 ISSN 1984-820X

Redes Neurais Artificiais na Melhoria de Desempenho de Métodos de Assimilação de Dados: Filtro de Kalman

Rosangela. S. Cintra, **Haroldo F. de Campos Velho**,
Laboratório de Computação e Matemática Aplicada, LAC,
Instituto Nacional de Pesquisas Espaciais, INPE
Av. dos Astronautas, 1758,
12.227-010, São José dos Campos, SP
E-mail: rosangela.cintra@lac.inpe.br, haroldo@lac.inpe.br,

Ricardo Todling
Global Modelling and Data Assimilation Office, GMAO,
NASA/Goddard Space Flight Center
27770, Greenbelt, MD/EUA
E-mail: todling@gmao.gsfc.nasa.gov

Resumo: *Assimilação de Dados é um método que combina dados de um modelo matemático e dados de observações, permitindo uma melhoria na previsão do modelo. Métodos seqüenciais ótimos são baseados em teoria de estimativa formal que minimiza os erros dos dados de acordo com a dinâmica do modelo. Métodos de assimilação de dados utilizando Redes Neurais Artificiais (RNA) vêm sendo propostos muito recentemente apresentando resultados consistentes: computacionalmente eficientes e eficazes quanto à aplicação. Este trabalho apresenta uma abordagem do método de assimilação por RNA, onde aplica-se uma RNA para substituir o cálculo da inversão de matrizes de erros constante do algoritmo de asssimilação baseado em filtro de Kalman. Para exemplo da aplicação desta abordagem, utilizou-se o Sistema de Lorenz e o Filtro de Kalman Estendido para obter parâmetros usados no treinamento da RNA e na comparação dos resultados.*

Palavras-chave: *assimilação de dados, redes neurais artificiais, sistemas não lineares*

1 Introdução

O problema de determinar as melhores condições iniciais para Previsão Numérica de Tempo (PNT) é de grande importância prática, e assunto de muitos estudos por pessoas de cenários diferentes. Em meteorologia e outros ramos de geofísica o processo de aproximar o ”verdadeiro” estado de um sistema físico em um determinado momento é chamado análise. Os modelos de PNT são simulações computacionais da atmosfera que tomam a análise como ponto de partida e desenvolvem o estado da atmosfera em um determinado instante de tempo, usando a compreensão de física e dinâmica do fluido atmosférico. A análise meteorológica é produzida pelo processo de assimilação de dados, onde a informação recebida da observação é usada em conjunto com a previsão mais recente de um modelo numérico no instante de tempo que a observação foi feita.

A qualidade da Previsão do Tempo é fruto da formulação dos modelos matemáticos e físicos e da quantidade e qualidade dos dados observados. A atmosfera se comporta como um sistema caótico, pois é muito sensível às pequenas variações nas condições iniciais. Por esta razão, existe a busca constante de obter a melhor análise, que é a melhor estimativa do estado atual da atmosfera para iniciar o ciclo de PNT. Para maior compreensão e maiores detalhes em assimilação de dados atmosféricos veja em [3].

Diferentes algoritmos de assimilação podem ser derivados de uma fonte comum, com aproximações próprias para cada algoritmo (Interpolação Ótima(OI), métodos variacionais (3D-VAR, 4D-VAR) e Filtro de Kalman). A Assimilação de dados pode ser descrita como um processo de dois passos:

Passo de previsão: $x_n^f = F[x_{n-1}^f]$

Passo de análise: $x_n^a = F[x_n^f] + d_n$

onde x_n representa o vetor de estado do modelo no passo de tempo t_n, $F[\cdot]$ é o modelo matemático de previsão, os sobrescritos f e a denotam respectivamente os valores preditos e de análise, finalmente d_n é o incremento da análise. O vetor incremento de análise é calculado como sendo um produto entre uma matriz de ponderação e uma função que mede a discrepância entre a previsão do modelo e as observações y^o:

$$d_n = M_n(y_n^o - x_n^f)$$

sendo M_n a matriz de ponderação ou matriz de ganho. Esta matriz pode ser calculada por um estimador de mínimos quadrados [3]. Para o sistema de análise, sabe-se que existem erros no modelo e nas observações. Os diversos métodos de assimilação existentes procuram uma estratégia que minimize a diferença entre a análise e a "verdade". Para projetar um algoritmo que faça isto automaticamente, é necessário representar matematicamente os erros dos estados, modelados estatisticamente usando conceitos probabilísticos. Assim, o algorítmo de análise pode ser formulado como um problema de otimização, onde se quer minimizar o erro médio quadrático.

O desafio computacional para a metodologia clássica de assimilação de dados reside na dimensão destas matrizes desenvolvidas em modelos de PNT, atualmente na ordem de um milhão de equações (o que equivale a matrizes cheias da ordem de 10^{12} elementos!). É neste cenário que se insere novas metodologias para assimilação de dados. A metodologia que utiliza Redes Neurais Artificiais(RNA) pode ser uma solução, veja em NOwosad [9], [10] e Harter [6], [5].

Com o enfoque na solução do algorítmo de análise formulado como um problema de otimização aplicou-se a técnica de Redes Neurais Artificiais (RNA) neste trabalho, desenvolvendo uma rede Perceptron de Múltiplas Camadas (PMC) com o algoritmo de treinamento *retropropagação*. Como exemplo de sistema dinâmico usou-se o Sistema de Lorenz devido a sua não linearidade e natureza caótico e como método de assimilação de dados, o Filtro de Kalman Estendido (EKF). A carga computacional do EKF reside na avaliação da evolução da matriz de covariância de erros do estado [12] e computação da matriz "ganho". Neste ponto, a nossa investigação na utilização da RNA foi para diminuição da carga computacional do EKF, uma vez que a avaliação da matriz de erro do estado não é computada. Os resultados do Filtro de Kalman com o Modelo de Lorenz foram as entradas da PMC no treinamento.

Neste artigo não será abordado a *complexidade de algorítmo*. Somente estamos voltados ao desempenho do sistema de assimilação a *per se*, ou seja, se as redes neurais são capazes de realizar com sucesso o processo de ingestão de dados de observação sem se desviar da dinâmica do processo.

2 Metodologia

2.1 Assimilação de Dados

Assimilação de dados é um meio manter o estado do modelo próximo à natureza pela assimilação de observações. Com base na formulação matemática do problema de análise faz-se a definição do espaço de trabalho. Técnicas clássicas de assimilação para reduzir a resolução ou domínio de análise utilizam o espaço da observação, ou seja, a localização da observação o mesmo espaço escolhido neste trabalho. Admite-se que este espaço em modelos mais realistas é composto ”somente” por observações convencionais.

Na formulação do problema, a representação do estado atmosférico, é uma matriz chamada vetor de estado x. A melhor representação da realidade é chamado x^t, o estado *verdadeiro* no instante da análise. O vetor de estado x^b, é um estado de *referência*, ou seja, uma previsão do modelo recente para o mesmo instante da análise, que representa um estado hipotético verdadeiro. Finalmente, o vetor análise chamado x^a, o estado que se quer encontrar. O problema da análise pode ser encontrar uma correção δx de x tal que $x^a = x^b + \delta x$, a análise x^a deve ser o mais próximo possível de x^t. Valores observados são reunidos em um vetor observação y, este vetor deve ser capaz de ser comparado com o vetor estado x^b. Uma função que ajusta o espaço do modelo para o espaço da observação e a unidade da variável analisada, é chamada de operador observação, de notação H. As diferenças entre as observações e o vetor de estado, no ponto da observação $y - H(x)$, é o vetor de partida e quando calculado com o modelo de referência x^b, é chamado *vetor inovação*, e calculado com a análise x^a de *incremento de análise* [3] A função H será usada também para que o incremento de análise volte para o espaço do modelo.

A equação fundamental para análise linear em uma forma algébrica geral aqui representada, será a estimação de mínimos quadrados [11]. Definida pelas seguintes equações:

$$x^a = x^b + K(y - H[x^b]) \tag{1}$$

$$K = BH^T(HBH^T + R)^{-1} \tag{2}$$

onde o operador linear K é chamado de ganho, ou matriz peso, da análise.

Para representar as incertezas do modelo, das observações e da análise, assumem-se alguns modelos de erros, calculados com uma função de densidade de probabilidade(FDP) para cada tipo de erro. As FDPs dos erros de observação e do modelo são ”gaussianas”, então x^a é também a estimação de variância mínima de x^t. $\mathbf{B}$ e $\mathbf{R}$ são matrizes de covariâncias de erros do modelo e da observação respectivamente, são *pré–determinadas estatisticamente.*[2] [11]

2.2 Filtro de Kalman

O Filtro de Kalman constitui um processo recursivo *eficiente* de estimação, uma vez que o erro quadrático médio é ”minimizado”, ou seja, é uma solução eficiente do método de *mínimos quadrados*. A implementação do filtro de Kalman padrão para um modelo numérico de PNT atual tem uma carga computacional inaceitável. Para obter um filtro computacionalmente eficiente em assimilação de dados, simplificações têm sido introduzidas.

O modelo propriamente dito é definido por duas equações: a equação do processo e a equação da medida. O Filtro de Kalman (KF) na sua versão estendida *Extended Kalman Filter (EKF)* desenvolve o método de mínimos quadrados para uma estrutura não linear, na qual o modelo de referência provêm de uma previsão que é iniciada por uma análise prévia. As equações do filtro de Kalman linear são exatamente as equações (1) e (2), descritas anteriormente, exceto que as Matrizes de covariância de erro do modelo de referência (de previsão) e da análise passam a ser chamadas P^f e P^a respectivamente, pois no *KF* de fato calcula-se a Matriz de covariância de erro da análise em tempo de execução. Podemos observar o algorítmo do Filtro de Kalman na (Figura 1).

2.3 Modelo de Lorenz

Edward Lorenz (1963) desenvolveu um modelo matemático do modo como o ar se move na atmosfera, e chegou à conclusão que com pequenas variações nos valores iniciais das variáveis do seu modelo, obteve resultados muito divergentes. Em sistemas dinâmicos complexos, estes resultados ”instáveis”dizem respeito à evolução temporal como função de seus parâmetros e variáveis. Lorenz em sua pesquisa de sistemas dinâmicos usou três equações para representar

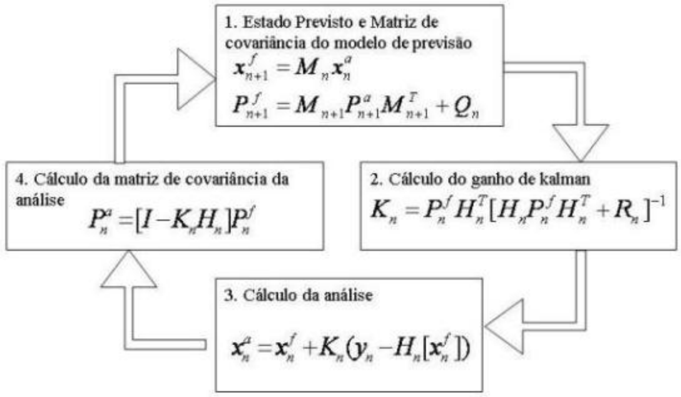

Figura 1: Algoritmo do Filtro de Kalman **Discreto**

graficamente o comportamento dinâmico através de computadores, descreveu um sistema relativamente simples com um padrão de complexidade infinita, onde se verificou que a partir de estados iniciais ligeiramente diferentes, o sistema de equações diferenciais resultava em soluções completamente diferentes entre si. O sistema de Lorenz consiste de três equações diferenciais ordinárias de primeira ordem, acopladas:

$$\frac{dx}{dt} = -\sigma x - y \tag{3}$$

$$\frac{dy}{dt} = -\rho x - y - xz \tag{4}$$

$$\frac{dz}{dt} = xy - \beta z \tag{5}$$

onde σ, ρ, β são parâmetros do modelo, com esta abordagem caótica, utilizou-se os valores 10, 28 e 8/3 , respectivamente. As variáveis x, y e z possuem uma interpretação espacial. A conseqüência da instabilidade dos resultados notada no Modelo de Lorenz é que mesmo em sistemas determinísticos existe uma grande sensibilidade a perturbações e erros. Assimilação de dados atmosféricos é uma das aplicações frequentes da utilização do sistema de Lorenz, por ser um modelo dinâmico simples com comportamento caótico, sensível às condições iniciais.[6]

2.4 Redes Neurais Artificiais

As RNA são sistemas paralelos distribuídos compostos por unidades de processamentos simples (nós ou neurônios artificiais) que computam determinadas funções matemáticas (normalmente não lineares). Tais unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões. Na maioria dos modelos estas conexões estão associadas a "pesos", que armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede. Os valores na saída da rede são denominados aqui como camada de saída. Um modelo de rede neural multicamadas simples e conhecido tem a denominação de Perceptron Multicamadas (PMC), Haykin (1991). As interconexões das entradas à camada de saída por pelo menos uma camada de neurônios intermediária, é denominada de camada escondida. [4].

As RNA possuem a capacidade de aprender através exemplos e fazer interpolações e extrapolações do que aprenderam. Um algoritmo de aprendizado é um conjunto de procedimentos bem definidos e são usados para adaptar os parâmetros de uma RNA, a fim de produzir uma saída esperada. Este é um processo de aprendizado (supervisionado) por correção de erros e procura minimizar a diferença entre a soma ponderada das entradas pelo pesos e a saída desejada. Este algoritmo é chamado de retroprogação do erro e foi o algoritmo utilizado neste trabalho. A forma genérica para alteração dos pesos por correção de erros do algoritmo de retropropagação, é definida por $w_i(t+1) = w_i(t) + \eta e(t) x_i(t)$, onde η é a taxa de aprendizado e $x_i(t)$ é a entrada para o neurônio i no instante de tempo t. O ajuste dos pesos $w_i(t)$ deve ser proporcional ao produto do erro ($e(t) = d(t) - y(t)$ onde $d(t)$ é o alvo ou saída esperada e $y(t)$ é a resposta atual (calculada pela RNA)), utilizando o valor de entrada $x_i(t)$ naquele instante de tempo t. Cada camada (escondida ou saída) tem um numero pré–determinado de neurônios e padrões de conectividade fixos.[1] [3]

3 Assimilação de Dados utilizando Redes Neurais Artificiais

A formulação matemática da equação da análise utilizando a RNA neste trabalho busca a melhoria de desempenho computacional, comparada à aplicação das equações da análise acima descritas. Esta formulação é equivalente no caso linear como:

$$x^a - x^b = BH^T(HBH^T + R)^{-1}(y - H(x^b)) \tag{6}$$

dividida em duas igualdades:

$$\lambda = (HBH^T + R)^{-1}(y - H(x^b)) \tag{7}$$

$$x^a - x^b = BH^T \lambda \tag{8}$$

onde λ tem a mesma dimensão que y e pode ser considerado o "incremento de análise" no espaço de observação, considerando que BH^T é um termo que mapeia este parâmetro para o espaço do modelo. O objetivo é resolver o problema de análise em termos de λ em lugar da *matriz produto da multiplicação das matrizes de covariâncias do erro*. Para resolver para λ o sistema linear é:

$$(HBHT + R)\lambda = (y - H(x^b)) \tag{9}$$

Na implementação da análise utilizando redes neurais artificiais, temos o seguinte algorítmo:

1. Cálculo do vetor de partidas $y - H(x^b)$;
2. Obtenção do vetor λ com a ativação da RNA;
3. Multiplicação do vetor λ estimado pelo BH^T para obter os incrementos de anállise;
4. Adicionar os incrementos ao modelo de referência x^b e o oter a análise x^a no espaço do modelo.

Neste trabalho utilizou-se um PMC para determinar o parâmetro λ após o treinamento com os dados de entrada: λ "alvo" e vetor inovação da execução do EKF com Modelo de Lorenz e observações sintéticas. A obtenção do parâmetro λ foi após o cálculo do termo $(HBHT + R)(y - H(x^b))$. No final da execução do EKF obteve-se também a Matriz de Covariância B para utilização na ativação da PMC. O treinamento da RNA foi feito para obter os pesos "ótimos" na sua convergência. Com os pesos adequados, a ativação da PMC determinou o λ para acada entrada do modelo de Lorenz em nova realização. O EKF foi executado também para comparação dos resultados.

4 Resultados

A rede possui três entradas relativas ao vetor inovação das variáveis, x, yez e o λ das respectivas variáveis utilizadas como "saída esperada" durante o treinamento. Seis neurônios foram utilizados na camada escondida e na camada de saída três neurônios referentes ao parâmetro λ de x, yez. Os treinamentos da RNA foram feitos após 32.000,16.000 e 2.000 passos de tempo de tamanho (0.00625) na execução do modelo. As informações de entrada foram inseridas em intervalos de 10 e 20 passos de tempo. A ativação da RNA foi feita com os pesos fixos obtidos após o treinamento e nova execução do modelo de Lorenz.

Os resultados apresentam os últimos passos de tempo das trajetórias das variaveis x, y, z na execução do modelo hipotético "verdadeiro", do modelo iniciado com a análise do *EKF*, as *observações* e o modelo iniciado com a análise *RNA*. As figuras abaixo indicam que a RNA acompanhou o modelo até o final. Verificou-se que a trajetória do estado com a condição inicial gerada pela RNA, acompanha a trajetória do estado verdadeiro, veja nas figuras 2 e 3.

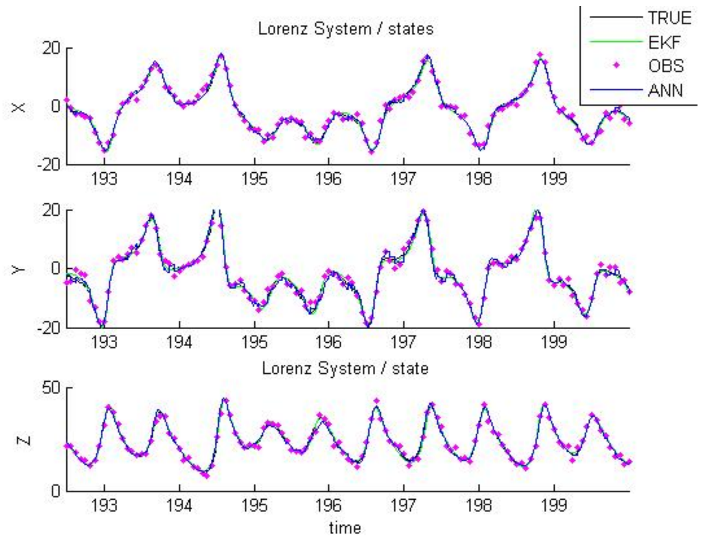

Figura 2: Sistema de Lorenz - observações a cada dez passos(pontilhado magenta), estado verdadeiro (linha preta), estado com EKF(linha verde), estado com RNA (linha azul)

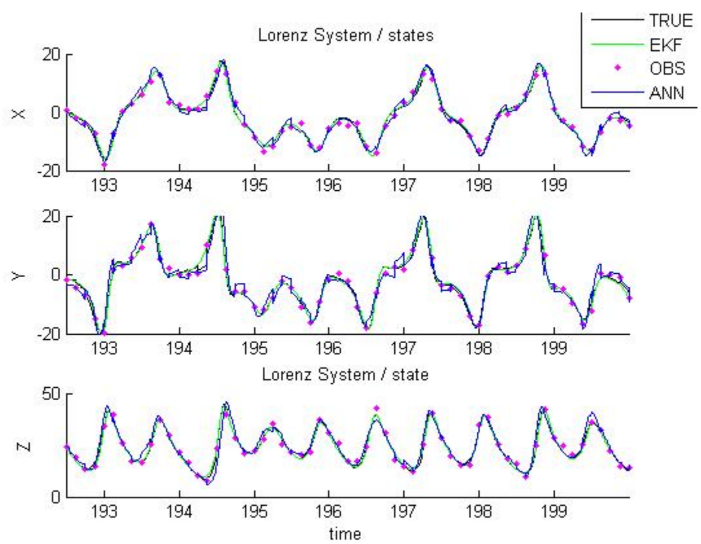

Figura 3: Sistema de Lorenz - observações a cada 20 passos (pontilhado magenta), estado verdadeiro (linha preta), estado com EKF(linha verde), estado com RNA (linha azul)

5 Conclusão

Neste trabalho implementou-se o Modelo de Lorenz caótico, o EKF e uma rede PMC com o algorítmo de treinamento *retropropagação* como melhoria de desempenho em métodos de assimilação. As metodologias aplicadas em centros de previsão, embora os resultados sejam satisfatórios, há dúvidas se os algoritmos são computacionalmente eficientes para realizar a tarefa de assimilação com o aumento exponencial de dados de observação, em tempo de uma previsão operacional. É importante investigar algorítmos que sejam computacionalmente mais eficientes. O processo de inversão de matrizes é um procedimento N^3, enquanto que a avaliação da rede é $N \times M$, onde N é a dimensão do sistema e M é o número de neurônios da rede.

Investigou-se uma abordagem da metodologia de assimilação de dados com Redes Neurais Artificiais em um modelo simples e verificou-se que com o obtenção do "pseudo" ganho (λ) pela RNA (já treinada) apresenta melhoria computacional em relação ao FK. Esta inversão é calculada em métodos tradicionais como Interpolação Ótima e FK para obtenção do campo inicial de modelos numéricos de previsão de tempo.

Referências

[1] A. Braga, A. de Carvalho, T. Ludermir; *Redes Neurais Artificiais: Teoria e Aplicações*, Editora LTC, (2000).

[2] F. Boutier and P. Courtier; Data Assimilation concepts and methods,*in: Meteorological Training Course*,Reading, UK, (1998).

[3] R. Daley (1991): *Atmospheric Data Analysis*, Cambridge University Press.

[4] M.W. Gardner, S.R. Dorling, Artificial Neural Networks (The Multilayer Perpectron) - A Review of Applications in the Atmospheric Sciences, *Atmospheric Environment*, 32(14/15),(1998) 2627-2636.

[5] F.P. Harter H.F. de Campos Velho, New Approach to Applying Neural Network in Nonlinear Dynamic Model, *Applied Mathematical Modelling* 32 (2008) 2621-2633. - DOI 10.1016/j.apm.2007.09.006 / ISSN: 0307-904X.

[6] F.P Hartër, "Redes Neurais Recorrentes Aplicadas à Assimilção de Dados em Dinâmica Não Linear", Tese de Doutorado, Computação Aplicada, CAP-INPE, São José dos Campos (SP), 2004.

[7] F.P. Harter, H.F. de Campos Velho, Recurrent and Feedforward Neural Networks Trained with Cross Correlation Applied to the Data Assimilation in Chaotic Dynamic, *Revista Brasileira de Meteorologia*, 20(3), (2005) 411-420.

[8] S. Haykin; " Neural Networks: A Comprehensive Foundation", Mcmillan,1994.

[9] A.G. Nowosad, A. Rios Neto, H.F. Campos Velho, Data Assimilation in Chaotic Dynamics Using Neural Networks, em "Third International Conference on Nonlinear Dynamics, Chaos, Control and Their Applications in Engineering Sciences", pp. 212-221, Campos do Jordão-SP, 2000.

[10] A. G. Nowosad, "Novas Abordagens para Assimilação de Dados Meteorológicos", Tese de Doutorado, CAP-INPE, 2001.

[11] O. Talagrand, Assimilation of observations, an introduction, *J. Meteor. Soc. Japan*, 75 (1997) 91-209.

[12] R. Todling, S. E. Cohn, Suboptimal Schemes for Atmospheric Data Assimilation Based on the Kalman Filter. *Mon. Wea.r Rev.*, 122 (1994) 2530-2557.