

# Assimilação de Dados Atmosféricos utilizando Redes Neurais Artificiais: Nova Abordagem.

Rosângela S. C. Cintra<sup>1</sup>, Haroldo F. de Campos Velho<sup>1</sup>, Ricardo Todling<sup>2</sup>

<sup>1</sup>Instituto Nacional de Pesquisas Espaciais (INPE/LAC)  
Av. Astronautas, 1758 – 12.227-101 -São José dos Campos – SP- Brazil

<sup>2</sup>Global Modelling and Data Assimilation Office, GMAO - NASA/GSFC 20770 -  
Greenbelt/Maryland -EUA

{Rosangela.cintra,haroldo}@lac.inpe.br, todling@gmao.gsfc.nasa.gov

**Abstract.** Atmospheric Data Assimilation combines model and observational data through methodologies that allow the data to constrain the model evolution. Several methods of data assimilation are in use in the atmospheric and ocean sciences. Methods using Artificial Neural Network (ANN) are being proposed, principally in the National Institute of Space Research, and they have been presenting efficient and solid results. In the publication of this previous workshop, an approach of computational complexity reduction for classic data assimilation algorithms was presented. In the present paper, a new approach was implemented, applying an ANN Multiple Layers Perceptron. This study was done with Kalman Filter for two non-linear dynamic models: Lorenz System (1963) and Dynamo System.

**Resumo.** Assimilação de Dados é a combinação de um modelo dinâmico e de observações por metodologias que permitem que os dados forcem a evolução do modelo. Diversos métodos de assimilação de dados estão atualmente em uso nas ciências atmosféricas e oceânicas. Métodos que utilizam Redes Neurais Artificiais (RNA) vêm sendo propostos, principalmente em pesquisas no Instituto Nacional de Pesquisas Espaciais, e têm apresentado resultados consistentes e eficazes quanto à aplicação. Em publicação do worcap2008, uma abordagem para redução da complexidade computacional de algoritmos clássicos de assimilação de dados foi apresentada. No presente artigo, a RNA foi aplicada com novo enfoque. Este estudo foi feito com o Filtro de Kalman e dois modelos dinâmicos não lineares: Sistema de Lorenz (1963) e Dynamo.

## 1. Introdução

O problema de determinar as melhores condições iniciais para os modelos de Previsão Numérica de Tempo (PNT) é de grande importância prática, e assunto de muitos estudos por pessoas de cenários diferentes. Os modelos de PNT são simulações computacionais da atmosfera que tomam a análise como ponto de partida e desenvolvem o estado da atmosfera no 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 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. Para maior compreensão e maiores detalhes em assimilação de dados atmosféricos, vejam em Daley (1991), Boutier and Courtier (1998).

Algoritmos de assimilação de dados possuem aproximações próprias, por exemplo: métodos sequenciais como Interpolação Ótima (OI) e Filtro de Kalman, métodos variacionais em três e quatro dimensões (3D-VAR/4D-VAR) e métodos que utiliza Redes Neurais Artificiais (Campos Velho et al., 2002). Atualmente nos grandes centros mais de um milhão de observações são utilizados nos horários de análise. Estas observações são selecionadas para que seja possível a execução de seus sistemas de assimilação de dados. Neste cenário se insere metodologias que utilizam Redes Neurais Artificiais (RNA) que podem ser uma solução, uma abordagem abrangente foi testada em Nowosad (2000) e em Härter (2004, 2005).

Nesta solução de algoritmo de análise foi implantada uma RNA para obter o incremento de análise a ser adicionado ao modelo previsto e obter a condição inicial para integrar um modelo numérico de PNT. Utilizou-se neste enfoque uma RNA Multicamadas. A nova metodologia foi aplicada ao Sistema de Lorenz (1963) e ao Sistema Dynamo (modelo com as equações de água rasa). O sistema de assimilação de dados utilizado no treinamento da RNA foi o Filtro de Kalman Estendido (FKE) por se tratarem de modelos não lineares. Jazswinski (1970) prevê uma introdução e discussão matemática mais precisa sobre o Filtro de Kalman e Boutier (1998) introduz o Filtro de Kalman no contexto de assimilação de dados atmosféricos.

Grande parte da carga computacional do FKE reside na avaliação da evolução de tempo da matriz de covariância de erros dos estados. O método que utiliza a RNA atua na diminuição da carga computacional comparada ao FKE. Neste artigo não será abordado a complexidade de software ou questões da validação algorítmica. O desempenho do sistema de assimilação será discutido, baseada na formulação matemática da assimilação de dados.

## 2. Assimilação de Dados

Assimilação de dados é um meio manter o estado do modelo próximo à natureza pela assimilação de observações. O modelo de previsão, a representação do estado atmosférico, é uma matriz coluna chamada estado  $\mathbf{x}$ . Alguns métodos de assimilação de dados são a estimação de máxima verossimilhança usando toda informação disponível: a previsão  $\mathbf{x}^f$  e as observações  $\mathbf{y}^o$ . Obtém-se o estado de análise  $\mathbf{x}^a$  pela média ponderada linear entre  $\mathbf{x}^f$  e  $\mathbf{y}^o$ , seguindo a formulação:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y}^o - \mathbf{H}(\mathbf{x}^f)) \quad (2.1)$$

onde  $\mathbf{K}$  denota a matriz peso ou ganho da análise:

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \quad (2.2)$$

Teoricamente, observações diretas não são feitas em todos os pontos de grade para todas as variáveis prognósticas, os vetores  $\mathbf{y}^o$  e  $\mathbf{x}$  estão em diferentes espaços. Quando o operador  $\mathbf{H}$  converte as variáveis prognósticas  $\mathbf{x}$  para o espaço dos dados de observação  $\mathbf{y}^o$  é chamado “operador de observação”.  $\mathbf{H}$  não é necessariamente um mapeamento linear. As diferenças entre observações e modelo de previsão (conhecidos *vetor inovação*), podem definir o espaço de trabalho nos pontos de observação, utilizando o operador  $\mathbf{H}$ , que também pode definir o espaço de trabalho do vetor de diferenças entre análise e modelo de previsão (conhecidos como *vetor correção*, ou *incremento de análise*) na grade do modelo de previsão.

A melhor representação possível da realidade é o vetor estado, chamado  $\mathbf{x}^t$  (um estado hipotético). As diferenças do estado  $\mathbf{x}^t$  com  $\mathbf{x}^a$ ,  $\mathbf{x}^f$  e  $\mathbf{y}^o$  definem os erros da análise, previsão, e de observação, respectivamente. Consequentemente tem-se as matrizes de covariância do erro da análise conhecida como  $\mathbf{P}^a$ , a matriz de covariância do erro da previsão

$\mathbf{P}^f$  e a matriz de covariância do erro da observação  $\mathbf{R}$ . Podem ser definidas estatisticamente por uma função de distribuição de probabilidade (FDP) de seus erros.

## 2.1 Filtro de Kalman Estendido

O Filtro de Kalman (FK) foi o método utilizado para emular as estratégias de análise utilizando RNA. O FK tem a formulação do problema de estimação de verossimilhança acima (equação 2.1), ou seja, estimação de mínimos quadrados com hipóteses gaussianas. Para sistemas não lineares o modelo dinâmico do sistema usa-se o Filtro de Kalman Estendido (FKE) onde tem-se um modelo "linearizado":  $\mathbf{x}_n = \tilde{\mathbf{x}}_n + \mathbf{A}(\mathbf{x}_{n-1} - \tilde{\mathbf{x}}_{n-1}) + \mathbf{W}(\eta_n)$  e o modelo de medição  $z_n = \tilde{z}_n + \mathbf{H}(\mathbf{x}_n - \tilde{\mathbf{x}}_n) + \mathbf{V}(\xi_n)$ , onde  $\mathbf{A}$  e  $\mathbf{H}$  são os Jacobianos em ordem a  $\mathbf{x}$ ; e  $\mathbf{W}$  e  $\mathbf{V}$  são os Jacobianos em ordem a  $\eta$  e  $\xi$  respectivamente. O EKF apresenta o seguinte algoritmo:

- Fase de previsão, a partir do modelo:

$$\text{Previsão de estado: } \mathbf{x}_n^f = \mathbf{f}_n[\mathbf{x}_{n-1}^f, 0] \quad (2.3)$$

$$\text{Previsão de covariância: } \mathbf{P}_n^f = \mathbf{A}_n \mathbf{P}_n^a \mathbf{A}_n^T + \mathbf{W}_n \mathbf{Q}_n \mathbf{W}_n^T \quad (2.4)$$

- Fase de atualização:

$$\text{Calculo da Matriz Ganho: } \mathbf{K}_n = \mathbf{P}_n^f \mathbf{H}_n^T [\mathbf{H}_n \mathbf{P}_n^f \mathbf{H}_n^T + \mathbf{V}_n \mathbf{R}_n \mathbf{V}_n^T]^{-1} \quad (2.5)$$

$$\text{Atualização do estado (análise): } \mathbf{x}_n^a = \mathbf{x}_n^f + \mathbf{K}_n (z_n - \mathbf{h}[\mathbf{x}_n^f, 0]) \quad (2.6)$$

$$\text{Atualização da covariância da análise } \mathbf{P}_n^a = [\mathbf{I} - \mathbf{K}_n \mathbf{H}_n] \mathbf{P}_n^f \quad (2.7)$$

O FKE constitui um processo recursivo eficiente, uma vez que o erro quadrático médio é "minimizado". O método FK tem as suas raízes no controle de sistemas, a sua aplicação em assimilação de dados já se faz há décadas. Jazswinski (1970) prevê uma introdução e discussão matemática mais precisa.

## 3. Nova Abordagem de Assimilação de Dados com Redes Neurais Artificiais

O processo de assimilação de dados em alguns métodos consiste em um algoritmo que combina previsões de um modelo numérico, para um determinado instante, com dados observacionais, para o mesmo instante, a fim de se obter uma melhor estimativa do estado da atmosfera. Tal processo pode ser expresso pelas equações (2.1) e (2.2), ou seja, a formulação matemática algébrica geral baseada na estimação de mínimos quadrados, Talagrand (1997). Nesta formulação, as incertezas do modelo, das observações e da análise, assumem modelos de erros com hipóteses "gaussianas", calculadas com uma FDP para cada tipo de erro. Quando estas matrizes são pré-determinando estatisticamente passam a ter a notação  $\mathbf{B}$  e  $\mathbf{R}$  para as matrizes de covariâncias de erros do modelo e da observação respectivamente. No FKE  $\mathbf{B}$  equivale a  $\mathbf{P}^f$  (matriz determinada em tempo de integração com o FKE).

A nova metodologia parte desta formulação matemática supra e utiliza uma Rede Neural Artificial (RNA) para armazenar o conhecimento e ponderar a entrada recebida por cada neurônio da rede até a saída. A RNA possui interconexões das entradas à camada de saída com uma camada de neurônios intermediária, denominada camada escondida. A rede utilizada é o Perceptron Multicamadas (PMC) que realiza um mapeamento complexo  $y = \Psi(w, x)$ , da entrada  $x$  à camada de saída  $y$ , parametrizado pelos pesos sinápticos  $w$ , Haykin (1991). Redes PMC são ferramentas que conseguem aproximar qualquer função linearmente contínua. Possui a capacidade de aprender através de exemplos e fazer interpolações e extrapolações do que

aprenderam utilizando um conjunto de procedimentos (chamado treinamento) bem definidos para adaptar seus parâmetros e produzir uma *saída desejada*. O treinamento supervisionado que consiste em minimizar o *erro* entre a saída da rede para um determinado padrão de entrada e a resposta esperada para aquele mesmo padrão é o algoritmo supervisionado de *retro-propagação do erro*, utilizado neste trabalho.

O método de assimilação de dados baseado em RNA foi apresentado no Brasil, por Nowosad et al. (2000) e Härter (2004) aplicados a modelos não-lineares onde as entradas são o modelo de previsão e as observações. A saída esperada apresentada no treinamento é a análise de outro método de assimilação. Neste trabalho, apresenta-se uma variante destes métodos:

### 3.1 Estimação do Incremento de análise

Esta variante do método de assimilação de dados com RNA, obtém a análise em dois passos:

$$\text{Passo de previsão: } x_n^f = F[x_{n-1}^f] \quad (3.1)$$

$$\text{Passo de análise: } x_n^a = x_n^f + d_n \quad (3.2)$$

O vetor *incremento de análise* é o produto entre uma matriz de ponderação e uma função que mede a discrepância entre a previsão do modelo e as observações:

$$d_n = M_{rna} (y_n^o - x_n^f) \quad (3.3)$$

sendo  $M_{rna}$  é a função de estimação do vetor *correção* calculada pela RNA, equivalente às equações 2.5 e 2.5 do FKE :

$$d = BH^T (HBH^T + R)^{-1} (y^o - H(x^f)) \quad (3.4)$$

A PMC teve como entrada o vetor de observações ( $y^o$ ), o vetor de previsão ( $x^f$ ) e o vetor *incremento de análise* ( $d$ ), para treinamento, obtido do cálculo da equação 2.5 do FKE. O  $d$  a ser adicionado ao modelo de previsão será obtido com a generalização da RNA no espaço do modelo. A assimilação de dados nesta abordagem apresenta o seguinte algoritmo:

- 1) Cálculo do *vetor inovação* ( $y^o - H(x^f)$ );
- 2) Obtenção do vetor incremento de análise  $d$  com a ativação da RNA, equação 3.3;
- 3) Adição do incremento de análise ao modelo de previsão  $x^f$  na obtenção da análise  $x^a$  (eq. 3.2).

A ativação da RNA substituiu o cálculo da equação 2.5, ou seja, a formulação da inversa da matriz ganho. O algoritmo de treinamento utilizou o *vetor inovação* (observação menos modelo) e o vetor  $d$  de incrementos de análise, obtidos da execução prévia do FKE com os modelos teste de uma realização. E a generalização foi feita em outra realização dos modelos.

## 4. RESULTADOS E CONCLUSÕES

Nesta seção será apresentada uma breve revisão dos modelos usados como teste para assimilação de dados com RNA e seus resultados.

### 4.1 Experimento 1: Sistema de Lorenz

Edward Lorenz em sua pesquisa dos sistemas dinâmicos escolheu três equações conhecidas como Equações de Lorenz que representam 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 acopladas equações diferenciais ordinárias de primeira ordem:

$$\frac{dX}{dt} = -\sigma(X - Y) \quad \frac{dY}{dt} = \rho X - Y - XZ \quad \frac{dZ}{dt} = XY - \beta Z$$

Para uma abordagem caótica  $\sigma$ ,  $\rho$  e  $\beta$  tem os valores 10, 28 e 8/3, respectivamente. As variáveis X, Y e Z possuem uma interpretação espacial. A consequência da instabilidade dos resultados notada no sistema de Lorenz é que, mesmo em sistemas determinístico, existe grande sensibilidade a perturbações e erros. Assimilação de dados atmosféricos é uma das aplicações freqüentes da utilização do sistema de Lorenz, por ser um modelo dinâmico simples com comportamento caótico, sensível às condições iniciais.

Assimilação de dados atmosféricos é uma das aplicações freqüentes da utilização do sistema de Lorenz, por ser um modelo dinâmico simples com comportamento caótico, sensível às condições iniciais. Nesta técnica, com o sistema de Lorenz utilizou-se uma RNA com três entradas relativas ao *vetor de inovação* das variáveis X, Y, Z e três neurônios de saída referente ao vetor *d* (*vetor incremento de análise*) das respectivas variáveis e oito neurônios na camada escondida. Para o treinamento, obteve-se os dados da execução prévia do FKE. O treinamento foi feito com dados coletados na execução do modelo com 32.000 passos de tempo de 0.00625. A assimilação de dados foi feita nos intervalos de 10, 20 e 40 passos de tempo. Os resultados mostram que com todos os tempos de espaçamento da assimilação de dados testados, as condições iniciais geradas pela ativação da RNA, acompanham a trajetória do modelo verdadeiro, veja a figura 4.1

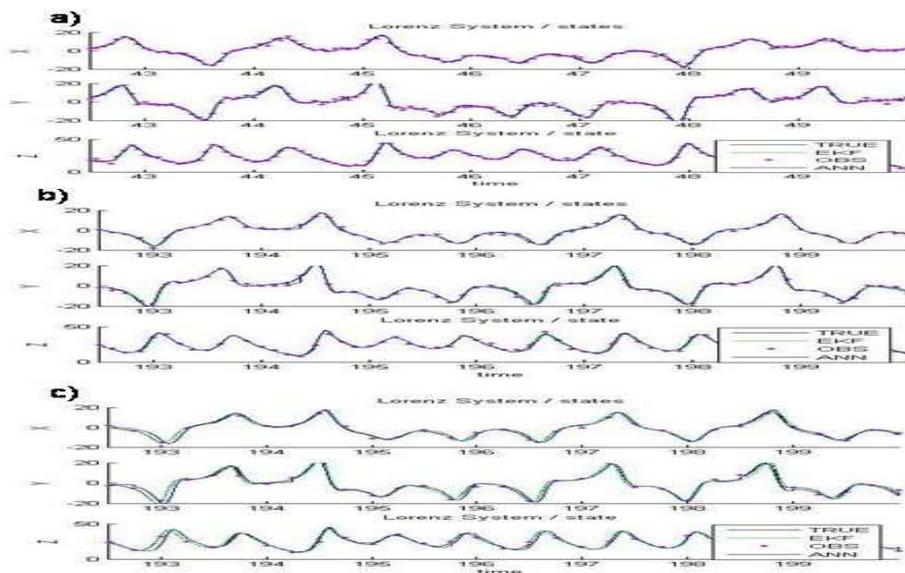


Figura 4.1 - Sistema de Lorenz: a) observações a cada dez passos de tempo; b) observações a 20 passos de tempo; c) observações a quarenta passos de tempo; observações (pontilhado magenta), estado verdadeiro (linha preta), estado com EKF (linha verde), estado RNA2 (linha azul).

## 4.2 Experimento 2: Modelo Meteorológico Dynamo

Segundo Campos Velho (1997) o modelo Dynamo foi concebido por Lynch (1984), para simular movimentos atmosféricos em grande escala (sinótica). Apesar de sua grande simplicidade, o código DYNAMO é capaz de reproduzir importantes fenômenos da dinâmica atmosférica. A superfície é considerada plana ( $z = 0$ ) e as condições de contorno são periódicas para todas as variáveis. Partindo das equações de água-rasa numa camada delgada de fluido homogêneo e incompressível acima de uma superfície plana sob a ação da gravidade, num sistema de referência não inercial, temos:

$$\frac{du}{dt} - fv + \frac{\partial \Phi}{\partial x} = 0 \quad \frac{dv}{dt} + fu + \frac{\partial \Phi}{\partial y} = 0 \quad \frac{d\Phi}{dt} + \Phi(u_x + v_y) = 0$$

onde  $d/dt$  é a derivada substantiva;  $x, y$  são as coordenadas no sentido oeste e norte;  $t$  é o tempo;  $\Phi = gh$  o geopotencial, sendo  $h$  a altura do fluido acima da superfície plana;  $u, v$  são as velocidades zonal e meridional. Considera-se o parâmetro de Coriolis na forma  $f = f_0 + \beta y$  com  $f_0, \beta$  constantes. Por último, os subscritos significam derivação parcial. Supondo que o vento zonal seja constante e esteja em balanço geostrófico com o geopotencial  $f\bar{u} = -\partial\bar{\Phi}/\partial y$  e os desvios sejam variáveis unidimensionais:  $u = \bar{u} + U(x, t)$ ;  $v = V(x, t)$ ;  $\Phi = \bar{\Phi}(y) + \phi(x, t)$ .

A importância de se avaliar o desempenho do método de assimilação com redes neurais com o modelo Dynamo-1D é devido a sua não-linearidade, por possuir ondas meteorológicas importantes (Rossby e gravidade), e possui dimensão espacial. Nos testes apresentados a seguir, a observação foi gerada artificialmente somando-se o ruído aleatório às variáveis previstas pelo modelo numérico. O modelo Dynamo foi integrado no passo de tempo de 60 segundos. Os dados foram inseridos a cada 360 passos de tempo.

A rede PMC para este modelo possui três entradas relativas ao *vetor de inovação* das variáveis  $U, V$  e  $\Phi$  e a camada de saída com três neurônios referentes aos vetores incremento  $d$  das respectivas variáveis. No treinamento os vetores *inovação e incremento* foram coletados da execução prévia do modelo Dynamo com FKE. A PMC foi implantada com quinze neurônios na camada escondida, e o treinamento utilizou 200 exemplos. Na generalização, a ativação ou assimilação foi feita a cada 360 passos de tempo (6 horas). O tempo total de integração foi de 1200 horas ou 50 dias. Os resultados mostram que com a assimilação de dados com a RNA a cada 6 horas, geraram condições iniciais para o modelo que acompanharam a curva do modelo de previsão e coincidissem com a curva do modelo verdadeiro, para as três variáveis, conforme figura 4.2. As diferenças entre as trajetórias são da ordem de  $10^{-6}$  entre os modelos, conforme o gráfico do erro logaritmo para cada figura correspondente.

Os resultados nos dois modelos mostram que o método de assimilação de dados por RNA pode ser competitivo em relação aos métodos hoje disponíveis, sendo este computacionalmente eficiente para utilizar toda a disponibilidade de dados existente, o que hoje não é realidade e mostram-se eficazes quanto às aplicações. Nesta abordagem, não existe dependência do conhecimento dos erros de modelagem e de observações. A generalização da RNA para obtenção do incremento de análise  $d$  substitui todo o cálculo da inversão das matrizes de covariância de erro para obter a matriz “ganho” da análise e substitui o cálculo do incremento no ponto de grade, reduzindo significativamente a complexidade do algoritmo.

## 5. Agradecimentos

Os autores agradecem ao Dr. Fabrício Harter pela concessão das rotinas referentes ao modelo Dynamo-1D, e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico(CNPq) pelo apoio financeiro à pesquisa.

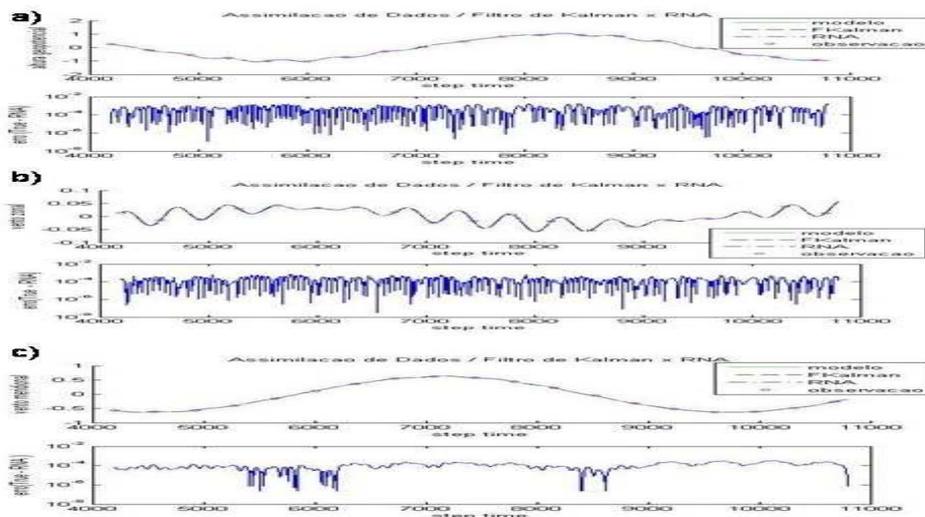


Figura 4.2 - Sistema Dynamo-1D: a) assimilação a cada 360 passos, para variável  $\phi$ , b) para a variável  $u$ ; c) para a variável  $v$ ; sendo observações (magenta), estado verdadeiro (linha preta), estado com EKF (linha verde), estado iniciado com RNA (linha azul). E respectivamente as diferenças entre o estado verdadeiro de o modelo com assimilação RNA.

## 6. Referências Bibliográficas

- Boutier, F.; Courtier, P. (1998) Data Assimilation concepts and methods. In: Meteorological Training Course}, Reading, UK, 53 pages.
- Campos Velho, H. F. C. V.; J.C.R., C.; (1997) A Comprehensive Analysis of a Barotropic Limited Area Model Using the Nonmodal Matrix Technique. Revista Brasileira de Meteorologia, v. 2, n. 12, p. 41–50.
- Campos Velho, H. F.; Vijaykumar, N.; Stephany, S.; Preto, A. J.; Nowosad, A. G. (2002) A Neural Network Implementation for Data Assimilation using MPI, Applications of High-Performance Computing in Engineering. United Kingdom: Southampton, WIT Press, 280 p.
- Daley, R. (1991) Atmospheric data analysis. Cambridge University Press, 457 p.
- Härter, F. P.; (2004) Redes neurais recorrentes aplicadas à assimilação de dados em dinâmica não linear, São José dos Campos, 127 p., Dissertação de Doutorado (INPE).
- Härter, F.P.; Campos Velho, H.F. (2005) Recurrent and feedforward neural networks trained with cross correlation applied to the data assimilation in chaotic dynamic, Revista Brasileira de Meteorologia, v.20, n.3, 411-420.
- Jazwinski, A. (1970) Stochastic processes and filtering theory. New York: Academic, 376 p.
- Lorenz, E. (1963) Deterministic nonperiodic flow. Journal of the Atmospheric Sciences, Boston, v. 20, n. 2, p. 130-141.
- Nowosad, A.G. (2001) Novas abordagens para assimilação de dados meteorológicos, São José dos Campos. Dissertação de Doutorado, INPE.
- Nowosad, A.; Campos Velho, H.; Rios Neto, A. (2000). Neural network as a new approach for data assimilation. In: Brazilian Congress On Meteorology, 11., 2000, Rio de Janeiro. Anais. Rio de Janeiro: SBMET, p. 3078-3086.