispersão de poluentes na camada limite convectiva. **Anais, IX Congresso Brasileiro de Meteorologia, Campos do Jordão, SP, nov. 1996. p. 4-9.**

Degrazia, G. A.; Moreira, D. M.; Friedrich, H. R. Uma comparação entre um modelo Gaussiano e a equação de difusão no cálculo da concentração de poluentes na camada limite convectiva. **Anais, IX Congresso Brasileiro de Meteorologia, Campos do Jordão, SP, nov. 1996. p. 86-89.**

Mendes, R.F.M. **Algoritmo de Box e Hill para discriminação entre modelos competitivos.** Relatório parcial de bolsista PIBIC/INPE - agosto de 1996 a fevereiro de 1997. São José dos Campos, SP, fev. 1997.

Paula, E. R. de; Souza, J. R. de; Abdul, N. A. ; Bailey, G. J.; Batista, I. S.; Bittencourt, J. A.; Bonelli, E. Ionospheric electron content over brazilian low latitude and its comparison with the IRI and SUPIM models. **Advanced Space Research, 18 (6):245-248, 1996.**

Weingang, L. **Dados observacionais e de simulações.** Comunicação pessoal, fev. 1997.

Weingang, L.; Nordemann, D. J. R. Study and prediction of the Paraguay river level by harmonic analysis and neural networks. **Revista Brasileira de Geofísica, 14(2):195-202, 1996.**

Weingang, L.; Sá, L. D. de A ; Galvão, G. P.; Bevilaqua, R. M. Prediction of the Paraguay river level using neural networks. Aceito para publicação no **Pesquisa Agropecuária Brasileira, 1996.**

Caso #1

Concentração de poluentes emitidos por fonte linha na camada limite convectiva, obtida pela solução da equação de dispersão turbulenta (Tabela 2 de Degrazia et al., 1966) utilizando:

Modelo 1 - Perfil Gaussiano

Modelo 2 - Difusão estatística de Taylor e teoria da similaridade convectiva para determinação do coeficiente de dispersão k_{zz} .

Modelo 3 - Simulação de grandes turbilhões para determinação de k_{zz} .

Dados observacionais: Laboratório; $\sigma = 0,1$

Resultado: Ver Tabela C.1.

$$\Pi_1 = 0,33 \quad \Pi_2 = 0,34 \quad \Pi_3 = 0,33$$

Empate técnico; modelo Gaussiano mais prático.

Tabela C.1 - Caso #1

Obs	Mod1	Mod2	Mod3
6,48	6,00	7,10	6,29
2,31	3,90	3,81	3,50
5,38	3,59	4,37	3,96
2,95	2,63	2,99	2,54
8,20	7,28	6,57	6,46
6,22	5,16	3,70	3,53
4,30	4,16	2,67	2,51
6,70	3,91	5,41	5,17
3,25	2,49	3,24	2,92
2,23	2,11	2,57	2,24
4,16	4,10	4,89	4,76
2,02	3,07	3,46	3,19
1,52	2,51	2,52	2,27
4,58	3,48	4,39	4,60
3,11	2,31	2,99	2,96
2,59	1,83	2,22	2,12

Caso #2

Concentração de poluentes emitidos por fonte linha na camada limite convectiva, obtida pela solução da equação de dispersão turbulenta (Tabela 1 de Carvalho et al., 1996) utilizando:

Modelo 1 - Difusão estatística de Taylor e Teoria da similaridade convectiva para determinação do coeficiente de dispersão k_{zz} .

Modelo 2 - Simulação de grandes turbilhões para determinação de k_{zz} .

Modelo 3 - Solução numérica de Misra (Atmospheric Environment 16:1239-243 (1982)).

Dados observacionais: Laboratório; $\sigma = 0,1$

Resultado: Ver Figura C.1

$$\Pi_1 = 0,30 \quad \Pi_2 = 0,30 \quad \Pi_3 = 0,40$$

Ligeira vantagem para o modelo 3.

Concentração de poluentes emitidos por fonte linha na camada limite convectiva

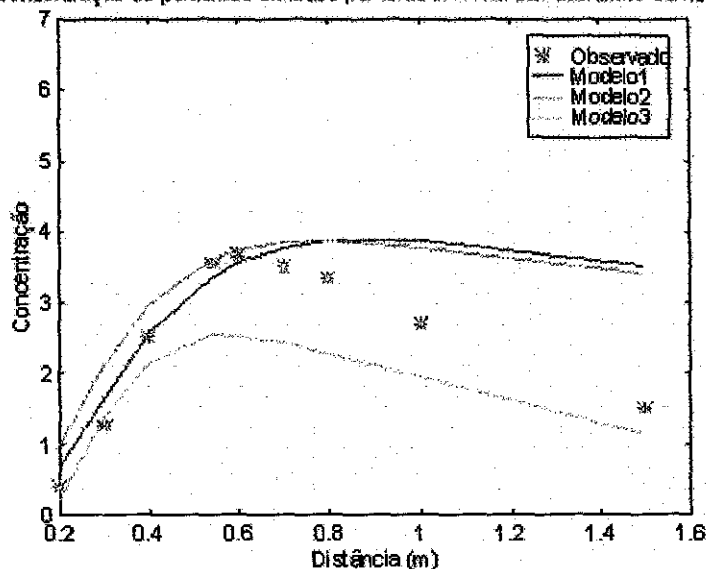


Figura C.1 Caso #2

Caso #3

Concentração de poluentes ao nível da superfície na camada limite convectiva, obtida pela solução da equação de dispersão turbulenta (Tabela 2 de Carvalho et al., 1996) utilizando:

Modelo 1 - Difusão estatística de Taylor e teoria da similaridade convectiva para determinação do coeficiente de dispersão k_{zz} .

Modelo 2 - Simulação de grandes turbilhões para determinação de k_{zz} .

Modelo 3 - Solução numérica de Misra (Atmospheric Environment 16:1239-243 (1982)).

Dados observacionais: Laboratório; $\sigma = 0,1$

Resultado: Ver Figura C.2.

$\Pi_1 = 0,24$ $\Pi_2 = 0,23$ $\Pi_3 = 0,53$
Modelo 3 mais provável.

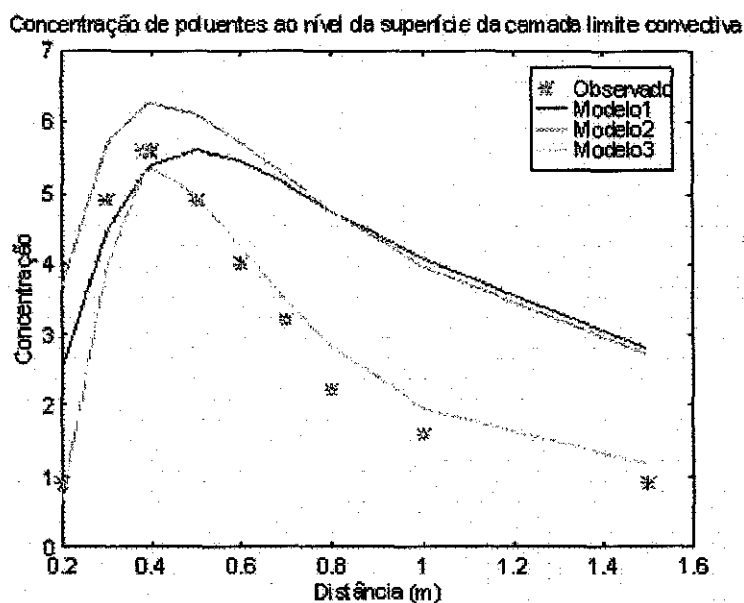


Figura C.2 Caso #3

Caso #4

Nível do Rio Paraguai em Ladário, MS, nos meses de março a junho de 1995 (curto período), obtida de série de longo período (1900 - 1995; - ver Figura C.4), conforme Weigang e Norderman (1996):

Modelo 1 - TDNN - Rede neural com retardo no tempo ("Time Delay Neural Network").

Modelo 2 - BPNN - Rede neural com alimentação "forward" com lei de aprendizado retro-propagada ("Feed Forward Neural Network with Back Propagation Learning Law").

Modelo 3 - Regressão iterativa por análise harmônica.

Dados Observacionais - Série Temporal de Ladário, MS, média mensal 1900-1995; $\sigma = 1$.

Resultado: Ver Figura C.3.

$\Pi_1 = 0,99$ $\Pi_2 = 0,01$ $\Pi_3 = 0,00$
Modelo 1 o melhor.

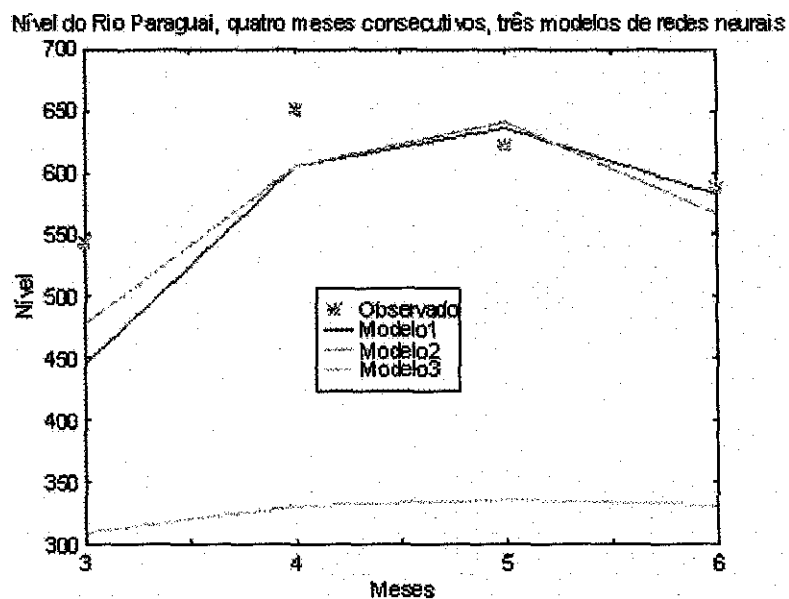
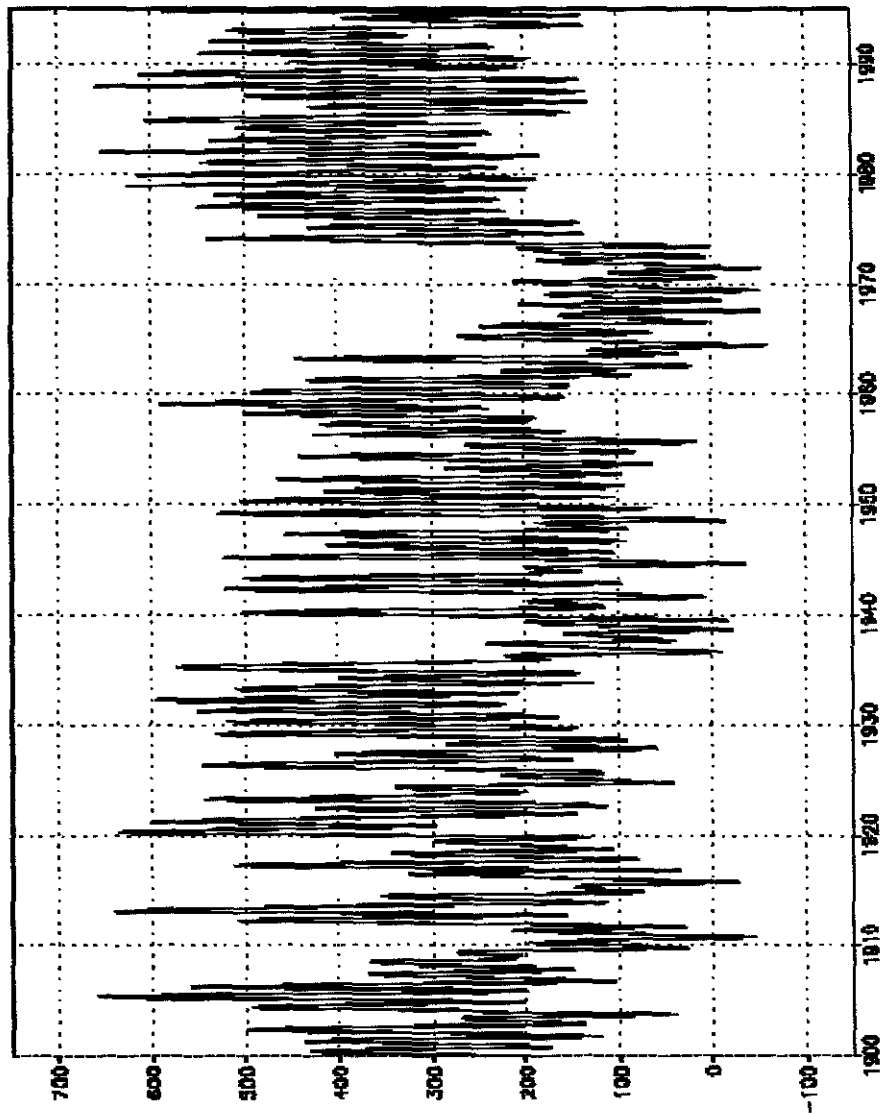


Figura C.3 Caso #4



SNAPS: COLAJ/RES

Figura C.4 - Nível médio mensal (cm) do Rio Paraguai em Ladário, MS - Período de 1900 a 1995.

Caso #5

Previsão do nível do Rio Paraguai em Ladário, MS, nos meses de março a junho de 1995 (curto período), obtida da série de longo período (1900 - 1995; ver Figura C.4), através de rede neural TDNN (com retardo no tempo), conforme Weigang et al. (1996):

Modelo 1 - com 135 observações.

Modelo 2 - com 274 observações.

Modelo 3 - com 571 observações.

Modelo 4 - com 1142 observações.

Dados observacionais: Série temporal de Ladário, MS, médias mensais 1900 - 1995; $\sigma = 1$.

Resultado: Ver Figura C.5.

$\Pi_1 = \Pi_2 = \Pi_3 = 0,00$ $\Pi_4 = 1,00$
Modelo 4 o melhor.

Nível do Rio Paraguai, quatro meses consecutivos, quatro modelos de redes neurais

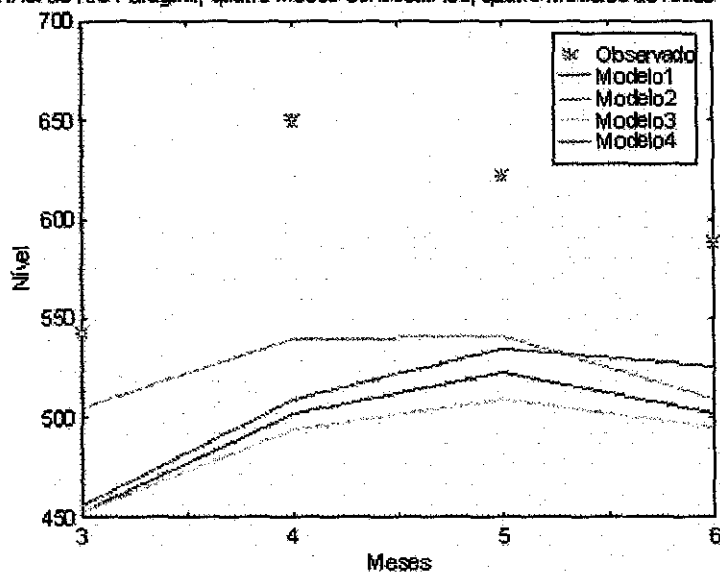


Figura C.5 - Caso #5

Caso #6

Previsão do nível do Rio Paraguai em Ladário, MS, nos meses de maio a agosto de 1959 (curto período), obtida através de rede neural TDNN (com retardo no tempo), utilizando períodos diversos, conforme Weigang et al.(1996):

Modelo 1: 344 observações entre 1900 e 1930.

Modelo 2: 700 observações entre 1900 e 1959.

Dados observacionais: Série temporal de Ladário, MS, médias mensais 1900-1959 - Figura C.4; $\sigma=1$

Resultados: Ver Figura C.6.

$\Pi_1 = 0,9999$ $\Pi_2 = 0,0001$
Modelo 1 o melhor.

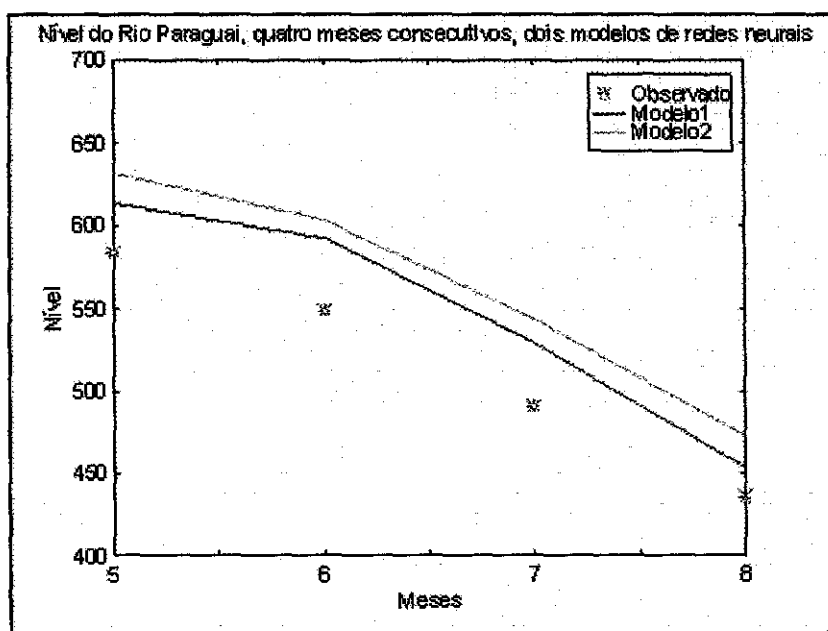


Figura C.6 - Caso #6

Caso #7

Conteúdo ionosférico de elétrons (CEI) sobre Cachoeira Paulista no Equinócio - médias horárias, estratificado de acordo com o nível de atividade solar (baixo, médio ou alto) nos anos de 1982, 1984, 1986 e 1989, período equinocial, conforme Paula et al. (1996):

Modelo 1: SUPIM - Plasmasfera e Ionosfera da Universidade de Sheffield ("Sheffield University Plasmasphere-Ionosphere"; modelo mecanístico).

Modelo 2: IRI90 - Ionosfera de Referência Global ("International Reference Ionosphere"; modelo empírico global).

Dados observacionais: CEI sobre Cachoeira Paulista (22°S, 45°W, 14°S latitude dipolar) obtida nos períodos equinociais de 1982, 1984, 1986 e 1989; $\sigma = 0,5$.

Resultado: Ver Figuras C.7, C.8 e C.9.

- a) Estudo agregado 0 - 24 HL: modelo SUPIM mais aceitável, com probabilidades 0,62, 0,79, e 0,86, respectivamente para atividade solar mínima, média e máxima.
- b) Estratificação em três fases: (i) nascimento - ocaso solar, (ii) intervalo ocaso solar-reversão noturna e (iii) reversão noturna: confirma a superioridade do modelo SUPIM, mas com situações de maior aproximação com o modelo IRI90.

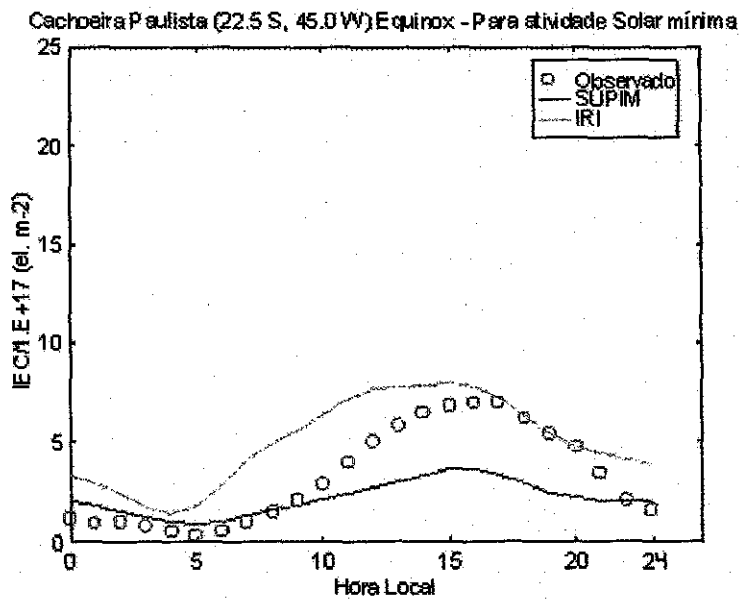


Figura C.7 - Caso #7 - Atividade solar mínima.

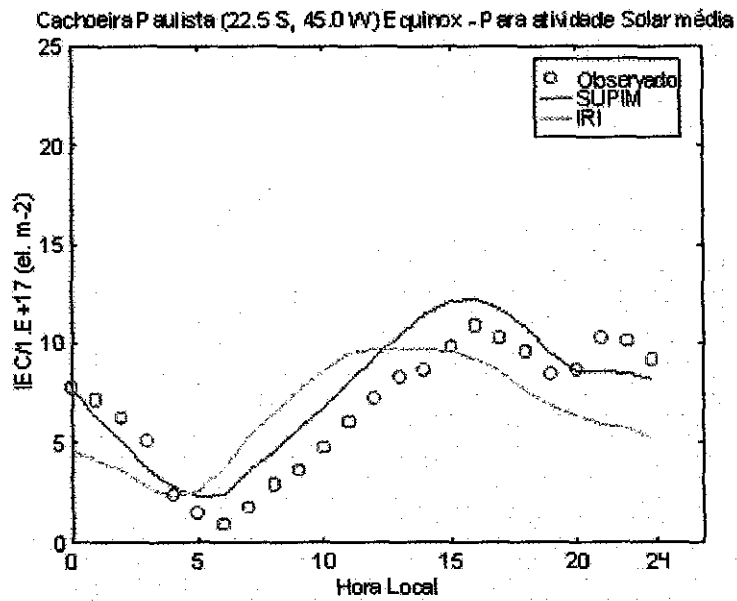


Figura C.8 - Caso #7- Atividade solar média.

Cachoeira Paulista (22.5 S, 45.0 W) Equinox - Para atividade Solar máxima

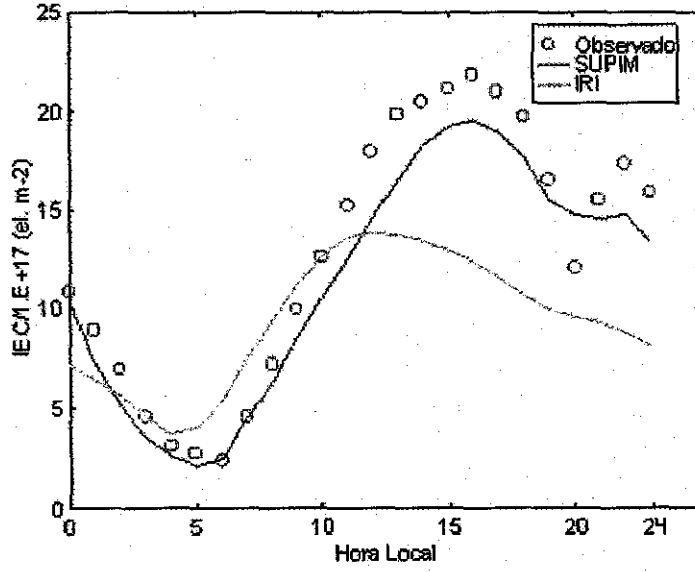


Figura C.9 - Caso #7 - Atividade solar máxima



IONOSPHERIC ELECTRON CONTENT OVER BRAZILIAN LOW LATITUDE AND ITS COMPARISON WITH THE IRI AND SUPIM MODELS

E. R. de Paula,* J. R. de Souza,* M. A. Abdu,* G. J. Bailey,**
I. S. Batista,* J. A. Bittencourt* and E. Bonelli***

* *Instituto Nacional de Pesquisas Espaciais - INPE, C.P. 515 - 12201-970,
São José dos Campos, SP, Brazil*

** *Department of Applied and Computational Mathematics, University of
Sheffield, Sheffield S10 2TN, U.K.*

*** *Universidade Federal do Rio Grande do Norte - UFRN, 59075-350,
DFTE, Natal, RN, Brazil*

ABSTRACT

Ionospheric Electron Content (IEC), measured at the Brazilian low latitude station of Cachoeira Paulista (22.5°S, 45°W, dip latitude 14°S) was compared with IEC calculated using the IRI-90 and Sheffield University Plasmaphere-Ionosphere (SUPIM) models for a large range of solar fluxes at 10.7 cm (F10.7). The analysis of the F10.7 influence on IEC showed that, for Cachoeira Paulista (CP) at 05 LT (around the diurnal minimum), there is a good agreement between the IEC measurements and the models. At 16 LT (around the diurnal maximum), substantial discrepancies were observed in the IRI-90 results while SUPIM presents a good agreement up to F10.7 ≈ 200. The measured IEC diurnal variation is better represented by SUPIM, mainly during higher F10.7. At low latitudes two critical parametric inputs for the SUPIM mathematical model are the equatorial ExB plasma drifts and thermospheric neutral winds. Therefore, updated values of these parameters should be introduced in the model calculations everytime they are available. At CP, a station located close to the southern Appleton anomaly peak due to the fountain effect, where large discrepancies between IRI-90 and observations were found mainly with high F10.7, this model should include adequately this effect, which we believe is the main reason for the discrepancies.

INTRODUCTION

The ionospheric electron content (IEC) is of considerable importance to study the upper atmosphere as well as for practical applications like space geodesy, satellite and space probe navigation. In the southern hemisphere there are few IEC measurements, so in many cases ionospheric models are used to calculate this quantity. Validation of numerical and theoretical ionospheric models is necessary to improve them.

The International Reference Ionosphere (IRI) /1/ is an empirical global reference ionospheric model which has been continuously updated /2, 3/. The Sheffield University Plasmaphere-Ionosphere Model (SUPIM) /4/ is a fully time-dependent mathematical model used in this work to calculate the IEC in order to compare it with the IRI-90 and measured IECs.

The purpose of this work is to test the reliability of the IRI-90 and SUPIM in predicting IEC at a low latitude station.

EXPERIMENTAL DATA AND MODEL CALCULATIONS

The Faraday rotation measured by VHF polarimeter during 1982, 1984, 1986, 1989, when F10.7 varied from 66 to 330, was used to calculate the IEC between CP and the GOES I, II and III satellites ($\sim 90^\circ\text{W}$) using the method described by de Souza /5/ which showed good agreement with the EIS *et al.* /6/ method. Only the days with $\Sigma Kp \leq 24$ were included in this study, and the data showed large variability for a given solar flux value. An example of this variability for 05 and 16 LT was presented by Batista *et al.* /7/.

The IRI-90 and SUPIM IEC's were calculated integrating the electron densities in the subionospheric points for each 20 km altitude steps along the line-of-sight from ground station to the satellite, however, for SUPIM we considered that the satellite was over the geographic equator in the same altitude as GOES and in the magnetic line passing over CP. We used this approximation for the SUPIM model, which simplifies the calculations, because the latitudinal electron density gradients are substantially larger than zonal gradients for this region. This oblique integration, used here to calculate IEC, gives reliable results that can be directly compared with the observations, with no need to convert the slant IEC in the vertical direction as was made by Batista *et al.* /7/.

SUPIM model /4/ was run with F10.7 values of 75, 125, 200 and 300. Two parametric entries for SUPIM are the thermospheric neutral wind /8/ and the equatorial vertical plasma drift /9/. As this drift presents longitudinal variation /10,11/ during the prereversal hours (around 18-22 LT), this parameter was represented by seasonal averages from the Jicamarca incoherent scatter radar /9/ combined with drifts inferred from Fortaleza (4°S , 38°W , magnetic dip 2°S) ionograms at 18-22 LT.

RESULTS

Figure 1 presents the IRI-90, SUPIM and observed IEC variation with solar flux during equinox (March, April, September and October) at CP at 5 and 16 LT. The IEC observed values were averaged for days with F10.7 around 75, 125, 200 and 300. At 5 LT the observed IEC agrees quite well with the IRI-90 and the SUPIM results. At 16 LT the SUPIM results are in reasonable agreement with the data but do not follow the saturation effect for F10.7 larger than about 200 that is seen in the data /7, 12/ and in IRI. IRI shows substantial discrepancies in absolute value especially at high solar activities.

Figure 2 shows the IRI-90, SUPIM and observed IEC diurnal variation during equinox for three F10.7 intervals at CP. For F10.7 < 100 SUPIM follows the observed diurnal variation but underestimates IEC between 9 and 21 LT, while IRI-90 overestimates the measured values from about 4 to 14 LT. For the other two higher F10.7 intervals the measured IEC diurnal variation is better reproduced by SUPIM during diurnal and prereversal (around 22 LT) peaks, while IRI-90 underestimates substantially the diurnal peak and does not reproduce the prereversal peak.

DISCUSSION

At CP, a station located close to the Appleton anomaly peak, there is good agreement between the IRI-90 results and the observed IEC for low levels of solar activity. Increased discrepancies are found with increasing F10.7. IRI-90 does not reproduce the observed prereversal IEC peak.

The SUPIM results agree well with the observations for F10.7 increasing up to 200. SUPIM overestimates the IEC for higher F10.7 values (see Figure 1 at 16 LT).

Although there are few works comparing IRI and observed IEC at low latitudes some of them /7, 13/ showed that IRI strongly underestimates the daytime IEC. Bittencourt and Chryssafidis /14, 15/ showed that IRI prediction discrepancies from measured electron density profiles, during high solar activity, are mainly due to the effects of the low latitude ExB plasma drifts and thermospheric neutral winds that are not correctly taken into account in the IRI profiles. It should be also kept in mind that the parametrization of these two dynamical parameters strongly affects the SUPIM model results.

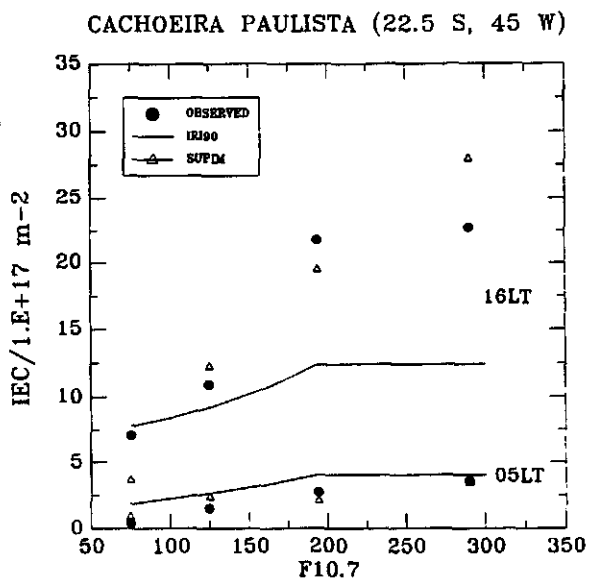


Fig. 1 - IEC variation with solar flux for equinox

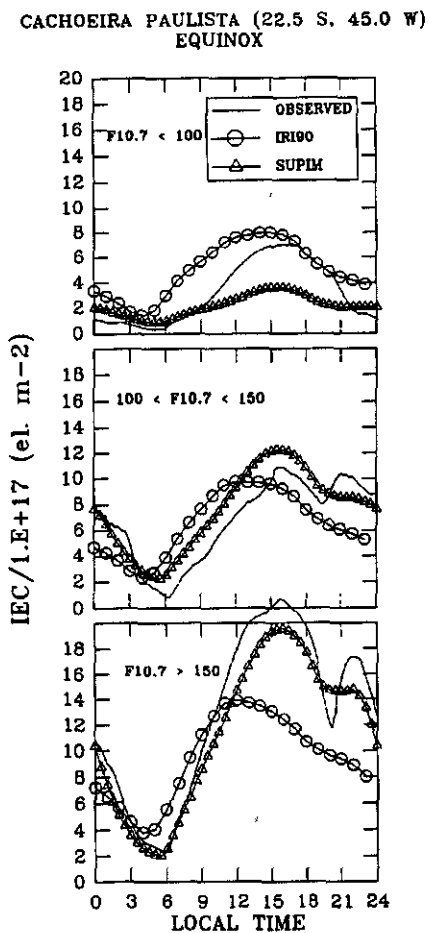


Fig. 2 - IEC diurnal variation for three solar activity levels

Efforts have been made to improve the neutral wind models /8/ and the equatorial plasma drifts /9, 10/, which should be updated and included in the mathematical models and their effects should also be considered in the empirical models, in order to get more realistic results.

REFERENCES

1. K. Rawer, D. Bilitza and S. Ramakrishnan, Goals and status of the International Reference Ionosphere, *Rev. Geophys. Space Phys.* 16, 177-181 (1978).
2. D. Bilitza, International Reference Ionosphere: Recent developments, *Radio Sci.* 21, 343-346 (1986).
3. D. Bilitza, International Reference Ionosphere 1990, *NSSDC/WDC-A-R&S, 90-22*, National Space Science Data Center, Greenbelt, Maryland (1990).
4. J. Bailey, R. Sellek, Y. Rippeth, A modelling study of the equatorial topside ionosphere, *An. Geophys.* 11, 267-272 (1993).
5. J.R. de Souza, Total Ionospheric Electron Content Studies, Master Dissertation (in portuguese), INPE-5417-TDI/481, INPE, São José dos Campos, São Paulo, Brasil, 1992.
6. K.E. Eis, J.A. Klobuchar, C. Malik, On the Installation, Operation, Data Reduction and Maintenance of VHF Electronic Polarimeters for Total Electron Content Measurements, Massachusetts, Hanscom AFB, AFGL (AFGL-TR-77-0130), 1977.
7. I.S. Batista, J.R. de Souza, M.A. Abdu and E.R. de Paula, Total electron content at low latitudes and its comparison with the IRI 90, *Adv. Space Res.* 14 # 12, 87-90 (1994).
8. A.E. Hedin, M.A. Biondi, R.G. Burnside, G. Hernandez, R.M. Johnson, T.L. Killen, C. Mazaudier, J.W. Meriwether, J.E. Salah, R.J. Sica, R.W. Smith, N.W. Spencer, V.B. Vickwar and T.S. Virdi, Revised global model of thermosphere winds using satellite and ground-based observations, *J. Geophys. Res.* 96, 7657-7688 (1991).
9. B.G. Fejer, E.R. de Paula, S.A. González and R.F. Woodman, Average vertical and zonal F region plasma drifts over Jicamarca, *J. Geophys. Res.* 96, 13901-13906 (1991).
10. B.G. Fejer, E.R. de Paula, R.A. Heelis and W.B. Hanson, Global equatorial ionospheric vertical plasma drifts measured by the AE-E satellite, *J. Geophys. Res.*, 100, 5769-5776 (1995).
11. I.S. Batista, M.A. Abdu and J.A. Bittencourt, Equatorial F-region vertical plasma drifts: Seasonal and longitudinal asymmetries in the American sector, *J. Geophys. Res.* 91, 12055-12064 (1986).
12. N. Balan, G.J. Bailey and B. Jayachandran, Ionospheric evidence for a non-linear relationship between the solar EUV and 10.7 cm fluxes during an intense solar cycle, *Plan. Space Sci.*, 41, 141-145 (1992).
13. L.F. McNamara, Predictions of total electron content using the International Reference Ionosphere, *Adv. Space Res.* 4, 25-50 (1984).
14. J.A. Bittencourt and M. Chryssafidis, Comparison of IRI model predictions with low latitude ionospheric observations, *Adv. Space Res.* 11 # 10, 97-100 (1991).
15. J.A. Bittencourt and M. Chryssafidis, On the IRI model predictions for the low-latitude ionosphere, *J. Atmos. Terr. Phys.*, 56, 995-1009 (1994).

Anexo D

Algoritmo de Box e Hill

Informações de ajuda "clicáveis"

Ajuda do Box e Hill

1 Algoritmo de Box e Hill (discriminação entre modelos competitivos)

O algoritmo de Box e Hill tem como objetivo principal permitir a determinação do Modelo mais provável que represente um sistema real, do qual se possuem e/ou se podem obter dados experimentais. Para tal, o processo de discriminação implementado permite a escolha entre modelos diferentes, usualmente não lineares, que estão competindo para representar o sistema em pauta.

A ferramenta para esta discriminação é o algoritmo de Box e Hill (1967). Este algoritmo associa probabilidades a cada um dos modelos que estão competindo entre si para representar um determinado sistema, bem como indica onde proceder um novo experimento para melhorar a discriminação entre os dois melhores, o que se obtém utilizando conceitos da Teoria da Informação (como a entropia) e da Estatística Bayesiana, de uma maneira sequencial.

Referência:

Box, G.E.; Hill, W.J. **Discrimination among mechanistic models. Technometrics**, 9(1):57-71, Feb. 1967.

2 Como usar o software

2.1 Maneiras de utilizar o software

Existem duas maneiras de utilizar este software, que são:

- Entrar com os modelos matemáticos.
- Entrar com os resultados dos modelos.

Para informar ao software a maneira que se deseja utilizá-lo deve-se entrar na opção de configuração do menu principal. Nesta opção aparecerão as seguintes opções de configuração.

- Probabilidade equiprovável.
- Entrada do resultado dos modelos.

2.2 Entrar com os modelos matemáticos

Quando a opção "Entrada do resultado dos modelos" não estiver selecionada será necessário entrar com os modelos matemáticos.

O usuário deverá escolher no menu de entradas a opção "modelos". Nesta opção aparecerão, para serem preenchidos, os seguintes campos :

"Modelo : " modelo matemático que descreve o fenômeno.

"Probabilidade a priori : " probabilidade associada ao modelo que está sendo introduzido. Este campo só será preenchido se no menu de configuração a opção de "Probabilidade equiprovável" não estiver selecionada.

"Modelo Atual : " Informação do sistema sobre o número do modelo atual.

Após o preenchimento de cada campo acima o usuário irá clicar no botão "Próximo" para inserir o modelo seguinte, e após inserir o último modelo deverá clicar o botão "finalizar".

Após a entrada dos modelos deverão ser informados os valores dos parâmetros e das variáveis de cada modelo digitado. Finalmente deve-se clicar no menu "Processamento" que executará o algoritmo de Box e Hill e determinará qual dos modelos é o mais provável.

2.2.1 Regras de sintaxe do modelo.

A utilização de expressões algébricas necessita de 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 a sua redigitação.

Regras de aceitação:

- 1 - Toda expressão deve possuir algum elemento, não sendo aceitas expressões vazias.
- 2 - Não são admitidos espaços em branco entre os componentes da expressão.
- 3 - Quando algum parêntese for aberto deve-se fechá-lo.
- 4 - A expressão deve ser matematicamente válida.
- 5 - Um modelo pode conter apenas caracteres alfabéticos, números e símbolos de operações aritméticas. Qualquer símbolo fora deste domínio é considerado erro.

Exemplo de modelo:

$$1/(1+A*x_i*\exp(-B/x_{ii}))$$

O analisador sintático considera como parâmetros qualquer sequência de caracteres alfabéticos e maiúsculos (A, B, TETA, BETA), enquanto as variáveis são sequências de minúsculos (x, var, t, z), admite-se também constantes numéricas e funções (sen, cos, exp,...).

2.3 Entrada dos resultados dos modelos

Quando a opção "Entrada do resultado dos modelos" estiver selecionada não será necessário a entrada dos modelos matemáticos. Mas será necessário informar ao sistema o nome do arquivo que contém os resultados experimentais e os resultados obtidos com cada modelo.

O usuário deverá escolher no menu de entradas a opção "modelos". Nesta opção aparecerão, para serem preenchidos, os seguintes campos :

"Arquivo:" nome e caminho do arquivo onde estão os pontos experimentais e os pontos dos modelos.

"Número de modelos:" a quantidade de modelos que estão competindo.

"Número de pontos experimentais:" quantidade de pontos experimentais e observacionais.

"Variância do erro experimental:" variância associada às medidas.

Após o preenchimento dos campos acima deve-se clicar o botão "finalizar" e depois o menu "Processamento", para subsequente determinação do modelo mais provável.

2.4 Opção de Probabilidade equiprovável

Caso esta opção esteja selecionada a probabilidade a priori associada a cada modelo será $1/M$ onde M é o número de modelos.

Se a opção não for selecionada deverá ser informada a probabilidade a priori associada a cada modelo.

2.5 Entrada do resultado dos modelos

Caso a opção "Entrada do resultado dos modelos" esteja selecionada não será necessária a entrada dos modelos matemáticos, porque o software lerá um arquivo que contém os valores obtidos experimentalmente e os obtidos com cada modelo, devendo-se informar ao sistema o nome e o caminho deste arquivo.

Caso a opção "Entrada do resultado dos modelos" não esteja selecionada será necessário entrar com os modelos matemáticos que representam o fenômeno.

3 Exemplos

3.1 Exemplo de entrada com os modelos matemáticos

Após a configuração do software para a opção de entrada dos modelos matemáticos, deve-se digitar as expressões destes seguindo as regras de sintaxe estabelecidas.

O primeiro passo para a introdução dos modelos é clicar no menu de "Entrada" e escolher "Modelos"; então, aparecerá uma tela com os seguintes campos:

Modelo : onde será digitada a expressão matemática do modelo.

Probabilidade a Priori: onde será digitada a probabilidade a priori associada ao modelo. Este campo aceitará entradas se no menu de configuração não estiver selecionada a opção "probabilidade equiprovável".

Modelo Atual: onde o software informa a quantidade de modelos inseridos até o momento.

Depois da digitação do modelo deve-se clicar no botão "Próximo" para inserir mais um modelo, ou então, no botão "finalizar" para terminar a entrada de modelos. Para que se possa editar modelos já digitados deve-se clicar no botão "Anterior", sendo então os modelos editados um-a-um na ordem inversa em que foram digitados. O botão "Cancelar" cancela as entradas de todos os modelos.

3.2 Exemplo de entrada dos resultados dos modelos.

No Exemplo abaixo observa-se a estrutura do arquivo que deve ser informado na opção "MODELOS" do menu "Entrada". Note-se que no arquivo deve haver apenas um espaço em branco entre cada coluna, e nenhum espaço entre as linhas.

A primeira coluna (da esquerda para a direita) contém os dados observados ou obtidos experimentalmente. A segunda coluna contém os resultados do primeiro modelo, a terceira coluna contém os resultados do segundo modelo e a quarta coluna contém os resultados do terceiro modelo. Então se Houver M modelos para competir, o arquivo deverá conter M+1 colunas. E os resultados do modelo M são colocados na coluna M+1.

0.90	2.54	3.71	0.66
4.90	4.49	5.7	4.02
5.59	5.29	6.23	5.3
5.59	5.4	6.27	5.37
4.90	5.61	6.12	5.00
4.00	5.45	5.70	4.23
3.2	5.13	5.22	3.47
2.20	4.77	4.75	2.84
1.60	4.07	3.96	1.95
0.90	2.81	2.72	1.18

Após o processamento obtém-se as seguintes probabilidades para cada um dos modelos.

modelo 1: 0.2425
modelo 2: 0.2326
modelo 3: 0.5249

Conclui-se que o modelo 3 é o mais provável, com probabilidade em torno de 0.5, frente aos modelos analíticos 1 e 2.