

**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 Weigang(1997), levando aos resultados mostrados na Tabela 5, referentes aos meses de março a junho de 1995.

**Tabela 5 - Caso #5**

<b>Obs</b>	<b>Mod1</b>	<b>Mod2</b>	<b>Mod3</b>	<b>Mod4</b>
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 Weigang 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**

---

# 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 alli (1994) and Holtslag-Moeng (1991).

With the utilization of the observations of ground-level concentrations, collected in Copenhagen (Gryning et alli, 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 obtêm-se a concentração superficial de contaminantes emitidos na CLC. Pelo emprego de observações de concentrações superficiais realizadas na cidade de Copenhagen (Gryning et alli, 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 obtêm-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:

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

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

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

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

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:

$$c_n^y(x, z) = \sum_{k=1}^8 A_k \cdot \frac{P_k}{x} \cdot \left[ A_n e^{-\sqrt{\frac{P_k U}{x K_n}}} + B_n e^{\sqrt{\frac{P_k U}{x K_n}}} + \frac{1}{2} \frac{Q}{\sqrt{\frac{P_k K_n U}{x}}} \left( e^{-\sqrt{\frac{P_k U}{x K_n}}} - e^{\sqrt{\frac{P_k U}{x K_n}}} \right) \right] \quad (2.6)$$

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 alii (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*.

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

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

### 3 Um Modelo Gaussiano Atualizado

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

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

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

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

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

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

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

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

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:



$$\frac{\sigma_a^2}{z_i^2} = \frac{15(z/z_i)^2}{16\pi(f_m)_i^2 q^2} \int_0^\infty \frac{\text{sen}^2 \left\{ \frac{4\pi(0.98c)^{1/2} [(f_m)_i^2 q^2 \psi]^{1/3} Xn'}{\sqrt{\pi} 1.5(z/z_i)^{2/3}} \right\}}{(1+n')^{5/3} n'^2} dn' \quad (3.5)$$

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  $\sigma'_z$  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:

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

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:

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

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:

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

## 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 ( $10^{-4}\text{sm}^{-2}$ )	Gaussiano ( $10^{-4}\text{sm}^{-2}$ )	mod. (2.6), (2.7) ( $10^{-4}\text{sm}^{-2}$ )	mod. (2.6), (2.8) ( $10^{-4}\text{sm}^{-2}$ )
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 alii, 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).

## 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  $z_c$  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 selecionadas 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

$$C_i(z, x) = Q \delta(z - z_i) \quad x = 0 \quad (6)$$

$$K \frac{\partial C_i}{\partial z} = 0 \quad z = 0, z = z_i, \quad (7)$$

onde  $C_i$  é a concentração integrada normal a direção do vento.

As variáveis tornam-se adimensionais como segue

$$C_i = (C_i z_i u) / Q \quad u = u / u \quad z = z / z_i \quad K = K' / (w_i z_i) \quad z_i = z_i / z_i \quad x = (x' w_i) / (z_i u)$$

onde as linhas representam variáveis dimensionais.

A equação (5) é integrada pelo método de Crank-Nicholson e as magnitudes selecionadas para  $u$ ,  $w$ , e  $z_i$  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,  $C_i(x, z = 0)$ , é possível calcular a concentração na superfície,  $C(x, y, 0)$ , em qualquer ponto, utilizando um modelo Gaussiano de concentrações laterais

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

onde  $y$  é distância lateral adimensional ( $y = y' / z_i$ ) e  $\sigma_y$  é o parâmetro de dispersão lateral adimensional

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

## 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

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

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

$$K_C = \frac{(1 - z/z_i + R_C z/z_i) K_b K_t}{(1 - z/z_i) K_t + R_C (z/z_i) K_b}, \quad (10)$$

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*

$$\frac{K_b}{w_* z_i} = \left(\frac{z}{z_i}\right)^{4/3} \left(1 - \frac{z}{z_i}\right)^2 \quad \frac{K_t}{w_* z_i} = 7 \left(\frac{z}{z_i}\right)^2 \left(1 - \frac{z}{z_i}\right)^3.$$

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.

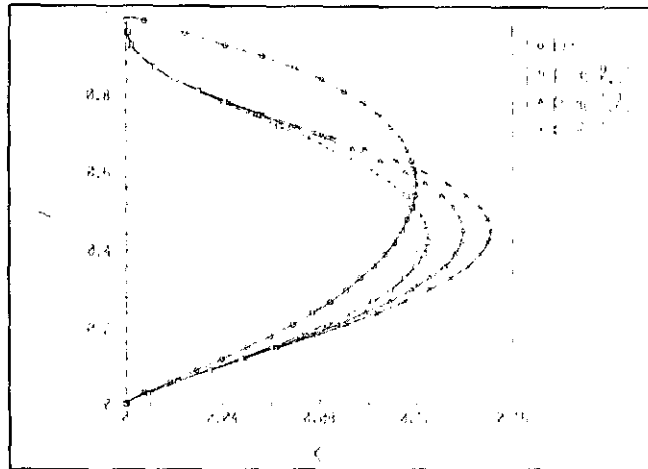


Figura 1 - Perfis verticais dos coeficientes de difusão (2)  $-K_{zz}$  e (2)  $-K_c$ .

## 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  $z_s = 0.025 z_c$ . 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,  $C_1(x,0)$ , para uma fonte linha localizada em  $z_s = 0.25 z_c$ . 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  $C_1$  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).

Holtslag, 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).

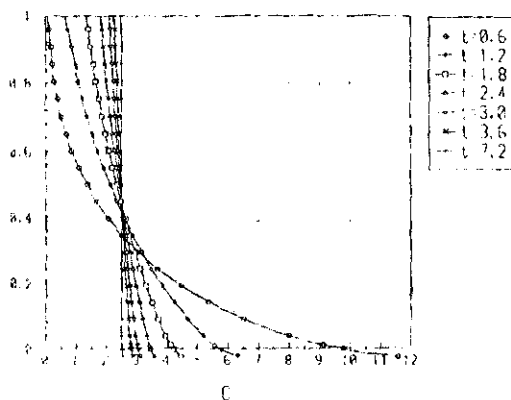


Figura 2 Perfil de concentração para diferentes tempos obtido por (2)- $K_u$ .



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

Tabela 1

x	(5)- $K_{u,}$	(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

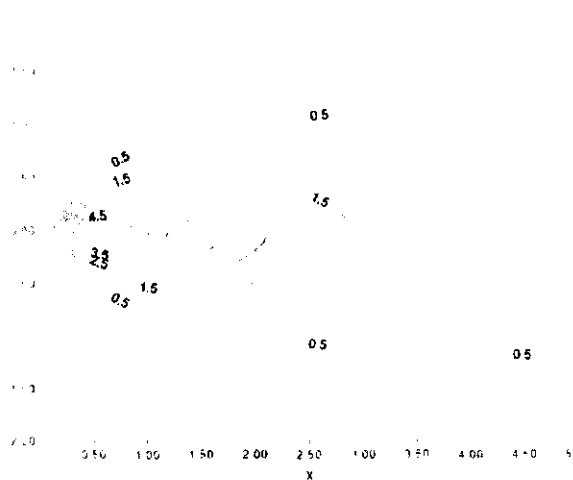


Figura 4 Isolinhas de concentração superficial no plano x-y obtidas pela expressão (8) com (5)-  $K_{//}$

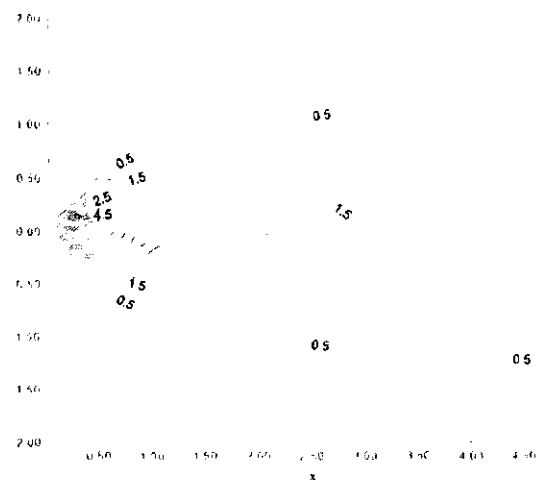


Figura 5 Isolinhas de concentração superficial no plano x-y obtidas pela expressão (8) com (5)-  $K_{\perp}$



# 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 km<sup>2</sup>, 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-PNUID, 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.

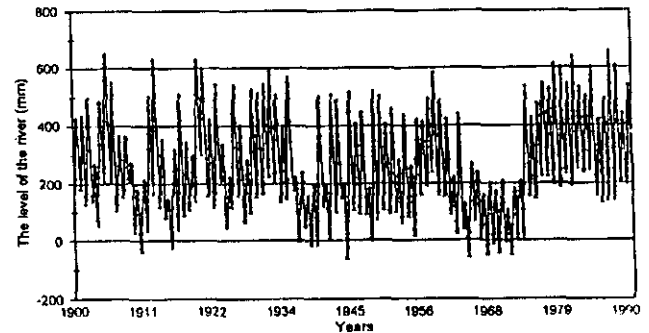


Figure 1 - The observed monthly means of the level of Paraguay River.

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.

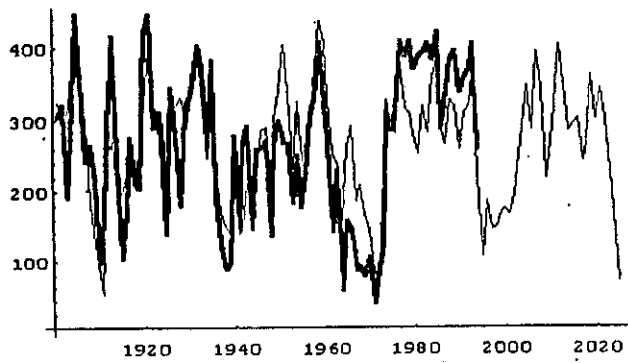


Figure 2 - Yearly means of the level of the Paraguay River at Ladário 1900 to 1994 (thick curve) and prediction based on 9 most important periodicities up to 2025 (thin curve). A rather good fit may be observed between observed and reconstructed values except in the case of several periods of flood or 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  $\Psi$  is the same for all units. Here,  $\Psi$  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  $\beta_j$ , 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.

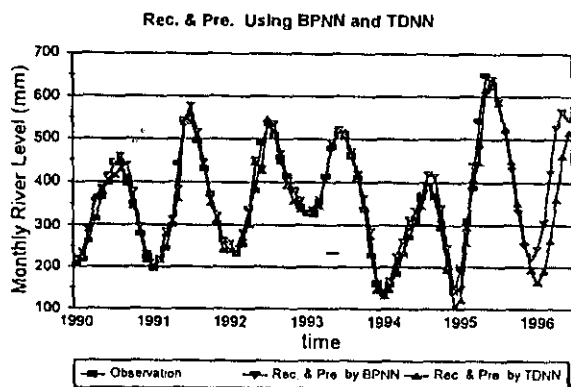


Figure 3 - Reconstruction (01/90-06/95) & prediction (07/95-06/96) of the monthly level of Paraguay River

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

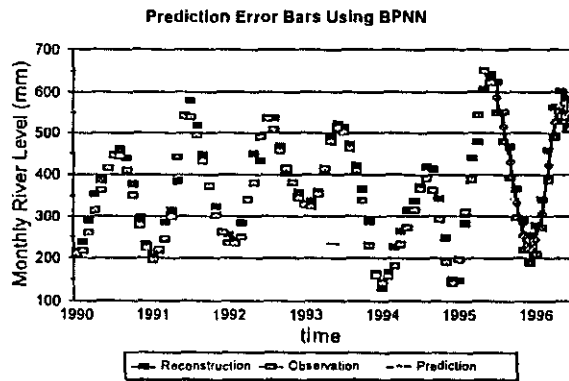


Figure 4 - Prediction error bars (06/95-07/96) using BPNN.  
Figura 4 - Barras de erro das previsões (06/95-07/96) utilizando BPNN.

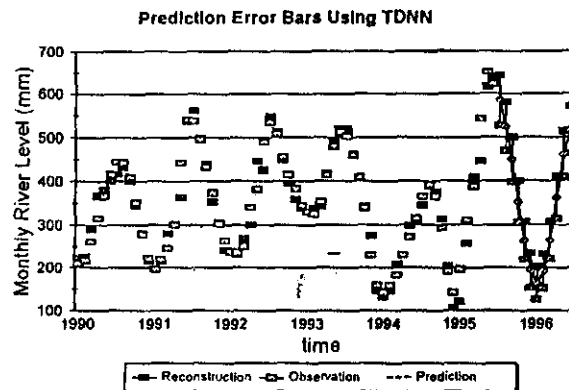


Figure 5 - Prediction error bars (06/95-07/96) using TDNN.  
Figura 5 - Barras de erro das previsões (06/95-07/96) utilizando TDNN.

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	nall1	nmse2	nall2
Backpropagation (BPNN)	2 hours	0.1114	9.5082	0.0584	8.5498
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 March/95 April/95 May/95 June/95  
(Months)

→ 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 1996—2001 2002—2023  
(Years)

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. I. 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 da 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. I.
- 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., 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.
- WOLFRAMS. - 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: 12/03/95

Revisado pelo(s) autor(es) em: 04/06/95

Aceito em: 10/06/95

## ESTUDO E PREVISÃO DO NÍVEL DO RIO PARAGUAI USANDO ANÁLISE HARMÔNICA E REDES NEURAIS

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 logarítmica 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)