



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-14222-TDI/1123

**REDES NEURAIIS RECORRENTES APLICADAS À
ASSIMILAÇÃO DE DADOS EM DINÂMICA NÃO-LINEAR**

Fabício Pereira Härter

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelo Dr. Haroldo Fraga de Campos Velho, aprovada em 16 de dezembro de 2004.

INPE
São José dos Campos
2007

681.3.019:681.511.4

Härter, F. P.

Redes neurais recorrentes aplicadas à assimilação de dados em dinâmica não-linear / Fabrício Pereira Härter. - São José dos Campos: INPE, 2004.

138p. ; – (INPE-14222-TDI/1123)

1. Assimilação de dados. 2. Dinâmica não-linear. 3. Redes neurais. 4. Recorrência. 5. Validação cruzada. I. Título.

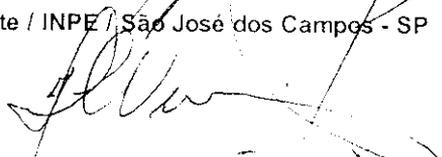
Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de **Doutor** em
Computação Aplicada

Dr. Fernando Manuel Ramos



Presidente / INPE / São José dos Campos - SP

Dr. Haroldo Fraga de Campos Velho



Orientador(a) / INPE / São José dos Campos - SP

Dr. Marco Antonio Chamon



Membro da Banca / INPE / São José dos Campos - SP

Dr. José Demisio Simões da Silva



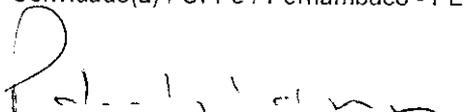
Membro da Banca / INPE / São José dos Campos - SP

Dra. Teresa Bernarda Ludemir



Convidado(a) / UFPE / Pernambuco - PE

Dr. Pedro Leite da Silva Dias



Convidado(a) / IAG/USP / São Paulo - SP

Aluno (a): Fabrício Pereira Härter

São José dos Campos, 16 de dezembro de 2004

“A América é potente não porque possua a Ford ou a Microsoft, mas porque possui universidades, laboratórios de pesquisa, o cinema e a CNN”.

DOMENICO DE MASI
em “O Ócio Criativo”, 2000

À minha namorada,
TALITHA CHAGAS DA SILVA.

AGRADECIMENTOS

Ao contribuinte do Estado de São Paulo, por ter financiado esta tese e à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), por gerir esse recurso de maneira exemplar.

Ao Instituto Nacional de Pesquisas Espaciais (INPE), em especial ao Programa de Pós-Graduação em Computação Aplicada (CAP), pela oportunidade de crescimento profissional.

Ao Dr. Haroldo Fraga de Campos Velho, pela orientação, entusiasmo, motivação, amizade e por estar sempre disponível, mesmo tendo inúmeros compromissos.

Aos colegas de curso, pela amizade e por muito colaborarem com o meu crescimento profissional. Agradeço em especial àqueles com que mais convivi, Alexandre Guirland Nowosad, Ana Paula Abrantes de Castro, Aritana Pinheiro Falconi, Douglas Rodrigues de Oliveira, Elcio Hideiti Shiguemori, Elias Canhadas Genvigir, Érico Rempel, Lúcio Henrique Franco, Luiz Gustavo Cunha Barbato, Maurício Pozzobon Martins, Leonardo Dagnino Chiwiakowski, Isabela Neves Drummond e Rosângela Saher Cintra.

Aos Professores José Demisio Simões da Silva, Fernando Manuel Ramos e Marco Antônio Chamon, aos quais recorri sempre que precisei.

Aos amigos Maria Cristina Maciel Lourenço, Sérgio Donizete Faria e Reinaldo Bonfim Silveira por me ajudarem no final desta etapa da minha vida, já em Brasília.

Ao amigo de longa data, Gilberto Caruccio Junior, com quem posso sempre contar.

A meus pais, Wilmar Antônio Härter e Leda Maria Pereira Härter, por me deixarem o estudo como herança.

RESUMO

Neste trabalho aplica-se as Redes Neurais (RN) Perceptron de Múltiplas Camadas (PMC), Funções de Base Radial (FBR), Elman (RN-E) e Jordan (RN-J) num contexto de Assimilação de Dados em dinâmica não-linear. Avalia-se a eficiência das RN em emular o filtro de Kalman (FK) e a possível aplicabilidade desta técnica a problemas de dimensão maior, como por exemplo, previsão de tempo operacional. A FBR, RN-E, RN-J e a técnica de treinamento conhecida como correlação cruzada até então nunca haviam sido testadas neste tipo de aplicação. Os resultados obtidos com os modelos simples (sistema de Lorenz em regime caótico e o modelo de Água Rasa - 1D) mostram que as redes foram eficientes na solução do problema, sendo que o principal resultado foi a redução do espaço de busca durante o treinamento das redes. Pela forma como os exemplos de treinamento são apresentados as RN, estima-se ter reduzido a dimensão do espaço de busca da ordem de 10^9 para 10^3 em problemas práticos de previsão de tempo e clima em relação ao trabalho pioneiro de Nowosad et al. (2001). Apresenta-se também, resultados da aplicação da PMC em Assimilação de Dados em Clima Espacial, o que abre uma nova linha de pesquisa em Geofísica Espacial.

RECURRENT NEURAL NETWORKS APPLIED TO DATA ASSIMILATION ON NON-LINEAR DYNAMIC

ABSTRACT

This work applies four neural network paradigms for data assimilation problem, focusing on non-linear dynamics. They are respectively named, according to literature, as Multilayer Perceptron (MLP), Radial Basis Function (RBF), Elman (E-NN) and Jordan (J-NN). The performance of these four NN on emulating the Kalman filter (KF) and the possibility of application of this technique at high dimension problems, as numerical weather forecasting, are analyzed. The RBF, E-NN, J-NN and the training technique knowledge, such as the cross validation have never been tested in this kind of application. Results with chaotic Lorenz system and Shallow Water Equations - 1D show that NN are an encouraging approach for data assimilation process. Thus the main result of this work is the improvement of the NN topologies in order to minimize the convergence rate, where a decreasing of 10^6 , if compared to the results given by Nowosad et al. (2001), could be achieved if this technique was applied to operational numerical weather forecasting. An application of NN techniques, for space geophysical data assimilation problem is also demonstrated and it was seen that it could open a new research line.

SUMÁRIO

Pág.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS E ABREVIATURAS

LISTA DE SÍMBOLOS

CAPÍTULO 1 - INTRODUÇÃO	27
CAPÍTULO 2 - FILTRO DE KALMAN	37
2.1 - Mínimos Quadrados	37
2.2 - Mínimos Quadrados Ponderados	38
2.3 - Escolha dos Pesos (θ)	39
2.4 - Mínimos Quadrados Recursivos	41
2.5 - Filtro de Kalman	43
2.6 - Um teste para o filtro de Kalman em Assimilação de Dados	47
2.6.1 - Descrição do Modelo de Advecção-Difusão	48
2.6.2 - Experimentos Numéricos	48
CAPÍTULO 3 - REDES NEURAIIS	55
3.1 - Resumo Histórico sobre a Evolução das Pesquisas em Redes Neurais	55
3.2 - Neurônio Biológico	57
3.3 - Neurônio Artificial	58
3.4 - Arquiteturas das Redes Neurais	59
3.5 - Funcionamento das Redes Neurais	60
3.6 - Função de Ativação	61
3.7 - Paradigmas e Regras de Aprendizagem	62
3.8 - Redes <i>feedforward</i> e Recorrentes	64
3.8.1 - Rede Perceptron de Múltiplas Camadas	65
3.8.2 - Rede Função de Base Radial	66
3.8.3 - Rede de Elman e Rede de Jordan	66
3.9 - Motivações e Limitações nos Estudos das Redes Neurais	68

CAPÍTULO 4 - SISTEMAS DE EVOLUÇÃO NÃO-LINEARES	71
4.1 - Sistema de Lorenz Caótico	71
4.2 - O Modelo Meteorológico DYNAMO	73
4.2.1 - Equações do Modelo	73
4.2.2 - Adimensionalização das Equações	75
4.2.3 - Discretização Espacial	77
4.3 - Modelo de Iteração Não-Linear entre Ondas Langmuir, Whistler e Alfvén - Modelo de 3 Ondas	80
 CAPÍTULO 5 - ASSIMILAÇÃO DE DADOS EM DINÂMICA NÃO-LINEAR COM O SISTEMA DE LORENZ E O MODELO DE 3 ONDAS	 85
5.1 - Assimilação de Dados	85
5.2 - Resultados com o sistema de Lorenz Caótico	86
5.2.1 - Assimilação: Perceptron de Múltiplas Camadas (PMC)	89
5.2.2 - Assimilação: Funções de Base Radial (RBF)	94
5.2.3 - Assimilação: Redes Recorrentes	97
5.3 - Resultados Numéricos com o Modelo de 3 Ondas: Aplicação em Clima Espacial	102
5.3.1 - Assimilação: Perceptron de Múltiplas Camadas (PMC) - Regime Pe- riódico - $\gamma = 28, 14$	104
5.3.2 - Assimilação: Perceptron de Múltiplas Camadas (PMC) - Regime Caó- tico - $\gamma = 28, 128$	106
 CAPÍTULO 6 - ASSIMILAÇÃO DE DADOS EM DINÂMICA NÃO-LINEAR COM O MODELO DYNAMO - 1D	 109
6.1 - Resultados com o Modelo DYNAMO - 1D	109
6.2 - Redução do Espaço de Busca	111
6.2.1 - Assimilação: Perceptron de Múltiplas Camadas (PMC)	112
6.2.2 - Assimilação: Funções de Base Radial (FBR)	116
6.2.3 - Assimilação: Redes Recorrentes	119
 CAPÍTULO 7 - CONCLUSÕES	 127
 REFERÊNCIAS BIBLIOGRÁFICAS	 133

LISTA DE FIGURAS

	<u>Pág.</u>
1.1 Diagrama de um ciclo de assimilação de dados intermitente de 6 horas. . .	32
2.1 Diagrama esquemático do filtro de Kalman linear	46
2.2 Efeito da frequência de observação no FK: (a) EXP3, (b) EXP5, (c) EXP6.	50
2.3 Assimilação com observações distribuídas em diferentes arranjos: (a) EXP10, (b) EXP11, (c) EXP12.	52
2.4 Ilustração do problema estacionário de difusão, que simula uma pluma liberada na origem do sistema de coordenadas	53
3.1 Representação do neurônio biológico.	58
3.2 Representação do neurônio artificial.	59
3.3 Função de ativação (a) sigmóide; (b) gaussiana.	61
3.4 Ilustração de Rede Perceptron de Múltiplas Camadas.	65
3.5 Ilustração das redes (a) Elman; (b) Jordan.	67
5.1 Ilustração da instabilidade causado no sistema de Lorenz ao se inserir dado sem um método de assimilação.	89
5.2 Erros com o sistema de Lorenz: (a) durante a fase de treinamento; (b) na validação cruzada.	90
5.3 Assimilação de dados com o sistema de Lorenz e PMC: de cima para baixo (a) componente- X , (b) componente- Y , (c) componente- Z	93
5.4 Erros com o sistema de Lorenz: (a) durante a fase de treinamento; (b) na validação cruzada.	94

5.5	Assimilação de dados com o sistema de Lorenz e FBR: de cima para baixo (a) componente- X , (b) componente- Y , (c) componente- Z	96
5.6	Erros com o sistema de Lorenz: (a) durante a fase de treinamento; (b) na validação cruzada.	97
5.7	Erros com o sistema de Lorenz: (a) durante a fase de treinamento; (b) na validação cruzada.	98
5.8	Evolução temporal do erro de previsão para (a) PMC, (b) RN-E.	99
5.9	Assimilação de dados com o sistema de Lorenz e RN-E: de cima para baixo (a) componente- X , (b) componente- Y , (c) componente- Z	100
5.10	Assimilação de dados com o sistema de Lorenz com RN-J: de cima para baixo (a) componente- X , (b) componente- Y , (c) componente- Z	101
5.11	Ilustração da instabilidade causado no Modelo de 3 Ondas ao se inserir dado sem um método de assimilação.	102
5.12	Erros do Modelo de 3 Ondas: (a) na fase de treinamento; (b) na validação cruzada.	104
5.13	Componente- A_L , (a) Modelo de referência, (b) assimilação por FK, e (c) assimilação pela PMC.	105
5.14	Erro para regime periódico: (a) filtro de Kalman; (b) rede neural.	106
5.15	Erros do Modelo de 3 Ondas: (a) treinamento; (b) validação.	106
5.16	Erro para regime periódico: (a) filtro de Kalman; (b) rede neural.	107
5.17	Componente- A_L , (a) Modelo de referência, (b) assimilação por FK, e (c) assimilação pela PMC.	108
6.1	Ilustração da instabilidade causado no modelo DYNAMO ao se inserir dado sem um método de assimilação.	109
6.2	Conjunto de dados usado no treinamento, validação cruzada e previsão com método de assimilação.	110

6.3	Erros com o modelo DYNAMO: (a) durante a fase de treinamento; (b) na validação cruzada.	112
6.4	Assimilação de dados com o modelo DYNAMO: campo de divergência no começo do domínio de integração. Linha azul - modelo de referência, linha vermelha - assimilação pela PMC.	114
6.5	Assimilação de dados com o modelo DYNAMO e PMC: Energia Cinética Divergente.	114
6.6	Assimilação de dados com o modelo DYNAMO e PMC: de cima para baixo, componentes (a) U , (b) V , (c) ϕ , (d) ζ , (e) δ	115
6.7	Erros com o modelo DYNAMO: (a) durante a fase de treinamento; (b) na validação cruzada.	116
6.8	Assimilação de dados com o modelo DYNAMO: campo de divergência no começo do domínio de integração. Linha azul - modelo de referência, linha vermelha - assimilação pela FBR.	117
6.9	Assimilação de dados com o modelo DYNAMO e FBR: de cima para baixo, componentes (a) U , (b) V , (c) ϕ , (d) ζ , (e) δ	118
6.10	Erros com o modelo DYNAMO: (a) durante a fase de treinamento; (b) na validação cruzada.	119
6.11	Assimilação de dados com o modelo DYNAMO: campo de divergência no começo do domínio de integração. Linha azul - modelo de referência, linha vermelha - assimilação pela RN-E.	120
6.12	Assimilação de dados com o modelo DYNAMO e RN-E: de cima para baixo, componentes (a) U , (b) V , (c) ϕ , (d) ζ , (e) δ	121
6.13	Erros com o modelo DYNAMO: (a) durante a fase de treinamento; (b) na validação cruzada.	122
6.14	Assimilação de dados com o modelo DYNAMO e RN-J: de cima para baixo, componentes (a) U , (b) V , (c) ϕ , (d) ζ , (e) δ	123

6.15	Assimilação de dados com o modelo DYNAMO: campo de divergência no começo do domínio de integração. Linha azul - modelo de referência, linha vermelha - assimilação pela RN-J.	124
6.16	Evolução temporal do erro de previsão para (a) PMC, (b) RN-E.	125

LISTA DE TABELAS

	<u>Pág.</u>
2.1 Erro Relativo aos Experimentos 1 2 3 4 5 6.	49
2.2 Erro para diferentes números de sensores na direção - z.	51
2.3 Erro para três arranjos de observações na direção- z.	51
5.1 Parâmetros das RN com o sistema de Lorenz: 1 camada escondida. . . .	87
5.2 Parâmetros das RN com o sistema de Lorenz: 2 camadas escondidas. . .	88
5.3 Erros com o PMC.	91
5.4 Erros com a FBR.	95
5.5 Tempo de Treinamento (s) com o sistema de Lorenz.	95
5.6 Erros com a RN-E.	98
5.7 Erros com a RN-J.	98
6.1 Parâmetros das redes RN com o modelo DYNAMO: 1 camada escondida.	110
6.2 Erros com a PMC.	113
6.3 Erros com a FBR.	117
6.4 Erros com a RN-E.	120
6.5 Erros com a RN-J.	122
6.6 Tempo de Treinamento (s) com o DYNAMO-1D.	124

LISTA DE SIGLAS E ABREVIATURAS

CPTEC	–	Centro de Previsão de Tempo e Estudos Climáticos
DAO	–	<i>Data Assimilation Office</i>
FBR	–	Rede Neural Funções de Base Radial
FK	–	Filtro de Kalman
FKE	–	Filtro de Kalman Estendido
INMET	–	Instituto Nacional de Meteorologia
IO	–	Interpolação Ótima
LAPS	–	Local Analysis and Prediction System
MBAR	–	Modelo Brasileiro de Alta Resolução
MQ	–	Mínimos Quadrados
MQP	–	Mínimos Quadrados Ponderados
MQR	–	Mínimos Quadrados Ponderados com Recursividade
NASA	–	<i>Nacional Aeronautic and Space Administration</i>
NCEP	–	<i>National Center for Environmental Prediction</i>
PMC	–	Rede Neural Perceptron de Múltiplas Camadas
PNT	–	Previsão Numérica de Tempo
PSAS	–	<i>Physical-space Statistical Analysis System</i>
RH	–	Rede de Hopfield
RN	–	Redes Neurais Artificiais
RN-E	–	Rede Neural de Elman
RN-J	–	Rede Neural de Jordan
RPSAS	–	<i>Regional Physical-space Statistical Analysis System</i>

LISTA DE SÍMBOLOS

a	–	Constante responsável pela inclinação da tangente hiperbólica
A_1	–	Amplitude das ondas Langmuir
A_2	–	Amplitude das ondas Whistler
A_3	–	Amplitude das ondas Alfvén
b_k	–	Viés
d	–	Vetor alvo (desejado) na saída das RN
$\mathbf{E}[\cdot]$	–	Valor esperado
EP	–	Erro de Previsão com assimilação
ET	–	Erro de Treinamento
ETA	–	Modelo Regional de Previsão Numérica de Tempo
EV	–	Erro de Validação
\mathbf{F}	–	Matriz que representa o sistema dinâmico
\mathbf{G}	–	Ganho de Kalman
\mathbf{H}	–	Matriz que representa o sistema de observação
\mathbf{I}	–	Matriz Identidade
$J(\cdot)$	–	Função custo
\mathbf{k}	–	k-ésimo instante de tempo discreto
\mathbf{K}_{zz}	–	Difusividade vertical turbulenta para a camada limite estável
l	–	Índice referente a camada oculta das RN
L	–	Índice referente a camada de saída das RN
o	–	Vetor obtido na saída das RN
\mathbf{P}_a	–	Covariância do erro de estimativa
\mathbf{P}_f	–	Covariância do erro de previsão
\mathbf{Q}	–	Covariância do erro de modelagem
r	–	Número de Rayleigh
\mathbf{R}	–	Matriz de covariância dos erros de observação
R_c	–	Número de Rayleigh crítico
R_F	–	Número de Froude
R_o	–	Número de Rossby
\mathbf{u}	–	Velocidade do vento zonal
\mathbf{v}	–	Velocidade do vento meridional
\mathbf{w}^a	–	Vetor de estado de análise
\mathbf{w}^f	–	Vetor de estado previsto pelo modelo numérico (sem assimilação)
X	–	Proporcional a intensidade do movimento convectivo
y_k	–	Sinal na saída do neurônio
Y	–	Proporcional a diferença de temperatura entre correntes ascendentes e descendentes
z_k	–	k-ésima componente do vetor de observações
\mathbf{z}	–	Vetor de observações
z	–	Altura

Z	–	Proporcional a perturbação do perfil de temperatura vertical
α	–	Constante de momento
δ	–	Gradiente local
δ	–	Divergência
Δx	–	Espaçamento de grade na direção X
Δz	–	Espaçamento de grade na direção Z
ζ	–	Vorticidade
θ	–	Vetor de pesos
ν	–	Viscosidade cinemática do fluido
σ	–	Número de Prandtl
τ	–	Tempo adimensional
ϕ	–	Velocidade do geopotencial
$\varphi(\cdot)$	–	Função de ativação
$\hat{(\cdot)}$	–	Refere-se a estimativa

CAPÍTULO 1

INTRODUÇÃO

Historicamente, a análise de dados meteorológicos começou com as primeiras cartas sinóticas, nas quais eram grafadas variáveis meteorológicas, observadas em pontos esparsos. Os campos meteorológicos eram traçados manualmente, servindo de base para as previsões de tempo. Este tipo de procedimento, conhecido como análise subjetiva, depende da habilidade e da experiência do meteorologista.

No começo do século XX, a previsão de tempo foi abordada como um problema de valor inicial, onde o estado futuro da atmosfera pode ser previsto através do conjunto de equações que governam os movimentos atmosféricos, (Bjerknes, 1911). Com o mesmo intuito de Bjerknes, Richardson fez a primeira Previsão Numérica de Tempo (PNT), que apesar do insucesso no objetivo final, foi de grande valia para esta área da ciência, (Richardson, 1922).

Segundo Daley (1991), o trabalho de Richardson falhou por várias razões, discutidas em Platzman (1967), dentre as quais salienta-se a não representatividade dos dados (poucas estações, irregularmente espaçadas e nenhuma informação de ar superior), e o fato do método numérico não obedecer as restrições Courant, Friedrichs, Lewis (CFL) para a estabilidade, (Hoffman, 1993) . Outra razão foi que o problema de inicialização não era conhecido. Iniciar significa acelerar o processo de equilíbrio entre os campos de massa e velocidade nos dados iniciais, reduzindo o ruído gerado por ondas de gravidade de alta frequência, (Härter, 1999).

Ao passar dos anos, com o desenvolvimento dos métodos matemáticos e da informática, os modelos numéricos foram aperfeiçoados, sendo que atualmente os centros operacionais de previsão de tempo utilizam esta ferramenta diariamente, disponibilizando previsões com até 15 dias de antecedência nos Estados Unidos da América, na Europa e no Brasil. Todavia, o índice de acerto de 60 % (mínimo para que a previsão seja considerada válida) nem sempre é atingido a partir do quinto dia de previsão.

Assim como os modelos numéricos, a rede de observações de dados meteorológicos também tem evoluído. Surgem diferentes tipos de instrumentos que fornecem amostragens contínuas da atmosfera e superfície da terra de diferentes formas. Entre estes, pode-se citar instrumentos pontuais como termômetros, barômetros e pluviô-

metros, irregularmente espaçados e cujos dados são coletados em horários sinóticos. Há também radiossondas, fornecendo perfis verticais da atmosfera em um determinado local. Há ainda, instrumentos que fazem medidas remotas como radares e satélites, estes últimos com boa resolução espacial e temporal. A grande vantagem na utilização dos dados de satélite é que estes cobrem áreas com baixa densidade de informações meteorológicas, tais como oceanos e florestas tropicais.

Com a evolução dos computadores, dos modelos numéricos e da rede de observação, a previsão numérica passa a ter melhor qualidade, desde que os dados sejam inseridos nos modelos de maneira adequada, o que não é uma tarefa trivial. Ressalta-se que teoricamente, mesmo numa situação hipotética, onde tanto os modelos como as observações são perfeitas, a natureza caótica da atmosfera limitaria a previsibilidade para alguns dias, segundo Kalnay (2004), duas semanas. Isto ocorre porque pequenas variações nas condições iniciais, após um determinado período de integração resulta em estados completamente diferentes, (Lorenz, 1963).

Uma maneira mais rápida e consistente do que a análise subjetiva de se produzir mapas sinóticos, conhecida como análise objetiva, foi produzida por Panofski (1949), na qual foi utilizada a técnica de ajustes por mínimos quadrados em duas dimensões.

A evolução histórica dos métodos de assimilação de dados meteorológicos passa pelo Ajuste de Funções, Métodos das Correções Sucessivas, Interpolação Ótima, Métodos Variacionais em 3 e 4 dimensões e Filtro de Kalman (FK), (Daley, 1991). Todos estes métodos são diferentes formas de combinar uma previsão de “background” com observações resultando na análise, o que matematicamente pode ser apresentado da seguinte forma

$$\mathbf{x}^a = \mathbf{x}^f + \theta[\mathbf{z} - \mathbf{H}(\mathbf{x}^f)] \quad (1.1)$$

onde \mathbf{x}^a é variável estimada, \mathbf{x}^f refere-se a previsão do modelo (ou campo de “background”), θ são pesos determinados com base na estatística de covariância dos erros de previsão e observação, \mathbf{z} refere-se as observações, \mathbf{H} é o operador que representa o sistema de observação, $\mathbf{z} - \mathbf{H}(\mathbf{x}^f)$ é a inovação e $\mathbf{H}(\mathbf{x}^f)$ é a estimativa inicial (“first guess”).

Algoritmos de interpolação em duas dimensões também foram elaborados por

Berghthorsson e Döös (1955) e Cressman (1959). Campos de “background” definidos nos pontos de grade do modelo de previsão são interpolados para os locais de observação. A diferença entre as observações e os valores interpolados são novamente interpolados para os pontos de grade do modelo com o intuito de definir uma correção que será aplicada ao “first-guess”. Atualmente as idéias básicas destes algoritmos ainda são aplicadas.

Apesar dos avanços descritos acima, a rede de observações nunca representará com perfeição o estado da atmosfera e os modelos numéricos sempre serão aproximações da natureza. Alguns erros de observação podem ser tratados com técnicas estatísticas, como erros de instrumento e erros de representatividade (erros de representatividade são próprios da amostragem discreta do meio contínuo). Estes erros podem ter viés, podem ser espacialmente e temporalmente correlacionados com o sinal. Em assimilação de dados o conhecimento estatístico destes erros é de grande valia.

Eliansen (1954) e Gandin (1963), deram base estatística ao processo de análise, definindo o que veio a ser conhecido em meteorologia como Interpolação Ótima (IO), criando a função de covariância estatística dos campos meteorológicos.

Progressos em análise objetiva continuaram na década de 60. O uso de modelos de previsão numérica para gerar a condição inicial tornou-se universal. Na década de 70, poderosas técnicas de interpolação estatística ganharam aceitação (Daley, 1991). Nesta época também tornaram-se habituais previsões numéricas com modelos de equações primitivas. A partir de então, a denominação Ciclo de Assimilação de Dados passou a ser usado para designar o complexo processo de criar uma condição inicial “balanceada” e incorporar novos dados observados durante a integração do modelo de previsão. Daley (1991) divide o ciclo de Assimilação de Dados nas seguintes quatro componentes:

- a) controle de qualidade dos dados: algoritmos de controle de qualidade são projetados para rejeitar ou corrigir dados ruins. Primeiro é verificado se há erros de codificação e se há problemas de localização de sensores. Em seguida se compara cada observação com a sua vizinha, requerendo-se consistência espacial e temporal. Verifica-se também relações dinâmicas, com o equilíbrio hidrostático e geostrófico para latitudes médias. Faz-se também a verificação contra o campo “background”. O controle de qualidade pode ainda atingir um nível de sofisticação mais alto, com o uso da estrutura da

análise objetiva, como processo consistente;

- b) análise objetiva: procedimento automático para estimar as variáveis atmosféricas dependentes numa grade regular bidimensional ou tridimensional usando os dados disponíveis oriundos da rede de observações irregularmente espaçadas;
- c) inicialização: iniciar significa acelerar o processo de equilíbrio entre os campos de massa e velocidade nos dados iniciais, reduzindo o ruído gerado por ondas de gravidade de alta frequência. Esta etapa é realizada junto com a previsão final (não se deve confundir com a previsão para obtenção da estimativa *a priori*);
- d) previsão curta para preparar a estimativa *a priori*: nesta etapa usa-se o modelo numérico para preparar o campo de “background” (estimativa *a priori* do próximo tempo de observação). Este modelo deve incluir as parametrizações necessárias para assegurar que, se não for atualizado com novas observações, o clima por ele gerado aproximar-se-á do verdadeiro. Isso assegura que na falta persistente de dados a estimativa *a priori* produzida pelo modelo assimilador permaneça fisicamente plausível.

Os métodos de assimilação de dados podem ser classificados em determinísticos (Transformada de Laplace, Nudging) ou estocásticos (variacionais, filtro de Kalman), métodos estes que encontram fundamentação matemática em teoria de estimação. Entre os métodos estocásticos destaca-se o filtro de Kalman (FK), que apresenta solução ótima quando aplicado a problemas em que a estatística dos erros de modelagem e de observação seguem uma distribuição gaussiana. Como a maioria dos problemas práticos de interesse são não-lineares, aplica-se a estes o filtro de Kalman Estendido, versão não-linear do filtro original, que por sua vez é um filtro sub-ótimo. Outra vantagem do FK é que o mesmo calcula não apenas uma estimativa do estado do escoamento, mas também uma estimativa da incerteza associada. Este texto fica restrito ao FK, pois aqui, as Rede Neurais Recorrentes (RNR), nova técnica em assimilação de dados proposta nesta tese, emulam um filtro de Kalman Estendido, abreviado a partir daqui como FK.

Apesar do avanço dos métodos de assimilação de dados nas últimas décadas, há uma diferença substancial entre teoria e prática, devido basicamente a dois problemas relacionados entre si, que são a dimensão e o custo computacional da aplicação.

Modelos de equações primitivas tem graus de liberdade da ordem de 10^7 . Por exemplo, um modelo com resolução de 1 grau (latitude-longitude) com 20 níveis verticais tem $360 \times 180 \times 20 = 1.3 \times 10^6$ pontos de grade. Para cada ponto de grade tem-se no mínimo 4 variáveis prognósticas (2 componentes horizontais do vento, temperatura e umidade), mais a pressão à superfície para cada coluna, resultando num total de 5 milhões de variáveis a serem inicializadas. Para um ciclo de assimilação de dados de 6 horas (com janela de assimilação de ± 3 horas), há tipicamente 10 – 100 mil observações a serem assimiladas, duas ordens de magnitude a menos do que o número de graus de liberdade do modelo (Kalnay, 2004). Em situações nas quais se tem mais variáveis a serem assimiladas e modelos com melhor resolução vertical, a dimensão do vetor de observações pode aumentar para ordem de 10^8 . Deve-se também considerar que regiões como a América do Norte e Eurásia têm uma rede relativamente densa de dados, enquanto há outras áreas com deficiência de dados, principalmente sobre os oceanos e florestas tropicais. Por isto são necessárias informações adicionais como campo de “background”, “first guess” ou informação *a priori* para preparar a análise (condição inicial para a previsão).

Ressalta-se que em um ciclo de assimilação de dados intermitente de 6 horas, usual na prática operacional, Figura 1.1, a inovação ($\mathbf{y}^o - \mathbf{H}(\mathbf{x}^f)$, ver equação 1.1), é calculada após o campo de background ser interpolado para os pontos de observação [$\mathbf{H}(\mathbf{x}^f)$]. No caso em que as variáveis do modelo não são as mesmas medidas diretamente pelo sensor, como no caso de satélites que fornecem radiâncias, as variáveis do modelo devem ser transformadas para as variáveis medidas pelo sensor do instrumento.

No começo da PNT o custo computacional da análise era negligenciada em relação ao custo de uma previsão de 24 horas. Atualmente, o custo computacional da assimilação (somada ao custo da integração do modelo de “background” por um período de 24 horas) é tipicamente o custo da previsão de 24 horas. E a maioria dos centros de previsão estão considerando alocar diariamente, para 24 horas de assimilação, o equivalente a 10 dias de previsão ou mais, (Talagrand, 1997).

Devido a este alto custo computacional das técnicas de assimilação de dados, a bibliografia tem mostrado tentativas por parte dos cientistas em desenvolver métodos que amenizem este problema. A maioria das soluções encontradas na literatura são as aproximações do filtro de Kalman.

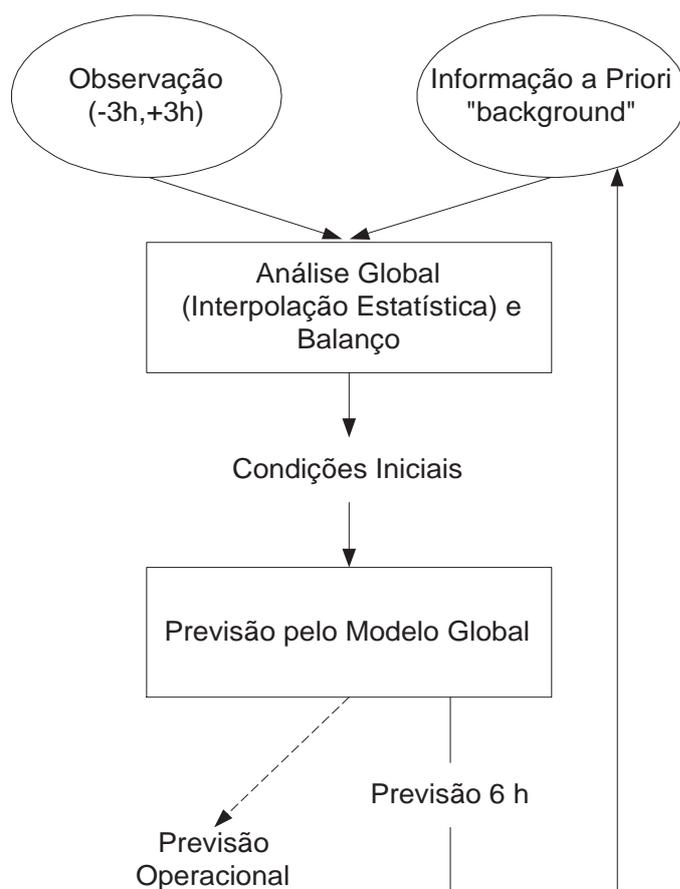


FIGURA 1.1 - Diagrama de um ciclo de assimilação de dados intermitente de 6 horas.

Em colaboração com o *Data Assimilation Office-National Aeronautic and Space Administration* (DAO-NASA), foi implementado no CPTEC um sistema de assimilação de dados regional para a América do Sul para uso com o modelo regional ETA. Neste sistema, denominado *Regional Physical-space Statistical Analysis System* (RPSAS), o filtro é aplicado no espaço físico (pontos de observação) ao invés de trabalhar no espaço do modelo, como no método IO, por exemplo. O RPSAS é um sistema estável e com erro aceitável quando comparado com as análises do *National Center for Environmental Prediction* (NCEP). Uma descrição detalhada do sistema de assimilação de dados do CPTEC encontra-se em Cintra (2004).

Em Belyaev et al. (2000) é testado um método de assimilação de dados baseado na teoria de Kalman, cujo desempenho é avaliado com o Modelo Oceânico Modular (*MOM2*). Experimentos nos quais os autores assimilam a temperatura do Oceano Atlântico tropical, mostram que este método funciona e tem as seguintes vantagens

em relação ao filtro de Kalman padrão: apresenta melhor concordância com a variabilidade intrasazonal de temperatura observada e não requer a linearização do modelo numérico, (Belyaev et al., 2002).

Na última década tem surgido uma grande quantidade de aplicações de Redes Neurais (RN) em meteorologia¹. Em Hsieh e Tang (1998), esta técnica é sugerida como um possível método de assimilação, onde os autores chamam a atenção para as seguintes dificuldades em aplicar RN em meteorologia e oceanografia: (a) instabilidade não-linear, (b) grande dimensão da aplicação, (c) dificuldade em interpretar os resultados das RN. Porém no mesmo trabalho são feitas sugestões para se contornar estas dificuldades, são elas: controlar a instabilidade não-linear com médias de previsões por conjunto (“ensemble”), aplicar análise de componentes principais para reduzir a dimensão do problema e interpretar o sinal na saída dos neurônios das camadas não-lineares através de análise espectral.

Em Nowosad et al. (2000), são apresentados os primeiros resultados de RN como método de assimilação de dados aplicadas a modelos não-lineares. A Rede Neural Perceptron de Múltiplas Camadas (PMC) mostrou-se uma boa técnica para assimilação de dados nos testes realizados com os modelos Henon, Lorenz e DYNAMO (equações de água rasa 1D). Com esta nova metodologia, uma vez que a rede esteja treinada, o problema de custo computacional é significativamente reduzido. O algoritmo foi paralelizado em Campos Velho et al. (2002), sendo que os autores do trabalho reduziram de maneira significativa o tempo de processamento durante o treinamento do PMC.

Em Tang e Hsieh (2001), é aplicado um modelo híbrido dinâmico-neural, assim denominado pelos autores por utilizarem a rede PMC para substituir uma das equações do sistema de Lorenz, num contexto de assimilação de dados, com o método variacional em 4 dimensões (4-DVAR). Neste trabalho os autores estimam parâmetros do modelo neural, parâmetros do modelo dinâmico e condições iniciais para o modelo híbrido. Foi verificado que em casos de não-linearidade fraca a metodologia apresenta bons resultados. Já no caso de não-linearidade forte, as estimativas falham nos casos em que a rede foi treinada com dados de uma das asas do atrator de Lorenz e a assimilação foi feita em outra asa do atrator. Modelos híbridos são investigados também em Liaqat et al. (2003).

¹Uma revisão destas aplicações é encontrada em Gardner and Dorling (1998)

Neste trabalho avalia-se a eficiência das redes PMC, Funções de Base Radial (FBR), Rede de Elman (RN-E) e Rede de Jordan (RN-J), aplicadas ao problema de assimilação de dados. Para tal utiliza-se dois modelos, o sistema de Lorenz em regime caótico e o modelo de Água Rasa não-linear (DYNAMO-1D). O primeiro tem sido utilizado na literatura, por exemplo (Miller, 1994) para testar técnicas de assimilação de dados por ser caótico, regime encontrado na atmosfera real. O modelo DYNAMO - 1D), embora tenha sido concebido para testar esquemas de inicialização, por exemplo (Heemink et al., 1997) também vem sendo empregado em testes de novas técnicas de assimilação porque possui em sua solução ondas de Rossby e ondas de gravidade, responsáveis pelos fenômenos atmosféricos mais importantes. O modelo DYNAMO-1D também permite que se avalie o desempenho do método ao se acrescentar a dimensão espacial ao problema não-linear.

Este trabalho apresenta quatro pontos inéditos em relação aos trabalhos encontrados na literatura, são eles:

- a) as RN-E e RN-J são recorrentes, ou seja, acrescenta-se memória ao sistema, o que se supõe melhorar as previsões à medida que se aumenta o alcance da mesma;
- b) no treinamento das RN implementou-se a técnica conhecida como validação cruzada (*cross-validation*), o que permite que se conheça totalmente a superfície de erros de treinamento e de validação, permitindo que se obtenha o melhor conjunto de pesos para o problema estudado na implementação feita;
- c) diminuição do espaço de busca, o que vem a ser o principal resultado deste trabalho, pois em problemas de minimização de gradiente, em aplicações de grande dimensão, diminuir o espaço de busca pode ser a diferença entre tornar a aplicação operacionalmente viável ou não;
- d) aplicação em clima espacial.

O texto é estruturado da seguinte forma: no Capítulo 2 revisa-se as técnicas de Teoria de Estimção por Mínimos Quadrados, base para o desenvolvimento do FK. Descreve-se também o FK, pois as estimativas por RN emulam este filtro. No Capítulo 3 apresenta-se uma evolução histórica sobre as RN, apresenta-se o modelo neuronal e se descreve o funcionamento e arquitetura das redes estudadas. Neste

Capítulo também se descreve o paradigma e a regra de aprendizagem implementada nos algoritmos avaliados nesta tese. No Capítulo 4 são descritos os sistemas não-lineares nos quais as novas abordagens são testadas. Os Capítulos 5 e 6 são reservados para os resultados e no Capítulo 7 são apresentadas as conclusões e considerações finais.

CAPÍTULO 2

FILTRO DE KALMAN

Segundo Sorenson (1970) o conceito de Gauss de Teoria de Estimação por Mínimos Quadrados, originalmente estimulado por estudos astronômicos, tem sido a base para o estudo de técnicas de teoria de estimação nos últimos 170 anos, provavelmente nenhuma tão útil em relação às necessidades atuais como o FK.

O FK, embora tenha sua origem nos Mínimos Quadrados (MQ), como verifica-se na afirmação de Sorenson, tem sido utilizado em estudos de poluição atmosférica (Zhang et al., 1997), Oceanografia (Miller et al., 1994) e Meteorologia (Daley, 1991), no contexto de Assimilação de Dados.

Como esta tese trata de Assimilação de Dados, que por as vez é um problema de estimação, reserva-se este capítulo ao FK, técnica emulada pelas RN. Por motivos evidentes na afirmação de Sorenson, revisa-se os Mínimos (MQ), Mínimos Quadrados Ponderados (MQP) e Mínimos Quadrados Ponderados com Recursividade (MQR). Maneiras alternativas de dedução das equações do FK são obtidas em Anderson e Moore (1979).

2.1 Mínimos Quadrados

Um sistema $\mathbf{H}\mathbf{w} = \mathbf{z}$ mal condicionado não pode ser resolvido diretamente por $\mathbf{w} = \mathbf{H}^{-1}\mathbf{z}$, portanto uma alternativa para estimar os parâmetros \mathbf{w} com as observações \mathbf{z} é minimizar $\|\mathbf{H}\mathbf{w} - \mathbf{z}\|$, onde H é o operador que transforma \mathbf{z} em \mathbf{w} . Para tal, considera-se o seguinte modelo linear e erro de observação

$$\mathbf{z} = \mathbf{H}\mathbf{w} + \varepsilon \quad (2.1)$$

$$\varepsilon = \mathbf{z} - \mathbf{H}\mathbf{w} \quad (2.2)$$

A solução deste sistema consiste em minimizar o erro ε de forma a se obter a melhor estimativa ($\hat{\mathbf{w}}$) do estado \mathbf{z} , segundo o critério quadrático, o que pode ser obtido minimizando-se o seguinte funcional $J(\mathbf{w})$ em relação a \mathbf{w}

$$J(\mathbf{w}) = \varepsilon^T \varepsilon = \|\varepsilon^2\| \quad (2.3)$$

onde ε^T é dado pela equação (2.2) e usando a propriedade $(AB)^T = B^T A^T$, obtém-se

$$J(\mathbf{w}) = \mathbf{w}^T \mathbf{H}^T \mathbf{H} \mathbf{w} - \mathbf{w}^T \mathbf{H}^T \mathbf{z} - \mathbf{z}^T \mathbf{H} \mathbf{w} + \mathbf{z}^T \mathbf{z}. \quad (2.4)$$

Como $\mathbf{z}^T \mathbf{H} \mathbf{w}$ é igual a $\mathbf{w}^T \mathbf{H}^T \mathbf{z}$, pois neste caso a ordem dos vetores é irrelevante para o cálculo do produto interno, a derivada primeira de $J(\mathbf{w})$ em relação a \mathbf{w} resulta na seguinte expressão

$$\hat{\mathbf{w}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{z} \quad (2.5)$$

onde o símbolo $(\hat{\cdot})$ foi introduzido para representar estimativa. Assim, verifica-se que o vetor \mathbf{w} que minimiza $\|\mathbf{H}\mathbf{w} - \mathbf{z}\|_2^2$, é a solução da equação normal $\mathbf{H}^T \mathbf{H} \mathbf{w} = \mathbf{H}^T \mathbf{z}$, sendo o vetor $\hat{\mathbf{w}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{z}$ a solução por mínimos quadrados para $\mathbf{H}\mathbf{w} = \mathbf{z}$

Pela interpretação geométrica, verifica-se que $\mathbf{H}\mathbf{w}$ é a projeção de \mathbf{z} , ou seja, o vetor mais próximo de \mathbf{z} entre todos os vetores $\mathbf{H}\mathbf{w}$ possíveis. Algebricamente a solução por mínimos quadrados pode ser obtida multiplicando-se $\mathbf{H}\mathbf{w} = \mathbf{z}$ por \mathbf{H}^T , resultando na matriz quadrática $\mathbf{H}^T \mathbf{H}$; obtendo-se n equações para os desconhecidos w_1, \dots, w_n . Assumindo-se que a matriz \mathbf{H} tem posto cheio (*full rank*, todas as suas colunas são independentes), tal que $\mathbf{H}^T \mathbf{H}$ possui inversa e \mathbf{z} é completamente determinado.

Em aplicações práticas não é usual estimar \mathbf{w} pela equação (2.5) devido às dificuldades numéricas na inversão de $\mathbf{H}^T \mathbf{H}$, e sim através de algoritmos estáveis (ver Golub e van Loan (1989)).

2.2 Mínimos Quadrados Ponderados

Para os MQP, considera-se o caso em que as medidas \mathbf{z} não têm a mesma precisão, ou seja, supondo-se que a medida z_1 seja mais precisa do que a medida z_2 , é natural atribuir maior peso para z_1 . Assim, ao invés de minimizar o erro médio quadrático $\|\varepsilon^2\|$, minimiza-se o erro médio quadrático dado por $\|\theta \varepsilon^2\|$, ou seja, encontra-se a melhor solução por mínimos quadrados que minimiza $\theta \mathbf{H}\mathbf{w} = \theta \mathbf{z}$, obtendo-se

$$\hat{\mathbf{w}} = (\mathbf{H}^T \mathbf{C} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C} \mathbf{z} \quad (2.6)$$

onde $\mathbf{C} = \theta^T \theta$. Fazendo-se

$$\mathbf{L} = (\mathbf{H}^T \mathbf{C} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}, \quad (2.7)$$

tem-se

$$\hat{\mathbf{w}} = \mathbf{Lz} \quad (2.8)$$

Para $\theta = \mathbf{I}$, tem-se os MQ padrão. Quando as medidas não são correlacionadas, a matriz de pesos θ é diagonal, caso contrário é uma matriz cheia.

2.3 Escolha dos Pesos (θ)

Aqui, tratar-se-á (θ) como propriedades estatísticas do experimento. Para calcular (θ), primeiro assume-se que o ruído de observação $\varepsilon = \mathbf{z} - \mathbf{Hw}$ é não tendencioso (sem viés). O valor esperado ¹ é dado por $\mathbf{E}[\varepsilon_i]$ e a variância por $\mathbf{E}[\varepsilon_i^2]$. Apenas o desvio padrão (raiz quadrada da variância) é necessário para determinar os pesos nos mínimos quadrados. Em outras palavras, tendo apenas uma medida e conhecendo apenas o primeiro e o segundo momentos, média e variância respectivamente, determinam-se as propriedades dos pesos nos mínimos quadrados, ou seja, assumindo uma distribuição gaussiana, não é necessário conhecer a distribuição de probabilidade completa. Havendo mais do que uma medida, é necessário conhecer a dependência mútua entre os ruídos, a qual é medida pela sua covariância.

Para ruídos independentes, o que é mais comum na prática (Strang, 1986), a covariância é nula. Os pesos θ são dados por $\theta_i = 1/\sigma_i$ (o que significa que quanto menor for a variância σ_i , mais confiáveis são as observações e maiores os pesos θ_i). Observa-se que θ e $\mathbf{C} = \theta^T \theta$, são diagonais, \mathbf{C} contém os números $1/\sigma_i$ e no caso em que as variâncias forem todas iguais, tem-se os mínimos quadrados.

Como \mathbf{C} torna o estimador por mínimos quadrados ponderados o “melhor estimador”, pode-se definir uma matriz de covariância dos erros de observação (nas medidas) \mathbf{R} , tal que $\mathbf{C} = \mathbf{R}^{-1}$.

Os elementos r_{ii} de \mathbf{R} são dados pelas variâncias $\mathbf{E}[\varepsilon_i^2]$ e os elementos fora da diagonal são dados por $\mathbf{E}[\varepsilon_i \varepsilon_j]$. Como os produtos ε_i^2 e $\varepsilon_i \varepsilon_j$ aparecem na matriz quadrada $\varepsilon \varepsilon^T$, pode-se abreviar a matriz de covariância \mathbf{R} como o valor esperado de $\varepsilon \varepsilon^T$

$$\mathbf{R} = \mathbf{E}[\varepsilon \varepsilon^T]. \quad (2.9)$$

Procura-se \mathbf{C} tal que $\mathbf{E}[\varepsilon_i] = \mathbf{0}$ (o melhor possível). Para tanto, o estimador $\hat{\mathbf{w}}_w =$

¹ $E[g(x)] = \int_{-\infty}^{+\infty} g(x) p(x) dx$, sendo $p(x)$ a função de densidade de probabilidade associada à variável x .

\mathbf{Lz} , que estima o verdadeiro, mas desconhecido parâmetro \mathbf{w} das medidas \mathbf{z} , deve ser linear e não tendencioso. Será linear se \mathbf{L} for uma matriz e não tendencioso se o valor esperado do erro de estimativa for zero, ou seja:

$$\mathbf{E}[\mathbf{w} - \hat{\mathbf{w}}] = \mathbf{E}[\mathbf{w} - \mathbf{Lz}] = \mathbf{E}[\mathbf{w} - \mathbf{L}(\mathbf{Hw} + \epsilon)] = \mathbf{E}[(\mathbf{I} - \mathbf{LH})\mathbf{w}] = 0. \quad (2.10)$$

de onde se deduz que \mathbf{L} é não tendencioso se $\mathbf{LH} = \mathbf{I}$.

Usando a suposição, primeiramente feita por Gauss de que a matriz \mathbf{C} deve ser o inverso da matriz de covariância \mathbf{R} ($\mathbf{C} = \mathbf{R}^{-1}$) e fazendo $\mathbf{L} = \mathbf{L}_0$, tem-se da equação (2.7) que

$$\mathbf{L}_0 = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \quad (2.11)$$

o qual se supõe ser a escolha ótima, pois satisfaz $\mathbf{L}_0 \mathbf{H} = \mathbf{z}$.

O melhor estimador linear não tendencioso (BLUE), é tal que $\mathbf{C} = \mathbf{R}^{-1}$. O estimador $\hat{\mathbf{w}}$ e a matriz ótima \mathbf{L}_0 satisfazem

$$\hat{\mathbf{w}} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z} = \mathbf{L}_0 \mathbf{z}. \quad (2.12)$$

Esta escolha minimiza o erro esperado na estimativa, medida pela matriz de covariância do erro de estimativa \mathbf{P} , dado por

$$\mathbf{P} = \mathbf{E}[(\mathbf{w} - \hat{\mathbf{w}})(\mathbf{w} - \hat{\mathbf{w}})^T] \quad (2.13)$$

Estimar \mathbf{w} ótimo é equivalente a obter \mathbf{P} mínimo, portanto minimiza-se a expressão (2.13), que substituindo o valor de $\hat{\mathbf{w}}$ fica

$$\mathbf{E}[(\mathbf{w} - \mathbf{LHw} - \mathbf{L}\epsilon)(\mathbf{w} - \mathbf{LHw} - \mathbf{L}\epsilon)^T]. \quad (2.14)$$

Desde que $\mathbf{LH} = \mathbf{I}$ e \mathbf{L} seja linear, usando a equação (2.9) obtém-se

$$\mathbf{P} = \mathbf{E}[(\mathbf{L}\epsilon)(\mathbf{L}\epsilon)^T] = \mathbf{L}\mathbf{E}[\epsilon\epsilon^T]\mathbf{L}^T = \mathbf{LRL}^T \quad (2.15)$$

Para mostrar que \mathbf{L}_0 é a escolha ótima, escreve-se \mathbf{L} como $\mathbf{L}_0 + (\mathbf{L} - \mathbf{L}_0)$ que substituindo em $\mathbf{P} = \mathbf{LRL}^T$, resulta em

$$\mathbf{P} = \mathbf{L}_0 \mathbf{R} \mathbf{L}_0^T + (\mathbf{L} - \mathbf{L}_0) \mathbf{R} \mathbf{L}_0^T + \mathbf{L}_0 \mathbf{V} (\mathbf{L} - \mathbf{L}_0)^T + (\mathbf{L} - \mathbf{L}_0) \mathbf{R} (\mathbf{L} - \mathbf{L}_0)^T. \quad (2.16)$$

Com algumas manipulações algébricas verifica-se que a segunda e a terceira parcelas da equação (2.16) são nulas, portanto (2.16) pode ser escrita como

$$\mathbf{P} = \mathbf{L}_0 \mathbf{R} \mathbf{L}_0^T + (\mathbf{L} - \mathbf{L}_0) \mathbf{R} (\mathbf{L} - \mathbf{L}_0)^T. \quad (2.17)$$

Esta expressão é mínima para $\mathbf{L} = \mathbf{L}_0$, então a equação (2.17) se reduz a $\mathbf{P} = \mathbf{L}_0 \mathbf{R} \mathbf{L}_0^T$. Substituindo-se \mathbf{L}_0 por \mathbf{L} em $\mathbf{P} = \mathbf{L} \mathbf{R} \mathbf{L}^T$ obtém-se a seguinte expressão para a covariância do erro de estimativa

$$\mathbf{P} = (\mathbf{H} \mathbf{R}^{-1} \mathbf{H})^{-1}. \quad (2.18)$$

Uma forma alternativa de derivação da matriz de covariância dos erros de estimativa pode ser encontrada em Mendel (1995), bem como a análise das vantagens e desvantagens em cada alternativa de derivação.

2.4 Mínimos Quadrados Recursivos

A seguir deduz-se a expressão do mínimos quadrados, para estimar um estado, ao passo que se dispões de novas medidas, sem a necessidade de se refazer os cálculos como todo o conjunto de dados já disponíveis. Esta é a idéia de recursividade.

Uma outra maneira de expressar a idéia dos MQR é considerar que se pretende estimar w_1 a partir da estimativa w_0 , agregando a nova observação z_1 .

De 2.12 $\hat{\mathbf{w}} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z} = \mathbf{L}_0 \mathbf{z}$, sendo que o erro de estimativa $(\mathbf{w} - \hat{\mathbf{w}})$ tem média zero e matriz de covariância $\mathbf{P} = (\mathbf{H} \mathbf{R}^{-1} \mathbf{H})^{-1}$, segundo a Equação (2.18).

Para calcular w_1 faz-se a mesma suposição feita para calcular w_0 , ou seja, a matriz \mathbf{C} que torna \hat{w}_1 a melhor estimativa é $\mathbf{C} = \mathbf{R}^{-1}$, onde

$$\mathbf{R} = \begin{bmatrix} R_0 & 0 \\ 0 & R_1 \end{bmatrix}, \quad (2.19)$$

é a matriz de covariância do erro de observação (erro nas medidas)

$$\epsilon = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \end{bmatrix}. \quad (2.20)$$

\mathbf{R} é diagonal por bloco porque ϵ_1 é independente de ϵ_0 . Desta forma, a matriz

$\mathbf{H}^T \mathbf{C} \mathbf{H}$, no cálculo de w_1 , será

$$P_1^{-1} = \begin{bmatrix} H_0 & H_1 \end{bmatrix} \begin{bmatrix} R_0 & 0 \\ 0 & R_1 \end{bmatrix} \begin{bmatrix} H_0 & H_1 \end{bmatrix} \quad (2.21)$$

que desenvolvendo os produtos matriciais e usando $P_0^{-1} = H_0^T R_0^{-1} H_0$, resultando em $P_1^{-1} = P_0^{-1} H_1^T R_1^{-1} H_1$. Generalizando, obtém-se a expressão para a covariância do erro para os MQR.

$$\mathbf{P}_k^{-1} = \mathbf{P}_{k-1}^{-1} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \quad (2.22)$$

O próximo passo é obter a equação de atualização para a estimativa $\hat{\mathbf{w}}_{k/k}$ (estimativa no instante k dadas medidas disponíveis até o instante k).

Lembrando que \hat{w}_1 não baseia-se apenas em z_1 e é o melhor estimador do sistema $H_0 w_0 z_0$ e $H_1 w_0 z_1$. Desde que a equação normal seja $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{w} = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z}$ e $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} = \mathbf{P}^{-1}$, com algumas manipulações algébricas verifica-se que $\hat{\mathbf{w}}$ é dado por

$$\hat{w} = P_1 (H_0^T R_0^{-1} z_0 + H_1^T R_1^{-1} z_1) \quad (2.23)$$

Com a chegada da medida z_1 há um aumento de informação. Isto faz com que \mathbf{P} diminua (\mathbf{P} é uma medida da incerteza). Nota-se que P_k^{-1} não depende das medidas z_0 e z_1 e refere-se apenas às propriedades estatísticas de w_1

Já a estimativa atual w_1 baseia-se em z_0 e z_1 . Rearranjando a equação (2.23), obtém-se

$$\hat{w}_1 = \hat{w}_1 + G_1 (z_1 - z_0) \quad (2.24)$$

Generalizando a equação (2.24) obtém-se a expressão de atualização por MQR para $\hat{\mathbf{w}}$ dadas \mathbf{z}_k medidas

$$w_k = w_{k-1} + G_k (z_k - H_k w_{k-1}). \quad (2.25)$$

onde $G_k = P_k H_k^T R_k^{-1}$ é a matriz ganho, $z_k - H_k w_{k-1}$ é a inovação e $G_k (z_k - H_k w_{k-1})$, refere-se a correção.

A equação (2.25) é uma forma eficiente da estimativa do estado \mathbf{w} , no instante k , dadas as z_{k-1} medidas, ser atualizada para uma estimativa \mathbf{w} , no instante k , dadas as z_k medidas.

Nesta seção derivou-se a expressão dos MQR (equação (2.25)), com a qual a estimativa de um determinado parâmetro pode ser atualizada de maneira ótima e

recursiva, ao passo que se dispõe de novas medidas. Derivou-se também uma expressão (equação (2.22)) que quantifica a incerteza associada a esta estimativa, ou seja, a covariância do erro de estimação.

Na próxima Seção deriva-se as expressões para o FK, que fornece uma estimativa recursiva ótima para estados que mudam de acordo com a evolução de um sistema dinâmico.

2.5 Filtro de Kalman

Dada a equação linear de medidas

$$\mathbf{z} = \mathbf{H}\mathbf{w} + \varepsilon \quad (2.26)$$

onde ε é um vetor gaussiano aleatório com média zero e matriz de covariância \mathbf{R} ($\varepsilon = \mathbf{N}(\mathbf{0}, \mathbf{R})$).

O melhor estimador $\hat{\mathbf{w}}$ linear não tendencioso de \mathbf{w} é $\hat{\mathbf{w}} = \mathbf{P}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z}$, onde $\mathbf{P} = \mathbf{E}[(\mathbf{w} - \hat{\mathbf{w}})(\mathbf{w} - \hat{\mathbf{w}})^T] = (\mathbf{H}\mathbf{R}^{-1}\mathbf{H})^{-1}$ é a covariância do erro de estimação.

Dado o seguinte sistema dinâmico com as seguintes equações para o sistema e medidas

$$w_{k+1} = F_k w_k + \eta_k \quad (2.27)$$

$$z_k = H_k w_k + \varepsilon_k \quad (2.28)$$

onde o ruído no sistema dinâmico (η_k) e o ruído nas medidas (ε_k) são vetores aleatórios gaussianos ($\eta_k \approx N(0, Q_k)$ e ($\varepsilon_k \approx N(0, R_k)$). O subíndice k refere-se ao k -ésimo instante de tempo discreto.

\hat{w}_0 é o melhor estimador linear não tendencioso de um estado inicial com matriz de covariância de erro P_0 . O filtro de Kalman calcula o melhor estimador linear não tendencioso no tempo k , dadas as medidas z_0, \dots, z_k ; ($w_{k/k}$). O filtro calcula também a matriz de covariância do erro $P_{k/k} (\hat{w}_{k/k} - \hat{w}_k)$ dadas estas medidas.

Os cálculos ocorrem de acordo com as fases de atualização (estimativa de w no instante k dadas z_k medidas) e propagação (estimativa de w no instante $k + 1$ dadas z_k medidas). A seguir, aplica-se os resultados do estimador ótimo para estimar

um estado e calcular a covariância do erro de estimação nas fases de atualização e propagação.

Atualização

$\hat{w}_{k/k}$ é o estimador que difere do estado real w_k por um termo de erro e_k com covariância ($P_{k/k-1}$)

$$w_{k/k+1} = w_k + e_k \quad (2.29)$$

$$E[e_k e_k^T] = P_{k/k-1} \quad (2.30)$$

A nova medida z_k é relacionada com o estado w_k pela equação

$$z_k = H_k w_k + \varepsilon_k \quad (2.31)$$

com covariância do erro $E[\varepsilon_k \varepsilon_k^T] = R_k$

As equações (2.30) e (2.31) podem ser agrupadas no seguinte sistema

$$\mathbf{z} = \mathbf{H}\mathbf{w} + \boldsymbol{\eta} \quad (2.32)$$

onde

$$\mathbf{z} = \begin{bmatrix} \hat{w}_{k/k-1} \\ z_k \end{bmatrix} \quad (2.33)$$

$$\mathbf{H} = \begin{bmatrix} I \\ H_k \end{bmatrix} \quad (2.34)$$

$$\boldsymbol{\eta} = \begin{bmatrix} e_k \\ \eta_k \end{bmatrix} \quad (2.35)$$

e covariância de $\boldsymbol{\eta}$

$$\mathbf{R} = \begin{bmatrix} P_{k/k-1} & 0 \\ 0 & R_k \end{bmatrix} \quad (2.36)$$

A solução deste sistema é

$$\hat{\mathbf{w}} = \mathbf{P}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z} \quad (2.37)$$

$$\mathbf{P} = (\mathbf{H}\mathbf{R}^{-1}\mathbf{H})^{-1} \quad (2.38)$$

As equações (2.37) e (2.38) representam o estágio de atualização do FK; que são as mesmas equações dos mínimos quadrados ponderados, que na forma recursiva resultam em

$$\hat{w}_{k/k} = \hat{w}_{k/k-1} + P_{k/k}H_k^T R_k^{-1}(z_k - H_k w_{k/k-1}) \quad (2.39)$$

$$P_{k/k}^{-1} = P_{k/k-1}^{-1}H_k^T R_k^{-1}H_k \quad (2.40)$$

onde

$$G_k = P_{k/k}H_k^T R_k^{-1} \quad (2.41)$$

é o ganho de Kalman.

O ganho de Kalman especifica a quantidade pela qual a inovação deve ser multiplicada (ou amplificada) para se obter o termo de correção e transformar a estimativa \hat{w}_k na nova estimativa $\hat{w}_{k/k}$.

Propagação

Desde que o novo estado seja relacionado com o estado anterior do FK, a matriz de covariância propagada é

$$P_{k+1/k} = E[(w_{k+1} - \hat{w}_{k+1/k})(w_{k+1} - \hat{w}_{k+1/k})^T]. \quad (2.42)$$

Abrindo a expressão acima, tem-se

$$P_{k+1/k} = E[(F_k w_k + G_k w_k - F_k \hat{w}_{k/k})(F_k w_k + G_k w_k - F_k \hat{w}_{k/k})^T]. \quad (2.43)$$

Desenvolvendo os produtos, usando a propriedade $(AB)^T = B^T A^T$, considerando F_k e G_k conhecidos e não aleatórios, lembrando que os ruídos são gaussianos não correlacionados de média nula e matriz de covariância conhecida, tem-se que

$$P_{k+1/k} = F_k P_k F_k^T + G_k Q_k G_k^T. \quad (2.44)$$

O FK usa uma estimativa inicial e matriz de covariância do erro inicial $\hat{w}_{0/k-1} \cong \hat{w}_0$ e $\hat{P}_{0/k-1} \cong \hat{P}_0$, ambos dados como parâmetros iniciais do problema.

O algoritmo do filtro de Kalman linear, num contexto de assimilação de dados, segundo Nowosad (2001), é apresentado a seguir e ilustrado na Figura 2.1:

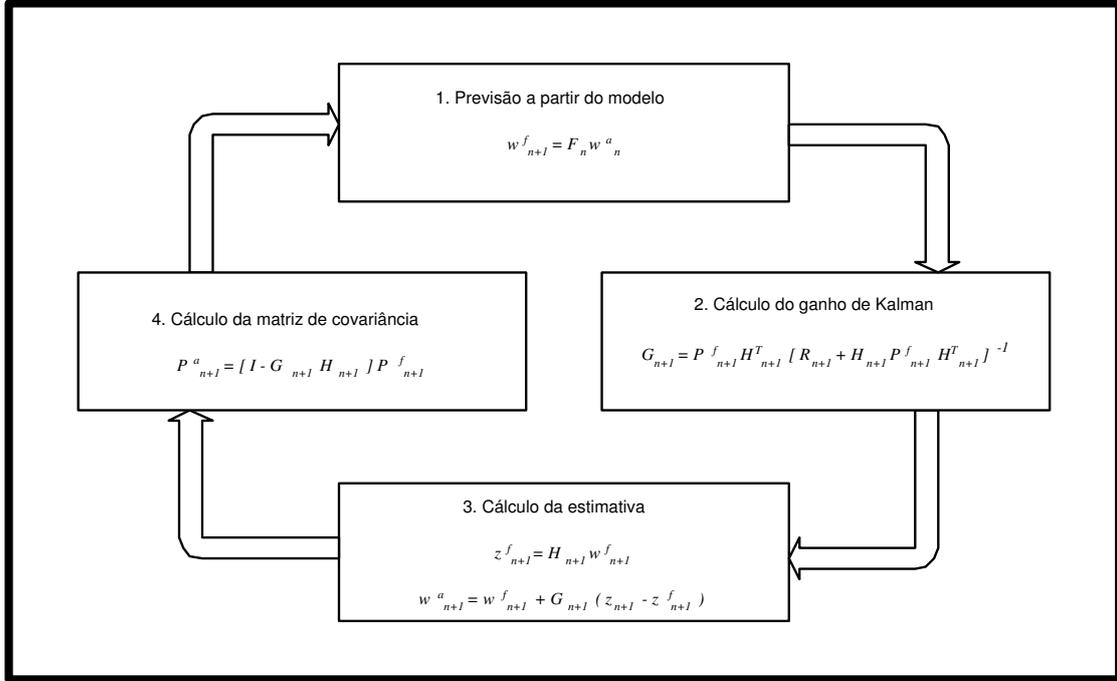


FIGURA 2.1 - Diagrama esquemático do filtro de Kalman linear.

a) Previsão a partir do modelo

$$w_{k+1}^f = F_{k+1} w_k^a \quad (2.45)$$

$$P_{k+1}^f = F_k P_k^a F_k^T + Q_k \quad (2.46)$$

b) Cálculo do ganho de Kalman

$$G_{k+1} = P_{k+1}^f H_{k+1}^T [R_{k+1} + H_{k+1} P_{k+1}^f H_{k+1}^T]^{-1} \quad (2.47)$$

c) Cálculo da estimativa

$$z_{k+1}^f = H_{k+1} w_{k+1}^f \quad (2.48)$$

$$w_{k+1}^a = w_{k+1}^f + G_{k+1} (z_{k+1} - z_{k+1}^f) \quad (2.49)$$

d) Cálculo da matriz de covariância

$$P_{k+1}^a = [I - G_{k+1}H_{k+1}]P_{k+1}^f \quad (2.50)$$

O sobreíndice **f** (equivalente à fase de propagação) refere-se a previsão (forecasting) e **a** (referente a fase de atualização) refere-se a assimilação (assimilation).

Nas expressões acima, para aliviar a notação substituiu-se $k+1/k$ por $k+1$ e k/k por k . A expressão (2.47) é equivalente a (2.41), sendo que \mathbf{R}^{-1} é da ordem do vetor de medidas e $[\mathbf{R} + \mathbf{HPH}]^{-1}$ é da ordem do vetor de estados. Deve-se optar por inverter a matriz de menor ordem. Para passar da expressão (2.41) para a expressão (2.47) usa-se o lema da inversão matricial. A equação (2.49) é obtida substituindo-se o valor do ganho de Kalman na equação (2.39) e a expressão (2.50) é obtida substituindo-se o valor de P^f em (2.40).

Quando aplicado a modelos não-lineares $F_{k+1} = F_k(w_k^a)$ (a matriz \mathbf{F} é função das variáveis de estado), lineariza-se \mathbf{F} e tem-se o filtro de Kalman Estendido; modificação mais usada do filtro de Kalman para problemas não-lineares.

A principal diferença entre o MQPR para o FK é que os MQPR foi desenvolvido para parâmetros e o FK é mais genérico. Kalman trata o ruído aleatório e prova que é o melhor estimador linear não tendencioso.

Em operações práticas de previsão de tempo e clima para a atmosfera e oceano, a matriz P_k^f não é calculada por $P_{k+1}^f = F_k P_k^a F_k^T + Q_k$, mas parametrizada baseando-se em avaliações estatísticas e vínculos dinâmicos (Nowosad, 2001). P_k^f e P_k^a também podem ser estimadas por médias de conjuntos (ensemble), Keppene (2000) .

2.6 Um teste para o filtro de Kalman em Assimilação de Dados

Apresenta-se uma aplicação do FK, num problema de assimilação de dados em dispersão de poluentes, governado pela equação de advecção-difusão. Por se tratar de um problema linear, observa-se claramente o funcionamento do filtro e o impacto causado pela inserção do dado. Esta aplicação pode ser encontrada com mais detalhes em Härter et al. (2002).

2.6.1 Descrição do Modelo de Advecção-Difusão

Considerando o transporte de poluentes atmosféricos na direção x , o fenômeno pode ser modelado pela seguinte equação de advecção-difusão e condições de contorno para $z = 0$ e $z = h$

$$U \frac{\partial \mathbf{w}}{\partial \mathbf{x}} = \frac{\partial}{\partial z} \left(K_{zz} \frac{\partial \mathbf{w}}{\partial z} \right), \quad (2.51)$$

$$K_{zz} \frac{\partial \mathbf{w}}{\partial z} = 0, \quad em \quad z = 0 \quad e \quad z = h \quad (2.52)$$

onde \mathbf{w} é a concentração de poluente, \mathbf{U} é a velocidade média do vento, \mathbf{K}_{zz} é a difusividade vertical turbulenta para a camada limite estável, segundo a modelagem de Degrazia e Moraes (1992), h é a altura da camada limite. A equação diferencial parcial foi aproximada em diferenças finitas pelo método de Euler explícito na horizontal e centrado na vertical (operador de difusão) (Hoffman, 1993). Definindo $d = \Delta x / U \Delta z^2$ a equação (2.51) em diferenças finitas pode ser expressa como

$$\mathbf{w}_{n+1} = F_n \mathbf{w}_n \quad (2.53)$$

com a seguinte matriz de dinâmica do sistema $F_n =$

$$\begin{bmatrix} [1 - d(k_{i+1/2} + k_{i-1/2})] & 2dk_{i+1/2} & & & \\ dk_{i-1/2} & [1 - d(k_{i+1/2} + k_{i-1/2})] & dk_{i+1/2} & & \\ & dk_{i-1/2} & [1 - d(k_{i+1/2} + k_{i-1/2})] & dk_{i+1/2} & \\ & \vdots & \vdots & \vdots & \\ & & 2dk_{i-1/2} & & [1 - d(k_{i+1/2} + k_{i-1/2})] \end{bmatrix} \quad (2.54)$$

com grade vertical $i = 1, \dots, N_z$. O índice n refere-se a posição na direção x , começando em $n = 0$. A aproximação por diferenças finitas é estável para $d \leq 0.3$.

2.6.2 Experimentos Numéricos

Com o propósito de testar o esquema de assimilação, os seguintes parâmetros foram usados: $\Delta \mathbf{x} = \mathbf{6}$ m, $\Delta z = \mathbf{10}$ m, $\mathbf{U} = \mathbf{0.31}$ ms^{-1} e $\mathbf{h} = \mathbf{400}$ m. A matriz que representa o sistema de observações é $\mathbf{H} = \mathbf{I}$, e as matrizes de covariância dos ruídos de observação e modelagem, são respectivamente $\mathbf{R}_n = \mathbf{2I}$ e $\mathbf{Q}_n = \mathbf{0.5I}$, sendo que

\mathbf{I} é a matriz identidade da ordem de $N_z = 41$. O número de pontos na direção x é 2000. Assume-se que os valores das concentrações observadas são dados pela equação (2.53) adicionados a uma fonte estocástica constante de poluição (com média zero).

Este problema de difusão simula uma pluma liberada na origem do sistema de coordenadas, sendo que esta condição é modelada pela seguinte função delta para $\mathbf{x} = \mathbf{0}$:

$$\mathbf{w}(\mathbf{x}, z) = \mathbf{Q}\Delta(z). \quad (2.55)$$

Com o objetivo de analisar a influência da frequência das observações na performance do FK integrou-se a equação (2.51) e assimilou-se observações nos 41 pontos de grade. Ressalta-se que o problema estudado é estacionário, com integração na direção x , conforme ilustrado na Figura 2.4, anexada ao texto na última página deste capítulo. Fez-se os seguintes experimentos: observações inseridas a cada $1 \Delta x$ (EXP1), $100 \Delta x$ (EXP2), $200 \Delta x$ (EXP3), $300 \Delta x$ (EXP4), $500 \Delta x$ (EXP5) e $700 \Delta x$ (EXP6). Na Tabela 2.1 constam os erros relativos a cada experimento e nas Figuras 2.2 (a)-(c), apresenta-se respectivamente os EXP3, EXP5, EXP6 para $Z = 150\text{m}$. O cálculo do erro é dado pela equação (2.56) e os símbolos \mathbf{w}_f , \mathbf{z} e \mathbf{w}_a , referem-se respectivamente à concentração predita pelo modelo, observada e estimada pelo FK.

$$Erro = \frac{1}{N_x N_z} \sum_{n=1}^{N_x} \sum_{i=0}^{N_z} (\mathbf{w}_{i,n}^z - \mathbf{w}_{i,n}^{w_a})^2 \quad (2.56)$$

TABELA 2.1 - Erro Relativo aos Experimentos 1 2 3 4 5 6.

EXP	Freq. de OBS	ERRO
1	cada $1 \Delta x$	0,328
2	cada $100 \Delta x$	7,236
3	cada $200 \Delta x$	11,880
4	cada $300 \Delta x$	13,877
5	cada $500 \Delta x$	21,627
6	cada $700 \Delta x$	31,779

Claramente, a estimativa feita pelo FK é melhor quando o dado é assimilado com maior frequência. Contudo, o custo computacional é maior para dados melhor amostrados.

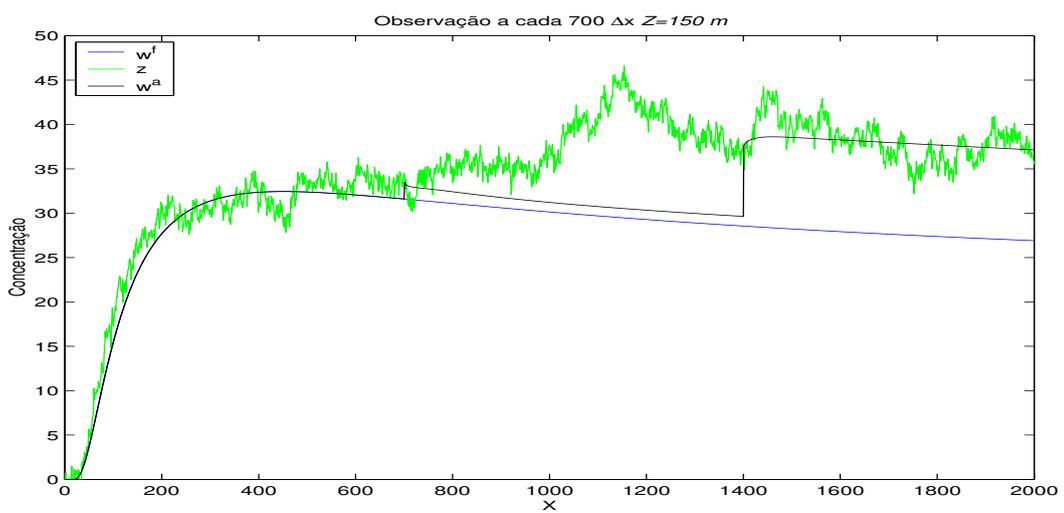
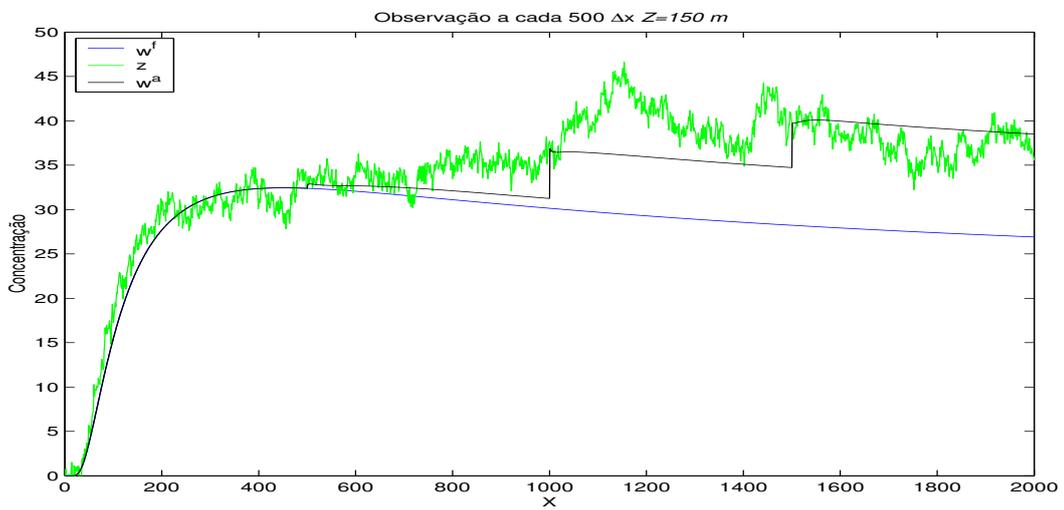
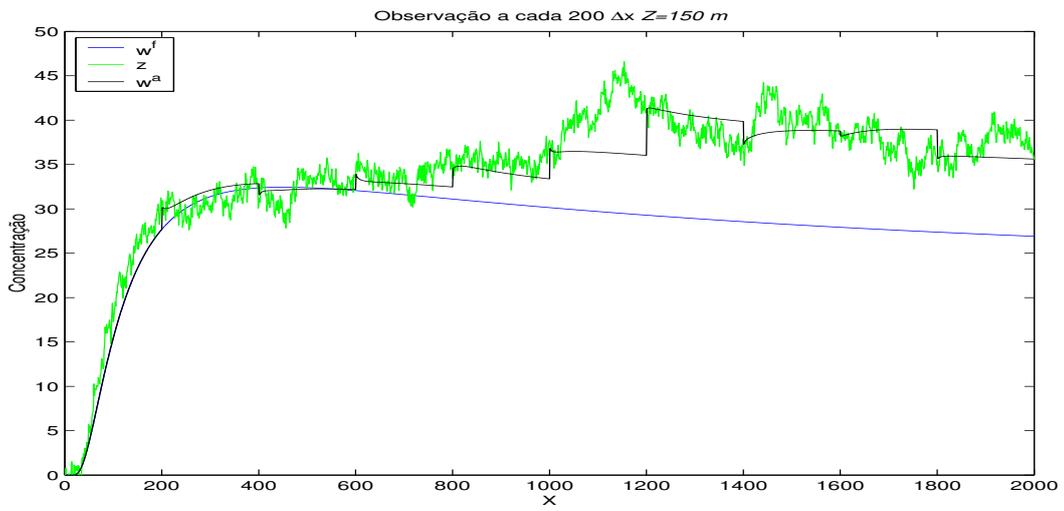


FIGURA 2.2 - Efeito da frequência de observação no FK: (a) EXP3, (b) EXP5, (c) EXP6.

O processo de assimilação também foi avaliado em relação ao número de pontos de medidas na coordenada vertical. Nesta classe de experimentos, as observações foram amostradas a cada $100\Delta x$. No EXP7 foi usado $N_m = 9$ para as seguintes posições: $z=0, 50, 100, 150, 200, 250, 300, 350, 400$ m; para o EXP8 foi usado $N_m = 7$ para as seguintes posições: $z=0, 100, 150, 200, 250, 300, 400$ m; e no EXP9 foi usado $N_m = 5$ para as seguintes posições: $z=0, 100, 200, 300, 400$ m. Os resultados são apresentados na Tabela 2.2 e Figura 2.3.

TABELA 2.2 - Erro para diferentes números de sensores na direção - z .

EXP	Número de sensores na direção- z	ERRO
7	$N_m = 9$	22,377
8	$N_m = 7$	44,650
9	$N_m = 5$	77,933

Na Tabela 2.2 e Figura 2.3 observa-se que o erro diminui à medida que o número de níveis com observações aumenta, para observações uniformemente distribuídas em x .

TABELA 2.3 - Erro para três arranjos de observações na direção- z .

EXP	Tipo de Grade	ERRO
10	Grade-1	77,933
11	Grade-2	355,897
12	Grade-3	394,200

Por fim analisa-se a performance do filtro para dados amostrados a todo Δx , com cinco pontos de observação em z ($N_m = 5$). Três diferentes arranjos são considerados: Grade-1, sensores uniformemente distribuídos (EXP10), Grade-2 com sensores posicionados próximos ao chão ($z=0, 20, 40, 60, 80$) m (EXP11) e Grade-3 com sensores posicionados próximos ao topo da camada limite ($z=320, 340, 360, 380, 400$) m (EXP12). Na Tabela 2.3 apresenta-se os erros destes diferentes arranjos.

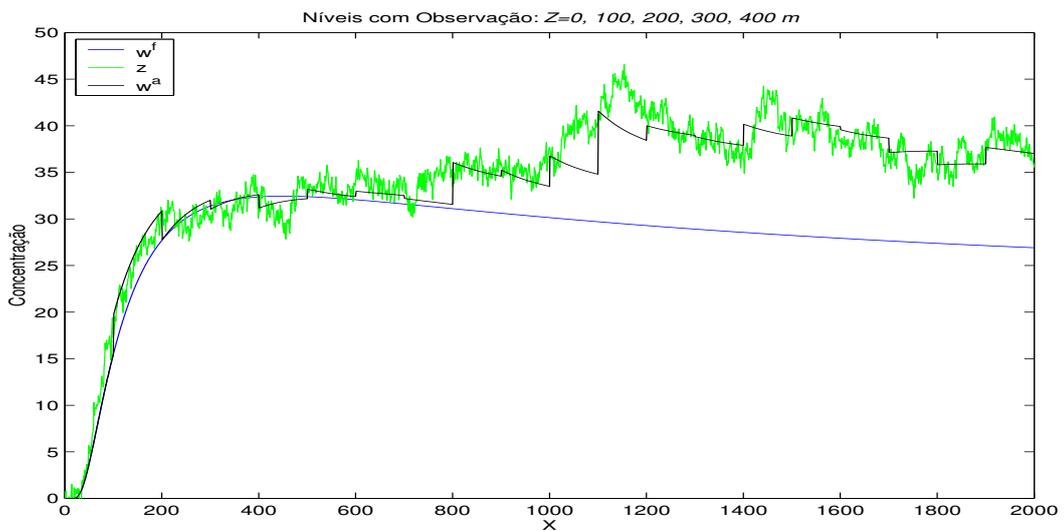
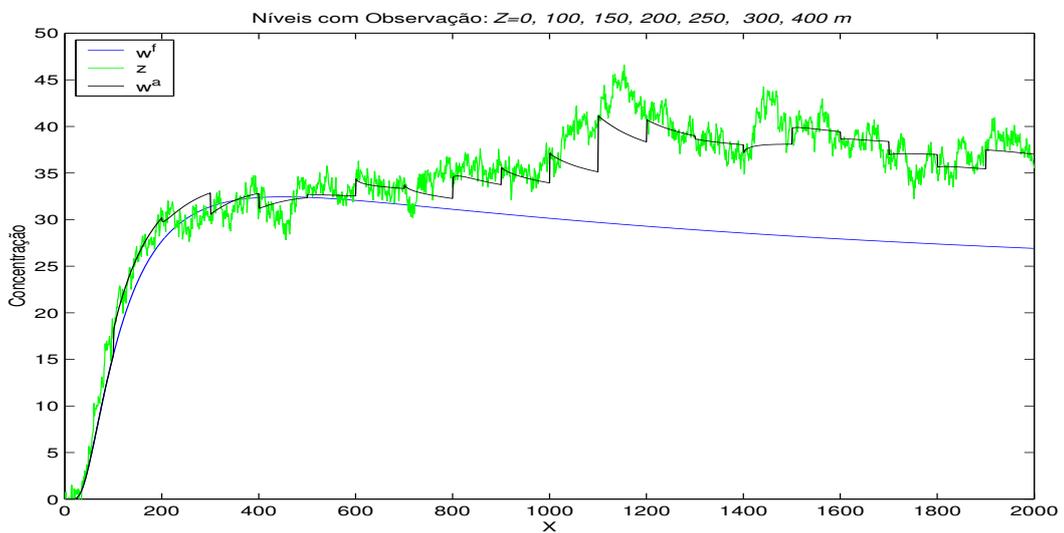
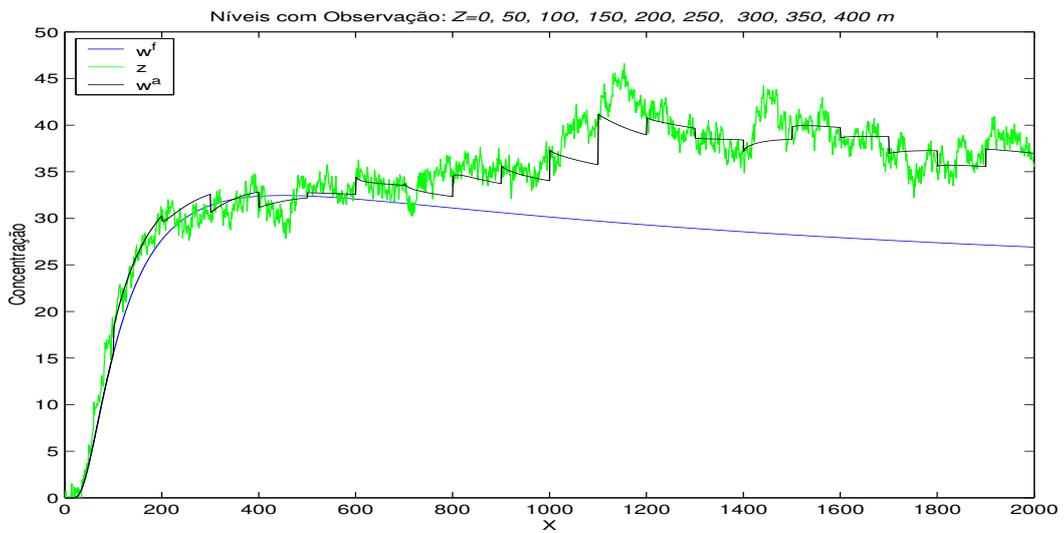


FIGURA 2.3 - Assimilação com observações distribuídas em diferentes arranjos: (a) EXP10, (b) EXP11, (c) EXP12.

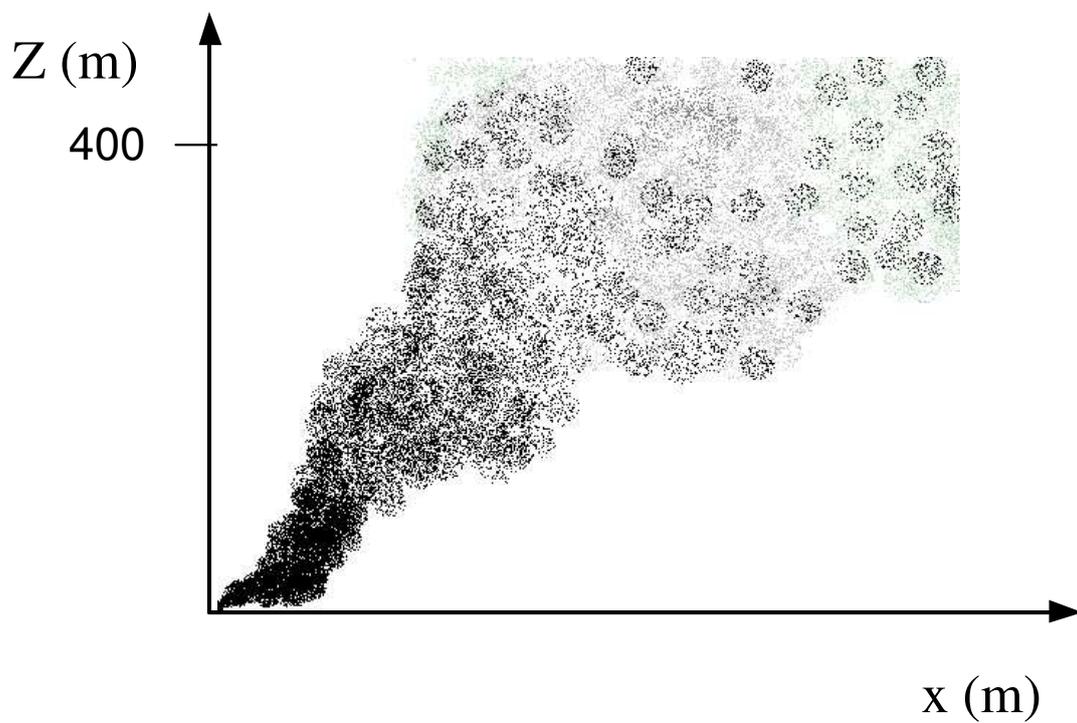


FIGURA 2.4 - Ilustração do problema estacionário de difusão, que simula uma pluma liberada na origem do sistema de coordenadas.

CAPÍTULO 3

REDES NEURAIS

Neste capítulo é descrito a teoria das Redes Neurais Artificiais, tratadas ao longo do texto apenas como Redes Neurais, inspiradas nas Redes Neurais Biológicas. Apresenta-se também uma breve revisão sobre o desenvolvimento desta área da ciência ao longo das últimas décadas. Na seqüência é apresentado o modelo do neurônio biológico e artificial. Discute-se as arquiteturas das RN exploradas neste texto. Descreve-se o funcionamento das RN, onde se aborda as fases de treinamento e ativação, bem como os paradigmas e regras de aprendizagem que aparecem na literatura com maior freqüência. Uma seção é reservada à discríção das redes PMC, FBR, RN-E e RN-J. Neste capítulo também se discute algumas razões que motivam o emprego de RN e algumas dificuldades encontradas ao se aplicar esta ferramenta.

3.1 Resumo Histórico sobre a Evolução das Pesquisas em Redes Neurais

McCulloch e Pitts (1943) introduziram a idéia de redes neurais como máquinas computacionais e o artigo destes autores intitulado “*A Logical Calculus Of the Ideas Immanent in Nervous Activity*”, passou a ser a primeira referência em redes neurais artificiais. O neurônio de MacCulloch e Pitts é um dispositivo binário, no qual a saída é pulso ou não pulso, sendo que suas entradas tem ganho arbitrário, podendo ser excitatório ou inibitório. Para determinar a saída do neurônio, calcula-se a soma ponderada das entradas com os respectivos pesos como fatores de ponderação, positivos nos casos excitatórios e negativos nos casos inibitórios. Se este resultado for maior ou igual a certo limiar então a saída do neurônio é pulso, e caso contrário é não pulso.

Segundo Kovacs (1996), o Modelo de McCulloch e Pitts, embora seja rudimentar quando comparado ao potencial dos modelos hoje disponíveis, foi inovador e de natureza seminal. Registra-se nos anais da ciência, que vários fundadores de algumas das assim chamadas modernas áreas de conhecimento, como Marvim Minski em inteligência artificial, John von Neumann em ciência da computação e Norbert Wiener em cibernética, tiveram inspiração em algum momento de suas carreiras no trabalho de McCulloch e Pitts.

O próximo desenvolvimento significativo em RN foi dado em Hebb (1949), onde é postulada a primeira regra de aprendizagem auto-organizada, baseada na seguinte

premissa: se dois neurônios em ambos os lados de uma sinapse (conexão) são ativados simultaneamente então a sinapse deve ser fortalecida, senão a sinapse deve ser enfraquecida ou eliminada.

Outro passo importante no desenvolvimento de RN foi dado em Rosenblatt (1958), que propôs o perceptron como o primeiro modelo para o aprendizado com um professor (aprendizagem supervisionada). Rosenblatt demonstrou com o perceptron, que se as RN com nodo fossem acrescidas de sinapses ajustáveis, poderiam ser treinadas para classificar padrões. Rosenblatt descreveu uma topologia de RN, estrutura de ligações entre os nodos e propôs um novo algoritmo para treinar a rede para executar determinados tipos de funções. O perceptron simples descrito por Rosenblatt possui três camadas: a primeira recebe as entradas do exterior e possui conexões fixas, a segunda recebe impulsos da primeira através de conexões cuja eficiência de transmissão (peso) é ajustável e, por sua vez, envia saída para a terceira camada (resposta). Este tipo elementar de perceptron comporta-se como um classificador de padrões, dividindo o espaço de entradas em regiões distintas para cada uma das classes existentes. O perceptron é capaz apenas de separar padrões que sejam linearmente separáveis.

Simultaneamente ao trabalho de Rosenblatt, Widrow (1962) desenvolveu um modelo linear conceitualmente simples batizado de ADALINE (*ADaptive LINear Element*) e mais tarde sua generalização multidimensional, o MADALINE (Múltipla Adaline). Apesar disso, a contribuição realmente importante do trabalho de Widrow foi a criação de um princípio de treinamento extremamente eficiente para as redes Adaline, conhecido como Regra Delta, que foi mais tarde generalizado para redes com modelos neurais mais elaborados.

Minski e Papert (1969) , afirmaram que não havia razão para supor que qualquer uma das limitações do perceptron de camada única poderia ser superada na versão de múltiplas camadas. Isto fez com que os estudos em RNs passassem por uma década de letargia.

Hopfield (1982) formulou um modelo de rede recorrente com conexões sinápticas simétricas, supondo que os sistemas devam convergir para um estado de mínima energia. O modelo de Hopfield é usualmente referenciado na literatura como sinônimo de rede recorrente. Foi neste trabalho que pela primeira vez o princípio do armazenamento de informação em redes dinamicamente estáveis foi explícito. Assim, o trabalho de Hopfield dá um novo impulso ao estudo de RN.

Em 1983 foi desenvolvido um novo procedimento denominado recozimento simulado para resolver problemas de otimização combinatória. Esta idéia levou ao desenvolvimento da máquina de Boltzman (Ackley et al., 1985), que foi a primeira realização bem sucedida de uma rede neural de múltiplas camadas. Como destacado por Haykin (2001), apesar de o algoritmo de aprendizagem da máquina de Boltzman não ter se mostrado tão eficiente do ponto de vista computacional como o algoritmo de retropropagação, ele superou o impasse psicológico, mostrando que a especulação de Minski e Papert não estava corretamente embasada.

Rumelhart et al. (1986) propuseram a utilização do algoritmo de retropropagação em aprendizagem de máquina, que tornou-se o algoritmo de aprendizagem mais popular para o treinamento do PMC.

3.2 Neurônio Biológico

Nesta seção identifica-se algumas unidades básicas do neurônio artificial e descreve-se sua funcionalidade com o intuito de mostrar a analogia do neurônio artificial, descrito na próxima seção, com este modelo biológico.

O neurônio ou célula nervosa, ilustrado na Figura 3.1, aparece em diferentes formas e tamanhos no cérebro humano, sendo composto basicamente por dendritos, axiônio, sinapses e núcleo ou soma. A partir do corpo celular ou soma, centro dos processos metabólicos da célula nervosa, projetam-se extensões filamentosas, os dendritos e o axiônio. Os dendritos muitas vezes cobrem um volume maior do que o corpo celular. O axiônio ou fibra nervosa, serve para conectar a célula nervosa a outras células do sistema nervoso. O neurônio possui geralmente um único axiônio, embora possa apresentar algumas ramificações e em alguns casos, estendem-se por distâncias comparáveis as dimensões do organismo, podendo chegar a vários metros. Alguns tipos de neurônios possuem uma capa segmentada de mielina, que serve para acelerar a transmissão da informação pelo axiônio.

A atividade do neurônio biológico é assim descrita por Kovacs (1996): As sinapses são regiões eletroquimicamente ativas, compreendidas entre duas membranas celulares: a membrana pré-sináptica, por onde chega o estímulo proveniente de outra célula, e a membrana pós-sináptica, que é a membrana do dendrito. Nesta região intersináptica, o estímulo nervoso que chega a sinapse é transferido à membrana dendrital através de substâncias conhecidas como neurotransmissores. O resultado desta transferência

é uma alteração no potencial elétrico da membrana pós-sináptica. Dependendo do tipo do neurotransmissor, a conexão sináptica será excitatória ou enibitória. Uma conexão excitatória provoca uma alteração no potencial da membrana que contribui para a formação de um impulso nervoso no axiônio de saída, enquanto que uma conexão inibitória age no sentido oposto.

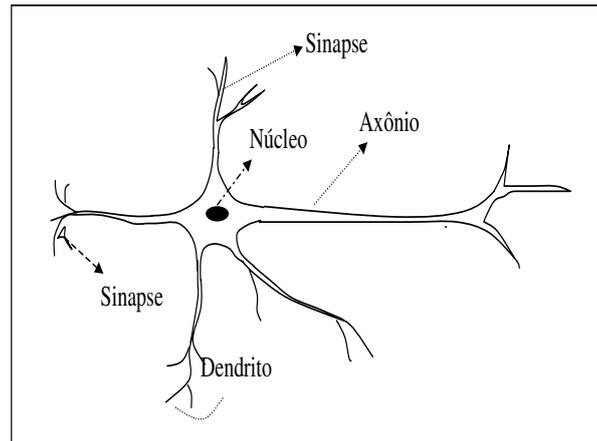


FIGURA 3.1 - Representação do neurônio biológico.

3.3 Neurônio Artificial

Um neurônio é uma unidade de processamento de informação fundamental para a operação de uma RN (Haykin, 2001). Identifica-se três elementos básicos do modelo neural:

- um conjunto de sinapses ou elos de conexão, cada uma caracterizada por um peso próprio. Especificamente, um sinal x_j na entrada da sinapse j conectada ao neurônio k é multiplicada pelo peso sináptico θ_{kj} ,
- um somador para somar os sinais de entrada, ponderados pelas respectivas sinapses do neurônio,
- uma função de ativação para limitar a amplitude da saída de um neurônio. Tipicamente a faixa de amplitude da saída de um neurônio é o intervalo $[0, 1]$ ou alternativamente $[-1, 1]$.

No modelo de neurônio artificial mostrado na Figura 3.2, pode-se opcionalmente, incluir um limiar b_k aplicado externamente que tem o efeito de acrescentar um grau

de liberdade a cada neurônio.

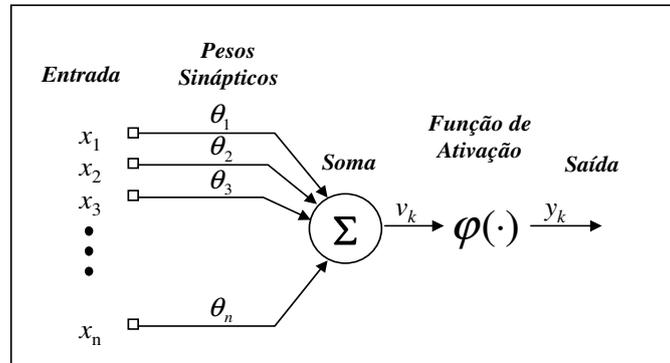


FIGURA 3.2 - Representação do neurônio artificial.

Matematicamente pode-se escrever um neurônio k através do seguinte par de equações

$$u_k = \sum_{j=1}^m \theta_{kj} x_j, \quad (3.1)$$

$$y_k = \varphi(u_k + b_k) \quad (3.2)$$

onde x_1, \dots, x_m são sinais de entrada; $\theta_{k1}, \dots, \theta_{km}$ são os pesos sinápticos do neurônio k ; u_k é a saída do combinador linear devido aos sinais de entrada; b_k é o viés; $\varphi(\cdot)$, é a função de ativação e y_k é o sinal de saída do neurônio. O uso do viés ou limiar b_k tem o efeito de aplicar uma transformação afim à saída u_k do combinador linear do modelo da Figura 3.2, conforme

$$v_k = u_k + b_k. \quad (3.3)$$

3.4 Arquiteturas das Redes Neurais

As diferentes arquiteturas de RN são formadas pela combinação de neurônios artificiais. Os neurônios podem ser dispostos de maneira a formar uma ou mais camadas. Em parte da literatura as entradas são consideradas a primeira camada da

rede. Neste texto as entradas serão denominadas de camada de entrada, a primeira camada sem contato com o meio externo de primeira camada escondida e assim sucessivamente para as demais camadas escondidas. Os valores na saída da rede são denominados de camada de saída.

Nas redes em que as entradas projetam-se diretamente sobre a camada de saída, como na Figura 3.2, a rede é chamada de “feed-forward” (alimentação para frente). Quando há realimentação entre uma das camadas da RN, a rede é chamada de recorrente.

Destaca-se ainda na arquitetura das RN, os tipos de conexão entre os neurônios, ou seja, diz-se que a rede é totalmente conectada quando todas as saídas de uma camada estão conectadas em todos os neurônios da próxima camada. Na falta de alguma dessas conexões se diz que a rede é parcialmente conectada.

A definição da arquitetura de uma RN é um parâmetro importante na sua concepção, uma vez que ela restringe o tipo de problema que pode ser tratado pela rede. RN com uma única camada de nodos, por exemplo, só consegue resolver problemas linearmente separáveis. Redes recorrentes, por sua vez, são mais apropriadas para resolver problemas que envolvam processamento temporal.

3.5 Funcionamento das Redes Neurais

Em RN, o procedimento usual na solução de problemas passa inicialmente por uma fase de aprendizado, também conhecida como treinamento, em que um conjunto de exemplos é apresentado para a rede, a qual extrai as características necessárias para representar a informação fornecida. Estas características, armazenadas nos pesos sinápticos, são utilizadas posteriormente na fase conhecida como ativação, para gerar respostas para o problema. Neste trabalho a resposta fornecida pela rede é o sinal estimado ou variável assimilada.

O treinamento pode ser feito por lote, onde os pesos são atualizados após todos os padrões de treinamento serem apresentados a rede, ou pode ser por padrão, onde os pesos são atualizados à medida que cada padrão de treinamento (também denominado exemplo de treinamento) é apresentado a rede.

Um ponto crucial no funcionamento das RN é adotar um critério de parada para o treinamento. Geralmente a rede é treinada até que um número máximo de épocas

de treinamento, definido *a priori*, seja atingido ou que um erro objetivo seja alcançado. Define-se uma época de treinamento a apresentação de todo o conjunto (ou padrões) de treinamento a rede. O critério de parada em função do erro objetivo pode ser aplicado com paradigma de aprendizagem supervisionada, definido na próxima seção. Na apresentação dos resultados (capítulo 5) será discutido o critério de parada adotado na implementação das RN estudadas neste trabalho.

3.6 Função de Ativação

A função de ativação restringe a amplitude do sinal na saída de um neurônio. Geralmente se restringe o sinal ao intervalo $[0,1]$ ou ao intervalo $[-1,1]$. A seguir apresentase as funções de ativação aplicadas na camada oculta das RN exploradas neste trabalho. A função de ativação é representada por $\varphi(v)$, onde v é campo local induzido, (Figura 3.3),

Função Tangente Hiperbólica - implementada na camada oculta da rede PMC, RN-E e RN-J com $a = 1$

$$\varphi(v_j) = \tanh\left(\frac{av_j}{2}\right) \quad (3.4)$$

Função Exponencial - implementada na camada oculta da rede FBR com $\sigma = 1$ e $\mu = 0$

$$\varphi(v_j) = \exp\left[-\frac{(v_j - \mu)^2}{2\sigma^2}\right] \quad (3.5)$$

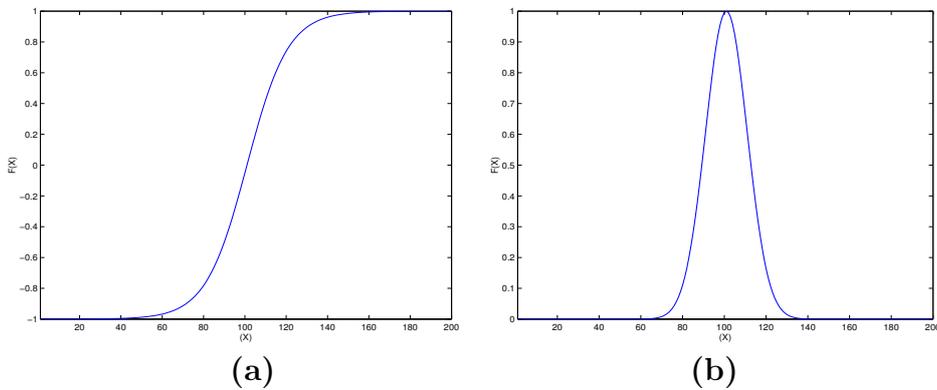


FIGURA 3.3 - Função de ativação (a) sigmóide; (b) gaussiana.

Os valores na entrada devem ser normalizadas para valores restritos ao intervalo da função de ativação utilizada e, posteriormente, deve ser feito o processo inverso para que as variáveis estimadas voltem ao intervalo da função original.

3.7 Paradigmas e Regras de Aprendizagem

Um conjunto de procedimentos bem definidos para adaptar os parâmetros de uma rede neural para que a mesma possa aprender uma determinada função é chamado de algoritmo de aprendizagem. Os algoritmos de aprendizagem seguem basicamente dois paradigmas, o supervisionado e o não-supervisionado. No paradigma supervisionado, tanto a entrada como a saída da RN são conhecidas, sendo que a rede opera no sentido de encontrar pesos que minimizem a diferença entre a entrada e a resposta desejada num sentido estatístico, geralmente o erro médio quadrático.

Na aprendizagem não-supervisionada somente os padrões de entrada estão disponíveis para a rede, ao contrário da aprendizagem supervisionada, cujo conjunto de treinamento possui pares de entrada-saída. A partir do momento em que a rede estabelece uma harmonia com as regularidades estatísticas da entrada de dados, desenvolve-se nela uma habilidade de formar representações internas para codificar características da entrada e criar novas classes ou grupos automaticamente (Braga et al., 2000).

As regras comumente aplicadas com o paradigma supervisionado são correção de erros e a regra delta, que foi generalizada para o treinamento do perceptron de múltiplas camadas, conhecido como algoritmo de retropropagação do erro. O algoritmo de retropropagação é a mais poderosa técnica para o aprendizado do perceptron de múltiplas camadas.

A seguir resume-se o algoritmo de retropropagação, segundo Haykin (2001), utilizado no treinamento das redes PMC, FBR, RN-E e RN-J.

1 - Iniciar os pesos: assumindo nenhuma informação prévia disponível, os pesos sinápticos e limiares são iniciados com números aleatórios que seguem uma distribuição uniforme.

2 - Apresentação dos exemplos de treinamento: apresenta-se uma época de exemplos de treinamento à rede. Para cada exemplo do conjunto de treinamento executa-se os passos 3 e 4 apresentados a seguir.

3 - Cálculo para Frente (Propagação): sendo um exemplo do conjunto de treinamento representado por $(x(n), d(n))$, com o vetor de entrada $x(n)$ aplicado a camada de entrada de nós sensoriais e o vetor resposta desejada $d(n)$ apresentado à camada de saída de nós computacionais, calcula-se os campos locais induzidos e os sinais funcionais camada por camada da rede. O campo local induzido $v_j^{(l)}(n)$ para o neurônio j na camada l é

$$v_j^l(n) = \sum_{i=0}^{m_0} \theta_{ji}^{(l)}(n) y_i^{l-1}(n), \quad (3.6)$$

onde $y_i^{(l-1)}(n)$ é o sinal de saída do neurônio i na camada interior $l - 1$, na iteração n , e $\theta_{ji}^{(l)}(n)$ é o peso sináptico do neurônio j da camada l , que é alimentado pelo neurônio i da camada $l - 1$. Para $i = 0$, temos $y_0^{(l-1)}(n) = +1$ e $\theta_{j0}^{(l)}(n) = b_j^{(l)}(n)$ é o viés aplicado ao neurônio j na camada l . O sinal de saída do neurônio j na camada l é

$$y_j^l = \varphi_j(v_j(n)). \quad (3.7)$$

Se o neurônio j está na primeira camada oculta (i.e., $l = 1$), faz-se $y_j^{(0)}(n) = x_j(n)$, $x_j(n)$ é o j -ésimo elemento do vetor de entrada $\mathbf{x}(\mathbf{n})$. Se o neurônio j está na camada de saída (i.e., $l = L$ onde L é denominado a profundidade da rede), se faz $y_j^{(L)} = o_j(n)$.

Calcula-se o sinal de erro

$$e_j^{(n)} = d_j^{(n)} - o_j^{(n)} \quad (3.8)$$

onde $d_j^{(n)}$ é o j -ésimo elemento do vetor resposta desejada $\mathbf{d}(\mathbf{n})$.

4 - Cálculo para Trás (Retropropagação): Calcula-se os gradientes locais da rede definidos por

$$\delta_j^{(l)}(n) = \begin{cases} e_j^{(L)}(n)\varphi_j^{(l)}(v_j^{(L)}(n)) & \text{para o neurônio } j \text{ na camada de saída } L, \\ \varphi_j^{(l)}(v_j^{(l)}(n)) \sum_k \delta_k^{(l+1)}(n)\theta_{kj}^{(l+1)}(n) & \text{para o neurônio } j \text{ na camada oculta } l. \end{cases} \quad (3.9)$$

onde o apóstrofo em $\varphi_j^{(l)}(\cdot)$ representa a diferenciação em relação ao argumento. Assim, ajusta-se os pesos sinápticos da rede na camada l de acordo com seguinte regra, conhecida como regra delta generalizada:

$$\theta_{ji}^{(n+1)} = \theta_{ji}^{(n)} + \alpha[\theta_{ji}^{(n)} - \theta_{ji}^{(n-1)}] + \eta\delta_j^{(l)}(n)y_i^{(l-1)}(n) \quad (3.10)$$

onde η é a taxa de aprendizagem e α é a constante de momento.

5 - Iteração: repete-se os itens 3 e 4 , apresentando a rede novos exemplos de treinamento até que o critério de parada seja satisfeito.

Neste trabalho, a ordem de apresentação dos exemplos de treinamento seguem sempre a mesma seqüência, ao invés de serem apresentados de maneira aleatória, de época para época, como sugerido em Haykin (2001). Os parâmetros de momento e a taxa de aprendizagem são constantes, mas diferem para cada experimento.

O paradigma não-supervisionado é comumente aplicado com os algoritmos de aprendizagem Hebbiana e competitivo. Durante o desenvolvimento deste trabalho, explorou-se a aprendizagem competitiva na implementação de Rede de Kohonen, ao se testar formas de calcular o centro das funções gaussianas para a FBR. Como os resultados foram insatisfatórios, esta regra de aprendizagem não será apresentada neste texto.

3.8 Redes *feedforward* e Recorrentes

Como mencionado na seção 3.4, as redes em que as entradas projetam-se diretamente sobre a camada de saída, são chamadas de *feedforward* (alimentação para frente) e quando há realimentação entre uma das camadas da RN, a rede é chamada de recorrente. Neste trabalho, cujo um dos objetivos é testar assimilação de dados em dinâmica não-linear com redes recorrentes, além do uso de redes *feedforward*, implementou-se em linguagem de programação FORTRAN77, as redes PMC, FBR,

3.8.2 Rede Função de Base Radial

Redes com uma única camada oculta são capazes de aproximar com uma precisão pré-definida qualquer função contínua, pois uma função contínua e limitada sobre um certo intervalo, pode ser considerada como uma superposição linear de gaussianas. Estas gaussianas por sua vez, podem ser implementadas na camada oculta da RN. A precisão da aproximação dependerá apenas do número de funções gaussianas, ou seja, do número de neurônios da camada oculta, Kovacs (1996).

Redes Neurais representando funções sintonizadas em torno de uma região e cujas funções de ativação, implementadas nos neurônios da camada oculta, são funções gaussianas, denominam-se Funções de Base Radial (FBR).

Do ponto de vista de uma FBR, aprender é equivalente a encontrar uma superfície, em um espaço multidimensional que forneça o melhor ajuste para os dados de treinamento, com o critério de melhor ajuste sendo medido num sentido estatístico (Bennett, 1992).

Neste trabalho, tanto a FBR como a PMC foram implementadas com aprendizagem por correção de erro (também conhecido como algoritmo de retropropagação do erro ou Regra Delta de Widrow) e paradigma de aprendizagem supervisionada. As redes são totalmente conectadas, apresentam uma camada de entrada, apenas uma camada escondida e a camada de saída. A rede FBR explorada neste trabalho difere da rede PMC apenas pelo fato da função de ativação ser gaussiana (com centros zero e variâncias 1) e não tangente hiperbólica. Tentativas de selecionar os centros das gaussianas pelo algoritmo de clusterização K-means e pela rede de Kohonem não foram bem sucedidas.

3.8.3 Rede de Elman e Rede de Jordan

A seguir descreve-se as RN-E e RN-J, conforme Braga et al. (2000). Nas RN-E, Figura 3.5 (a), além das unidades de entrada, intermediária e de saída há também unidades de contexto, como nas redes recorrentes em geral. As unidades de entrada e de saída interagem com o ambiente externo, enquanto as unidades intermediárias e de contexto não o fazem. As unidades de entrada são apenas unidades de armazenamento (buffer) que passam os sinais sem modificá-los. As unidades de saída são unidades lineares que somam os sinais que recebem. Neste trabalho as unidades intermediárias tem funções de ativação não-lineares, embora segundo a teoria

possam ser lineares. As unidades de contexto são usadas apenas para memorizar as ativações anteriores das unidades intermediárias e podem ser consideradas como atraso no tempo em um passo. As conexões *feedforward* são modificáveis, e as conexões recorrentes podem ser fixas, motivo pelo qual a RN-E é apenas parcialmente recorrente. Neste trabalho as conexões recorrentes também são modificáveis.

Em um intervalo de tempo específico k , as ativações das unidades intermediárias (em $k-1$) e as entradas correntes (em k) são utilizadas como entradas da rede. Em um primeiro estágio, *feedforward*, estas entradas são propagadas para frente para produzir as saídas. Posteriormente, a rede é treinada como o algoritmo de aprendizagem de retropropagação padrão. Após este passo de treinamento, as ativações das unidades intermediárias no tempo k são introduzidas, através das ligações recorrentes nas unidades de contexto, sendo salvas nestas unidades para o próximo passo de treinamento ($k+1$).

Na RN-J, Figura 3.5 (b), a saída da rede é copiada para a unidade de contexto. Além disso, as unidades de contexto são localmente recorrentes. A diferença em termos de topologia entre as duas redes é que a recorrência na RN-E é feita da camada escondida para as entradas, enquanto na RN-J a recorrência é feita das saídas para as entradas.

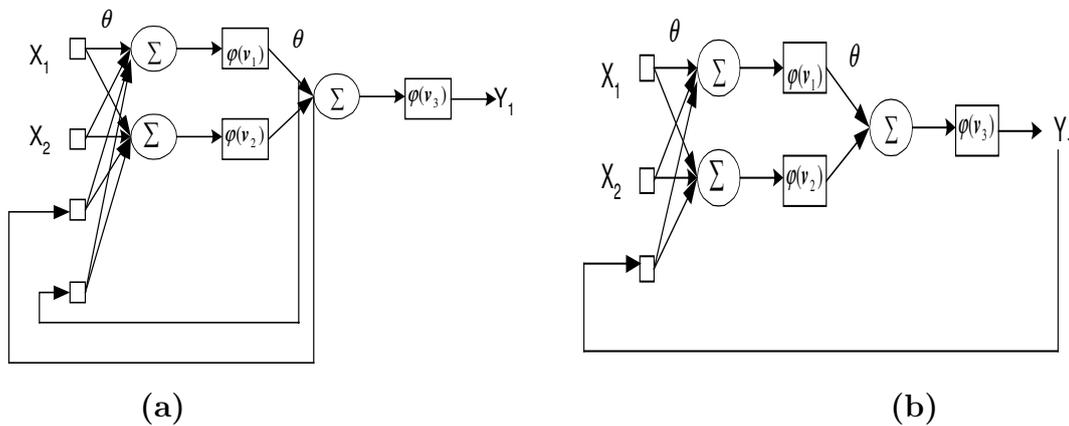


FIGURA 3.5 - Ilustração das redes (a) Elman; (b) Jordan.

3.9 Motivações e Limitações nos Estudos das Redes Neurais

Tem havido cada vez mais produção científica associada a pesquisas que utilizam RN, pois estas encontram aplicações em diversas áreas da ciência, tais como, associação e reconhecimento de padrões, aproximação de funções, controle e filtragem, entre outras.

Esta ampla aplicabilidade das RN, deve-se ao fato, pelo menos em parte, das redes serem uma boa alternativa à solução de problemas nos quais não se conhece, ou pouco se conhece, da física do problema. A literatura refere-se a esta característica das RN relatando que estas são um bom substituto a ignorância. Os métodos clássicos partem de um modelo numérico baseado na física do problema, que posteriormente é validado com dados observados, enquanto as RN são baseadas diretamente nos dados observados, sendo que no seu projeto não se faz necessário o conhecimento físico do problema e tampouco do modelo numérico. Todavia, anexar o conhecimento físico e um modelo numérico a um projeto de RN, o que acontece em neurodinâmica, como neste trabalho, faz das RN uma técnica de maior aplicabilidade.

Entre as principais propriedades e capacidades desta técnica, estão: não-linearidade, mapeamento de entrada-saída, adaptabilidade, aprender com exemplos, processamento paralelo e distribuído, capacidade de generalização e tolerância à falhas (Haikyn, 2001).

A capacidade de aprender através de exemplos e de generalizar a informação aprendida é, sem dúvida, o atrativo principal da solução do problema através de RN. A generalização, que esta associada a capacidade da rede aprender através de um conjunto relativamente reduzido de exemplos e posteriormente dar respostas coerentes para dados não conhecidos, é uma demonstração de que a capacidade das RN vai muito além do que simplesmente mapear relações de entrada e saída. As RN são capazes de extrair informações não apresentadas de forma explícita através de exemplos.

Não obstante, as RN são capazes de atuar como mapeadores universais de funções multivariáveis com custo computacional que cresce apenas linearmente com o número de variáveis. Outra característica importante é a capacidade de auto-organização e de processamento temporal que aliada aquelas citadas anteriormente, faz das redes uma ferramenta computacional extremamente poderosa e atrativa para

a solução de problemas complexos (Braga et al., 2000).

Apesar da ampla aplicabilidade das RN, deve-se levar em consideração que esta técnica não resolve todo o tipo de problema. Kovacs (1996) adverte que para fenômenos essencialmente aleatórios, fenômenos para os quais não é possível encontrar nenhum modelo que permita descrever seu comportamento sob condições arbitrárias, a única modelagem possível é a probabilística. Nesse sentido, quando se procura modelar com redes neurais, inadvertidamente, um processo essencialmente aleatório, como por exemplo os índices do mercado especulativo, pode-se até conseguir algum desempenho temporário, enquanto alguma estrutura oculta perdurar, mas o modelo será imprevisivelmente instável.

Uma dificuldade encontrada no uso de RN é definir o número de neurônios, de camadas escondidas, tamanho do conjunto de treinamento, taxa de aprendizagem, que pode ser fixa ou variável e constante de momento. A rede de Correlação em Cascata (Fahlman, 1991) propõe uma alternativa para definir o número de neurônios e camadas mais adequado para o problema a ser resolvido. Nesta rede o número de neurônios aumentam e diminuem conforme uma regra onde os neurônios competem entre si. Todavia sua implementação é mais difícil do que a implementação das redes exploradas neste trabalho e resultados parciais em problemas mais simples do que esta aplicação foram inferiores aos resultados obtidos com as redes PMC e FBR.

CAPÍTULO 4

SISTEMAS DE EVOLUÇÃO NÃO-LINEARES

A seguir descreve-se os modelos numéricos com os quais se explora a capacidade das redes em emularem o FK num sistema de assimilação de dados. O modelo de Lorenz em regime caótico é apresentado na seção 4.1, o modelo meteorológico DYNAMO - 1 D na seção 4.2 e o modelo de 3-ondas, que será explorado em regime periódico e caótico na seção 4.3.

4.1 Sistema de Lorenz Caótico

Saltzman (1962) interessado em movimentos atmosféricos de origem convectiva, desenvolveu um modelo simplificado (baseado nas equações do movimento, continuidade, e termodinâmica) para estudar o escoamento em uma camada de fluido com profundidade constante. Quando a diferença de temperatura entre as superfícies inferior e superior do fluido for mantida constante, o sistema permanece em equilíbrio. Neste caso se diz que o sistema possui solução estável. No entanto, se o sistema em questão for o sistema terra-atmosfera, pode-se descrever o processo que ocorre dentro da camada de fluido da seguinte maneira: da radiação incidente sobre a terra, emitida pelo sol, uma parte é absorvida e outra parte é refletida para a atmosfera. As camadas mais altas da atmosfera vão resfriando por perda radiativa para o espaço, de forma que na parte superior da atmosfera o ar fica mais frio e mais pesado do que nas camadas próximas a superfície, onde concentra-se o ar mais quente e mais leve. Este ar mais leve tende a subir, enquanto o mais pesado tende a descer na atmosfera. Esse problema de transporte, com camadas oscilantes de ar frio e quente, gera movimentos convectivos em diferentes escalas e características, podendo ser desde a turbulência térmica caótica até sistemas altamente organizados como furacões. Neste caso se diz que o sistema possui solução instável.

Lorenz (1963), mais interessando na natureza não periódica das soluções do modelo de Saltzman do que no problema convectivo, expandiu as variáveis de estado do modelo em série de Fourier e, retendo apenas os termos de baixa ordem, obteve o seguinte sistema acoplado de três equações diferenciais ordinárias de primeira ordem não-linear

$$dX/d\tau = -\sigma(X - Y), \quad (4.1)$$

$$dY/d\tau = rX - Y - XZ, \quad (4.2)$$

$$dZ/d\tau = XY - bZ \quad (4.3)$$

onde $\tau = \pi H^{-2}(1 + f^2)kt$ é o tempo adimensional, com H , f , k e t sendo respectivamente altura (ou profundidade) da camada, número de onda, condutividade térmica e tempo. $\sigma = k^{-1}\nu$ é o número de Prandtl, uma quantidade adimensional que depende da natureza do fluido e, em menor extensão, de sua temperatura; ν é a viscosidade cinemática do fluido. O parâmetro $r = \frac{R}{R_c}$ é o número de Rayleigh (representa a diferença de temperatura entre as superfícies), com $R = g\alpha H^3 \delta T \nu^{-1} k^{-1}$, sendo $R_c = \pi_4 f^{-2}(1 + f^2)^3$, onde g é a aceleração da gravidade e R_c o número de Rayleigh crítico. O mínimo valor de R_c é $27\frac{\pi^4}{4}$, ocorrendo quando $f^2 = 0.5$.

Nestas equações X é proporcional a intensidade do movimento convectivo, Y é proporcional a diferença de temperatura entre as correntes ascendentes e descendentes. Os sinais similares de X e Y significa que o fluido quente esta em movimento ascendente e o fluido frio em movimento descendente. A variável Z é proporcional a perturbação do perfil de temperatura vertical, sendo que valores positivos indicam forte gradiente próximo a fronteira.

As equações (4.1)-(4.3) apresentam resultados realísticos para r próximo a unidade, mas à medida que r difere da unidade, situações de forte movimento convectivo, as soluções afastam-se da realidade, devido ao erro de truncamento imposto por Lorenz.

O sistema de Lorenz foi integrado com o método preditor-corretor com $\Delta t = 10^{-3}$, resultado no seguinte conjunto de equações

$$W_{n+1}^p = W_n + \Delta t f_{n,w}, \quad (4.4)$$

$$W_{n+1}^c = W_n^c + \frac{1}{2}\Delta t(f_{n,w} + f_{n+1,w}^p) \quad (4.5)$$

onde os sobrescritos p e c referem-se as fases de previsão e correção e $W = [X, Y, Z]$,

com

$$f_{n,x} = -\sigma(X - Y), \quad (4.6)$$

$$f_{n+1,x}^p = -\sigma(X_{n+1}^p - Y), \quad (4.7)$$

$$f_{n,y} = rX - Y - XZ, \quad (4.8)$$

$$f_{n+1,y}^p = rX - Y_{n+1}^p - XZ, \quad (4.9)$$

$$f_{n,z} = XY - bZ, \quad (4.10)$$

$$f_{n+1,z}^p = XY - bZ_{n+1}^p. \quad (4.11)$$

4.2 O Modelo Meteorológico DYNAMO

O modelo descrito abaixo (Lynch (1984), Campos Velho (1997)) simula movimentos atmosféricos em grande escala (a escala sinótica). Apesar de sua grande simplicidade, o código DYNAMO é capaz de reproduzir importantes fenômenos da dinâmica atmosférica, embora tenha sido imaginado para testar esquemas de inicialização.

Algumas hipóteses do modelo são:

- aproximação de água-rasa (*shallow-water equations*): considera-se a atmosfera como um fluido incompressível e em equilíbrio hidrostático na vertical, com o campo de velocidades independente da componente z ;
- admite-se atmosfera seca;
- fronteira livre para a altura geopotencial (contorno superior);
- a superfície (fronteira no nível $z = 0$) é considerada plana;
- Assumem-se condições de contorno periódicas para todas as variáveis dependentes perturbadas.

4.2.1 Equações do Modelo

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,

$$\frac{du}{dt} - fv + \frac{\partial \Phi}{\partial x} = 0, \quad (4.12)$$

$$\frac{dv}{dt} + fu + \frac{\partial \Phi}{\partial y} = 0, \quad (4.13)$$

$$\frac{d\Phi}{dt} + \Phi(u_x + v_y) = 0 \quad (4.14)$$

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, β constantes. Por último, os subscritos significam derivação parcial.

Supondo que o vento básico (zonal) seja constante e esteja em balanço geostrófico com o geopotencial ($f\bar{u} = -\partial\bar{\Phi}/\partial y$) e os desvios (perturbações) sejam variáveis unidimensionais

$$u = \bar{u} + U(x, t), \quad (4.15)$$

$$v = V(x, t), \quad (4.16)$$

$$\Phi = \bar{\Phi}(y) + \phi(x, t). \quad (4.17)$$

Elimina-se a dependência em y subtraindo-se a relação geostrófica da eq. (4.13), obtendo-se a equação (4.13'). As equações da vorticidade ($\zeta = v_x$) e da divergência ($\delta = u_x$) são obtidas das combinações [(4.13')_x - (4.12)_y] e [(4.12)_x + (4.13')_y], respectivamente. Ou seja,

$$\zeta_t + (u\zeta)_x + \delta f + \beta V = 0 \quad (4.18)$$

$$\delta_t + (u\delta)_x - \zeta f + \beta U + \Phi_{xx} = 0 \quad (4.19)$$

$$\Phi_t + (u\Phi)_x - f\bar{u}V + \bar{\Phi}\delta = 0 \quad (4.20)$$

a equação (4.20) é obtida incorporando-se o balanço geostrófico à equação da continuidade (4.14).

O sistema (4.18)—(4.20) forma o conjunto de equações básicas do modelo, onde as variáveis prognósticas (ζ, δ, Φ) dependem somente de x e t (independentes). A dependência em y ocorre somente no parâmetro f e no geopotencial médio $\bar{\Phi}$;

porém, neste modelo f e $\bar{\Phi}$ são considerados constantes onde eles não aparecem diferenciados.

O teorema de Helmholtz permite a partição do campo de vento horizontal $\vec{V}_H = U\vec{i} + V\vec{j}$ em componentes não-divergentes e irrotacionais

$$\vec{V}_H = \vec{V}_\chi + \vec{V}_\psi = \vec{k} \times \vec{\nabla}\psi + \vec{\nabla}\chi \quad ; \quad (4.21)$$

onde ψ é a função corrente e χ o potencial de velocidade. Desta maneira tem-se

$$\nabla^2\psi = \zeta \quad , \quad (4.22)$$

$$\nabla^2\chi = \delta \quad . \quad (4.23)$$

Desta forma as componentes horizontais do vento são recuperadas de

$$v = \vec{\nabla}\psi \quad , \quad (4.24)$$

$$u = \vec{\nabla}\chi \quad . \quad (4.25)$$

Então, para o presente modelo deve-se resolver, em cada passo de integração, duas equações de Poisson.

4.2.2 Adimensionalização das Equações

A conveniência de tornar as equações adimensionais, é tornar clara a importância relativa de vários termos do sistema, definindo-se escalas características para cada variável. A variável espacial é adimensionalizada por L (comprimento do canal), as velocidades por uma velocidade característica V , o tempo por f^{-1} e o geopotencial por fLV .

O uso de uma adimensionalização deste tipo nas equações do modelo, faz sobressair alguns grupos adimensionais. Definem-se então

$$R_0 \equiv \frac{V}{fL} \quad ; \quad (4.26)$$

$$R_\beta \equiv \frac{\beta L}{f} \quad ; \quad (4.27)$$

$$R_F \equiv \frac{\bar{\Phi}}{(fL)^2} = \left(\frac{L_R}{L}\right)^2 \quad , \quad (4.28)$$

onde R_0 é chamado número de Rossby.

Deste modo as equações (4.18)—(4.20) podem ser reescritas na forma adimensional

$$\zeta_t + R_0 (u\zeta)_x + \delta + R_\beta V = 0 \quad (4.29)$$

$$\delta_t + R_0 (u\delta)_x - \zeta + R_\beta U + \phi_{xx} = 0 \quad (4.30)$$

$$\phi_t + R_0 (u\phi)_x - R_0 u_0 V + R_F \delta = 0 \quad . \quad (4.31)$$

Note-se que u_0 é o vento zonal médio (\bar{u}) adimensionalizado.

O sistema (4.29)—(4.31) pode ser escrito numa forma matricial prognóstica

$$\frac{\partial \vec{W}}{\partial t} + \mathbf{A} \vec{W} + \mathbf{B} \vec{X} = -R_0 \vec{N}_{\vec{W}}(\vec{W}, \vec{X}) \quad (4.32)$$

onde os vetores de estado são

$$\vec{W} = \begin{bmatrix} \zeta & \delta & \phi \end{bmatrix}^T, \quad (4.33)$$

$$\vec{X} = \begin{bmatrix} V & U & \phi \end{bmatrix}^T, \quad (4.34)$$

os coeficientes matriciais e o termo não linear são dados abaixo

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & \frac{\partial^2}{\partial x^2} \\ 0 & R_F & 0 \end{bmatrix}, \quad (4.35)$$

$$\mathbf{B} = \begin{bmatrix} R_\beta & 0 & 0 \\ 0 & R_\beta & 0 \\ -R_0 u_0 & 0 & 0 \end{bmatrix}, \quad (4.36)$$

$$\vec{N}_{\vec{W}}(\vec{W}, \vec{X}) = \begin{bmatrix} \vec{N}_{\vec{W}}^\zeta & \vec{N}_{\vec{W}}^\delta & \vec{N}_{\vec{W}}^\phi \end{bmatrix}^T = \frac{\partial(\vec{X}_2 \vec{W})}{\partial x}. \quad (4.37)$$

A relação entre o vetor das variáveis prognósticas (\vec{W}) e o vetor das variáveis primitivas (\vec{X}) é dada por

$$\vec{W} = \mathbf{M} \vec{X}, \quad (4.38)$$

no qual \mathbf{M} é um operador diferencial linear de primeira ordem:

$$\mathbf{M} = \begin{bmatrix} \frac{\partial}{\partial x} & 0 & 0 \\ 0 & \frac{\partial}{\partial x} & 0 \\ 0 & 0 & 1 \end{bmatrix} . \quad (4.39)$$

Aplicando a equação (4.38) na equação (4.32), obtém-se a formulação primitiva

$$\mathbf{M} \frac{\partial \vec{X}}{\partial t} + (\mathbf{A}\mathbf{M} + \mathbf{B})\vec{X} = -R_0 \vec{N}(\vec{X}) , \quad (4.40)$$

e o termo não linear associado é

$$\vec{N}(\vec{X}) = \frac{\partial(\vec{X}_2 \mathbf{M} \vec{X})}{\partial x} . \quad (4.41)$$

Doravante denotar-se-á a matriz $(\mathbf{A}\mathbf{M} + \mathbf{B})$ pela matriz \mathbf{H} .

4.2.3 Discretização Espacial

Para integrar o sistema (4.32) ou (4.40) usou-se uma aproximação por diferenças finitas, transformando a equação num sistema de equações diferenciais ordinárias.

Como as velocidades (V, U) e as variáveis prognósticas (ζ, δ) estão relacionadas por um operador diferencial de primeira ordem, é adotada uma grade alternada (Campos Velho e Clayssen, (1991)). As velocidades são especificadas para “semi-pontos” e as variáveis prognósticas e o geopotencial para “pontos inteiros”.

Os operadores de discretização são dados por

$$(q_m)_x = \frac{1}{\Delta x} (q_{m+1/2} - q_{m-1/2}) \quad (4.42)$$

$$(\bar{q}_m) = \frac{1}{2} (q_{m+1/2} + q_{m-1/2}) \quad (4.43)$$

$$(\bar{q}_m)_x = \frac{1}{\Delta x} (q_{m+1} - q_{m-1}) \quad (4.44)$$

$$(q_m)_{xx} = \frac{1}{\Delta x^2} (q_{m+1} - 2q_m + q_{m-1}) \quad (4.45)$$

$$(Uq)_x|_m = \frac{1}{2\Delta x} [U_{m+1/2} (q_{m+1} - q_m) - U_{m-1/2} (q_m - q_{m-1})] \quad (4.46)$$

Usando estes operadores na formulação prognóstica e primitiva e supondo condições de contorno periódicas para todas as variáveis dependentes perturbadas ($Q(x+L) = Q(x)$), obtém-se

$$\frac{d\vec{W}}{dt} + \mathbf{A}\vec{W} = -[1/2 \mathbf{B}\vec{X} + R_0 \vec{N}_W(\vec{W}, \vec{X})] = -R_0 \vec{N}_B(\vec{W}, \vec{X}) \quad (4.47)$$

$$\mathbf{M} \frac{d\vec{X}}{dt} + (\mathbf{A}\mathbf{M} + 1/2 \mathbf{B}) \vec{X} = -R_0 \vec{N}(\vec{X}) \quad (4.48)$$

onde os vetores de estado e os termos não lineares tornam-se

$$\vec{W} = [\zeta_1 \ \delta_1 \ \phi_1 \ \dots \ \zeta_{N_x} \ \delta_{N_x} \ \phi_{N_x}]^T \quad , \quad (4.49)$$

$$\vec{X} = [V_{-1/2} \ U_{-1/2} \ \phi_1 \ \dots \ V_{N_x-1/2} \ U_{N_x-1/2} \ \phi_{N_x}]^T \quad , \quad (4.50)$$

$$\vec{N}_W(\vec{W}, \vec{X}) = -\frac{1}{2\Delta x} \mathbf{C} \vec{D}(\vec{X}) \mathbf{R} \vec{W} \quad , \quad (4.51)$$

$$\vec{N}(\vec{X}) = -\frac{1}{2\Delta x} \mathbf{C} \vec{D}(\vec{X}) \mathbf{R} \mathbf{M} \vec{X} \quad , \quad (4.52)$$

e os coeficientes matriciais de (4.47) e (4.48) são dados na forma

$$\mathbf{A} = \begin{bmatrix} \mathbf{E} & \mathbf{F} & & & \mathbf{F} \\ \mathbf{F} & \mathbf{E} & \mathbf{F} & & \\ & \mathbf{F} & \mathbf{E} & \mathbf{F} & \\ & \vdots & \vdots & \vdots & \\ & & \mathbf{F} & \mathbf{E} & \mathbf{F} \\ \mathbf{F} & & & \mathbf{F} & \mathbf{E} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{G} & \mathbf{G} & & & \\ & \mathbf{G} & \mathbf{G} & & \\ & & \mathbf{G} & \mathbf{G} & \\ & & \vdots & \vdots & \\ & & & \mathbf{G} & \mathbf{G} \\ \mathbf{G} & & & & \mathbf{G} \end{bmatrix}. \quad (4.53)$$

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 & & & \\ & \mathbf{M}_1 & \mathbf{M}_2 & & \\ & & \mathbf{M}_1 & \mathbf{M}_2 & \\ & & \vdots & \vdots & \\ & & & \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_2 & & & & \mathbf{M}_1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} -\mathbf{I} & \mathbf{I} & & & \\ & -\mathbf{I} & \mathbf{I} & & \\ & & -\mathbf{I} & \mathbf{I} & \\ & & \vdots & \vdots & \\ & & & -\mathbf{I} & \mathbf{I} \\ \mathbf{I} & & & & -\mathbf{I} \end{bmatrix}. \quad (4.54)$$

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} & & & & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & & & \\ & \mathbf{I} & \mathbf{I} & & \\ & \vdots & \vdots & & \\ & & \mathbf{I} & \mathbf{I} & \\ & & & \mathbf{I} & \mathbf{I} \end{bmatrix}, \quad D(\vec{X}) = \text{diag} [U_{-1/2} \mathbf{I}, U_{1/2} \mathbf{I}, \dots, U_{N_x-1/2} \mathbf{I}]. \quad (4.55)$$

Os elementos (submatrizes) das matrizes acima são expressos por

$$\mathbf{E} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & -\frac{2}{\Delta x^2} \\ 0 & R_F & 0 \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\Delta x^2} \\ 0 & 0 & 0 \end{bmatrix}, \quad (4.56)$$

$$\mathbf{M}_1 = \begin{bmatrix} -\frac{1}{\Delta x} & 0 & 0 \\ 0 & -\frac{1}{\Delta x} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{M}_2 = \begin{bmatrix} \frac{1}{\Delta x} & 0 & 0 \\ 0 & \frac{1}{\Delta x} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (4.57)$$

$$\mathbf{G} = \begin{bmatrix} R_\beta & 0 & 0 \\ 0 & R_\beta & 0 \\ -R_0 u_0 & 0 & 0 \end{bmatrix}, \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.58)$$

Nota-se que apesar de as variáveis no lado direito da equação (4.47) (a parte for-

çante) serem \vec{W} e \vec{X} , na realidade elas dependem uma da outra, conforme (4.38). O código DYNAMO calcula $\vec{W}(t)$ por integração no tempo e a variável auxiliar $\vec{X}(t)$ resolvendo a equação $\vec{W}(t) = \mathbf{M}\vec{X}(t)$. A comunicação entre os filtros de Kalman e a subrotina de integração é exclusivamente através do vetor $\vec{W}_n = \vec{W}(t_n)$. O filtro calcula \vec{X}_n pelas equações (4.24) - (4.25).

Para recuperar os campos de vento deve-se primeiro resolver duas equações do tipo (4.23). Apresenta-se aqui somente a solução para a função de corrente, visto que o caso do potencial de velocidade é análogo. Para o modelo DYNAMO a equação de Poisson discretizada é como segue

$$A_{\Delta}\Psi = \Xi; \quad (4.59)$$

onde $\Psi = [\psi_1, \psi_2, \dots, \psi_{N_x}]^T$, $\Xi = \Delta x^2 [\zeta_1, \zeta_2, \dots, \zeta_{N_x}]^T$ e a matriz A_{Δ} é dada por

$$A_{\Delta} = \begin{bmatrix} -2 & 1 & & & 1 \\ 1 & -2 & 1 & & \\ & 1 & -2 & 1 & \\ & & \vdots & \vdots & \vdots \\ & & & 1 & -2 & 1 \\ 1 & & & & 1 & -2 \end{bmatrix}. \quad (4.60)$$

Como o problema para o potencial de velocidade é similar, a matriz do sistema é a mesma como dada em (4.59).

4.3 Modelo de Iteração Não-Linear entre Ondas Langmuir, Whistler e Alfvén - Modelo de 3 Ondas

Chávez (2000) , descreve a transição de ordem para o caos na iteração não-linear de ondas de Langmuir e ondas de Alfvén, impondo um descasamento das frequências na condição de casamento de fase e introduzindo amortecimento a fim de examinar a importância destes parâmetros no desenvolvimento de caos. Para tal, foi utilizado o Modelo de 3 Ondas, aproveitado nesta tese para testar o esquema de assimilação por RN. A seguir segue a derivação do modelo.

Considerando a iteração entre uma onda de Langmuir (L), uma onda eletromagnética de alta frequência polarizada circularmente à direita (w) e uma onda de Alfvén (A)

de baixa frequência e polarizada circularmente à esquerda, restringindo o estudo a uma análise unidimensional, onde todas as ondas que interagem entre si propagam-se ao longo do campo magnético $\mathbf{B} = B_0 \hat{\mathbf{z}}$. As interações coerentes, do tipo $L \rightleftharpoons w + A$ ocorrem quando os vetores de onda e as frequências satisfazem a condição de casamento de fase

$$\omega_1 \approx \omega_2 + \omega_3, \quad k_1 = k_2 + k_3, \quad (4.61)$$

sendo os índices (1, 2, 3) referentes às ondas (L,A,w), respectivamente. Na relação (4.61) tem-se um descasamento na frequência, mas se supos um casamento perfeito nos vetores de onda. Como as ondas de Langmuir não possuem polarização visto que são ondas longitudinais e considerando as ondas de Alfvén com polarização à esquerda, tem-se que a onda eletromagnética de alta frequência deve estar polarizada à direita.

Chávez (2000) , adotando o formalismo de dois fluídos, deduz que o sistema de equações que governa o processo de três ondas é dado por

$$D_1 \mathbf{E}_1 = -ic_{32} E_3 E_2 \hat{\mathbf{z}}, \quad (4.62)$$

$$D_2 \mathbf{E}_2 = -ic_{13} E_1 \mathbf{E}_3^*, \quad (4.63)$$

$$D_3 \mathbf{E}_3 = -ic_{12} E_1 \mathbf{E}_2^* \quad (4.64)$$

onde os parâmetros de dispersão D_i , são

$$D_1 = -\omega_1^2 + \omega_{pe}^2 + \gamma v_{th}^2 k_1^2 - iv_1 \omega_1, \quad (4.65)$$

$$D_2 = -\omega_2^2 + c_A^2 k_2^2 - iv_2 \omega_2, \quad (4.66)$$

$$D_3 = -\omega_3^2 + c^2 k_3^2 + \frac{\omega_{pe}^2 \omega_3}{\omega_3 - \omega_{ce}} - iv_3 \omega_3 \quad (4.67)$$

os coeficientes de acoplamento c_{ij} , são dados por

$$c_{32} = \frac{-e\omega_{pe}^2}{2m_e(\omega_3 - \omega_{ce})} \left[\frac{k_2}{\omega_2} + \frac{k_3(\omega_3 - \omega_{ce})}{\omega_3(\omega_2 - \omega_{ce})} \right], \quad (4.68)$$

$$c_{12} = \left(\frac{\omega_3^2}{\omega_1^2} \right) c_{32}, \quad (4.69)$$

$$c_{13} = \left(\frac{\omega_2^2 c_A^2}{\omega_1^2 c^2} \right) c_{32} \quad (4.70)$$

e os termos de amortecimento v_i , são expressos como

$$v_1 = \frac{\omega_{pe}^2}{\omega_1^2} v_E, \quad (4.71)$$

$$v_2 = \frac{\omega_{pe}^2 c_A^2 v_e}{c_2 (\omega_2 + \omega_{ce})^2} + \frac{\omega_{pi}^2 c_A^2 v_i}{c^2 (\omega_2 - \omega_i)^2}, \quad (4.72)$$

$$v_3 = \frac{\omega_{pe}^2 v_e}{c^2 (\omega_3 - \omega_{ce})^2}, \quad (4.73)$$

onde $\omega_{pe} = (n_0 e^2 / m_e \epsilon_0)^{1/2}$ é a frequência eletrônica de plasma, $\omega_{ce} = eB_0 / m_e$ é a frequência elétron-ciclotrônica, $\omega_{ci} = eB_0 / m_i$ é a frequência íon-ciclotrônica, $v_{th} = (Kt_e / m_e)^{1/2}$ é a velocidade térmica dos elétrons, $c_A = B_0 / (\mu_0 \rho_0)^{1/2}$ é a velocidade de Alfvén, $v_e(v_i)$ é a taxa de crescimento/amortecimento e γ é a razão entre os calores específicos.

Como resultado da interação não-linear entre as ondas, surge uma lenta modulação espaço-temporal nos campos, a qual pode ser representada por

$$E_\alpha(z, t) = \frac{1}{2} \varepsilon_\alpha(z, t) \exp(i\theta_\alpha) + c.c., \quad (4.74)$$

onde $\varepsilon_\alpha(z, t)$ representa uma variação lenta do envelope tal que $|\partial_t^2 \varepsilon_\alpha| \ll |\omega_\alpha \partial_t \varepsilon_\alpha|$ e $|\partial_z^2 \varepsilon_\alpha| \ll |k_\alpha \partial_z \varepsilon_\alpha|$, $\theta_\alpha = k_\alpha z - \omega_\alpha t$ é a variação rápida da fase e $\alpha = (1, 2, 3)$. A aproximação aqui adotada, de variação lenta na amplitude, implica em fraca não-linearidade, Chávez (2000). Substituindo 4.74 e aplicando os operadores espacial e temporal $\partial_t \rightarrow ik$ e $\partial_z \rightarrow -i\omega$ nas equações 4.64 obtiveram-se

$$(\partial_t v_{g1} \partial_z + v_1') \varepsilon_1 = -\frac{c_{32}}{2\partial D_1 / \partial \omega_1} \varepsilon_3 \varepsilon_2 \exp(i\Delta t), \quad (4.75)$$

$$(\partial_t v_{g2} \partial_z + v_2') \varepsilon_2 = \frac{c_{13}}{2\partial D_2 / \partial \omega_2} \varepsilon_1 \varepsilon_3 \exp(-i\Delta t) \quad (4.76)$$

$$(\partial_t v_{g3} \partial_z + v_3') \varepsilon_3 = \frac{c_{12}}{2\partial D_3 / \partial \omega_3} \varepsilon_1 \varepsilon_2 \exp(-i\Delta t) \quad (4.77)$$

sendo $v_{g\alpha} = \partial \omega_\alpha / \partial k_\alpha$ a velocidade de grupo, $v_\alpha' = -v\omega / \partial_{\omega_\alpha} D_\alpha$ e $\Delta = \omega_1 - \omega_3 - \omega_2$ o descasamento da frequência. O sistema de equações (4.75) - (4.77), após alguma

manipulações algébricas pode ser obtido na seguinte forma normalizada

$$\dot{A}_1 = \nu_1'' A_1 + A_3 A_2, \quad (4.78)$$

$$\dot{A}_2 = i\delta A_2 + \nu_2'' A_2 - A_1 A_3^*, \quad (4.79)$$

$$\dot{A}_3 = \nu_3'' A_3 - A_1 A_2^*, \quad (4.80)$$

com

$$A_1 = \left[\frac{c_{13}c_{12}}{4k^2(v_{g2} - v)(v_{g3} - v)(\partial D_2 \partial \omega_2)(\partial D_3 \partial \omega_3)} \right]^{\frac{1}{2}} \varepsilon_1 \quad (4.81)$$

$$A_2 = \left[\frac{c_{32}c_{12}}{4k^2(v_{g1} - v)(v_{g3} - v)(\partial D_1 \partial \omega_1)(\partial D_3 \partial \omega_3)} \right]^{\frac{1}{2}} \varepsilon_2 \exp(i\Delta t) \quad (4.82)$$

$$A_3 = \left[\frac{c_{32}c_{13}}{4k^2(v_{g1} - v)(v_{g2} - v)(\partial D_1 \partial \omega_1)(\partial D_2 \partial \omega_2)} \right]^{\frac{1}{2}} \varepsilon_3 \quad (4.83)$$

onde o ponto denota derivada em relação a variável temporal $\tau = k(z - vt)$, v e k são velocidades e vetores de onda arbitrários, respectivamente, $\nu_\alpha'' = \nu_\alpha' / [k(v_{g\alpha} - v)]$ e $\delta = \Delta / [k(v_{g2} - v)]$.

A seguir apresenta-se os resultados obtidos com as RN e o FK com os sistema de Lorenz, o Modelo de 3 Ondas e o modelo DYNAMO - 1D.

CAPÍTULO 5

ASSIMILAÇÃO DE DADOS EM DINÂMICA NÃO-LINEAR COM O SISTEMA DE LORENZ E O MODELO DE 3 ONDAS

Neste capítulo apresenta-se os resultados obtidos com o método de assimilação proposto. Na seção 5.1 revisa-se o problema de assimilação de dados e se ressalta a importância do sistema de assimilação. Para tal mostra-se o impacto da inserção de dado sem uma técnica adequada nos sistemas dinâmicos avaliados. Reserva-se as seções 5.2 e 5.3 respectivamente para os resultados com o sistema de Lorenz e com o Modelo de 3 Ondas (aplicação da metodologia em clima espacial). Explora-se a inserção de memória no sistema através das redes recorrentes e o treinamento com a técnica de correlação cruzada. Para uniformizar a apresentação dos resultados, em todos os gráficos onde se avalia o desempenho do método de assimilação contra a previsão sem assimilação e contra as observações, as cores verde, azul, vermelha e preta, referem-se respectivamente a observação, modelo numérico, assimilação com RN e assimilação com FK. Nos gráficos dos erros, as cores azul, vermelha verde e preta, referem-se respectivamente a média, primeira componente do modelo (X no sistema de Lorenz, L no Modelo de 3 Ondas), segunda componente do modelo (Y no sistema de Lorenz, W no Modelo de 3 Ondas) e terceira componente do modelo (Z no sistema de Lorenz, A no Modelo de 3 Ondas).

5.1 Assimilação de Dados

Os métodos de assimilação de dados como Relaxação Newtoniana, Interpolação Ótima, métodos variacionais e filtro de Kalman por conjuntos (Daley, 1990), podem ser descritos como procedimentos que usam dados observacionais para melhorar uma previsão feita por um modelo matemático impreciso. A assimilação pode ser descrita como um processo de dois passos:

$$\begin{aligned} \text{Passo de previsão:} \quad & w_n^p = F[w_{n-1}^a] \\ \text{Passo de análise:} \quad & w_n^a = w_n^p + \mathbf{d}_n \end{aligned}$$

onde w_n representa o vetor de estado do modelo no passo de tempo t_n , $F[.]$ é o modelo matemático de previsão, os sobre índices p e a denotam respectivamente os valores preditos e analisados, finalmente \mathbf{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:

$$\mathbf{d}_n = -G_n f(w_n^o - w_n^p)$$

sendo G_n a matriz de ponderação (ou matriz de ganho, como é chamado em filtros de Kalman). Esta matriz é calculada por um estimador na técnica de IO, o estimador poderá ser mínimos quadrados ou máxima variância; ou por uma fórmula envolvendo a matriz de covariância, como na técnica do FK; ou pode ser parametrizada, por exemplo $G_n = \alpha_n I$, onde α_n é um parâmetro que varia com o tempo. Esta parametrização é usada, por exemplo, em métodos variacionais de assimilação.

A função de discrepância assume formas diferentes para cada método de assimilação:

$$f(w_n^o - w_n^p) = \begin{cases} w_n^o - H_n w_n^p & \text{(Filtro de Kalman e Interpolação ótima)} \\ \nabla J(w_n^p - w_n^o) & \text{(Método variacional)} \end{cases}$$

onde a função custo $J(w_n^p - w_n^o)$ é dada por:

$$J(w_n^p - w_n^o) = \frac{1}{2} (w_n^p - w_n^o)^T \theta (w_n^p - w_n^o) .$$

O novo método de assimilação de dados baseado em RN pode ser descrito como

$$w_n^a = F_{RN}(w_n^p, w_n^o)$$

onde F_{RN} representa a RN descrita anteriormente. Resultados com esta técnica são apresentados nas seções seguintes e em Härter e Campos Velho (2004) , Härter et al. (2004) e Härter e Campos Velho (2004) .

5.2 Resultados com o sistema de Lorenz Caótico

O FK, alvo das redes, foi implementado com as seguintes matrizes:

$$\mathbf{Q}_n = 0.1 \mathbf{I} ; \quad \mathbf{R}_n = 2 \mathbf{I} ; \quad \mathbf{P}_0^f = \begin{cases} 10 (\mathbf{w}_0^f)_i^2 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} . \quad (5.1)$$

O sistema de Lorenz foi integrado com o método preditor-corretor de primeira ordem, com $\Delta t = 10^{-3}$, para $X(0) = 1.508870$, $Y(0) = -1.531271$, $Z(0) = 25.46091$,

$\sigma = 10.0$, $r = 8.0/3.0$ e $b = 28.0$, ou seja, parâmetros com os quais a dinâmica do modelo segue uma trajetória caótica. Os dados são inseridos no sistema a cada 12 passos de tempo. As redes são treinadas com 2000 exemplos e 333 dados são usados na validação cruzada. Uma vez treinadas as RN, é feita uma previsão de 10^6 passos de tempo. 2000 é a quantidade de exemplos necessária e suficiente para a rede “aprender”, obtido de maneira empírica. Escolheu-se 333 dados para na validação cruzada, porque esta quantidade é suficiente para analisar graficamente se a rede consegue “generalizar”. A previsão é de longo prazo (10^6) para se verificar se as inserção de memória permite que se faça previsões de prazo maior do que as as feitas com redes sem memória.

Com o intuito de testar a capacidade das RN em emular o FK no contexto de assimilação de dados em regime caótico, as RN foram treinadas com uma camada escondida, com constantes de momento α_h na camada escondida e α_β ($\beta = X, Y, Z$) na camada de saída. A maneira análoga é usada para a taxa de aprendizagem, sendo η_h, η_β ($\beta = X, Y, Z$). Os valores numéricos destes parâmetros estão na Tabela 5.1

TABELA 5.1 - Parâmetros das RN com o sistema de Lorenz: 1 camada escondida.

	PMC	FBR	RN-E	RN-J
neurônios	2	40	2	2
α_h	0,6	–	0,51	0,4
α_X	0,6	–	–	0,4
α_Y	0,6	–	–	0,4
α_Z	0,6	–	–	0,4
η_h	0,001	0,0001	0,0001	0,0001
η_X	0,001	0,0001	0,00001	0,0001
η_Y	0,001	0,0001	0,0001	0,0001
η_Z	0,001	0,0001	0,0001	0,0001

Redes com duas camadas escondidas também foram testadas e os parâmetros empregados com estas topologias são apresentados na Tabela 5.2.

Um ponto crucial para o sucesso da aplicação de RN é saber o ponto em que o treinamento deve ser finalizado, pois um erro de treinamento menor não implica necessariamente em melhor generalização (melhor estimativa com dados não pertencentes ao conjunto de treinamento). Algumas vezes, durante o treinamento, o erro pode estar diminuindo pelo fato da rede estar se especializando no conjunto

de treinamento, o que pode implicar numa generalização ruim. Em outras, pode-se “cair” num mínimo local da superfície de erros. Taxas de aprendizagem e momento adequados podem evitar os mínimos locais e a validação cruzada é uma alternativa para se saber qual melhor conjunto de pesos implica em melhor generalização.

TABELA 5.2 - Parâmetros das RN com o sistema de Lorenz: 2 camadas escondidas.

	PMC (L1, L2)	FBR (L1, L2)	RN-J (L1, L2)
neurônios	(2, 6)	(11, 2)	(6, 2)
α_h	0,5	0,1	–
α_X	0,5	0,1	–
α_Y	0,5	0,0	–
α_Z	0,5	0,1	–
η_h	0,001	0,1	0,01
η_X	0,001	0,1	0,001
η_Y	0,1	0,1	0,01
η_Z	0,001	0,1	0,01

A validação cruzada consiste em dividir o conjunto de dados disponíveis em dois subconjuntos, um para o treinamento e outro para validação ou teste. Assim, para cada conjunto de pesos obtido com os dados de treinamento, é feita uma estimativa com o conjunto de teste (validação). Fixou-se o treinamento em 1000 épocas e obteve-se o conjunto de pesos que resultou em melhor generalização. A vantagem desta técnica é que se conhece a superfície dos erros de treinamento e de estimativa de todo o período de integração.

Os erros de treinamento, validação cruzada ou erro com o conjunto de teste (calculados em cada época de treinamento) e erro de previsão com o método de assimilação (calculado sobre o tempo total de previsão) são obtidos como segue

$$ET_{\beta} = \frac{1}{2000} \sum_{i=1}^{2000} \frac{1}{2} (\beta_i^{FK} - \beta_i^{RN})^2 ; \quad (5.2)$$

$$EV_{\beta} = \frac{1}{333} \sum_{k=1}^{333} \sqrt{(\beta_k^{FK} - \beta_k^{RN})^2} ; \quad (5.3)$$

$$EP_{\beta} = \frac{1}{10^5} \sum_{k=1}^{10^5} \sqrt{(\beta_k^{Obs} - \beta_k^{RN})^2} . \quad (5.4)$$

Os resultados obtidos com as redes PMC, FBR, RN-E e RN-J com uma camada escondida são mostrados a seguir. Ressalta-se que resultados com as redes treinadas com duas camadas escondidas são similares a estes e não serão discutidos neste texto.

5.2.1 Assimilação: Perceptron de Múltiplas Camadas (PMC)

Os dados assimilados durante a integração dos modelos testados neste trabalho são gerados artificialmente. Com os testes feitos com o modelo de Lorenz, o dado foi gerado somando-se ruído aleatório de variância 2 ao método de integração do modelo, o que faz com que a curva de observação siga uma trajetória desacoplada da curva gerada pelo sistema de Lorenz. Assim, a curva de referência para o método de assimilação é a curva de observação sintética, ou seja, quanto mais próxima as curvas geradas pela RN estiverem da curva de dados observados, melhor é a estimativa fornecida pela RN. Na Figura 5.1, ilustra-se o desequilíbrio provocado no sistema de Lorenz ao se inserir observação sem um método de assimilação.

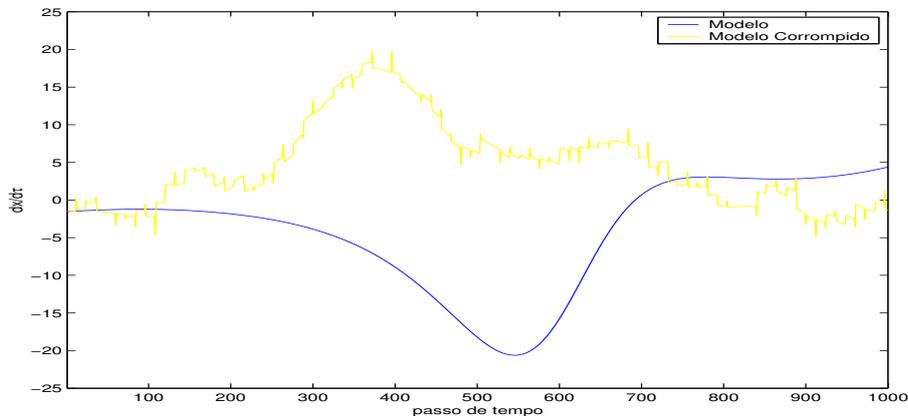


FIGURA 5.1 - Ilustração da instabilidade causado no sistema de Lorenz ao se inserir dado sem um método de assimilação.

As Figuras 5.2 (a) - (b) mostram respectivamente as curvas de Erro de Treinamento (ET) e Validação Cruzada (EV). Os erros são apresentados para cada componente do sistema de Lorenz e para a média das três componentes ($[\text{erro-}X + \text{erro-}Y + \text{erro-}Z]/3$). Uma vez obtido o melhor conjunto de pesos para o problema e implementação estudada é feita a previsão para o período de 10^6 passos de tempo, com dados assimilados a cada 12 passos.

É importante frisar, que se considera o melhor conjunto de pesos, aquele que implica em menor erro com o conjunto de testes. Considera-se o menor erro com o conjunto de testes, o menor erro médio na época em que o erro das três componentes diminuem em relação a época anterior. Ou seja, se o erro médio com o conjunto de testes diminuir em relação a época anterior e o erro de uma das componentes não diminuir, o conjunto de pesos desta época de treinamento não é considerado o melhor. Isto evita que a rede se especialize em apenas uma ou duas componentes dos modelos.

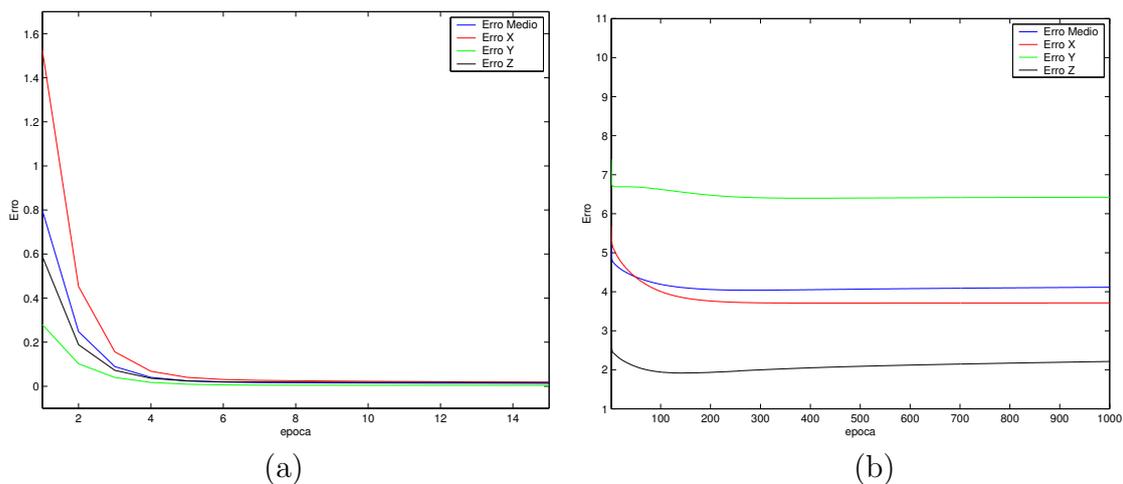


FIGURA 5.2 - Erros com o sistema de Lorenz: (a) durante a fase de treinamento; (b) na validação cruzada.

A PMC treinada com 2 neurônios na camada oculta e as taxas de aprendizagem descritas na Tabela 5.3, atingiu o melhor conjunto de pesos em 144 épocas. Na Figura 5.2 (a) mostra-se as primeiras 14 épocas e não até a 144-ésima época para que o leitor possa perceber que o decaimento abrupto dos erros ocorreram nas primeiras 4 épocas. Caso o critério de parada utilizado fosse o erro de treinamento e não a validação cruzada, ter-se-ia parado o treinamento em poucas épocas, pois conforme as Figuras 5.2 (a) - (b) os erros tornaram-se quase estáveis após a quarta época de

treinamento, o que neste caso resultaria num treinamento mais rápido com pequena perda de generalização. Todavia, o uso da correlação cruzada permite afirmar que os conjunto de pesos obtido foi o melhor conjunto de pesos para o problema estudado e implementação avaliada. A Tabela 5.3 resume os ET, EV e (EP) com o PMC como método de assimilação.

TABELA 5.3 - Erros com o PMC.

	ET	EV	EP
$(X + Y + Z)/3$	0,003014	4,106462	3,212483
X	0,000891	3,851951	3,099526
Y	0,001548	6,552105	3,321165
Z	0,006003	1,915329	3,216759

Pelos erros ET, a variável que menos aprendeu foi a variável Z , porém foi a que apresentou menor erro com o conjunto de teste. A componente Y apresentou o pior resultado com o conjunto de teste, o que se confirmou na previsão, com menor imprecisão. A variável que mais aprendeu foi a X , resultando no menor erro de previsão, embora não tenha sido a variável de menor erro com o conjunto de testes.

Nas Figuras 5.3 (a) - (c) são grafados os últimos 10^3 passos de tempo referente às variáveis X , Y e Z do sistema de Lorenz, experimento cuja previsão foi feita com a rede PMC com uma camada oculta como método de assimilação. Observa-se que a rede PMC foi eficiente, pois evitou o choque mostrado na Figura 5.1.

A variável Z indica que o domínio de integração apresentado nas Figuras 5.3(a)-(c) é uma região de forte gradiente vertical de temperatura entre as camadas fria e quente do fluido, pois em todo o domínio os valores de Z são positivos. Isto indica que o método de assimilação foi eficiente em regiões de inversão térmica, embora se deva considerar que o sistema de Lorenz é bastante simplificado em relação aos modelos de equações primitivas usados na previsão de tempo operacional.

A região analisada apresenta, além de forte gradiente vertical de temperatura, forte movimento convectivo, já que o fluido quente está em movimento ascendente e o fluido frio em movimento descendente, pois X e Y tem o mesmo sinal. Os processos convectivos que ocorrem na atmosfera são uma das principais fontes de energia dentro do sistema terra-atmosfera, sendo responsáveis por chuvas fortes, denominadas no jargão meteorológico de precipitação convectiva. Por sua vez os processos

atmosféricos que envolvem precipitação estão associados a mudanças de fase, com consequente liberação de calor latente e implicações diretas nos perfis de temperatura e umidade previstos por modelos numéricos. A precipitação influencia também a circulação geral da atmosfera. Daí a importância de se avaliar a performance do método de assimilação em situações de forte movimento convectivo.

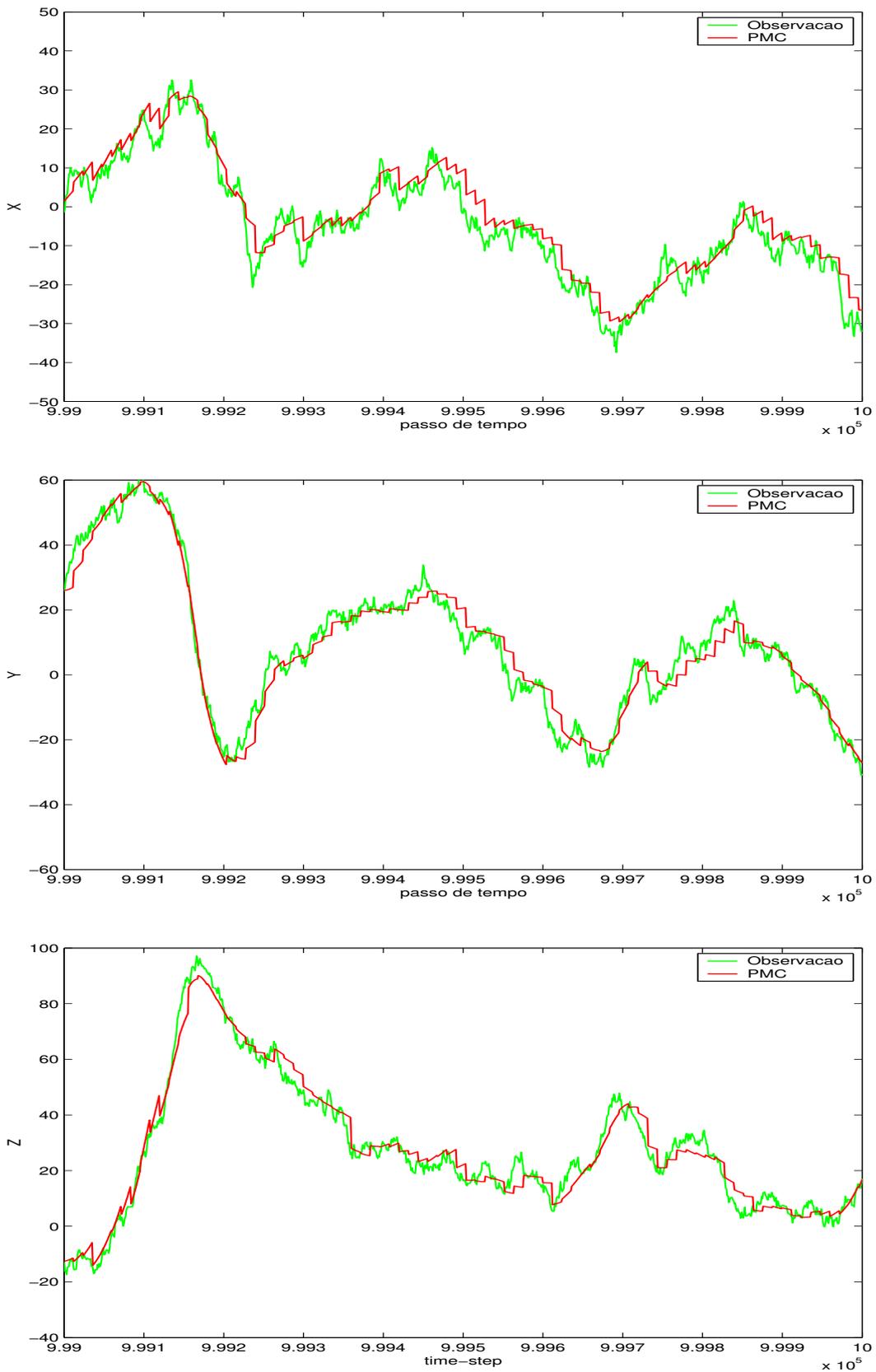


FIGURA 5.3 - Assimilação de dados com o sistema de Lorenz e PMC: de cima para baixo (a) componente- X , (b) componente- Y , (c) componente- Z .

5.2.2 Assimilação: Funções de Base Radial (RBF)

Os experimentos acima foram refeitos para a rede FBR com uma camada escondida. Os ET e EV são mostrados nas Figuras 5.4 (a) - (b) e os últimos 10^3 passos de tempo de um total de uma previsão de 10^6 passos de tempo são grafados nas Figuras 5.5 (a) - (c). A melhor arquitetura desta rede foi obtida com 2 neurônios na camada escondida, com as taxas de aprendizagem e momento descritas na Tabela 5.1.

O melhor conjunto de pesos foi obtido na primeira época de treinamento, o que significa que os pesos foram atualizados 2.000 vezes, pois o treinamento é por padrão (tem-se 2.000 padrões de treinamento). Tanto os erros de treinamento como os erros com o conjunto de testes oscilaram nas primeiras épocas e diminuíram assintoticamente após a décima quinta época. Porém, lembra-se que para um conjunto de pesos ser considerado ótimo, é necessário que o EV das três componentes assimiladas diminuam na mesma época. Por esta razão o conjunto de pesos ótimo foi obtido na primeira época. Os ET, EV e EP para os melhor conjunto de pesos são apresentados na Tabela 5.4.

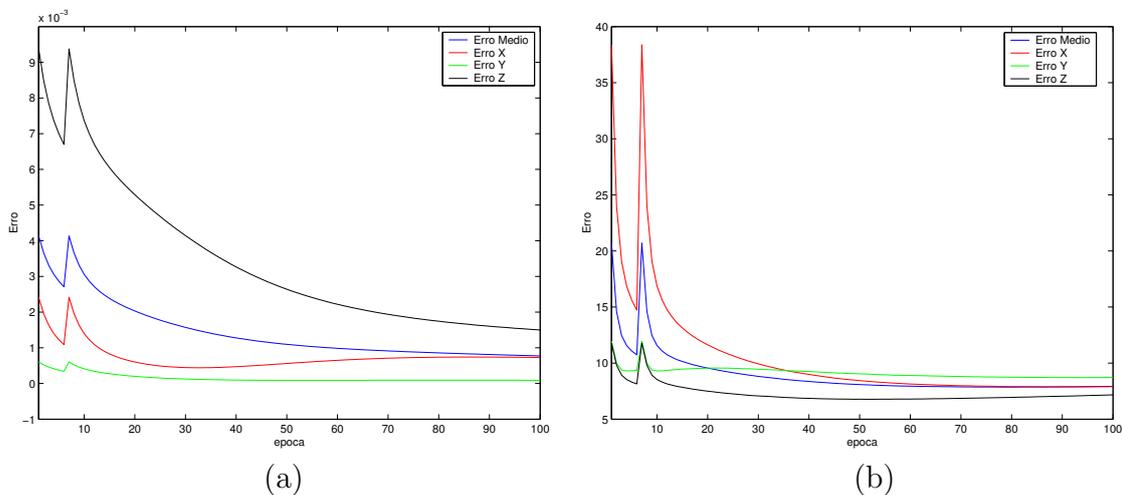


FIGURA 5.4 - Erros com o sistema de Lorenz: (a) durante a fase de treinamento; (b) na validação cruzada.

TABELA 5.4 - Erros com a FBR.

	ET	EV	EP
$(X + Y + Z)/3$	5,606438	4,620559	2,809641
X	4,377834	3,426893	2,355447
Y	6,997956	7,988125	4,983895
Z	5,443536	2,446659	1,089582

As Figuras 5.5 (a) - (c) (próxima página) mostram que a FBR também resolveu o problema de assimilação de dados com este modelo, capturando bem a diferença de temperatura entre as correntes ascendentes e descendentes de fluido e também apresentando uma boa resposta, tanto quanto a intensidade do movimento convectivo, quanto ao perfil vertical de temperatura.

Na Tabela 5.5 verifica-se o custo computacional, em segundos, do treinamento de cada RN. Como esperado, o custo computacional está diretamente relacionado ao número de épocas de treinamento, que por sua vez é função do critério de parada adotado. Nos experimentos realizados, a PMC atingiu o melhor conjunto de pesos em 102,6 segundos e a RN-J em 0,6 segundos. Entretanto, ressalta-se que o treinamento é finalizado, quando o erro com o conjunto de validação diminui em relação a época anterior, para as três componentes do sistema de Lorenz, de maneira concomitante. Esta tabela evidencia a importância de investigar diferentes redes neurais antes de se operacionalizar uma aplicação.

TABELA 5.5 - Tempo de Treinamento (s) com o sistema de Lorenz.

<i>RedeNeural</i>	Tempo (s)	Épocas de Treinamento
<i>PMC</i>	102,6	144
<i>FBR</i>	0,7	1
<i>RN - E</i>	5,0	5
<i>RN - J</i>	0,6	1

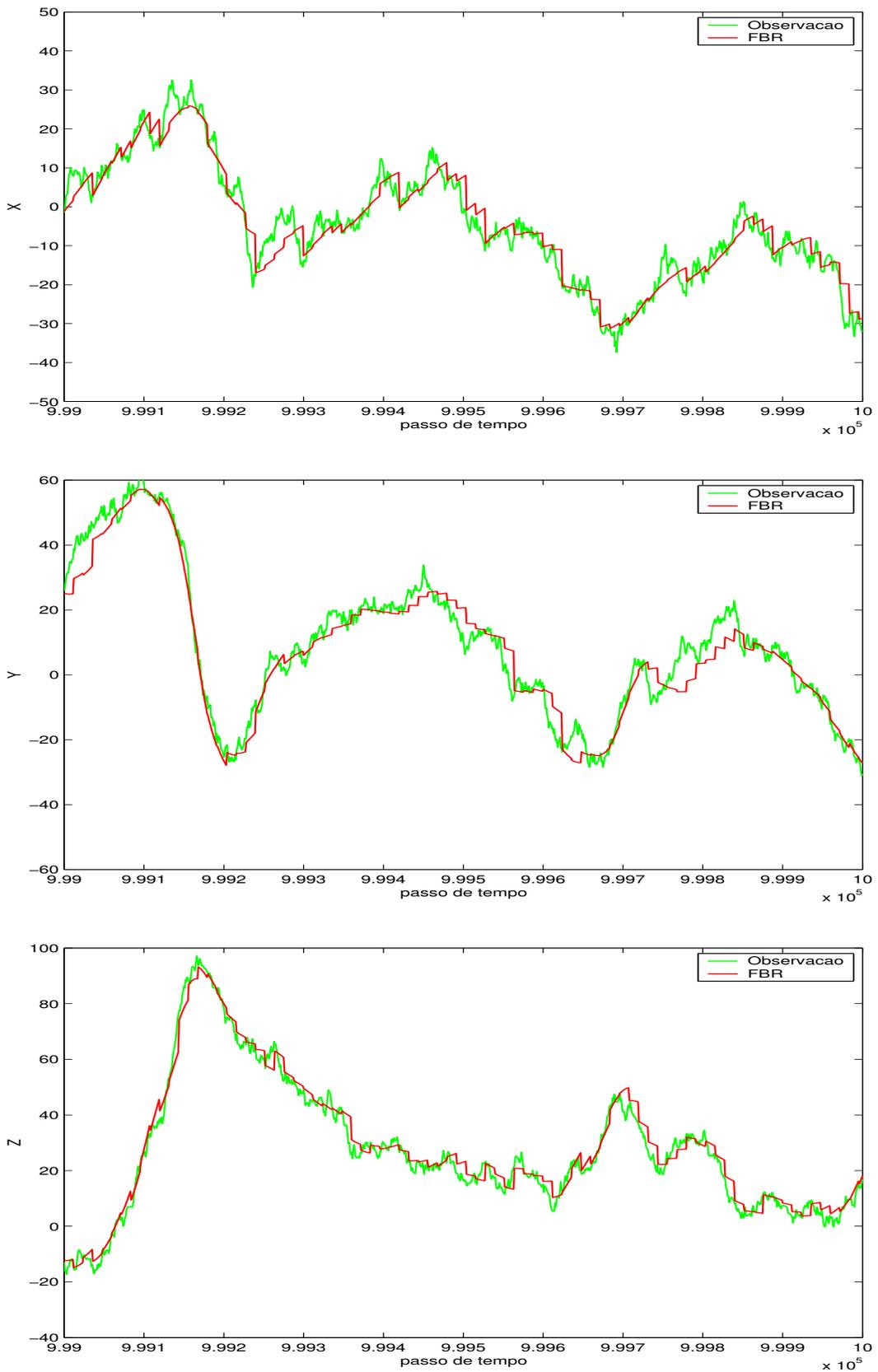


FIGURA 5.5 - Assimilação de dados com o sistema de Lorenz e FBR: de cima para baixo (a) componente- X , (b) componente- Y , (c) componente- Z .

5.2.3 Assimilação: Redes Recorrentes

Nesta seção explora-se os resultados com a RN-E (recorrência da camada oculta para a camada de entrada) e a RN-J (recorrência da camada de saída para a camada de entrada). O objetivo no estudo de redes recorrentes é verificar se a inserção de memória permite que se aumente o alcance da previsão sem perda de precisão.

Nas Figuras 5.6 (a) - (b) são grafados os erros de treinamento e validação cruzada para a RN-E. Os resultados referem-se a implementação de dois neurônios na camada escondida, com taxas de aprendizagem e momento dadas na Tabela 5.1. As figuras dos ET e EV para a RN-E evidenciam a vantagem na utilização da validação cruzada, pois no treinamento desta rede, à medida que o ET começa a diminuir o EV aumenta, ou seja, a rede em poucas épocas se especializa no conjunto de treinamento, levando a uma generalização ruim. Numa implementação onde se considera o erro de treinamento como critério de parada, certamente se optaria por considerar um bom conjunto de pesos, um conjunto obtido após a 60-ésima época, o que causaria perda de generalização. O melhor conjunto de pesos (obtido na quinta época) resultou nos erros apresentados na Tabela 5.6

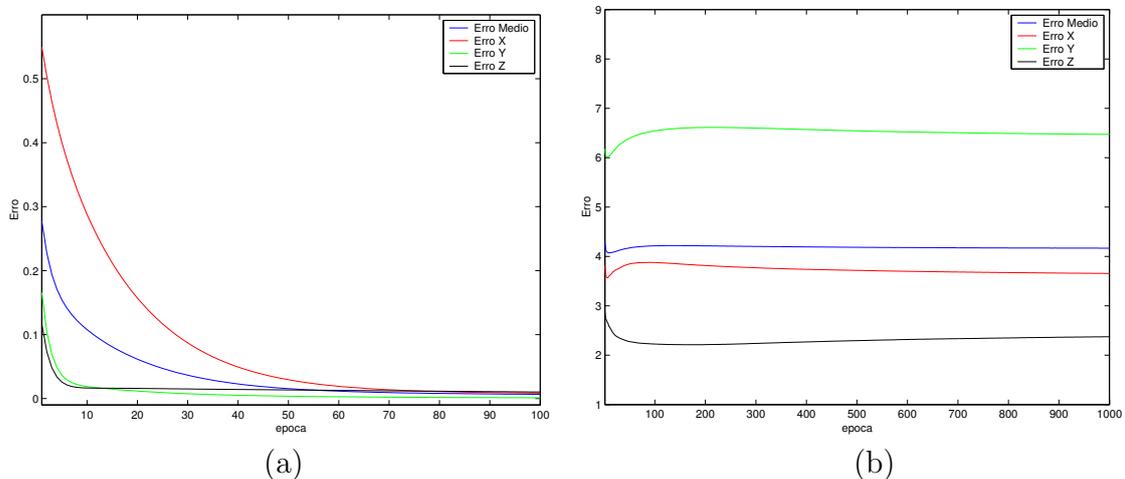


FIGURA 5.6 - Erros com o sistema de Lorenz: (a) durante a fase de treinamento; (b) na validação cruzada.

TABELA 5.6 - Erros com a RN-E.

	ET	EV	EP
$(X + Y + Z)/3$	0,154134	4,091083	3,640524
X	0,400177	3,572664	3,840467
Y	0,036458	6,015399	3,468690
Z	0,025767	2,685187	3,612416

Com a RN-J, bem como a FBR, o melhor conjunto de pesos foi obtido na primeira época, pois nas primeiras trezentas épocas o EV diminui nas componentes X e Z e aumenta na componentes Y . Após a 300-ésima época os erros com o conjunto de teste estabilizaram, não havendo portanto, uma época em que os EV das três componentes diminuam concomitantemente de uma época para a seguinte, (ver Figuras 5.7 (a) - (b) e Tabela 5.7, abaixo).

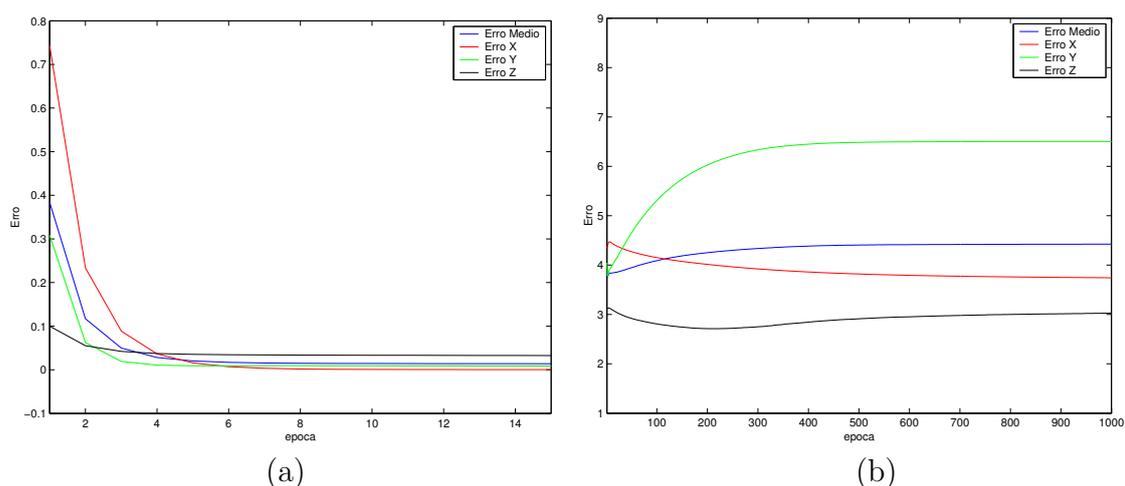


FIGURA 5.7 - Erros com o sistema de Lorenz: (a) durante a fase de treinamento; (b) na validação cruzada.

TABELA 5.7 - Erros com a RN-J.

	ET	EV	EP
$(X + Y + Z)/3$	0,383787	3,842690	3,043522
X	0,742981	4,348868	2,133769
Y	0,308399	4,046068	4,609145
Z	0,099979	3,133135	2,387650

A RN-E também resultou em um bom método de assimilação, sendo que as três componentes do sistema de Lorenz resultaram em EP próximos à média, ou seja, a rede não se especializou em nenhuma das componentes, embora o conjunto de teste tenha resultado num erro alto na previsão de diferença de temperatura entre as camadas ascendentes e descendentes de fluido, componente Y do sistema de Lorenz (ver Figuras 5.9 (a) - (c), previsão feita com a RN-E como método de assimilação).

Previsões com a RN-J como método de assimilação também se mostraram eficientes, com poucas diferenças em relação as demais redes neurais (Figuras 5.10 (a) - (c)). Na Tabela 5.7 constam os erros para o melhor conjunto de pesos, obtidos na primeira época de treinamento para a implementação com dois neurônios na camada escondida.

Embora a assimilação feita pelas redes recorrentes tenham apresentado bons resultados, não houve ganho significativo de precisão em relação as redes estáticas, na revisão com alcance de 10^6 passos de tempo. Entretanto, as redes recorrentes resolveram o problema em menos épocas do que a PMC e com menos neurônios do que a FBR. Na Figura 5.8 é grafada a evolução temporal dos últimos 10^3 passos de tempo de um total de 10^6 do erro de previsão com as redes PMC e RN-E como método de assimilação. O objetivo é mostrar que os erros não tendem a aumentar com o tempo.

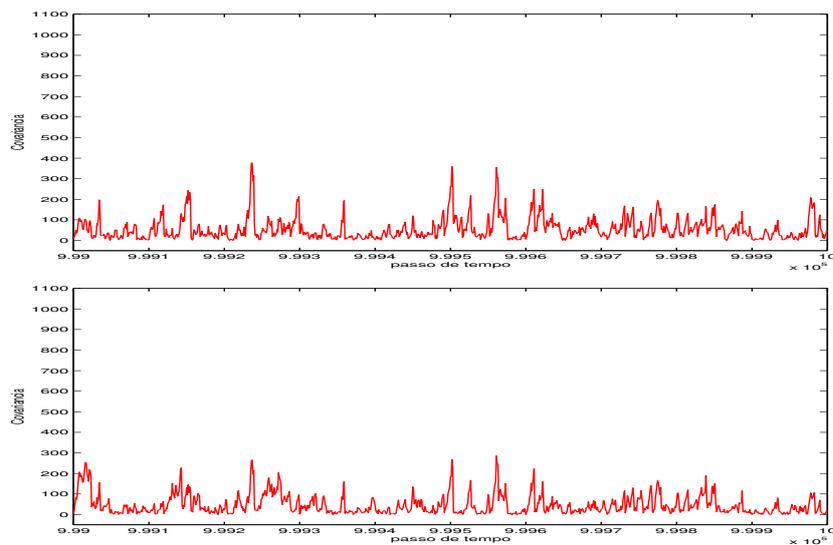


FIGURA 5.8 - Evolução temporal do erro de previsão para (a) PMC, (b) RN-E.

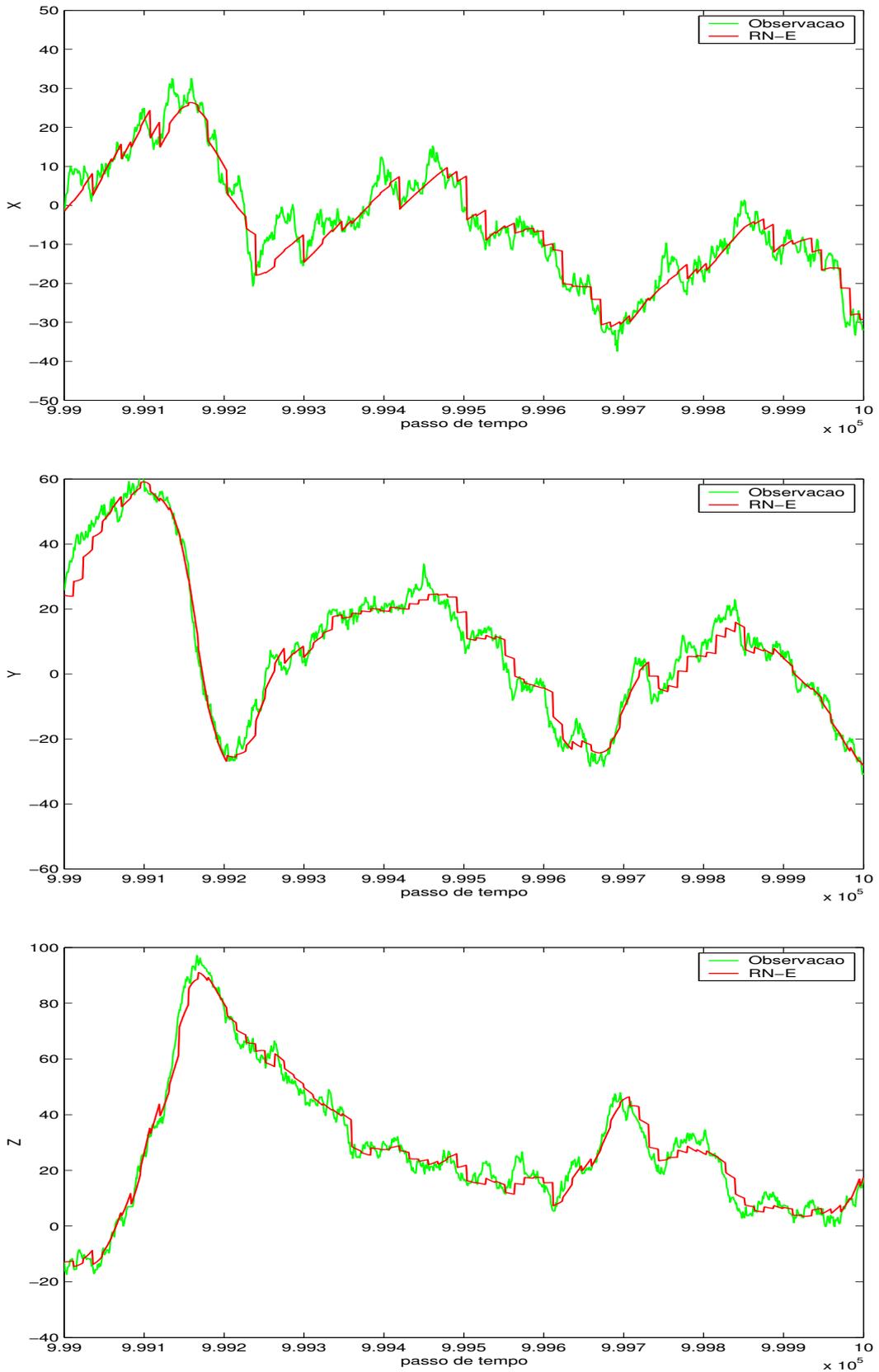


FIGURA 5.9 - Assimilação de dados com o sistema de Lorenz e RN-E: de cima para baixo (a) componente-X, (b) componente-Y, (c) componente-Z.

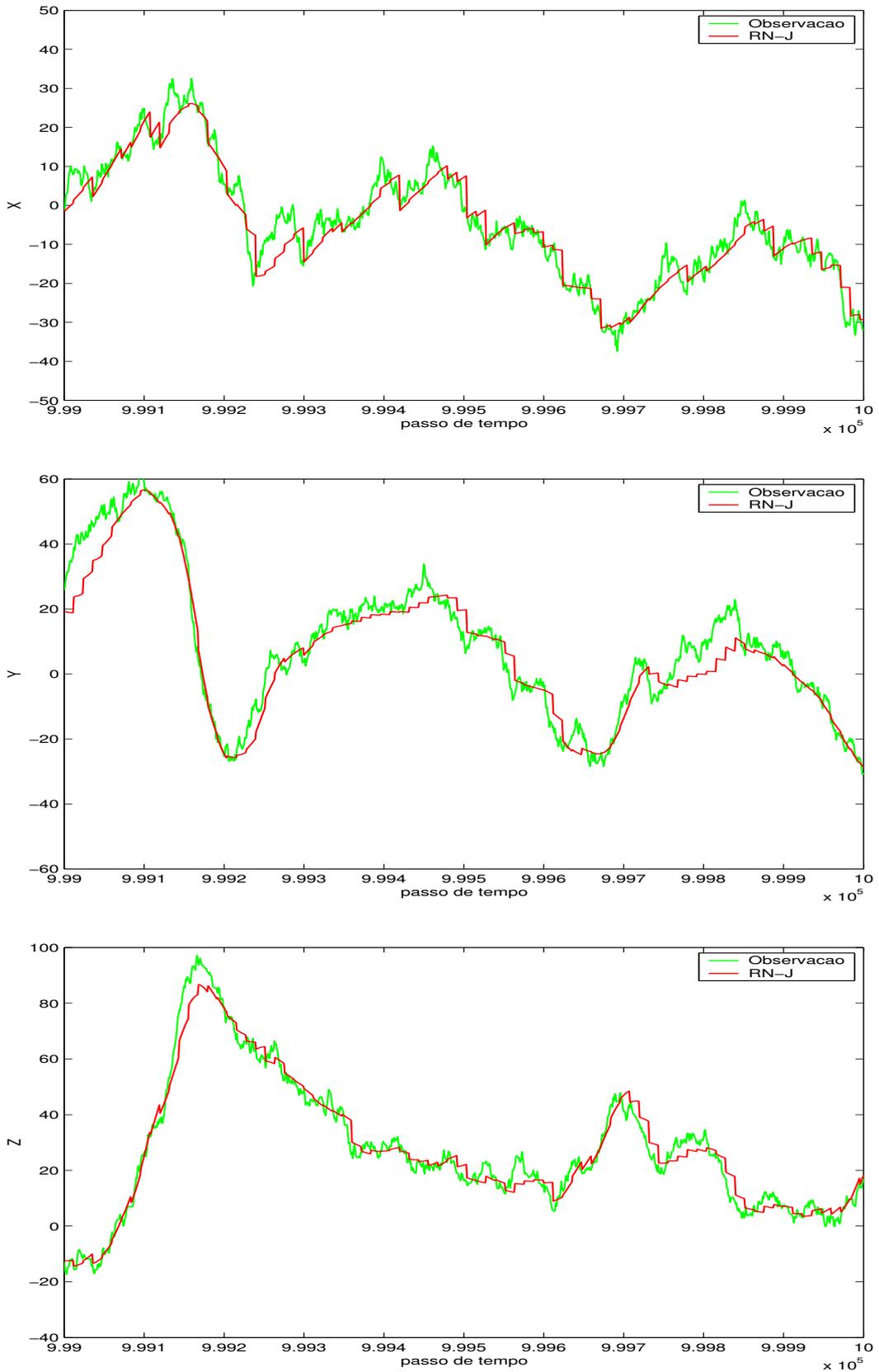


FIGURA 5.10 - Assimilação de dados com o sistema de Lorenz com RN-J: de cima para baixo (a) componente- X , (b) componente- Y , (c) componente- Z .

5.3 Resultados Numéricos com o Modelo de 3 Ondas: Aplicação em Clima Espacial

Pequenas variações nas condições iniciais de um modelo dinâmico em regime caótico, fazem com que a dinâmica prevista pelo modelo atinja estados completamente diferentes para cada uma das condições iniciais. No entanto, surge a necessidade de se corrigir estes modelos com dados observados. Corrigir um modelo dinâmico com dados observados tem aplicação em diversas áreas da ciência, tais como geofísica espacial. Portanto, testa-se a assimilação pela rede PMC com o Modelo de 3 Ondas, aplicado a previsão de clima espacial. Testa-se a rede PMC e o FK nos regimes periódico e caótico.

O Modelo de 3 Ondas é bastante sensível a ruído, tal que para ilustrar o impacto da observação no modelo, integrou-se as equações em regime periódico e se fez inserção do dado sem assimilação em apenas um passo de tempo (ver Figura 5.11).

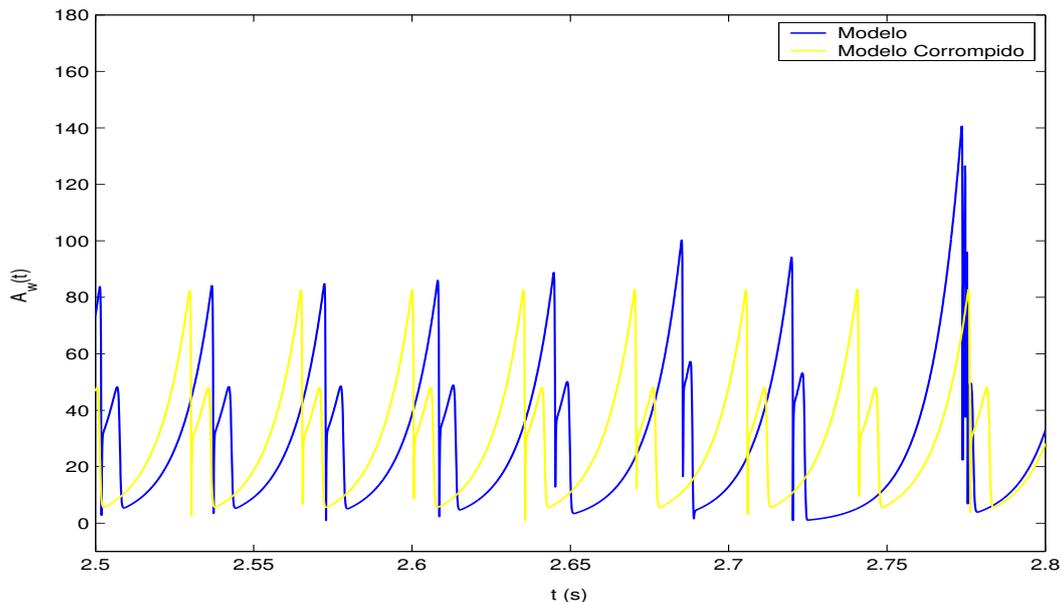


FIGURA 5.11 - Ilustração da instabilidade causado no Modelo de 3 Ondas ao se inserir dado sem um método de assimilação.

Os experimentos apresentados nesta seção são similares aos feitos com o modelo DYNAMO, ou seja, a observação é gerada de maneira artificial com ruído aleatório somado a saída do Modelo de 3 Ondas, tal que a curva de referência para a estimativa pela RN e pelo FK é o Modelo de 3 Ondas. Os experimentos foram feitos com as

seguintes matrizes para o FK:

$$\mathbf{Q}_n = 0.1 \mathbf{I} ; \mathbf{R}_n = 2 \mathbf{I} ; \mathbf{P}_0^f = \begin{cases} 10 (\mathbf{w}_0^f)_i^2 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} . \quad (5.5)$$

O Modelo de 3 Ondas foi integrado com o método de Runge Kutta de quarta ordem, com $\Delta t = 10^{-2}$. A inserção de dados é feita a cada 5 passos de tempo, pois na prática estes dados são disponibilizados com alta frequência. Para o treinamento das RN utiliza-se 2.000 exemplos e 3.000 dados são utilizados na validação cruzada. Após o treinamento, é feita uma previsão de 10^4 passos de tempo.

Com o objetivo de emular o FK, a PMC foi treinada com 2 camadas escondidas, taxa de aprendizagem 0.1, 2 neurônios na primeira camada escondida, 10 neurônios na segunda camada escondida, sem constante de momento e com a técnica de validação cruzada.

Os erros de treinamento, validação e previsão para as componentes do Modelo de 3 Ondas são calculados através das seguintes equações, onde $\beta = L, W, A$. O erro médio é calculado pela média aritmética simples dos erros destas três ondas ([erro- L + erro- W + erro- A]/3).

$$ET_\beta = \frac{1}{2000} \sum_{i=1}^{2000} \frac{1}{2} (\beta_i^{FK} - \beta_i^{RN})^2 ; \quad (5.6)$$

$$EV_\beta = \frac{1}{3000} \sum_{k=1}^{3000} \sqrt{(\beta_k^{FK} - \beta_k^{RN})^2} ; \quad (5.7)$$

$$EP_\beta = \frac{1}{10^4} \sum_{k=1}^{10^4} \sqrt{(\beta_k^P - \beta_k^{RN})^2} . \quad (5.8)$$

É importante notar que a forma de gerar observação exige que o EP (erro de estimativa ou previsão com método de assimilação) seja calculado em relação a previsão do modelo sem método de assimilação e não em relação a observação como no sistema de Lorenz. A seguir a rede PMC é testada nos regimes caótico e periódico.

5.3.1 Assimilação: Perceptron de Múltiplas Camadas (PMC) - Regime Periódico - $\gamma = 28, 14$

Nas Figuras 5.12 (a) - (b) são grafados os erros de treinamento e validação cruzada para L, W e A, bem como o erro médio ($[\text{erro-L} + \text{erro-W} + \text{erro-A}]/3$). Após se obter o melhor conjunto de pesos, o Modelo de 3 Ondas é integrado com a frequência de assimilação de 5 passos de tempo.

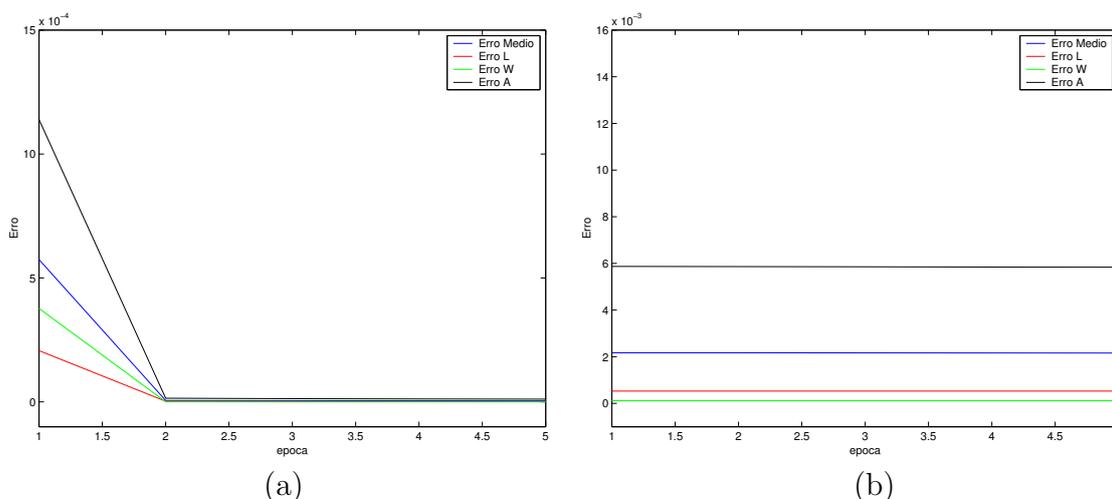


FIGURA 5.12 - Erros do Modelo de 3 Ondas: (a) na fase de treinamento; (b) na validação cruzada.

Durante o processo de aprendizagem o erro diminuiu de maneira abrupta nas primeiras duas épocas e permanece praticamente constante durante o restante do treinamento. A validação cruzada mais uma vez fornece o melhor conjunto de pesos para a implementação e aplicação em questão.

Nas Figuras 5.13 (a) - (c) são grafados os últimos 10^3 passos de tempo da integração de Modelo de 3 Ondas. A Figura 5.13 (a) representa o modelo numérico (referência - sem ruído), a Figura 5.13 (b) representa a integração do Modelo de 3 Ondas (componente - $|A_L|$), onde o processo de assimilação é feito pelo FK e a na Figura 5.13 (c) a assimilação é feita pela PMC, com 2 neurônios na primeira camada escondida e 10 neurônios na segunda camada escondida, em regime periódico. Observa-se, que tanto o FK como a PMC foram eficientes métodos de assimilação de dados na aplicação testada.

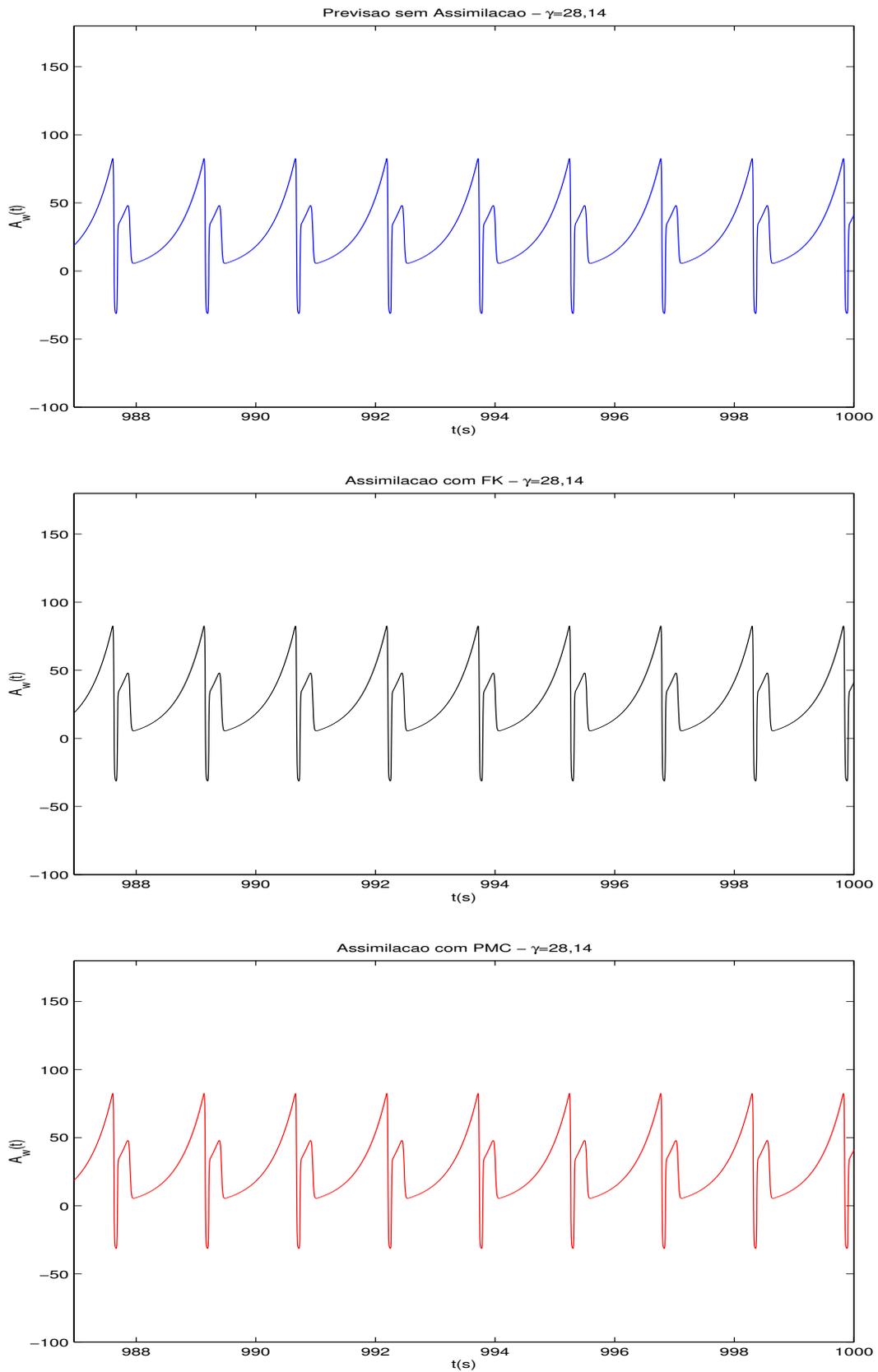


FIGURA 5.13 - Componente- A_L , (a) Modelo de referência, (b) assimilação por FK, e (c) assimilação pela PMC.

Nas Figuras 5.14 (a) - (b) são apresentados os erros do FK e da PMC, respectivamente, onde verifica-se que os métodos apresentam praticamente a mesma precisão no regime periódico.

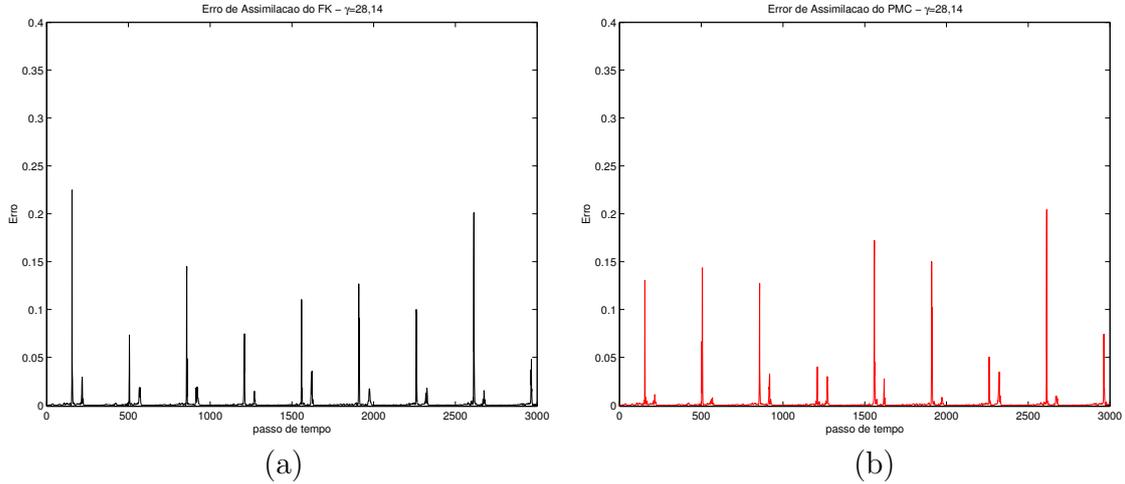


FIGURA 5.14 - Erro para regime periódico: (a) filtro de Kalman; (b) rede neural.

5.3.2 Assimilação: Perceptron de Multiplas Camadas (PMC) - Regime Caótico - $\gamma = 28, 128$

Nesta seção se analisa a eficiência do sistema de assimilação (PMC e FK) em dinâmica caótica. Assim, os experimentos analisados acima ($\gamma = 28, 4$), são refeitos para ($\gamma = 28, 128$). As curvas de erros obtidas no período de treinamento e validação cruzada são apresentadas nas Figuras 5.15 (a) - (b), bem como o erro médio.

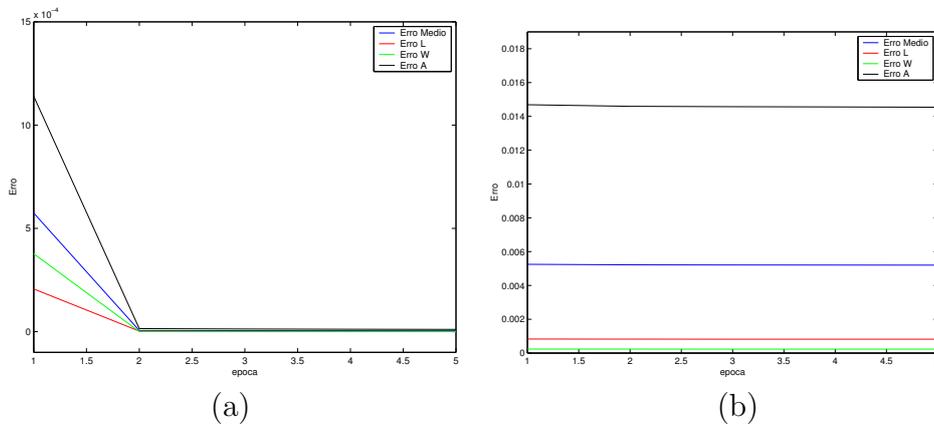


FIGURA 5.15 - Erros do Modelo de 3 Ondas: (a) treinamento; (b) validação.

Observa-se que o erro de treinamento decresce de forma abrupta nas primeiras duas iterações e permanece constante durante o restante do processo de aprendizagem, atingindo um erro médio mínimo (0,00518) de validação cruzada na terceira época.

Uma vez obtido o melhor conjunto de pesos, o sistema de 3-ondas é integrado com com dados sendo assimilados pelo FK e pela PMC a cada 5 passos de tempo. A PMC novamente é implementada com 2 neurônios na primeira camada escondida e 10 neurônios na segunda camada escondida

Os últimos 10^3 passos de integração são grafados na Figura 5.17 (a) - (c), respectivamente. Ambos os métodos de assimilação foram eficientes no problema de assimilação de dados com o Modelo de 3 Ondas em regime caótico, sendo que a PMC apresenta um erro menor do que o FK, como se verifica na Figura 5.16 -(a) - (b).

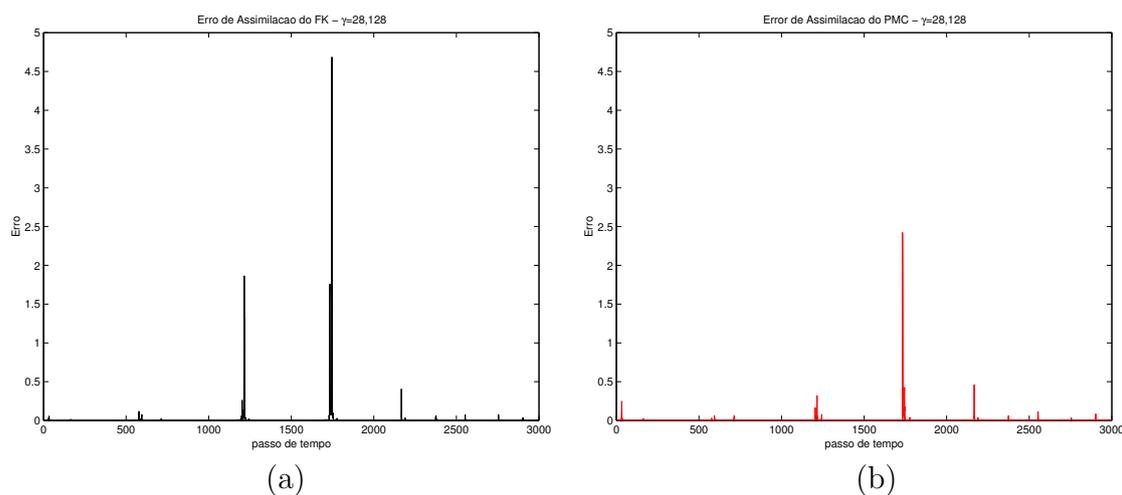


FIGURA 5.16 - Erro para regime periódico: (a) filtro de Kalman; (b) rede neural.

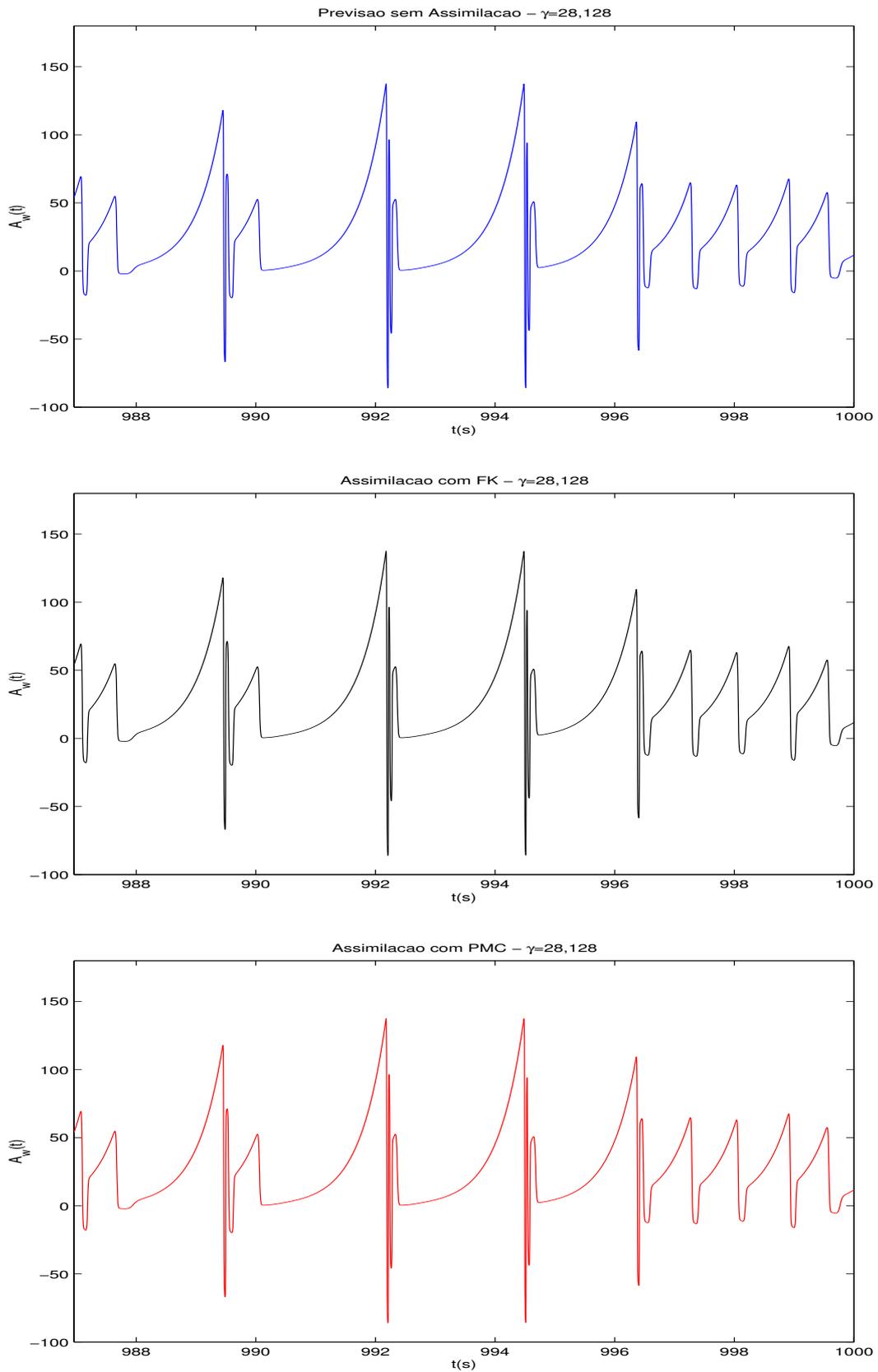


FIGURA 5.17 - Componente- A_L , (a) Modelo de referência, (b) assimilação por FK, e (c) assimilação pela PMC.

CAPÍTULO 6

ASSIMILAÇÃO DE DADOS EM DINÂMICA NÃO-LINEAR COM O MODELO DYNAMO - 1D

Neste capítulo apresenta-se os resultados obtidos com o método de assimilação proposto com o modelo DYNAMO - 1D. Explora-se a inserção de memória no sistema através das redes recorrentes, o treinamento com a técnica de correlação cruzada e a diminuição do espaço de busca, pois o DYNAMO apresenta dimensão espacial. As convenções de cores são mantidas, ou seja, as cores verde, azul, vermelha e preta, referem-se respectivamente a observação, modelo numérico, assimilação com RN e assimilação com FK. Nos gráficos dos erros, as cores azul, vermelha verde e preta, referem-se respectivamente a média, componente U , V e ϕ do modelo DYNAMO-1D.

6.1 Resultados com o Modelo DYNAMO - 1D

A seguir avalia-se os resultados do método de assimilação proposto com o modelo DYNAMO-1D. A importância de se avaliar o desempenho do método com este modelo é que além de ser não-linear, possui as ondas meteorológicas mais importantes (Rossby e gravidade), possui dimensão espacial.

A Figura 6.1 ilustra o impacto negativo causado na variável ζ do DYNAMO-1D ao se inserir dados a cada 6 horas sem um método de assimilação. Nos testes apresentados a seguir, a observação foi gerada artificialmente somando-se ruído aleatório as variáveis previstas pelo modelo numérico e não ao método integrador, ou seja, não há o desacoplamento da curva de observação em relação a dinâmica seguida pelo modelo numérico. Isto significa que após o processo de filtragem, a curva estimada pelas RN e pelo FK devem coincidir com a curva resultante do modelo numérico.

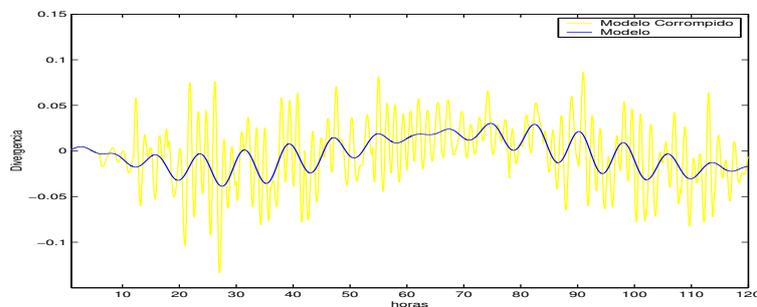


FIGURA 6.1 - Ilustração da instabilidade causado no modelo DYNAMO ao se inserir dado sem um método de assimilação.

O DYNAMO foi integrado com $\Delta t = 60s$, fronteira cíclica e CI não inicializada. Os dados são inseridos no sistema a cada 6 horas (360 passos de tempo). As redes são treinadas com 600 exemplos (-3700 horas até -100 horas) e 6000 dados são usados na validação cruzada (-100 até 0 horas). Uma vez treinadas as RN, é feita uma previsão de 180 dias (0 até 4320 horas). Este esquema é grafado na Figura 6.2.

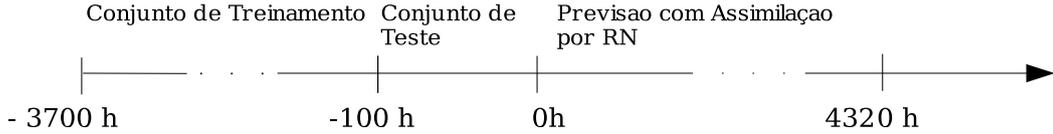


FIGURA 6.2 - Conjunto de dados usado no treinamento, validação cruzada e previsão com método de assimilação.

Para testar a eficiência das RN em emular o FK com o modelo DYNAMO, num contexto de assimilação de dados, as redes foram treinadas com uma camada escondida, com constantes de momento α_h na camada escondida e α_n ($n = U, V, \phi$) na camada de saída. Maneira análoga é usada para a taxa de aprendizagem, sendo η_h , η_n ($n = X, Y, Z$). Os valores numéricos destes parâmetros estão na Tabela 6.1.

TABELA 6.1 - Parâmetros das redes RN com o modelo DYNAMO: 1 camada escondida.

	PMC	FBR	RN-E	RN-J
neurons	15	5	5	2
α_h	–	–	0,3	–
α_X	–	–	0,3	–
α_Y	–	–	0,3	–
α_Z	–	–	0,3	–
η_h	0,1	0,1	0,00001	0,0001
η_X	0,1	0,1	0,00001	0,0001
η_Y	0,1	0,1	0,00001	0,0001
η_Z	0,1	0,1	0,00001	0,0001

O FK, técnica a ser emulada pelas RN, foi implementado com as seguintes matrizes

$\mathbf{H}_n = I, \mathbf{Q}_n = (0.1\iota)^2 I, \mathbf{P}_0^f = 0$ e $\mathbf{F}_n = \mathbf{I}$. $\mathbf{R}_n = \mathbf{r}_{\max} \mathbf{I}$, onde $\mathbf{r}_{\max} = \max_i[\mathbf{r}_i^2(\mathbf{n})]$ and $\mathbf{r}_n = \mathbf{z}_n - \mathbf{z}_n^f$. Onde ι é um número aleatório com distribuição uniforme.

Os ET, EV e EP seguem as equações de erros aplicadas no Modelo de 3 Ondas com

a diferença de que $\alpha = U, V, \phi$; referem-se as velocidades zonal e meridional do vento e ao geopotencial.

6.2 Redução do Espaço de Busca

Embora a inserção de memória e a validação cruzada sejam um diferencial em relação aos trabalhos encontrados na literatura, o principal contribuição desta pesquisa é a redução do espaço de busca no treinamento das RN. Esta redução na dimensão do gradiente a ser minimizado significa a diferença entre tornar a metodologia aplicável ou não na prática operacional de previsão de tempo. Esta redução do espaço de busca foi obtida apresentando-se à rede cada ponto de grade do modelo separadamente, ao invés de apresentar todos os pontos de grade de uma única vez, como em Nowassad et al. (2000).

Considerando-se o modelo DYNAMO unidimensional com 20 pontos de grade, verifica-se que na versão de Nowosad et al. (2000), a rede é composta por 120 variáveis de entrada (previsão e observação de U, V e ϕ em 20 pontos de grade, 80 neurônios na camada escondida e 60 variáveis na saída (U, V , e ϕ estimados pelo FK em 20 pontos de grade). Isto equivale a 14800 pesos a serem identificados durante a fase de treinamento. Na versão atual tem-se para este mesmo modelo, 6 variáveis de entrada (previsão e observação de U, V , e ϕ em 1 ponto de grade) 15 neurônios na camada escondida para a rede PMC e 3 variáveis na saída (U, V , e ϕ estimados pelo FK em 1 ponto de grade), o que corresponde a 135 pesos de conexões.

Numa aplicação operacional, com por exemplo 10^6 pontos de grade, o tamanho da rede implementada conforme a estratégia adotada neste trabalho, permaneceria do mesmo tamanho, pois se treina um ponto de grade de cada vez, enquanto nos trabalhos anteriores, a rede cresceria proporcionalmente ao número de pontos de grade, podendo tornar impossível a minimização do gradiente.

A seguir apresenta-se somente os resultados nos quais as redes foram treinadas com uma única camada escondida, pois experimentos com duas camadas ocultas não resultaram em previsões melhores em relação a experimentos com redes com uma camada oculta.

6.2.1 Assimilação: Perceptron de Múltiplas Camadas (PMC)

Assim como foi feito para os sistemas anteriores, avalia-se os resultados obtidos com as equações de água rasa através dos gráficos de ET, EV, EP e evoluções temporais das variáveis do modelo em um ponto no interior da grade. À medida que se faz necessário, analisa-se também campos espaciais de uma determinada variável, bem como cálculos de energia e evolução temporal de erros.

As Figuras 6.3 (a) - (b) mostram respectivamente as curvas de erro de treinamento e validação cruzada. Os erros são apresentados para cada componente do DYNAMO assimilada (U , V e ϕ) e para a média das três componentes. Uma vez obtido o melhor conjunto de pesos para o problema e implementação estudada é feita a previsão para o período de 180 dias, com dados assimilados a cada 6 horas.

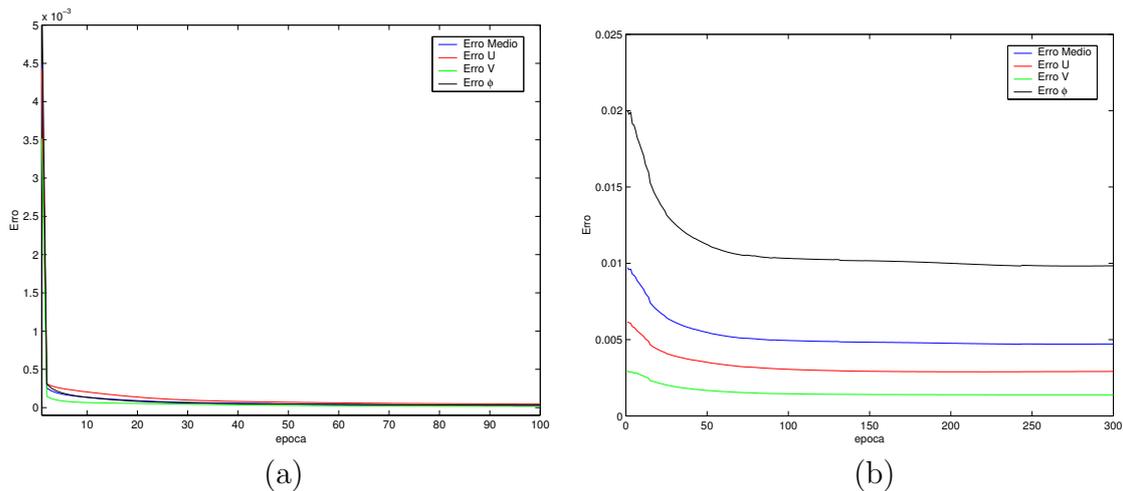


FIGURA 6.3 - Erros com o modelo DYNAMO: (a) durante a fase de treinamento; (b) na validação cruzada.

A arquitetura que resultou em menor erro com o conjunto de teste para o PMC tem 15 neurônios na camada escondida e taxas de aprendizagem descritas na Tabela 6.1. Embora o melhor conjunto de pesos tenha sido obtido na 220-ésima época de treinamento, o gráfico do ET mostra que estes diminuem abruptamente antes da décima época e estabilizam-se em torno da 40-ésima época de treinamento, correspondendo a EV também estáveis.

Com a PMC o vento zonal foi a variável que menos aprendeu, porém os EV mostram

uma boa generalização para todas as variáveis do modelo DYNAMO-1D. Fazer a previsão com o melhor conjunto de pesos nem sempre resulta em uma previsão significativamente mais precisa do que a previsão feita como um conjunto obtido após o erro de treinamento ter atingido valores assintóticos, porém a correlação cruzada fornece o melhor conjunto de pesos para o problema estudado na implementação em uso. Os ET, EV e EP para a PMC constam na Tabela 6.2.

TABELA 6.2 - Erros com a PMC.

	ET	EV	EP
<i>Medio</i>	0,000023	0,004726	0,008488
<i>U</i>	0,000032	0,002887	0,005154
<i>V</i>	0,000001	0,001384	0,002974
ϕ	0,000002	0,009907	0,019424
ζ	—	—	0,003416
δ	—	—	0,011473

Nas Figuras 6.6 (a) - (e) são grafadas a evolução temporal dos últimos 5 dias de previsão de um período total de 180 dias de U , V , ϕ , ζ e δ , num ponto no interior da grade, onde o método de assimilação empregado foi a PMC com uma camada escondida.

O vento zonal, o geopotencial e principalmente a divergência, que por sua vez é derivada do vento, são as variáveis que sofrem maior impacto pela inserção do dado observado. Embora as Figuras 6.6 (a) - (e) mostrem que a rede PMC foi eficiente no processo de assimilação, verifica-se pequenas imperfeições entre o 176-ésimo e 177-ésimo e entre o 179-ésimo e 180-ésimo dia de previsão. Para uma melhor análise destas imperfeições, grafa-se na Figura 6.4 o campo espacial de divergência para o primeiro dia de previsão do DYNAMO, onde a assimilação é feita pela rede PMC.

No modelo DYNAMO, o escoamento é dividido numa parte puramente rotacional e outra puramente divergente. Ondas de Rossby são responsáveis pela parte rotacional do escoamento e ondas de gravidade pela parte divergente. Como o campo de divergência mostra algumas ondas curtas, há indicativo de que sejam ondas de gravidade de alta frequência, causada pelo desequilíbrio entre os campos de massa e velocidade após a inserção do dado observado. Este é um problema de inicialização, uma das componentes do ciclo de assimilação de dados, importante tema de estudo em meteorologia e oceanografia. Usualmente, este problema é resolvido por

métodos de inicialização, como Modos Normais Não-Linear quando se conhece os modos normais do modelo (modelos globais) e como o Filtro Digital (muito eficiente quando aplicado a modelos de área limitada), (Daley, 1991) . Para uma a análise destas ondas curtas, grafa-se na Figura 6.5 a evolução temporal de Energia Cinética Divergente. Conservando-se a notação de cores adotada, a curva em vermelho refere-se a energia cinética divergente resultante da previsão feita com a PMC como método de assimilação e sob esta, em azul, foi grafada a curva de previsão sem assimilação. Este gráfico mostra, que após a filtragem não houve ganho ou perda de energia cinética proveniente de ondas de gravidade. Todavia, como mencionado, a inicialização é uma das componentes do ciclo de assimilação de dados, ou seja, na prática, a assimilação de dados deve ser acompanhada de inicialização.

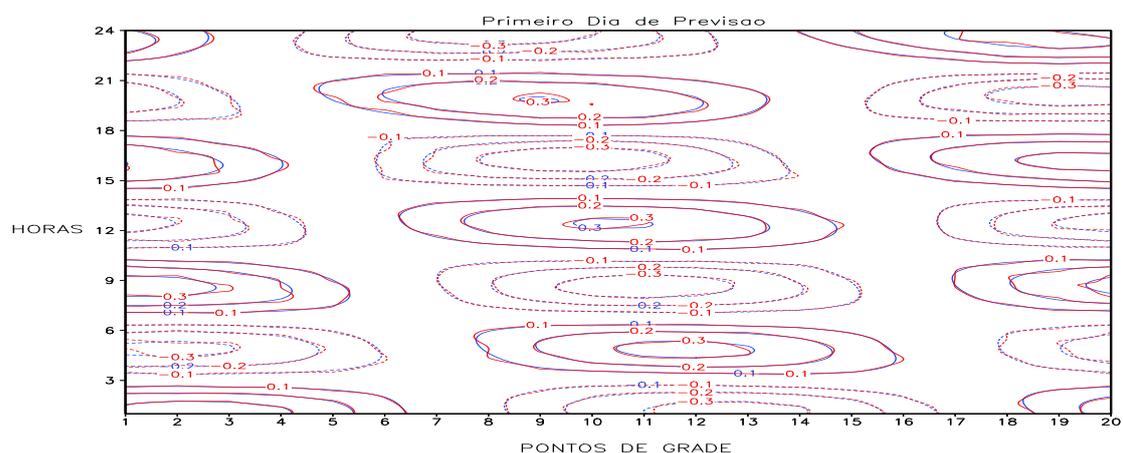


FIGURA 6.4 - Assimilação de dados com o modelo DYNAMO: campo de divergência no começo do domínio de integração. Linha azul - modelo de referência, linha vermelha - assimilação pela PMC.

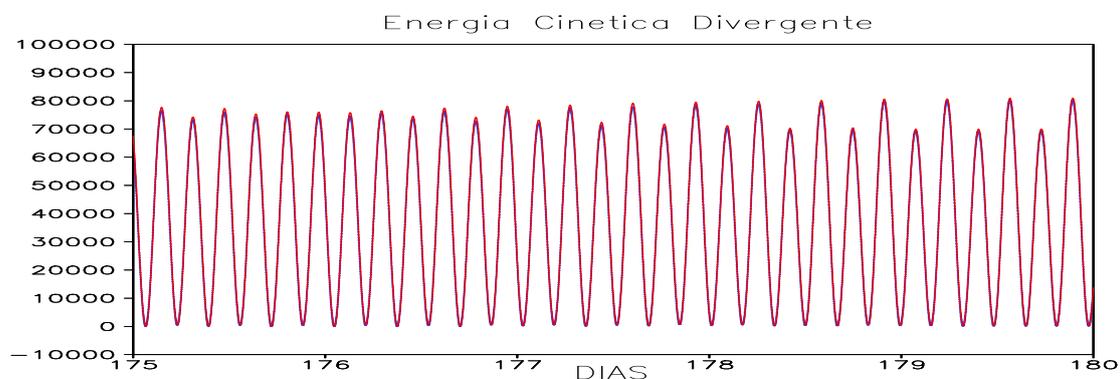


FIGURA 6.5 - Assimilação de dados com o modelo DYNAMO e PMC: Energia Cinética Divergente.

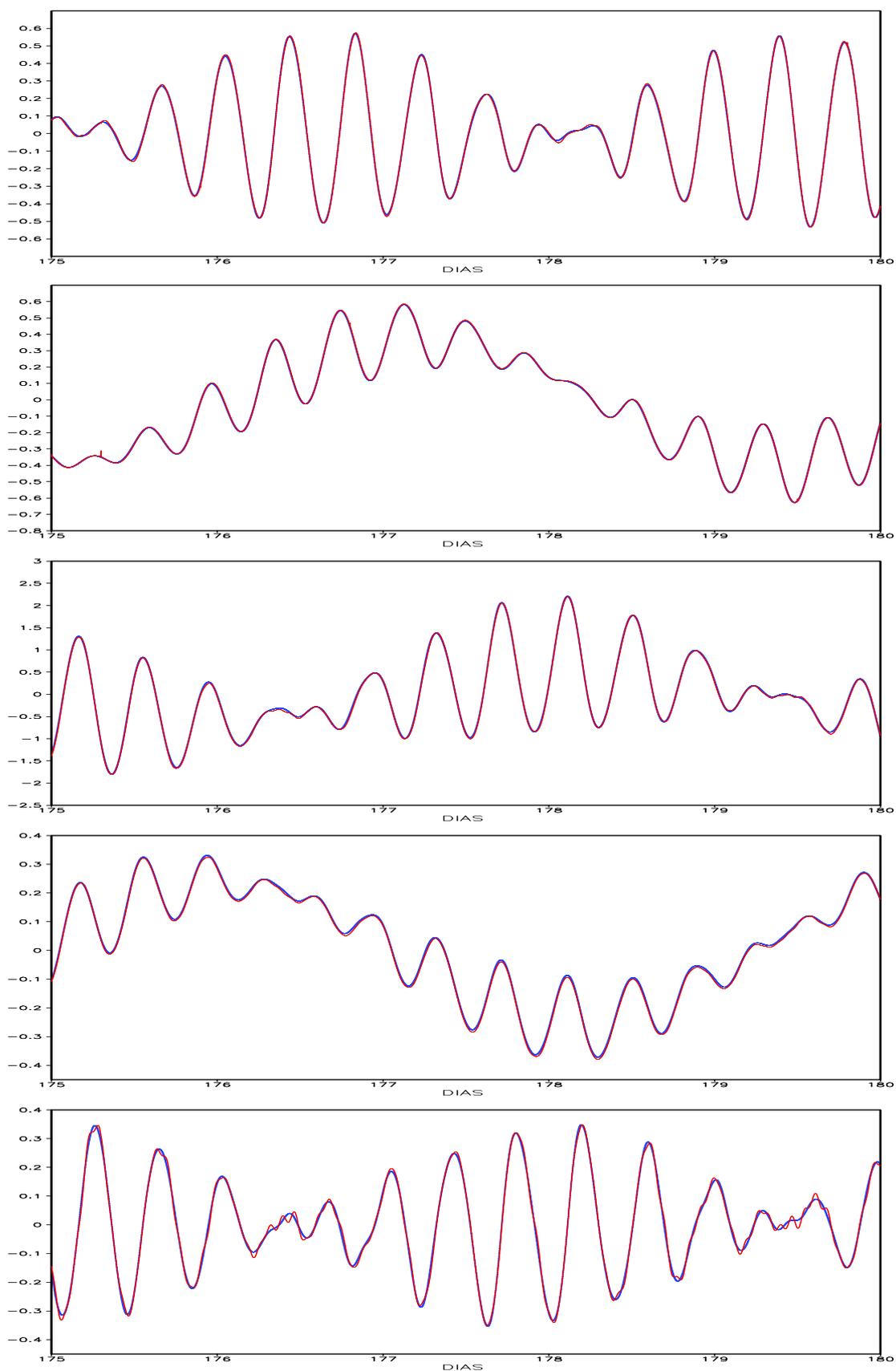


FIGURA 6.6 - Assimilação de dados com o modelo DYNAMO e PMC: de cima para baixo, componentes (a) U , (b) V , (c) ϕ , (d) ζ , (e) δ .

6.2.2 Assimilação: Funções de Base Radial (FBR)

Basicamente, a diferença entre a FBR e a PMC implementadas neste trabalho, está na função de ativação dos neurônios da camada interna das redes. Na FBR, a função de ativação calcula a distância euclidiana entre os vetores de entrada da rede e o centro de cada neurônio, enquanto a função de ativação da PMC calcula o produto interno do vetor de entrada da rede pelo vetor de pesos de cada neurônio.

Há diferentes estratégias para o cálculo dos centros da FBR. Tentou-se calcular os centros através de algoritmos de classificação como k-médias e com a rede da Kohonen (Haykin, 2001), mas a FBR que melhor resolveu o problema de assimilação proposto foi a FBR com função de ativação gaussiana de média zero e variância um na camada oculta.

A seguir, os mesmos experimentos explorados com a rede PMC são refeitos para a rede FBR com uma camada oculta. Apresenta-se os ET e EV nas Figuras 6.7 (a) - (b) e os erros de previsão nas Figuras 6.9 (a) - (c).

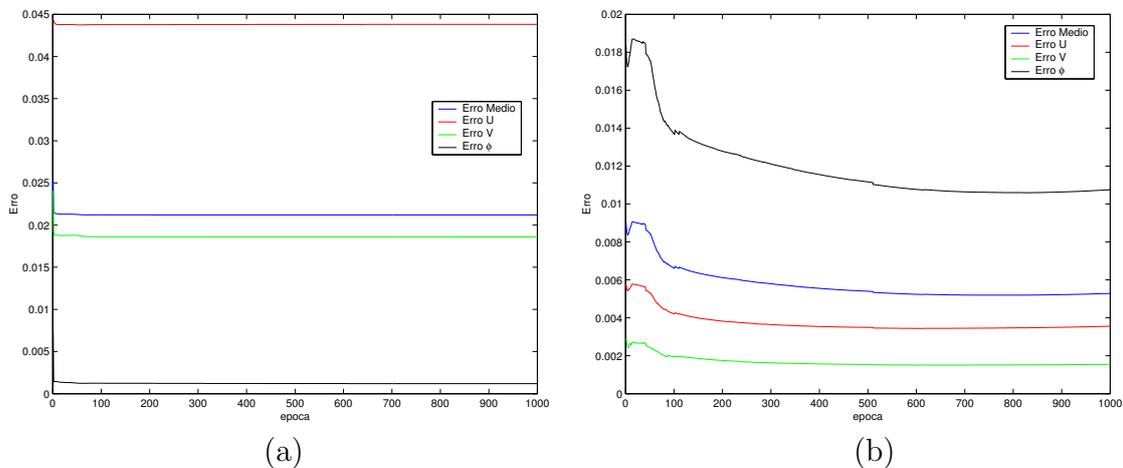


FIGURA 6.7 - Erros com o modelo DYNAMO: (a) durante a fase de treinamento; (b) na validação cruzada.

A melhor arquitetura para a rede FBR foi obtida com 5 neurônios na camada escondida com taxa de aprendizagem e momento descritas na Tabela 6.1. Com esta rede o menor erro de validação foi obtido na 612-ésima época, embora o erro de treinamento diminua abruptamente na 2-ésima época, o que mais uma vez mostra a vantagem na utilização da validação cruzada.

TABELA 6.3 - Erros com a FBR.

	ET	EV	EP
<i>Medio</i>	0,021200	0,005230	0,006754
<i>U</i>	0,043806	0,003438	0,003680
<i>V</i>	0,018600	0,001507	0,002725
ϕ	0,001193	0,010744	0,013535
ζ	—	—	0,005049
δ	—	—	0,008784

Da Tabela 6.3, verifica-se que tanto U , V e ϕ extraíram conhecimento do conjunto de treinamento, porém a generalização do ϕ foi a pior entre todas as variáveis, tanto para com o conjunto de teste como para a previsão. Inclusive o erro de previsão (escalar) de ϕ é maior do que o erro de δ , porém numa análise qualitativa, nota-se que a previsão da vorticidade é inferior a previsão de geopotencial (comparar a Figura 6.9 - (c) com a Figura 6.9 - (e)).

As Figuras 6.9 (a) - (e) mostram que a FBR também apresenta boa performance na assimilação, sendo que são percebidas pequenas oscilações no campo de divergência, como na estimativa com a PMC, as quais podem ser notadas também na Figura 6.8 (campo de divergência para os primeiros 5 dias de integração do DYNAMO, onde a assimilação de dados é feita pela FBR). A FBR foi um pouco mais precisa do que a PMC e exigiu menos neurônios no seu treinamento.

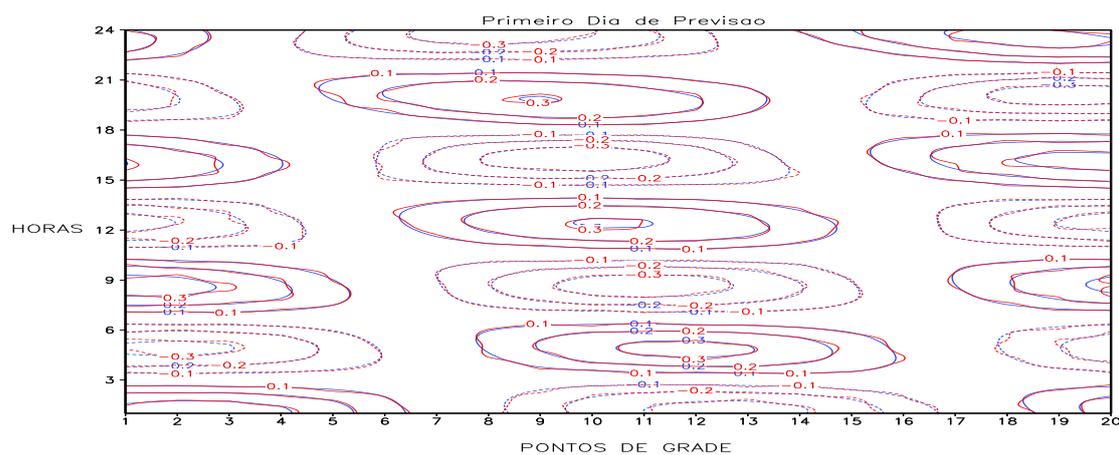


FIGURA 6.8 - Assimilação de dados com o modelo DYNAMO: campo de divergência no começo do domínio de integração. Linha azul - modelo de referência, linha vermelha - assimilação pela FBR.

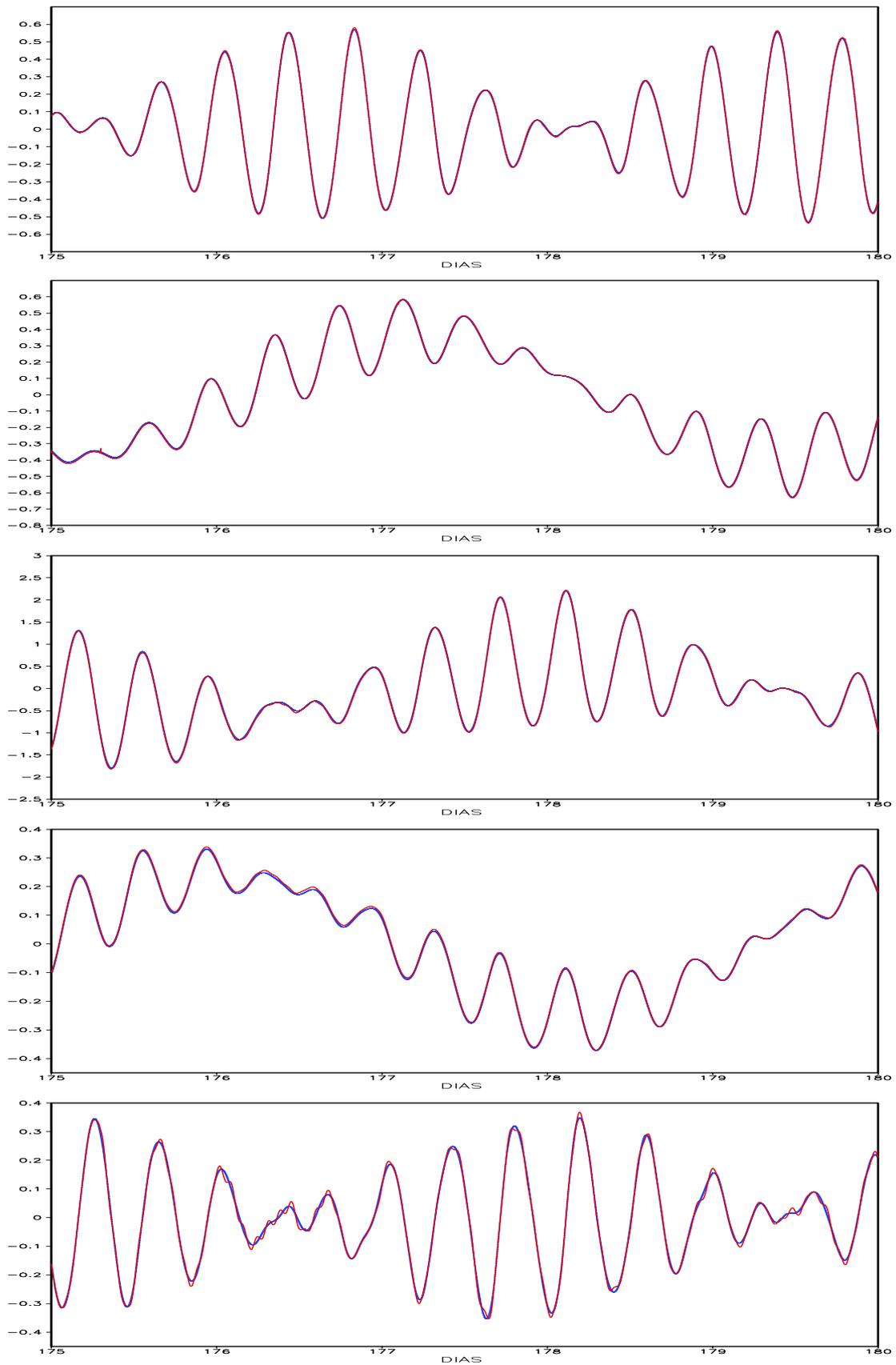


FIGURA 6.9 - Assimilação de dados com o modelo DYNAMO e FBR: de cima para baixo, componentes (a) U , (b) V , (c) ϕ , (d) ζ , (e) δ .

6.2.3 Assimilação: Redes Recorrentes

Nesta Seção se explora os resultados com a RN-E (recorrência da camada escondida para a camada de entrada e com a RN-J (recorrência da camada de saída para a camada de entrada). O estudo das redes recorrentes tem por objetivo verificar se as redes com memória permitem que se aumente o tempo de alcance das previsões sem perda de precisão.

Como nas redes “feedforward” analisa-se os ET, EV e EP, pois estas curvas indicam a viabilidade do método. Nas Figuras 6.10 (a) - (b) são grafados os erros calculados com o conjunto de treinamento e teste para a RN-E. Os resultados apresentados referem-se a implementação da RN-E com uma camada oculta, com taxas de aprendizagem e momento dadas na Tabela 6.1.

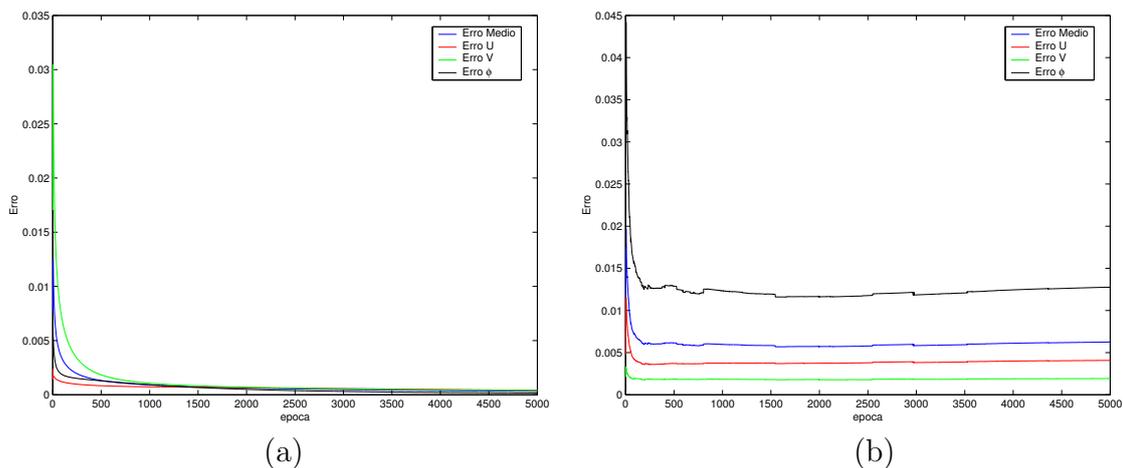


FIGURA 6.10 - Erros com o modelo DYNAMO: (a) durante a fase de treinamento; (b) na validação cruzada.

Para esta rede, o melhor conjunto de pesos foi obtido na 186-ésima época de treinamento, com 5 neurônios na camada oculta. As três componentes assimiladas do modelo DYNAMO apresentam erro de mesma magnitude, ou seja, a rede aprendeu as três componentes sem se especializar numa componente em específico. Este experimento mostra muito bem a importância da validação cruzada, haja vista que sem esta técnica facilmente se incorreria no engano de parar o treinamento após 2000 épocas, causando uma perda de generalização. Os EV mostram generalização menos eficiente no geopotencial (Tabela 6.4), o que se confirma na previsão com a RN-E como método de assimilação, (Figura 6.12 (c)).

TABELA 6.4 - Erros com a RN-E.

	ET	EV	EP
<i>Medio</i>	0,002311	0,006024	0,013286
<i>U</i>	0,001042	0,003626	0,006733
<i>V</i>	0,004290	0,001769	0,004047
ϕ	0,001601	0,012676	0,029939
ζ	—	—	0,006666
δ	—	—	0,019047

Da mesma forma que se fez com as redes PMC e FBR, analisa-se os últimos 5 dias de um total de 180 dias de previsão, Figuras 6.12 (a) - (e). Observa-se que a RN-E apresenta pequeno erro de fase na previsão de vorticidade. No evolução temporal do campo de divergência as ondas curtas são mais evidentes do que nas redes PMC e FBR (Figura 6.12 (e)), fato este que se confirma no campo de divergência, Figura 6.11. Apesar destas ondas, pode-se dizer que a RN-E também resolveu o problema de assimilação com poucos (5) neurônios.

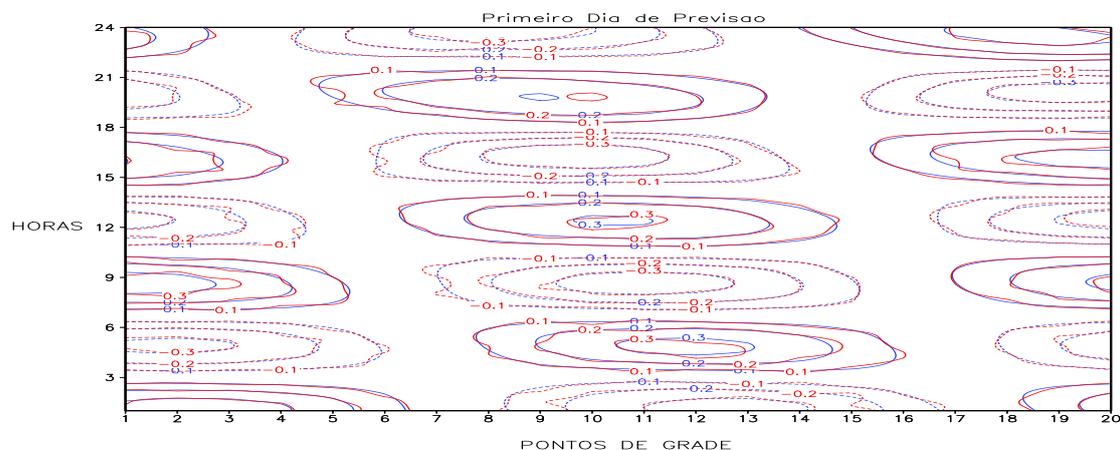


FIGURA 6.11 - Assimilação de dados com o modelo DYNAMO: campo de divergência no começo do domínio de integração. Linha azul - modelo de referência, linha vermelha - assimilação pela RN-E.

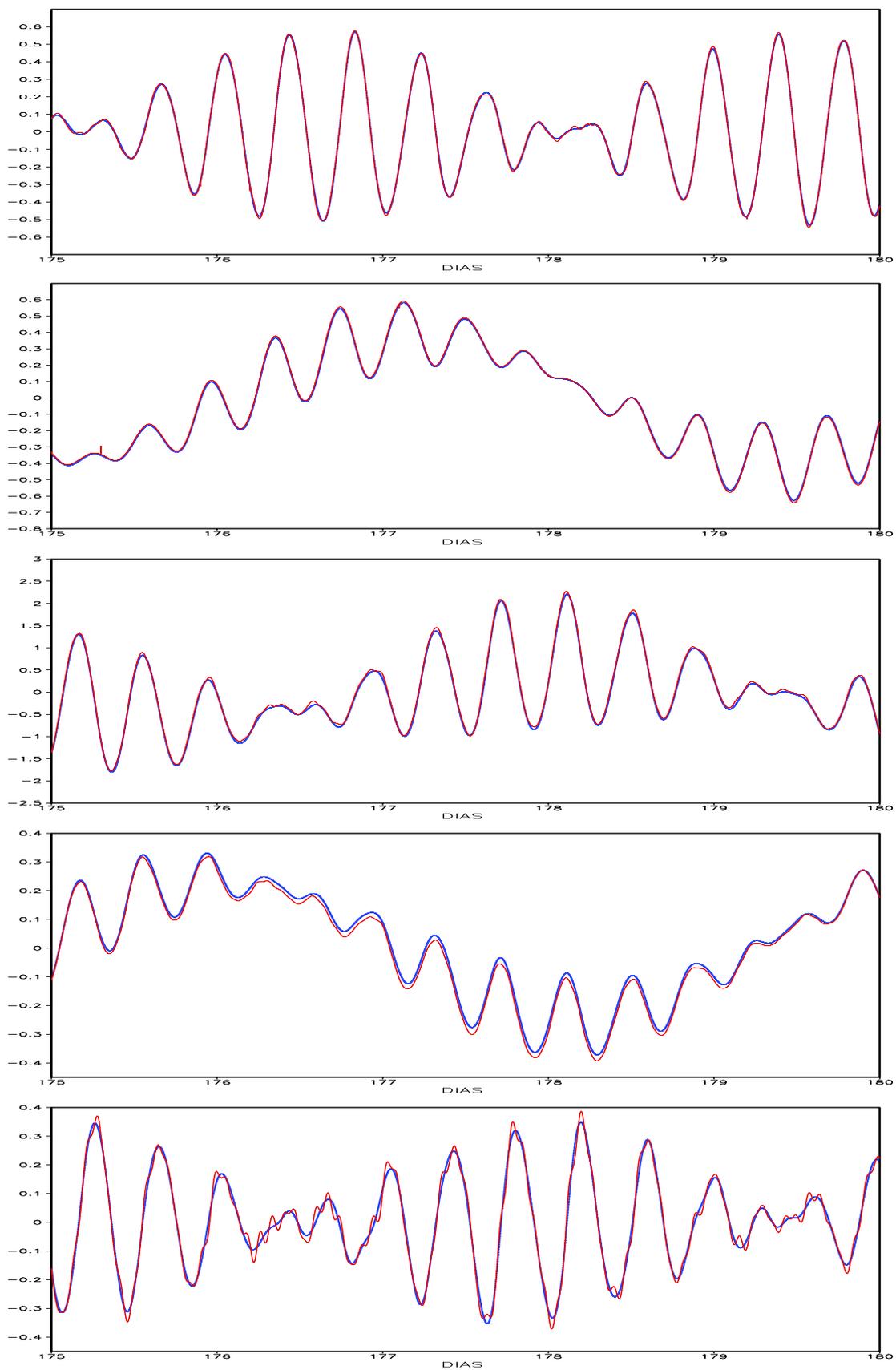


FIGURA 6.12 - Assimilação de dados com o modelo DYNAMO e RN-E: de cima para baixo, componentes (a) U , (b) V , (c) ϕ , (d) ζ , (e) δ .

Entre as redes testadas, a RN-J foi que se revelou menos eficiente no contexto de assimilação de dados com o modelo DYNAMO. A causa desta imprecisão foi o pequeno número de neurônios implementados na camada oculta, o que dificultou o aprendizado. Tentativas de treinamento desta rede com mais neurônios foram mal sucedidas, pois o ET não convergiu. O melhor conjunto de pesos foi obtido após 20-ésima época (Figuras 6.13). Na Tabela 6.5, são mostrados os ET, EV e EP.

TABELA 6.5 - Erros com a RN-J.

	ET	EV	EP
<i>Medio</i>	0,160839	0,009669	0,017384
<i>U</i>	0,105533	0,006170	0,007689
<i>V</i>	0,201924	0,002769	0,008293
ϕ	0,175059	0,020068	0,024584
ζ	—	—	0,025427
δ	—	—	0,020926

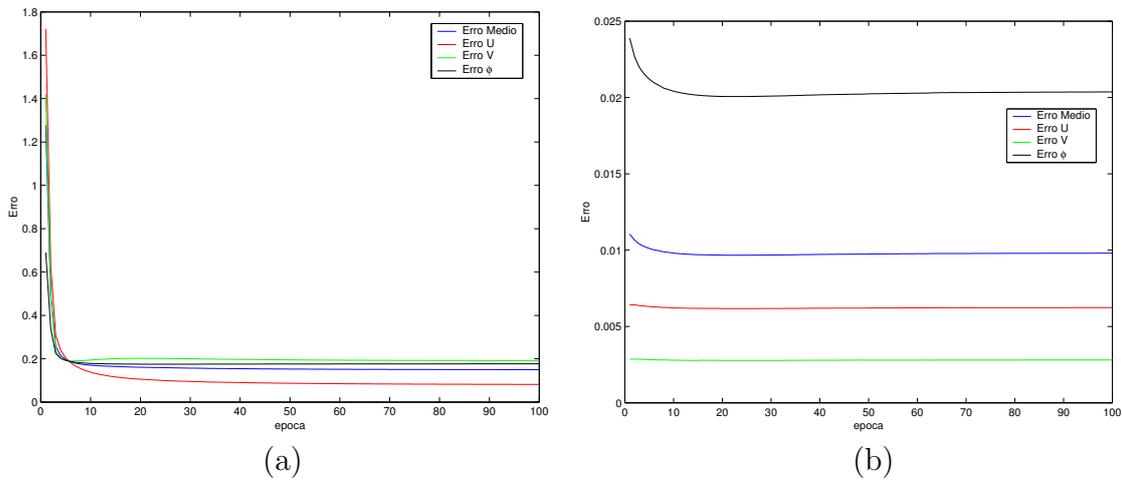


FIGURA 6.13 - Erros com o modelo DYNAMO: (a) durante a fase de treinamento; (b) na validação cruzada.

O ϕ apresentou maior ET e por conseqüência a pior generalização com o conjunto de testes. Quantitativamente, o EP para o ϕ tem a mesma magnitude dos EP para a ζ e para a δ , porém uma análise qualitativa das Figuras 6.14 (a) - (e), revela que os EP em ζ e δ são mais evidentes, havendo erro de amplitude considerável na ζ e ondas curtas na δ (bem evidentes na Figura 6.15). Entretanto, a RN-J resolveu relativamente bem o problema de assimilação com este modelo unidimensional não-linear.

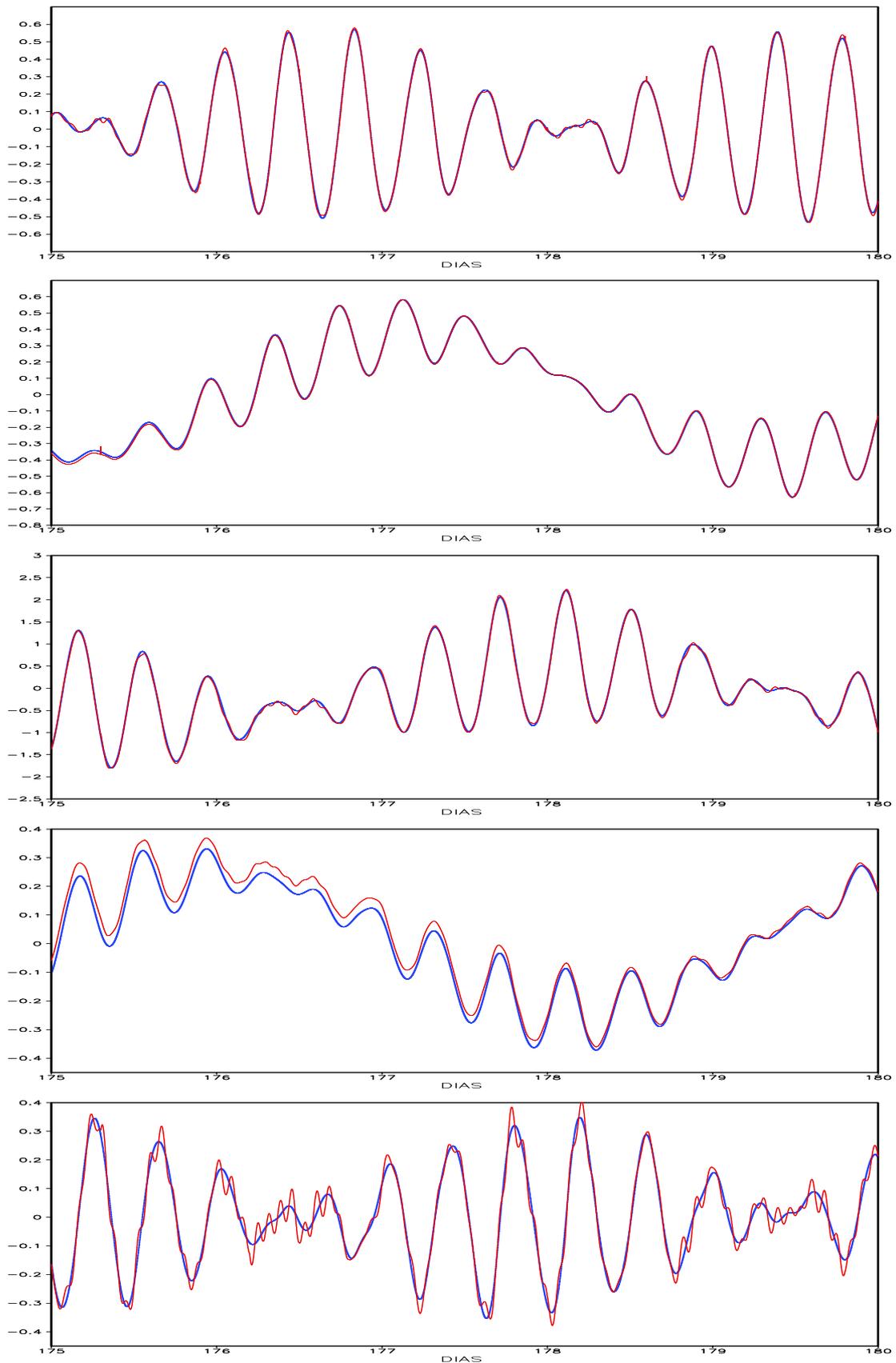


FIGURA 6.14 - Assimilação de dados com o modelo DYNAMO e RN-J: de cima para baixo, componentes (a) U , (b) V , (c) ϕ , (d) ζ , (e) δ .

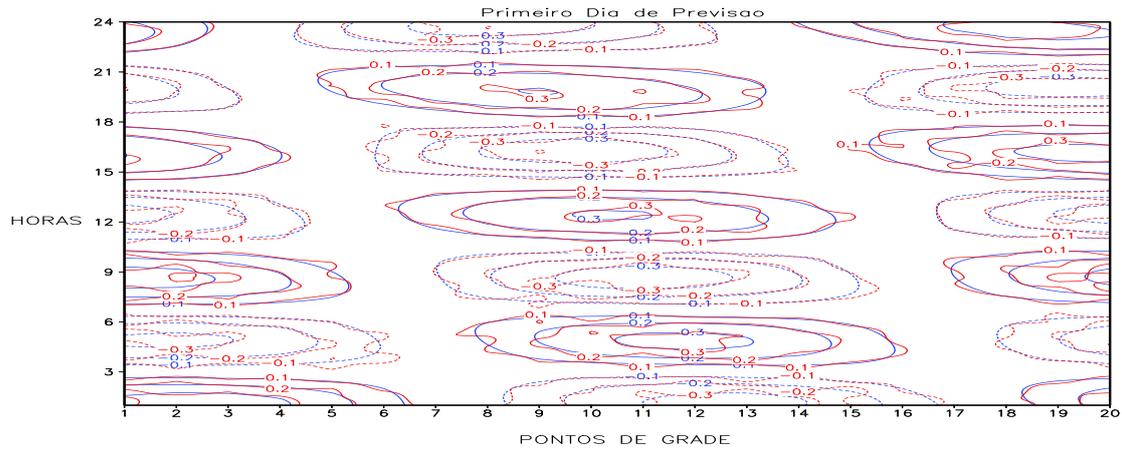


FIGURA 6.15 - Assimilação de dados com o modelo DYNAMO: campo de divergência no começo do domínio de integração. Linha azul - modelo de referência, linha vermelha - assimilação pela RN-J.

Na Figura 6.16 é grafada a evolução temporal do erro de 180 dias previsão com as redes PMC (linha vermelha) e RN-E (linha vermelha) e com o FK (linha preta) como método de assimilação. Observa-se que, embora a assimilação por RN resulte num erro com uma ordem de magnitude maior do que o erro resultante da asssimilação feita com o FK, os erros não tendem a aumentar com o tempo.

Na Tabela 6.6 verifica-se o custo computacional, em segundos, do treinamento de cada RN com o modelo DYNAMO-1D. Faz-se aqui, as mesmas considerações feitas para o sistema de Lorenz, ou seja, o custo computacional esta diretamente relacionado com número de épocas de treinamento, que por sua vez é função do critério de parada adotado. Nos experimentos realizados, a PMC atingiu o melhor conjunto de pesos em 651,6 segundos e a RN-J em 52,7 segundos. Entretanto, ressalta-se que o treinamento é finalizado, quando o erro com o conjunto de validação diminui em relação a época anterior, para as três componentes do modelo, de maneira concomitante.

TABELA 6.6 - Tempo de Treinamento (s) com o DYNAMO-1D.

<i>RedeNeural</i>	Tempo (s)	Épocas de Treinamento
<i>PMC</i>	651,6	220
<i>FBR</i>	1741,0	612
<i>RN - E</i>	402,9	186
<i>RN - J</i>	52,7	20

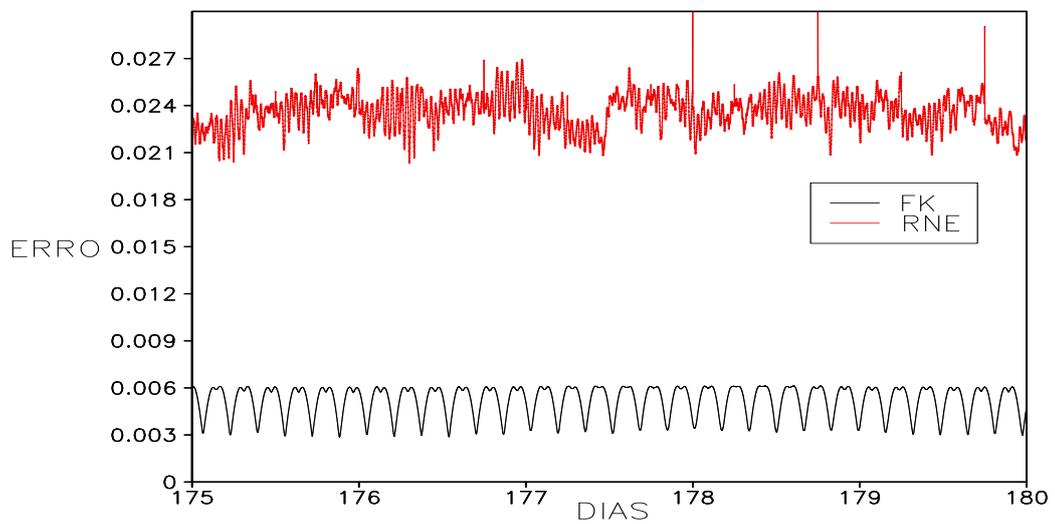
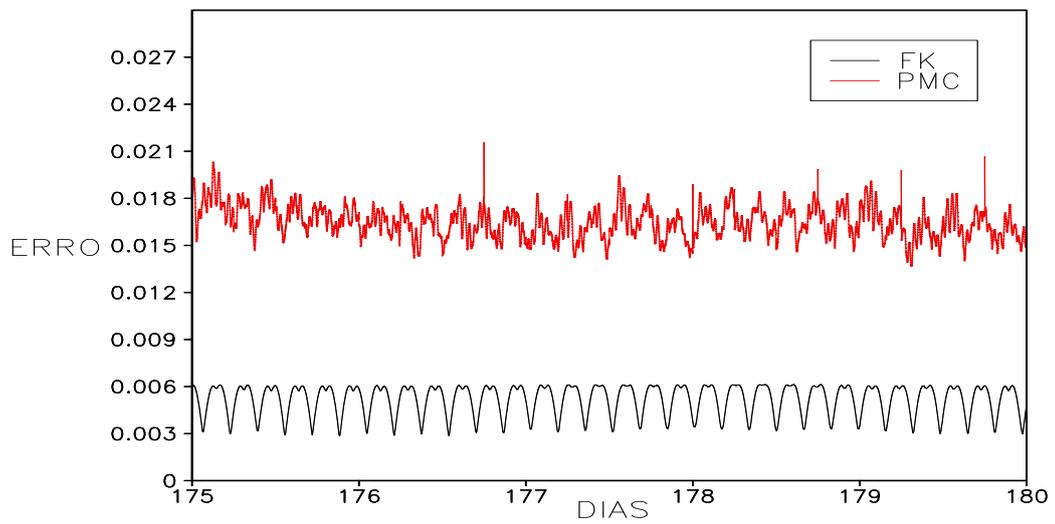


FIGURA 6.16 - Evolução temporal do erro de previsão para (a) PMC, (b) RN-E.

CAPÍTULO 7

CONCLUSÕES

Assimilação de Dados talvez seja o mais importante tema de pesquisa em previsão numérica de tempo da atualidade. Nesta tese avalia-se a eficiência das redes Perceptron de Múltiplas Camadas (PMC), Funções de Base Radial (FBR), Rede de Elman (RN-E) e Rede de Jordan (RN-J), num contexto de assimilação de dados em dinâmica não-linear. A rede FBR (“feedforward padrão”), a RN-E e a RN-J são inéditas neste tipo de aplicação.

Na avaliação das redes utiliza-se dois modelos, o sistema de Lorenz em regime caótico e o modelo de Água Rasa não-linear (DYNAMO-1D). O primeiro tem sido utilizado na literatura (por exemplo, Miller (1994)) para testar técnicas de assimilação de dados. O modelo DYNAMO - 1D, embora tenha sido concebido para testar esquemas de inicialização (por exemplo Heemink et al. (1997)) também vem sendo empregado em testes de novas técnicas de assimilação, pois mesmo sendo um sistema simplificado, possui em sua solução ondas de Rossby e ondas de Gravidade, dimensão espacial e não-linearidade.

Nesta tese explorou-se os seguintes pontos inéditos em relação aos trabalhos encontrados na literatura:

- a) as redes FBR, RN-E e RN-J até então não haviam sido exploradas neste tipo de aplicação. A FBR foi investigada por representar uma alternativa a redes *feedforward* padrão e as RN-E e RN-J por serem recorrentes. Há especial interesse na RN-E e na RN-J porque as conexões recorrentes acrescentam memória ao sistema, o que se supõe melhorar as previsões à medida que se aumenta o alcance das mesmas. Todavia, verificou-se que as redes resolveram o problema de assimilação de dados com menos neurônios comparadas com a PMC, assim como a rede FBR. Isto representa uma redução na complexidade dos algoritmos, o que é desejável neste tipo de aplicação. Nos testes aqui apresentados (com modelo simples) não foi verificado melhora nas previsões de mais longo alcance com o uso das redes recorrentes;
- b) no treinamento das RN implementou-se a técnica conhecida como correlação cruzada (*cross-validation*), o que permite que se conheça a superfície

de erros de treinamento e de validação, permitindo que se obtenha o melhor conjunto de pesos para o problema estudado na implementação feita. Esta técnica explorada no treinamento das redes representa um avanço em relação a metodologia de treinamento de RN em Assimilação de Dados;

- c) diminuição do espaço de busca, o que vem a ser o principal resultado deste trabalho, pois em problemas de minimização de gradiente, em aplicações de grande dimensão, diminuir o espaço de busca pode ser a diferença entre tornar a aplicação operacionalmente viável ou não. Supondo um modelo operacional com 10^6 pontos de grade, partindo da hipótese de que utilizaríamos 10^5 neurônios na camada interna de uma PMC, como na versão de Nowasad (2001), o número de conexões a serem identificadas na rede seria

$$2 \times 10^6 \times 10^5 + 10^6 \times 10^5 = 3 \times 10^{11}, \quad (7.1)$$

onde 2×10^6 refere-se ao número de variáveis em pontos de grade na entrada da RN, 10^5 refere-se ao número de neurônios, pois nesta forma de implementação o número de neurônios aumenta com o aumento do número de pontos de grade. 10^6 na segunda parcela da soma refere-se ao número de variáveis em pontos de grade na saída da RN.

Na atual versão do algoritmo, para uma RPM ter-se-ia

$$6 \times 15 + 3 \times 15 = 135, \quad (7.2)$$

onde 6 refere-se ao número de variáveis em pontos de grade na entrada da RN, 15 refere-se ao número de neurônios na camada interna e 3, na segunda parcela da soma, refere-se ao número de variáveis em pontos de grade na saída da RN. Ou seja, na versão do algoritmo explorado neste trabalho, a redução do número de conexões entre os pesos diminui na ordem de 10^9 ;

- d) aplicação em clima espacial, encontra boa aceitação por parte da comunidade de geofísica, o que abre uma nova linha de pesquisa nesta área, por isso se considera os resultados obtidos com a assimilação de dados em clima espacial, um resultado importante desta tese.

O grande desafio em Assimilação de Dados hoje, se dá pela dimensão do problema,

em termos de quantidade de pontos de grade nos modelos de previsão (10^6) e da quantidade de dados observados (que cresce vertiginosamente). Isto representa um desafio computacional que vem ocupando os cientistas da área.

Atualmente, o debate científico sobre o tema de Assimilação de Dados que vem ocupando a literatura científica recente é a aplicação do Filtro de Kalman por Conjuntos (Kalnay, 2004) e do Método Variacional em 4 Dimensões (Courtier et al., 1994).

Embora os resultados indiquem que estas metodologias possam ser efetivas para assimilação em relação a precisão, há dúvidas se estes algoritmos são computacionalmente eficientes para realizar a tarefa no tempo de uma previsão operacional, à medida que o número de pontos de grade cresçam em uma ordem de magnitude e os dados de observação aumentem em várias ordens de magnitude.

As RN com sua capacidade de emular sistemas, aliadas a duas características: (i) ser um algoritmo intrinsecamente paralelo; (ii) serem propícias a implementação em *hardware* (neuro-computadores), poderá ser uma solução ao desafio da Assimilação de Dados num cenário onde se tem uma quantidade crescente de observações disponíveis e onde os modelos de previsão estão sempre evoluindo em termos de quantidade de pontos de grade computacional, pois hoje em dia já se pretende trabalhar com modelos globais em que, para determinadas regiões específicas do globo, se tenha o refinamento e a física de modelos de mesoescala.

O trabalho aqui apresentado de aplicação de RN (com recorrência inclusive) como uma nova metodologia que permite reduzir drasticamente a complexidade do algoritmo (principalmente em relação ao tempo de treinamento) pretende ser uma contribuição ao desafio da assimilação de dados operacional.

Esta contribuição será efetiva, à medida que os resultados obtidos com o sistema de Lorenz e o modelo DYNAMO-1D sejam reproduzidos com modelos de 3-dimensões. Entretanto, os testes apresentados levam a crer que, a reprodução dos resultados desta metodologia para assimilação de dados, seja obtida em sistemas de dimensão maior, pois RN são algoritmos capazes de aprender através de “exemplos”, sejam eles oriundos de um modelo de 1, 2 ou 3 dimensões (ou sinais de diversas outras fontes, com uma infinidade de aplicações, ver (Haykin, 2001)). Por isso, para avaliar se o sistema de assimilação proposto é viável em aplicações práticas, é necessário responder as seguintes questões: 1) o método é eficiente em situações de não-linearidade

forte ?, 2) é possível minimizar o gradiente quando se tem 10^6 ou mais pontos de grade, aproveitando-se ao máximo a quantidade de observações disponíveis ? Os resultados obtidos com o sistema de Lorenz e o Modelo de 3-Ondas em regime caótico, levam a crer que o método é robusto quando submetido a regime fortemente não-linear, respondendo a primeira questão. Neste trabalho, as redes foram treinadas, apresentando-se um ponto de grade por época de treinamento, o que reduz a dimensão do gradiente a ser minimizado, como comentado acima. Esta abordagem adotada no treinamento das RN, faz com que o gradiente a ser minimizado tenha a mesma ordem, seja para sistemas de 1 ou 3 dimensões, o que responde a segunda questão. A diferença é que em sistemas de 3 dimensões, o gradiente deverá ser minimizado mais vezes. Isto é possível, porque o conhecimento adquirido pelas RN, armazenado nos pesos, independe da correlação entre os pontos. Assim, o fato das RN serem robustas a regimes fortemente não-lineares e por ser possível o seu treinamento ponto a ponto, permite a suposição de que esta técnica seja uma alternativa promissora na solução do problema de assimilação de dados em aplicações operacionais de previsão de tempo.

Como trabalhos futuros, um caminho natural é o uso das técnicas de RN para assimilação de dados em modelos de dimensão maior, como o DYNAMO-2D, chegando-se até a PNT operacional. As RN devem emular, por exemplo, o Local Analysis and Prediction System (LAPS)(Albers, 1995), sistema de assimilação do Modelo Brasileiro de Alta Resolução (MBAR) (Bernardet et al., 2000), utilizado no Instituto Nacional de Meteorologia (INMET), ou o PSAS e RPSAS, sistema de assimilação implementado pelo CPTEC nos modelos global e regional, respectivamente, referenciados no primeiro capítulo desta tese, ou qualquer outro sistema de assimilação usado por outros centros operacionais de previsão.

Um tema de pesquisa básica na área, seria a investigação da aplicação da rede recorrente de Hopfield (RH), (Hopfield, 1982). A RH representa uma mudança de paradigma em relação as redes apresentadas, pois não se tem uma fase de ativação e o treinamento é feito à medida que a equação (diferencial no caso de assimilação de dados) evolui no tempo. Observa-se também, que esta rede é não-supervisionada, ou seja, não há um alvo que deva ser atingido durante a fase de treinamento. Na aplicação da RH o sistema evolui no sentido de se encontrar o estado de energia mínima, como descrito a seguir.

Considerando que se esta resolvendo um problema inverso do tipo $\mathbf{z}_n = \mathbf{H}\mathbf{w}_n^a + \varepsilon$,

expressa-se o erro ε como nos mínimos quadrados ponderados

$$\varepsilon = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{w}_n - \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z}_n, \quad (7.3)$$

onde $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} = \mathbf{A}$ e $\mathbf{H}^T \mathbf{R}^{-1} = \mathbf{B}$, de forma que, para um sistema 3×3 (como Lorenz) \mathbf{A} e \mathbf{B} são matrizes IXJ , para $I = 3$ e $J = 3$ independentes do tempo e \mathbf{w}^f e \mathbf{z} são vetores $JX1$ dependentes do tempo n .

Este erro (função de energia que deve ser mínima) é expresso na seguinte forma quadrática como em Vemuri e Gyu-Sang (1992)

$$\mathbf{E} = \frac{1}{2} \left[\left(\sum_{j=1}^J \mathbf{A}_{1j} \mathbf{w}_{j,n} - \mathbf{B}_{1j} \mathbf{z}_{1,n} \right)^2 + \left(\sum_{j=1}^J \mathbf{A}_{2j} \mathbf{w}_{j,n} - \mathbf{B}_{2j} \mathbf{z}_{2,n} \right)^2 + \left(\sum_{j=1}^J \mathbf{A}_{3j} \mathbf{w}_{j,n} - \mathbf{B}_{3j} \mathbf{z}_{3,n} \right)^2 \right]. \quad (7.4)$$

Desta função de energia, a RH (Hopfield, 1984) é formulada através da seguinte relação Hamiltoniana

$$\frac{dw_i}{dt} = - \frac{\partial \varepsilon}{\partial w_i}, \quad (7.5)$$

que derivando resulta em

$$\begin{aligned} \frac{dw_i}{dt} = & - \left(\sum_{j=1}^J A_{1j} w_{j,n} - B_{1j} z_{1,n} \right) A_{1i} \\ & - \left(\sum_{j=1}^J A_{2j} w_{j,n} - B_{2j} z_{2,n} \right) A_{2i} \\ & - \left(\sum_{j=1}^J A_{3j} w_{j,n} - B_{3j} z_{3,n} \right) A_{3i}, \end{aligned} \quad (7.6)$$

que pode ser escrita como

$$\frac{dw_i}{dt} = \sum_{j=1}^J T_{i,j} w_{j,n} + I_{i,n}, \quad (7.7)$$

com os pesos T_{ij} e viés $I_{i,n}$ dados pelas expressões abaixo

$$T_{ij} = - \sum_{l=1}^L A_{li} A_{lj}, \quad (7.8)$$

$$I_i = \sum_{j=1}^J C_{ji} z_{j,n}, \quad (7.9)$$

sendo que $C = BA$ e a estimativa é obtida por

$$w_n^a = g(W_n). \quad (7.10)$$

Resultados parciais (não apresentados nesta tese) devem se aprofundados futuramente.

REFERÊNCIAS BIBLIOGRÁFICAS

ACKLEY, D.; HINTON, G.; SEJNOWSKI, T. A learning algorithm for boltzman machines. **Cognitive Science**, v. 9, n. 1, p. 147–169, 1985.

ALBERS, S. The laps wind analysis. **Weather and Forecasting**, v. 10, n. 2, p. 342–352, 1995.

ANDERSON, B.; MOORE, J. **Optimal filtering**. Englewood Cliffs: Prentice-Hall, 1979. 357 p.

BELYAEV, K.; TANAJURA, C. An extension of a data assimilation method based on the application of the Fokker-Planck equation. **Applied Mathematical Modelling**, v. 26, n. 11, p. 1019–1027, 2002.

BELYAEV, K.; TANAJURA, C.; O'BRIEN, J. **A data assimilation method used with an ocean circulation model and its application to the tropical atlantic**. Petrópolis: Gráfica do LNCC, 2000. 357 p.

BENNET, A. **Inverse methods in physical oceanography**. Cambridge: University Press, 1992. 346 p.

BERGTHORSSON, P.; DÖÖS, B. Numerical weather map analysis. **Tellus**, v. 7, n. 3, p. 329–340, 1955.

BERNARDET, L.; SILVEIRA, R.; EDWARDS, J. Verificação objetiva das previsões numéricas do modelo brasileiro de alta resolução. In: SOCIEDADE BRASILEIRA DE METEOROLOGIA, 11., 2000, Rio de Janeiro. **Anais...** Rio de Janeiro: SBMET, 2000. p. 3557–3565. 1 CD-ROM.

BJERKNES, V. **Dynamic meteorological and hidrography**. Cargnegie Institute, New York: Gibson Bros., 1911.

BRAGA, A.; LUDERMIR, T.; CARVALHO, A. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC - Livros Técnicos e Científicos, 2000. 262 p.

CAMPOS VELHO, H.; CLAEYSSSEN, J. Initialization of a barotropic model using dynamical matrix solution approach. In: INTERNATIONAL CONFERENCE ON INDUSTRIAL AND APPLIED MATHEMATICS, 1991, Washington. **Proceedings...** Washington, D.C., 1991. p. 8–12.

_____. A comprehensive analysis of a barotropic limited area model using the nonmodal matrix technique. **Revista Brasileira de Meteorologia**, v. 12, n. 2, p. 41–50, 1997.

CAMPOS VELHO, H.; VIJAYKUMAR, N.; STEPHANY, S.; PRETO, A.; NOWOSAD, A. **A neural network implementation for data assimilation using MPI, applications of high-performance computing in engineering**. United Kingdom: Southampton: WIT Press, 2002.

CHÁVEZ, F. **Dinâmica caótica de ondas de Alfvén no vento solar**. 2000. 134 p. (INPE-8161-TDI/755). Dissertação (Mestrado em Geofísica Espacial) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2000.

CINTRA, R. **Implementação do sistema estatístico de assimilação de dados em espaço físico para o modelo global do CPTEC**. 2004. Dissertação (Mestrado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2004.

COURTIER, P.; J., T.; HOLLINGSWORTH, A. A strategy for operational implementation of 4d-var, using an incremental approach. **Quarterly Journal Royal Meteorological Society**, v. 120, n. 519 B, p. 1367–1387, 1994.

CRESSMAN, G. An operational objective analysis system. **Monthly Weather Review**, v. 87, n. 10, p. 367–374, 1959.

DALEY, R. **Atmospheric data analysis**. Cambridge: University Press, 1991. 457 p.

DEGRAZIA, G.; MORAES, O. A model for eddy diffusivity in a stable boundary layer. **Boundary Layer Meteorology**, v. 58, n. 3, p. 205–214, 1992.

ELIASSEN, A. **Provisional report on calculation of spatial covariance and autocorrelation of the pressure field**. Oslo: [S.n], 1954. 11 p.

FAHLMAN, S.; LEBIERE, C. The cascade-correlation learning architecture. In: MORGAN KAUFMANN PUBLISHERS, 2., 1990, San Mateo. **Advances in Neural Information Processing Systems**. Denver 1989, 1990. p. 524–532.

Disponível em:

<http://citeseer.ist.psu.edu/fahlman90cascadecorrelation.html>. Acesso em: 29 jun. 2005.

GANDIN, L. **Objective analysis of meteorological fields**. Leningrad: Gibson, 1963. 240 p.

GARDNER, M. W.; DORLING, S. R. Artificial neural networks (the multilayer perceptron - a review of applications in the atmospheric sciences). **Atmospheric Environment**, v. 32, n. 14, p. 2627–2636, 1998.

GOLUB, G.; LOAN, V. **Matrix computations**. Baltimore: Johns Hopkins University Press, 1983. 476 p.

HÄRTER, F.; CAMPOS VELHO, H. Applying neural network in nonlinear model: an encouraging approach. In: WMO INTERNATIONAL SYMPOSIUM ON ASSIMILATION OF OBSERVATION IN METEOROLOGY AND OCEANOGRAPHY, 4., 2005, Praga. **Proceedings...** Praga: WMO, 2005. v. 1, p. 219.

_____. 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, p. 411–420, 2005.

HÄRTER, F.; CAMPOS VELHO, H.; CHAMON, M. Kalman filtering in the air quality monitoring. **Ciência e Natura**, p. 177–187, 2002.

HÄRTER, F.; CAMPOS VELHO, H.; REMPEL, E.; CHIAN, A. Neural networks in auroral data assimilation. **Journal of Atmospheric and Solar-Terrestrial Physics**. No prelo.

HÄRTER, F. P. **Uso de filtro digital para iniciar um modelo de área limitada**. 1999. 111 p. (INPE-7251-TDI/695). Dissertação (Mestrado em Meteorologia) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1999.

HAYKIN, S. **Redes neurais: princípios e práticas**. Porto Alegre: Bookman, 2001. 900 p.

HEBB, D. **The organization of behavior: a neuropsychological theory**. New York: Wiley-Interscience, 1949. 256 p.

HEEMINK, A.; BOLDING, K.; VERLAAN, M. Storm surge forecasting using kalman filtering. **Journal of the Meteorological Society of Japan**, v. 75, n. 1B, p. 305–318, 1997.

HOFFMAN, J. **Numerical methods for engineers and scientists**. Singapore: McGraw-Hill, 1993. 825 p.

HOPFIELD, J. Neural networks and physical systems with emergent collective computational abilities. **Biophysics**, National Academic Science, v. 79, n. 8, p. 2554–2558, 1982.

_____. Neurons with graded response have collective computational properties like those of two-state neurons. **Biophysics**, National Academic Science, v. 81, n. 10, p. 3088–3092, 1984.

HSIEH, W.; TANG, B. Applying neural network models top prediction and data analysis in meteorology and oceanography. **Bulletin of the American Meteorological Society**, v. 79, n. 9, p. 1855–1870, 1998.

KALNAY, E. **Atmospheric modeling, data assimilation and predictability**. United Kingdom: Cambridge University Press, 2004. 341 p.

KEPPENE, C. Data assimilation into a primitive-equation model with a parallel ensemble kalman filter. **Monthly Weather Review**, v. 128, n. 6, p. 1971–1981, 2000.

KOVACS, Z. **Redes neurais artificiais: princípio e aplicações**. Rio de Janeiro: Collegium Cognition, 1996. 173 p.

LIAQAT, A.; FUKUHARA, M.; TAKEDA, T. Applying a neural collocation method to an incompletely known dynamical system via weak constraint data assimilation. **Monthly Weather Review**, v. 131, n. 8, p. 1697–1714, 2003.

LORENZ, E. Deterministic nonperiodic flow. **Journal of the Atmospheric Sciences**, v. 20, n. 2, p. 130–141, 1963.

LYNCH, P. **The slow equations**. Glasnevin Hill, Dublin, 1987. 73 p.

McCULLOCH, W.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, n. 7, p. 115–133, 1943.

MENDEL, J. **Lessons in estimation theory for signal processing, communicatins, and control**. Englewood Cliffs: Prentice Hall, 1995. 561 p.

MILLER, R.; GUIL, M.; GAUTHIEZ, F. Advanced data assimilation in strongly nonlinear system. **Journal of the Atmospheric Sciences**, v. 51, n. 8, p. 1037–1056, 1994.

MINSKY, M.; PAPERT, S. **Perceptrons**. Cambridge: MIT Press, 1969. 258 p.

NOWOSAD, A. **Novas abordagens em assimilação de dados meteorológicos**. 2001. 120 p. (INPE-8481-TDI/779). Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2001.

NOWOSAD, A.; CAMPOSVELHO, H.; RIOSNETO, A. Neural network as a new approach for data assimilation. In: SOCIEDADE BRASILEIRA DE METEOROLOGIA, 2000, Brazil. **Anais...** Rio de Janeiro, 2000. p. 3078–3086. 1 CD-ROM.

PANOFSKY, H. Objective weather-map analysis. **Journal of Applied Meteorology**, v. 6, n. 6, p. 386–392, 1949.

PLATZMAN, G. A retrospective view of richardson's book on weather prediction. **Bulletin of the American Meteorological Society**, v. 48, n. 8, p. 514–550, 1967.

RICHARDSON, L. **Weather prediction by numerical process**. Cambridge: University Press, 1922.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, p. 386–408, 1958.

RUMELHART, D.; HINTON, G.; R.J., W. Learning representation of back-propagation errors. **Nature**, v. 323, n. 6088, p. 533–536, 1986.

SALTZMAN, B. Finite amplitude free convection as an initial value problem. **Journal of the Atmospheric Sciences**, v. 19, p. 329–341, 1962.

SHIGUEMORI, E.; HÄRTER, F.; CAMPOS VELHO, H.; SILVA, J. da. Estimation of boundary conditions in heat transfer by neural network. **TEMA: Tendências em Computação e Matemática Aplicada**, v. 3, n. 2, p. 189–195, 2001.

- SORENSEN, H. Least-square estimation: from Gauss to Kalman. **IEEE Spectrum**, v. 7, n. 7, p. 63–68, 1970.
- STRANG, G. **Introduction to applied mathematics**. Cambridge: Wellesley-Cambridge Press, 1986. 758 p.
- TALAGRAND, O. Assimilation of observations, an introduction. **Journal of the Meteorological Society of Japan**, v. 75, n. 1B, p. 191–209, 1997.
- TANG, Y.; HSIEH, W. Coupling neural networks to incomplete dynamical systems via variational data assimilation. **Monthly Weather Review**, v. 129, n. 4, p. 818–834, 2001.
- VEMURI, V.; GYU-SANG, J. Inversion of fredholm integral equations of the first kind with fully connected neural networks. **Journal of the Franklin Institute**, v. 329, n. 2, p. 241–257, 1992.
- WIDROW, B. **Generalization and information storage in networks of adalines neurons self-organizing systems**. Washington, DC: Spartan Books, 1962. 435-461 p.
- ZHANG, X.; HEEMINK, A.; EIJKEREN, J. Data assimilation in transport models. **Applied Mathematical Model**, v. 21, n. 1, p. 1–14, 1997.