

REDES NEURAIIS ARTIFICIAIS: UMA APLICAÇÃO NO ESTUDO DA QUALIDADE DO AR

SOUZA, Amaury

Físico, Mestre em Meteorologia e Doutor em Tecnologias Ambientais

PAVÃO, Hamilton Germano

Físico, Mestre em Meteorologia e Doutor em Tecnologias Ambientais

OLIVEIRA, Ana Paula Garcia

Bióloga, Mestre e Doutora em Tecnologias Ambientais

Universidade Federal de Mato Grosso do Sul, Programa de Pós-graduação em Tecnologias Ambientais, caixa postal 549, CEP-79090-900, Campo Grande, Brasil.

RESUMO: A estimativa da concentração do ozônio de superfície propicia a geração de dados para o planejamento de previsão da qualidade do ar, útil na gestão de saúde pública. O objetivo deste trabalho foi elaborar uma Rede Neural Artificial (RNAs) para estimar a concentração do ozônio de superfície em função de dados diários de clima. A RNA, do tipo FeedForward Multilayer Perceptron, foi treinada tomando-se por referência da concentração diária do ozônio medida. Nas camadas intermediárias e de saída foram utilizadas funções de ativação do tipo tan-sigmóide e lineares, respectivamente. O desempenho da RNA desenvolvida foi muito bom, podendo-se considerá-la como integrante do conjunto de métodos indiretos para estimativa da concentração do ozônio de superfície. O modelo proposto pode ser utilizado pelo governo público como ferramenta para ativar ações de ferramentas durante os períodos de estagnação atmosférica, quando os níveis de ozônio na atmosfera possam representar riscos à saúde pública.

Palavras-chave: redes neurais, clima, ozônio de superfície, saúde pública, qualidade do ar.

ARTIFICIAL NEURAL NETWORKS: AN APPLICATION IN THE STUDY OF AIR QUALITY

ABSTRACT: The estimation of the concentration of surface ozone promotes the generation of data for planning forecasting air quality, useful in the management of public health. The aim of this study was to develop an Artificial Neural Network (ANN) to estimate the concentration of surface ozone due to climate data daily. The RNA, feedforward Multilayer Perceptron was trained taking as reference the daily concentration of ozone measured. In the hidden layers and output functions were activation type tan-sigmoid and linear, respectively. The performance of the ANN developed was very good, and we can consider it as part of the set of indirect methods to estimate the concentration of surface ozone. The proposed model can be used by the government as a tool to enable the public actions of tools during the period of atmospheric stagnation, when ozone levels in the atmosphere may pose risks to public health.

Keywords: neural networks, climate, surface ozone, public health, air quality.

1. INTRODUÇÃO

A cada dia aumenta a preocupação do homem com a degradação ambiental, que se expressa através de mudanças climáticas, poluição do ar, contaminação das águas e do solo, com consequências desastrosas para a fauna e a flora, causando doenças e afetando negativamente a qualidade da vida humana (BATES *et al.*, 2008).

A poluição do ar tem sido apontada como um dos principais responsáveis por doenças relacionadas ao sistema respiratório, especialmente em crianças, idosos e pessoas com problemas respiratórios. A exposição aos poluentes pode provocar ou agravar doenças tais como asma, bronquite crônica, enfisema pulmonar, infecções pulmonares, rinofaringites e irritação nas vias respiratórias, entre outras, conduzindo a um elevado número de internações hospitalares, óbitos e aumentando a procura de atendimento nas salas de emergência (SINGER *et al.*, 2002; ŠRÁM *et al.*, 2005).

Vários episódios têm sido registrados em grandes centros urbanos em virtude das altas concentrações de poluentes na atmosfera, com trágicas consequências para a população

(BRAGA *et al.*, 2002; CONCEIÇÃO *et al.*, 2002). A preocupação causada pelos efeitos decorrentes da degradação do ar tem sido motivo de discussão entre líderes de diversos países, levando-os a firmarem acordos para controle e redução da emissão de gases de efeito estufa (UNFCCC, 1992; UNFCCC, 2007).

A poluição atmosférica é um problema de saúde pública, constituindo um grande desafio à gestão. A compreensão dos seus efeitos pode contribuir para o planejamento e nortear a adoção de medidas públicas visando proteger a população. O conhecimento científico é de grande utilidade no auxílio à gestão da informação relativa à saúde coletiva, subsidiando a elaboração de medidas voltadas à prevenção e à redução da poluição do ar.

O estudo foi realizado no município de Campo Grande, considerado de médio porte, com intensa atividade de queimadas. Esse município localiza-se no centro oeste brasileiro, com população de aproximadamente 724.000 habitantes, distribuídos numa área de 8.096 km², além de alto grau de desenvolvimento econômico, destacando-se a importância das agroindústrias.

Este trabalho tem como objetivo propor a utilização das redes neurais artificiais (HAYKIN, 2008; PEREIRA & RODRIGUES, 1998), como metodologia alternativa, no estudo da concentração de ozônio de superfície em função de dados meteorológicos.

2. BASE DE DADOS E MÉTODOS

Para estimar perfis da concentração de ozônio de superfície foi utilizada uma RNA, utilizou-se um conjunto de dados climáticos diários, medidos e armazenados através de uma estação meteorológica automática localizada em Campo Grande, MS (S 20°26'16" W 54°32'16" – 677 metros de altitude). A base de informação é composta por 10 variáveis explicativas (8 variáveis climáticas, 1 variável de número de veículos automotores e 1 de poluição atmosférica). Segundo a classificação climática de Köppen-Geiger, o clima da região de Campo Grande é classificado como tropical, com estação seca. Caracteriza-se por temperaturas elevadas 18°C a 28°C, com amplitude térmica de 5°C a 7°C e estações bem definidas, uma chuvosa e outra seca. Apresenta alto índice pluviométrico, em torno de 1500 mm/ano, com o período chuvoso começando em outubro e estendendo-se até abril e o período seco correspondendo aos demais meses do ano.

Foram armazenados dados diários entre os meses de janeiro de 2004 a dezembro de 2010, sendo o total dividido em um conjunto de treinamento (2/3) e um conjunto de teste (1/3). Para treinamento e validação dos resultados foram necessários dados observados de ozônio.

O programa para treinamento e teste da RNA foi desenvolvido no software Matlab (Mathworks, 2008). Visando o mapeamento desejado, foram testadas diversas topologias de redes do tipo FeedForward Multilayer Perceptron, com variações do número de neurônios das camadas intermediárias. Visto que, a temperatura do ar, umidade, precipitação, velocidade dos ventos e frota de veículos automotores é um dos principais fatores que influenciam na estimativa da concentração de ozônio, seus valores máximos, mínimos e médios foram utilizados como dados de entrada da RNA. Na camada intermediária foram usadas funções de ativação do tipo tan-sigmóide e na camada de saída foram usadas funções de ativação do tipo lineares, caracterizando esta rede neural como aproximador universal de funções. Foi feita a normalização dos dados, a depender do tipo de função de ativação na camada de saída da RNA, esse procedimento torna-se necessário. O software Matlab oferece duas formas de normalização dos dados no

intervalo [-1,1] e com média=0 e variância=1 e por fim os dados totais foram divididos em 2/3 consecutivos para treinamentos e 1/3 para validação.

Tendo em vista que, no início do treinamento, os parâmetros livres são gerados aleatoriamente e que esses valores iniciais podem influenciar no resultado final do treinamento, cada arquitetura de rede foi treinada dez vezes, sendo selecionado a que apresentou maior valor de coeficiente de determinação (r²). Esse coeficiente foi calculado a partir dos dados de observações da concentração de ozônio da amostra de teste e os respectivos valores estimados pela RNA.

Visando ao mapeamento desejado, foram treinadas diversas topologias de redes, variando-se o número de neurônios, funções de ativação nas camadas intermediárias, bem como o número de iterações (Tabela 1).

Tabela 1. Parâmetros testados no treinamento das RNAs

Parâmetro	Valores
Número de neurônios nas camadas intermediárias	1 a 5; 5 a 10 Logística Sigmóide; Tangente Hiperbólica
Funções de ativação nas camadas intermediárias	Sigmóide.
Número de ciclos	50; 100; 200; 500

Os valores do ozônio estimados pela RNA foram comparados com os calculados por, utilizando-se o erro percentual acumulado, o Erro Relativo Médio (ERM), o coeficiente de exatidão de Willmott (d) e o índice de desempenho (c).

O ERM foi calculado a partir da Equação 1.

$$ERM = \sum_{i=1}^n \frac{|O - E|}{O} \times \frac{1}{n} \quad (1)$$

Segundo CAMARGO & SENTELHAS (1997), os seguintes indicadores estatísticos são considerados para correlacionar os valores estimados com os medidos: exatidão - índice de Willmott "d"; e de confiança ou desempenho "c". A exatidão, relacionada ao afastamento dos valores estimados em relação aos observados, é dada estatisticamente pelo índice de concordância d proposto por WILLMOTT *et al.* (1985). Seus valores variam de zero, para nenhuma concordância, a 1, para a concordância perfeita. O índice é dado pela Equação 2:

$$d = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (|P_i - O| - |O_i - O|)^2} \quad (2)$$

sendo: P_i = valor estimado; O_i = valor observado; O = média dos valores observados.

O índice de desempenho c, apresentado por CAMARGO & SENTELHAS (1997), avalia o desempenho dos diferentes métodos de estimativa. Este índice reúne os índices de precisão, dado pelo coeficiente de correlação (r) que indica o grau de dispersão dos

dados obtidos em relação à média, ou seja, o erro aleatório e o de concordância d. O índice c é calculado conforme Equação 3.

$$C = r.d \quad (3)$$

CAMARGO & SENTELHAS (1997) propuseram um critério para interpretar o desempenho dos métodos de estimativa pelo índice c, apresentado na Tabela 1.

Após o desenvolvimento do algoritmo de treinamento da RNA e a realização de análises dos dados climáticos disponíveis e dos algoritmos de treinamento, foi obtida uma RNA capaz de estimar de modo satisfatório a concentração de ozônio de superfície. Esta estimativa é realizada por meio do mapeamento da relação entre dados de temperatura máxima, média e mínima, umidade relativa do ar máxima, média e mínima, velocidade dos ventos, precipitação e número de veículos automotores que foram tomados como entradas, e a concentração de ozônio de referência, que é a saída desejada.

Tabela 2- Critério de interpretação do desempenho de estimativa da concentração de ozônio de superfície.

Valor de c	desempenho
>0,85	ótimo
0,76 a 0,85	muito bom
0,66 a 0,75	bom
0,61 a 0,65	mediano
0,51 a 0,60	sofrível
0,41 a 0,50	mau
<0,40	péssimo

3. RESULTADOS E DISCUSSÃO

A RNA selecionada apresentou melhor desempenho com a mínima configuração possível. Essa configuração é composta de uma camada de entrada com três variáveis, duas camadas intermediárias cada uma com 4 e 2 neurônios artificiais, respectivamente, e um neurônio na camada de saída. Para os neurônios da camada intermediária foi adotada a função de ativação do tipo Tangente Hiperbólica Sigmóide. Em geral, as redes treinadas apresentaram melhores desempenhos com números menores de ciclos, sendo que a RNA selecionada atingiu maior eficiência em 200 ciclos. Além disso, verifica-se que redes com mais de 200 ciclos apresentam problemas de "memorização".

Tabela 2. Indicadores estatísticos do ajuste entre os valores observados do ozônio e os valores estimados pela RNA, erro relativo médio mensal, valores de c de janeiro de 2004 a dezembro de 2010.

MÊS	[O ₃] OBS	[O ₃] EST	ERM	C
JAN	10,32	13,07	-0,27	0,88
FEV	12,08	13,55	-0,12	0,88
MAR	13,28	14,34	-0,08	0,84
ABR	13,88	14,08	-0,01	0,83
MAI	14,66	11,40	0,22	0,73
JUN	14,93	15,54	-0,04	0,69
JUL	15,54	20,50	-0,32	0,56
AGO	24,29	24,57	-0,01	0,87
SET	29,69	26,30	0,11	0,86
OUT	26,79	21,19	0,21	0,79
NOV	21,20	19,84	0,06	0,82
DEZ	16,10	14,49	0,10	0,68

O valor médio anual foi $c=0,81$ com um desempenho muito ótimo e uma média mensal anual de desempenho igual a 0,79.

Nas Tabelas 2 apresentam-se, os valores do índice de desempenho (c) e do erro quadrático médio (ERM) para as RNA's. Menores valores de ERM associados a maiores valores de c indicam o desempenho da metodologia na estimativa da concentração do ozônio, a partir dos dados coletados.

As RNA's desenvolvidas, em geral, apresentaram bom desempenho, exceção feita no mes de Julho, que apresentaram índices estatísticos ERM de -0,32, apresentando valores de c com desempenho sofrível. As concentrações de ozônio apresentaram quatro meses que apresentaram desempenho ótimo, conforme ilustrado na tabela 2.

Nas RNA's seu desempenho foi muito bom principalmente devido à grande quantidade de dados utilizados em seu treinamento, o que facilita seu aprendizado. Também contribuíram para o desempenho muito bom o fato de terem sido testadas diferentes arquiteturas para as redes, isto é, diferentes números de camadas, algoritmos de aprendizagem, número de ciclos, etc.

Trabalhos como o de (MOREIRA *et al.*, 2005) e (MOREIRA *et al.*, 2007), avaliaram várias arquiteturas para as RNA's, obtendo desempenhos excepcionais. Ressalta-se que o número de ciclos utilizados no treinamento das RNA's foi alto, facilitando sua aprendizagem, reduzindo a possibilidade de ocorrência de memorização. A memorização leva as RNA's a apresentarem um bom desempenho estatístico (alto valor de c e baixo valor de ERM), pois este é calculado apenas com base na amostra dados disponíveis. Em contrapartida a memorização levaria a graves distorções na espacialização da concentração de ozônio extremamente altas ou extremamente baixas.

Analisando-se os valores com a utilização das RNA's (Tabela 2) verifica-se que não ocorreu memorização, pois não foram evidenciados severos desvios de concentração de ozônio estimada.

Analisando-se os dados da Tabela 2, verifica-se que as concentrações médias variaram entre 10,32 a 29,69 ppb (Tabela 2), com baixas nos meses janeiro, fevereiro março, estação chuvosa. As maiores valores foram evidenciadas nos meses de agosto, setembro, outubro, justificando limpeza da terra para o preparo do plantio.

Observa-se que altos valores de R^2 e d foram obtidos. Esses resultados foram comparados com os obtidos em outros estudos com previsões de concentrações diárias de ozônio (GRIVAS, CHALOULAKOU, 2006 (0,60 e 0,86); NAGENDRA, S. M. S & KHARE, M., 2005 (0,61 e 0,78)). Os valores médios anuais de R^2 e d deste estudo foram (0,8796 e 0,923798).

A Figura 1 exibe o gráfico que compara os valores observados e preditos pelo modelo na fase de validação. A Figura 2 apresenta os histogramas dos resíduos do modelo avaliado na fase de validação. Um bom modelo deve ter uma distribuição normal dos resíduos, ou seja, o histograma do resíduo deve ser simétrico, ter a forma de sino. Para visualizar a performance do modelo e da RNA, foram comparados os valores observados e as simulações, como ilustra a Figura 3. O gráfico mostra um bom ajuste do modelo aos dados observados, tanto na fase de estimação/treinamento quanto na fase de validação.

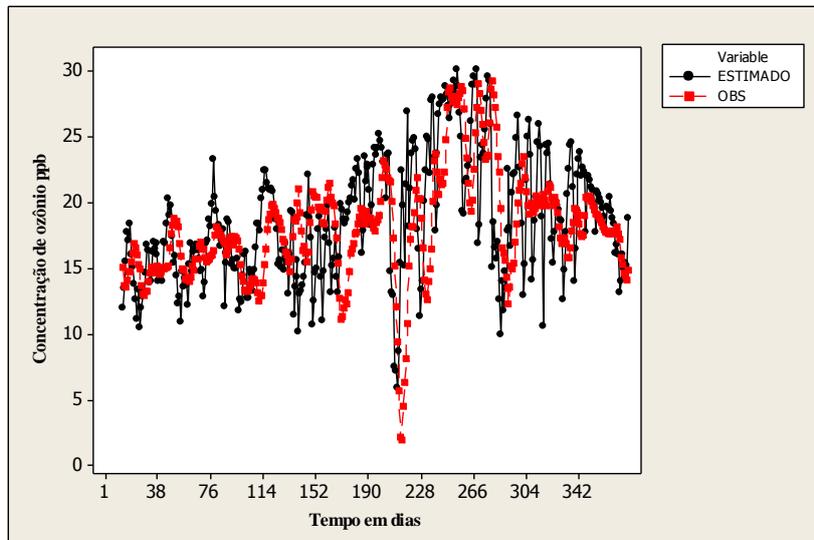


Figura 1- Concentração de ozônio observada e predita na fase de validação: RNA.

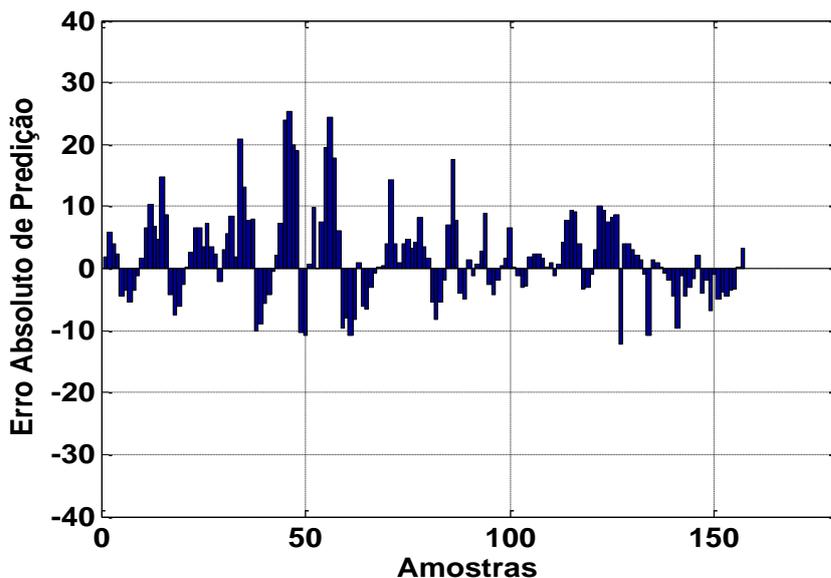


Figura 2 – Histograma dos erros de previsão da concentração de ozônio na fase de validação: RNA.

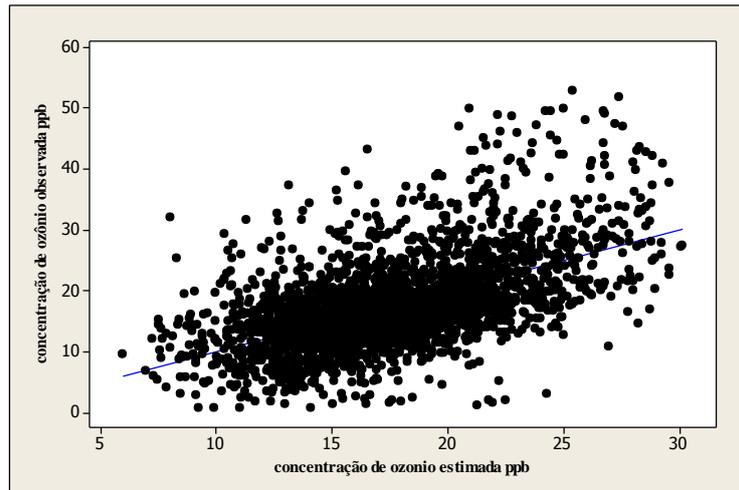


Figura 3 – Simulação dos valores observados-treinamento e validação.

CONCLUSÃO

O estudo de métodos para estimativa das concentrações do ozônio proporciona o comportamento médio dos parâmetros de estudo, em que pode ser úteis em trabalhos de prevenção da qualidade do ar, visando trabalho de modelagem. Com base nos resultados obtidos neste trabalho pode-se concluir que:

1. As RNA's desenvolvidas, na estimativa da concentração do ozônio, apresentaram um desempenho estatístico muito bom.
2. Existe necessidade de maior treinamento das RNA's e variação de sua arquitetura a fim de obterem-se melhores desempenhos estatísticos.
3. As RNA's desenvolvidas foram capazes de espacializar as concentrações do ozônio sem a presença de grandes desvios na estimativa do mesmo.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- BATES, B. C. *et al.* (Ed.). *Climate change and water. Technical paper of the Intergovernmental Panel on Climate Change*. IPCC Secretariat. Geneva, 2008. 210p.
Disponível em: <<http://www.ipcc.ch/pdf/technical-papers/climate-change-water-en.pdf>>. Acesso em: 19 mar. 2013.
- BRAGA, A.; PEREIRA, L. A. A.; SALDIVA, P. H. N. *Poluição atmosférica e seus efeitos na saúde humana*. In: **SUSTENTABILIDADE NA GERAÇÃO E USO DE ENERGIA NO BRASIL: OS PRÓXIMOS VINTE ANOS**. Campinas: UNICAMP, 2002.
- Chaloulakou A, Grivas G (2006) *Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece*. **Atmos Environ** 40:1216–1229.
- CAMARGO, A. P.; SENTELHAS, P. C. *Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no estado de São Paulo, Brasil*. **Revista Brasileira de Agrometeorologia**, v. 5, n. 1, p. 89-97, 1997.
- CONCEIÇÃO, G. M. S.; SALDIVA, P. H. N.; SINGER, J. M. *Associação entre séries de mortalidade / morbidade e concentrações de poluentes atmosféricos: uma estratégia para análise estatística*. **Rev. Bras. Estat.**, Rio de Janeiro, v.63, n.219, p.75-98, 2002.
- HAYKIN, S. **Neural networks and learning machines**. Prentice Hall, 2008.
- PEREIRA, B. B.; RODRIGUES, C. V. S. *Redes neurais em estatística*. In: **SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA - SINAPE**, 13., 1998, Caxambu. *Resumo...* p.13-20.

MOREIRA, M.C.; CECÍLIO, R.A.; SILVA JÚNIOR, J.L.C.; PINTO, F.A.C.; MEDEIROS, S.S. *Uso de redes neurais artificiais para a estimativa das normais das temperaturas médias mensais e anuais na região Nordeste*. In: Congresso Brasileiro de Agrometeorologia, 16, 2005, Campinas. **Anais**. Sociedade Brasileira de Agrometeorologia: Campinas, 2005. (CD-ROM) .

MOREIRA, M.C.; CECÍLIO, R.A.; SILVA, K.R. *Comparação de métodos para a estimativa das temperaturas do ar no Nordeste brasileiro*. In: Congresso Brasileiro de Agrometeorologia, 15, 2007, Aracaju. **Anais**. Aracaju : SB Agro, 2007. (CD-ROM).

NAGENDRA, S.M.S., KHARE, M. **Modelling urban air quality using artificial neural network. Clean Techn Environ Policy (2005) 7**: 116–126. DOI 10.1007/s10098-004-0267-6.

SINGER, J. M. et al. *Association between atmospheric pollution and mortality in São Paulo, Brazil: regression models and analysis strategy*. In: INTERNATIONAL CONFERENCE ON STATISTICAL DATA ANALYSIS BASED ON THE L1 NORM AND RELATED METHODS, 4., 2002, **Neuchâtel. Proceedings...** p.429-450.

ŠRÁM R. J. et al. *Intrauterine growth retardation, low birth weight, prematurity and infant mortality*. In: **OMS. Organização Mundial da Saúde. Special**. Effects of air pollution on children's health and development: a review of the evidence. Bonn: European Centre for Environment and Health, 2005. p.14-27.

UNFCCC - United Nations Framework Convention on Climate Change. 1992. Disponível em: <<http://unfccc.int/resource/docs/convkp/conveng.pdf>>. Acesso em: 12 junho. 2013.

UNFCCC - United Nations Framework Convention on Climate Change. *Kyoto protocol reference manual on accounting of emissions and assigned amounts*. 2007. Disponível em: <http://unfccc.int/files/national_reports/accounting_reporting_and_review_under_the_kyoto_protocol/application/pdf/rm_final.pdf>. Acesso em: 12 junho. 2013.

WILLMOTT, C. J.; DAVIS, R. E.; FEDDEMA, J. J.; JOHANNES, J.; KLINK, K. M.; LEGATES, D. R. et al. Statistics for the evaluation and comparison of models. **Journal of Geophysical Research Atmospheres**, v. 90, n. C5, p. 8995-9005, 1985.

<http://dx.doi.org/10.1029/JC090iC05p08995>.