



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-15235-TDI/1322

REDES NEURAIIS E DIFERENTES MÉTODOS DE ASSIMILAÇÃO DE DADOS EM DINÂMICA NÃO LINEAR

Helaine Cristina Moraes Furtado

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada,
orientada pelos Drs. Elbert Einstein Neher Macau e Haroldo Fraga de Campos Velho,
aprovada em 22 de fevereiro de 2008.

O original deste documento está disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m17@80/2008/02.12.12.07>>

INPE
São José dos Campos
2008

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3945-6911/6923

Fax: (012) 3945-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO:

Presidente:

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Membros:

Dr. Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Haroldo Fraga de Campos Velho - Centro de Tecnologias Especiais (CTE)

Dra. Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Dr. Ralf Gielow - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr. Wilson Yamaguti - Coordenação Engenharia e Tecnologia Espacial (ETE)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Jefferson Andrade Anselmo - Serviço de Informação e Documentação (SID)

Simone A. Del-Ducca Barbedo - Serviço de Informação e Documentação (SID)

Vinicius da Silva Vitor - Serviço de Informação e Documentação (SID) - bolsista

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Marilúcia Santos Melo Cid - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva e Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Viveca Sant´Ana Lemos - Serviço de Informação e Documentação (SID)



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-15235-TDI/1322

REDES NEURAIIS E DIFERENTES MÉTODOS DE ASSIMILAÇÃO DE DADOS EM DINÂMICA NÃO LINEAR

Helaine Cristina Moraes Furtado

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada,
orientada pelos Drs. Elbert Einstein Neher Macau e Haroldo Fraga de Campos Velho,
aprovada em 22 de fevereiro de 2008.

O original deste documento está disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m17@80/2008/02.12.12.07>>

INPE
São José dos Campos
2008

Dados Internacionais de Catalogação na Publicação (CIP)

F983r Furtado, Helaine Cristina Moraes.

Redes neurais e diferentes métodos de assimilação de dados em dinâmica não linear/ Helaine Cristina Moraes Furtado. – São José dos Campos: INPE, 2008.

123p. ; (INPE-15235-TDI/1322)

1. Assimilação de dados. 2. Redes neurais. 3. Filtro de partículas. 4. Método variacional. 5. Filtro de Kalman.
I. Título.

CDU 681.3.019

Copyright © 2008 do MCT/INPE. Nenhuma parte desta publicação pode ser reproduzida, armazenada em um sistema de recuperação, ou transmitida sob qualquer forma ou por qualquer meio, eletrônico, mecânico, fotográfico, microfilmico, reprográfico ou outros, sem a permissão escrita da Editora, com exceção de qualquer material fornecido especificamente no propósito de ser entrado e executado num sistema computacional, para o uso exclusivo do leitor da obra.

Copyright © 2008 by MCT/INPE. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use of the reader of the work.

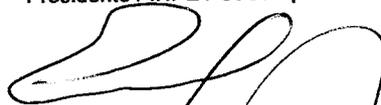
Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de Mestre em
Computação Aplicada

Dr. Fernando Manuel Ramos



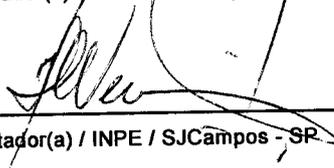
Presidente / INPE / SJC Campos - SP

Dr. Elbert Einstein Neher Macau



Orientador(a) / INPE / SJC Campos - SP

Dr. Haroldo Fraga de Campos Velho



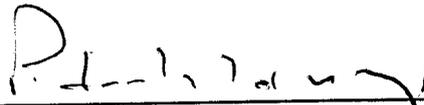
Orientador(a) / INPE / SJC Campos - SP

Dra. Margarete Oliveira Domingues



Membro da Banca / INPE / SJC Campos - SP

Dr. Pedro Leite da Silva Dias



Convidado(a) / LNCC / Petrópolis - RJ

Aluno (a): Helaine Cristina Morais Furtado

São José dos Campos, 22 de Fevereiro de 2008

Não Vou Sair

Celso Viáfora e Nilson Chaves

A geração da gente
Não teve muita chance
De se afirmar de arrasar de ser feliz

Sem nada pela frente pintou aquele lance
De se mudar de se mandar desse país

E aí você partiu pro Canadá
Mas eu fiquei no “já vou já”
Pois quando tava me arrumando pra ir

Bati com os olhos no luar
E a lua foi bater no mar
E eu fui que fui ficando...

Distante tantas milhas
São tristes os invernos
Não vou sair tá mal aqui mas vai mudar

Os velhos de Brasília

Não podem ser eternos

Pior que foi pior que tá não vai ficar

Não vou sair melhor você voltar pra cá
Não vou deixar esse lugar
Pois quando tava me arrumando pra ir
Bati com os olhos no luar
E a lua foi bater no mar
E eu fui que fui ficando ...

NILSON CHAVES
Cantor, compositor e violeiro amazônico.
CELSO VIÁFORA
Compositor amazônico.

A meus pais ...

AGRADECIMENTOS

Agradeço a Deus pela concretização de mais uma etapa em minha vida e a realização de mais um sonho. Obrigado Senhor por iluminar os meus caminhos e me proporcionar sabedoria e discernimento.

Aos meus orientadores. Ao professor Dr. Haroldo Fraga de Campos Velho, por fazer-me acreditar que todos somos capazes. Obrigado por seu otimismo, entusiasmo e paciência. Ao professor Dr. Elbert E. N. Macau, pela atenção e confiança em mim depositada.

Aos meus pais, João Medeiros Furtado e Maria Lucilene Moraes da Costa, pela confiança, credibilidade e incentivo. Aos meus avós, especialmente, à minha vó Claudia, que a considero como mãe, pois, o que sou hoje é resultado de sua educação a mim direcionada.

Aos meus irmãos e amigos que de alguma forma contribuíram para a concretização deste trabalho. Em especial quero agradecer à Rosângela Cintra e à Rosângela Follman, que acompanharam de perto todas as dificuldades enfrentadas. À Rosângela Cintra pelas discussões sobre Assimilação de Dados, à Rosângela Follman pela amizade, incentivo, apoio e disponibilidade para estar presente em todos os momentos, à você muito obrigado.

Agradeço ao Rodolfo pelo carinho, companheirismo, amor, amizade e atenção durante o desenvolvimento deste trabalho.

Finalmente, agradeço ao Instituto Nacional de Pesquisas Espaciais pela oportunidade em aprimorar a minha qualificação profissional, além dos recursos físicos necessários para o desenvolvimento deste trabalho. Ao Dr. Paul Krause, por ter cedido o código do Filtro de Partículas. Ao **CNPq**, Conselho Nacional de Desenvolvimento Científico e Tecnológico, pelo apoio financeiro.

RESUMO

Técnicas de assimilação de dados são essenciais para sistemas operacionais de previsão baseados em modelos físico-matemáticos. De modo simplificado pode-se dizer que a assimilação de dados é a ciência de ter uma adequada combinação de dados de um modelo matemático de previsão com dados de observação do sistema em estudo. Este trabalho propôs-se a avaliar métodos de assimilação de dados, que são: Filtro de Kalman (KF), Filtro de Partículas (FP), ou Método Seqüencial de Monte Carlo, Método Variacional (MV) e Redes Neurais Artificiais, utilizando a arquitetura de rede Perceptron de Múltiplas Camadas (PMC). Primeiramente, fez-se uma revisão de algumas técnicas tradicionais em assimilação, tais como: Correções Sucessivas, Análise de Correções e Interpolação Ótima. O objetivo principal desta pesquisa foi avaliar a capacidade da Rede Neural PMC “*emular*” o KF, FP e o MV. Avaliou-se também, a frequência com que as observações são inseridas no processo de assimilação. Os resultados obtidos com a rede PMC foram precisos para ambas as técnicas utilizadas no treinamento da rede. Sendo que, a RN obteve o resultado mais preciso que o Filtro de Kalman, com capacidade de fazer assimilação com frequência de observações a cada 500 passos de tempo. As técnicas foram aplicadas ao sistema de Lorenz em regime caótico.

NEURAL NETWORKS AND DIFFERENT METHODS OF DATA ASSIMILATION IN DYNAMIC NONLINEAR

ABSTRACT

Techniques of data assimilation are essential for operating systems of forecast based on physical and mathematical models. In simplified it can be said that the data assimilation is the science to have a adequate combination of data from a mathematical model of forecast with observation data from the system under study. This study proposed to evaluate methods of data assimilation, which are: Kalman Filter (KF), Particle Filter (PF), or Sequential Monte Carlo method, Variational Method (VM) and Neural Networks, using the network architecture of Perceptron Multiple Layers (PML). Firstly, made by a revisão of some technical tradicionias in assimilação, such as: Successive Correction, Analysis Correction and Optimo Interpolation. The main objective of this research was to evaluate the ability of Neural Network (PML) to emulate the KF, FP and VM. It also evaluated the frequency with which the observations are included in the process of assimilation. The results obtained with the network PML were accurate for both techniques used in the training of the network. The network obtained the result more accurate that the Kalman Filter, with capacity to make assimilation with frequency of observation to every 500 steps of time. The techniques were applied to the Lorenz system in chaotic regime.

SUMÁRIO

Pág.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

LISTA DE SÍMBOLOS

1	INTRODUÇÃO	27
2	TÉCNICAS TRADICIONAIS EM ASSIMILAÇÃO DE DADOS	33
2.1	Correções Sucessivas	36
2.2	Interpolação Ótima	36
2.3	Análise de Correções	39
3	FILTRO DE KALMAN E FILTRO DE PARTÍCULAS	43
3.1	Modelos de Sistemas Dinâmicos	43
3.1.1	Introdução ao estudo de Espaços de Estados de Modelos	45
3.1.2	Espaço de Estado de Modelos	46
3.1.2.1	Espaços de Estados de Modelos Não-Lineares	47
3.1.2.2	Espaço de Estado de Modelos Lineares	49
3.2	Estimação de Estado Não-Linear	49
3.2.1	Solução Conceitual	50
3.2.2	Estimativa Pontual	51
3.2.3	Sistemas Não-Lineares	52
3.2.3.1	Aproximação Local	53
3.2.3.2	Aproximação Global	54
3.2.4	Filtragem e Predição	54
3.3	Método Seqüencial de Monte Carlo (Filtro de Partículas)	56
3.3.1	Amostragem Perfeita	56
3.3.2	Geração de Números Aleatórios	58
3.3.2.1	Importância da Amostragem e Reamostragem	58
3.3.2.2	Aceitação e Rejeição de amostras	60
3.3.3	Filtro de Partículas	61
3.3.3.1	Algoritmo de Reamostragem	63

3.3.3.2	Um exemplo para aplicação do Filtro de Partículas e Filtro de Kalman . . .	65
4	MÉTODO VARIACIONAL	69
4.1	Procedimento Variacional	69
4.2	Valor estacionário de uma Integral Definida	72
4.3	Análise Variacional	75
4.4	Modelo Tangente Linear e Modelo Adjunto	77
5	REDES NEURAIS	81
5.1	Neurônio Biológico	82
5.2	Neurônio Artificial	83
5.2.1	Funções de ativação	84
5.3	Funcionamento das Redes Neurais	85
5.4	Regras de Aprendizagem	87
5.5	Algoritmo Backpropagation ou de Retropropagação	87
5.6	Perceptron de Múltiplas Camadas	88
6	RESULTADOS NUMÉRICOS	91
6.1	Sistema de Lorenz	91
6.2	Resultados com a Rede Neural emulando o Filtro de Kalman	94
6.3	Resultados com a Rede Neural emulando o Filtro de Partículas	102
6.4	Resultados com a Rede Neural emulando o Método Variacional	105
7	CONCLUSÕES	111
	REFERÊNCIAS BIBLIOGRÁFICAS	113
A	CONCEITOS BÁSICOS EM PROCESSOS ESTOCÁSTICOS	119
A.1	Probabilidade Condicional	119
A.2	Média e Covariância	119
A.3	Variáveis Gaussianas	120
A.4	Processo de Markov	120
A.5	Ruído Branco	121

LISTA DE FIGURAS

	<u>Pág.</u>
2.1 Ilustração da sensibilidade das condições iniciais. Para condições iniciais ligeiramente diferentes, o modelo produz trajetórias completamente diferente. . . .	33
2.2 Série temporal da componente x do sistema de Lorenz. Curva azul: referência de verdade; curva preta: estimativa “ <i>a priori</i> ”; quadrados verdes: observações; curva vermelha: estimativa.	34
2.3 Solução para a componente x . A janela de assimilação de intervalo $[0; 5]$, observações inseridas a cada 12 passos de tempo. Do intervalo $]5; 10]$ tem-se a previsão que foi gerada com a análise em $t = 5$, análise obtida com o método de correções sucessivas.	37
2.4 Erro de assimilação com o método de correções sucessivas do intervalo $[0; 5]$ e o erro de previsão do intervalo $]5; 10]$	38
2.5 A janela de assimilação de intervalo $[0; 5]$ com o método OI, observações inseridas a cada 12 passos de tempo. Do intervalo $]5; 10]$ tem-se a previsão que foi gerada com a análise em $t = 5$	38
2.6 O erro de assimilação do intervalo $[0; 5]$, com o método de Interpolação Ótima, e o erro de previsão do intervalo $]5; 10]$	39
2.7 Nessa figura tem-se duas janelas de assimilação com o método OI; uma de intervalo $[0; 5]$ e outra $[6; 10]$. Nota-se que o desacoplamento da análise (gráfico vermelho) se dá no tempo $t = 5, 4$, mas a medida em que volta-se a inserir observações no modelo no tempo $t = 6$ a dinâmica da análise volta a se ajustar ao gráfico azul (verdade).	40
2.8 O erro de assimilação nos intervalos $[0; 5]$ e $[6; 10]$, método OI.	41
2.9 A janela de assimilação de intervalo $[0; 5]$, com o método de Análise de Correções, observações inseridas a cada 12 passos de tempo. Do intervalo $]5; 10]$ tem-se a previsão que foi gerada com a análise em $t = 5$	41
2.10 O erro de assimilação do intervalo $[0; 5]$ com o método de Análise de Correções, e o erro de previsão do intervalo $]5; 10]$	42
3.1 Ilustração da amostragem por importância. $t(x)$ é a PDF verdadeira, representada pela linha cheia e $s(x)$ é a distribuição proposta (ou de importância) representada pela linha pontilhada.	59
3.2 Ilustração do passo de reamostragem do filtro de partículas.	64
3.3 (a): partículas iniciais (números aleatórios); (b): distribuição de probabilidade das partículas iniciais.	66
3.4 Ilustração do passo de reamostragem do filtro de partículas.	67

3.5	curva azul: verdade; curva vermelha: estimado.	67
3.6	Gráficos dos erros em escala semilogaritmica. Logaritmica no eixo y e linear no eixo x	68
4.1	Ilustração de uma bola (círculo vermelho) em estado de equilíbrio. Os círculos pontilhados representam o deslocamento virtual.	70
4.2	Gráfico da função seno.	71
4.3	Ilustração de $u(x), \tilde{u}(x)$ e $\varepsilon\eta(x)$ usado na derivação das equações de Euler-Lagrange.	73
4.4	Esquema do cálculo do gradiente da função custo relativo à observação para o período de 12 horas, observações a cada 3 horas e o modelo adjunto é integrado para trás dentro de cada intervalo.	77
5.1	Representação do diagrama em blocos do sistema nervoso.	81
5.2	Representação de um neurônio biológico.	83
5.3	Representação de um neurônio artificial.	84
5.4	Exemplos de funções de ativação mais utilizadas: (a) tangente hiperbólica; (b) exponencial; (c) função degrau.	86
5.5	Ilustração da Rede Perceptron de Múltiplas Camadas	89
6.1	Atrator de Lorenz caótico com condições iniciais $(-1, 5; 1, 5; 20)$ e suas projeções nos eixos xy , xz , e yz	92
6.2	Figuras exemplificando a sensibilidade das condições iniciais do modelo de Lorenz; figura (a): série temporal da componente x , curva contínua $x(0) = 1$ e a curva tracejada $x(0) = 1,05$; figura (b): série temporal da componente z , curva contínua $z(0) = 1$ e a curva tracejada $z(0) = 1,05$	93
6.3	Diagrama esquemático ilustrando a fase de treinamento da RNA. São apresentados os padrões de treinamento: dados do modelo matemático, dados de observação e a saída desejada, ou o “alvo” da rede, que é a estimativa obtida pelo FK, FP e o Método Variacional.	95
6.4	Figura esquemática ilustrando o modelo da rede neural com uma camada oculta, contendo 2 neurônios. Para a rede <i>feed-forward</i> , a informação é propagada dos neurônios da camada de entrada para os neurônios na camada de saída.	96
6.5	Diagrama esquemático do Filtro de Kalman Linear.	97
6.6	Assimilação de dados com o Filtro de Kalman e com RNA Perceptron de Múltiplas Camadas. As observações foram inseridas a cada 50 passos de tempo. Gráficos à direita: curva tracejada (vermelha) estimado pelo FK, curva cheia (azul) referência de verdade. Analogamente, gráficos à esquerda, curva tracejada (verde) estimado pela RNA, curva cheia (azul) referência de verdade.	98

6.7	Assimilação de dados a cada 500 passos de tempo, estimativa obtida com o FK: curva tracejada (vermelha) estimado, curva cheia (azul) referência de verdade.	99
6.8	Assimilação de dados a cada 500 passos de tempo, estimativa obtida com o Perceptron de Múltiplas Camadas: curva tracejada (verde) estimado, curva cheia (azul) referência de verdade.	100
6.9	Assimilação de dados a cada 500 passos de tempo, estimativa obtida com o Perceptron de Múltiplas Camadas: curva tracejada (verde) estimado, curva cheia (azul) referência de verdade.	100
6.10	Assimilação de dados a cada 500 passos de tempo, estimativa obtida com o Perceptron de Múltiplas Camadas: curva tracejada (verde) estimado, curva cheia (azul) referência de verdade.	101
6.11	Média global do erro $(x + y + z)/3$ de assimilação para as técnicas do FP e RNA Perceptron de Múltiplas camadas (PMC), eixo x é a frequência das observações e o eixo y é o erro.	101
6.12	Média global do erro $(x + y + z)/3$ de assimilação; eixo x é a frequência das observações e o eixo y é o erro.	103
6.13	Assimilação de dados com o Filtro de Partículas e com RNA Perceptron de Múltiplas Camadas. As observações foram inseridas a cada 500 passos de tempo. Gráficos à direita: curva tracejada (vermelha) estimado pelo FP, curva cheia (azul) referência de verdade. Analogamente, gráficos à esquerda, curva tracejada (magenta) estimado pela RNA, curva cheia (azul) referência de verdade.	104
6.14	Série temporal da componente x . Assimilação de dados a cada 12 passos de tempo, estimativa obtida com o método variacional: curva vermelha estimado, curva azul referência de verdade.	105
6.15	Série temporal da componente z . Assimilação de dados a cada 12 passos de tempo, estimativa obtida com o método variacional; curva vermelha: estimado; curva azul: referência de verdade.	106
6.16	(a) Gráfico do erro das componentes x e y do modelo de Lorenz, para assimilação realizada a cada 12 passos de tempo; (b) Gráficos de convergência da função custo e da norma do gradiente.	106
6.17	Série temporal da componente x . Assimilação de dados a cada 500 passos de tempo, estimativa obtida com o método variacional; curva vermelha: estimado; curva azul: referência de verdade.	107
6.18	(a) Gráfico do erro das componentes x e y do modelo de Lorenz, para assimilação realizada a cada 500 passos de tempo; (b) Gráficos de convergência da função custo e da norma do gradiente.	107

6.19	Série temporal da componente x . Assimilação de dados a cada 12 passos de tempo. fig. à direita: estimativa do método variacional; fig. à esquerda: estimativa da RN.	108
6.20	Gráfico do erro da componentes x do modelo de Lorenz, para assimilação realizada a cada 12 passos de tempo; fig. à direita: variacional; fig. à esquerda: RN.	108
6.21	Erros em escala logaritmica; fig. à direita: variacional; fig. à esquerda: RN. . .	109
6.22	Gráfico dos erros, com observações a cada 12 passos de tempo.	109
A.1	Função densidade de probabilidade gaussiana de média zero; ilustra-se distribuições alterando o valor da variância; $\sigma^2 = 0.2$ (azul), $\sigma^2 = 1$ (verde), $\sigma^2 = 5$ (vermelho), $\sigma^2 = 0.5$ (azul esverdeado).	120

LISTA DE TABELAS

	<u>Pág.</u>
3.1 Algoritmo: Filtro de Kalman Estendido	54
3.2 Aceitação e Rejeição de Amostras	60
3.3 Algoritmo do Filtro de Partículas	63
3.4 Amostragem Sistemática	64
6.1 Parâmetros da RNA	96
6.2 Erros de Assimilação: Filtro de Kalman e RNA Perceptron.	101
6.3 Erros de Assimilação: Filtro de Partículas e RNA Perceptron.	102

LISTA DE ABREVIATURAS E SIGLAS

CA	–	Análise de Correções (<i>Analysis Correction</i>)
CPTEC	–	Centro de Previsão de Tempo e Estudos Climáticos
EKF	–	Filtro de Kalman Extendido (<i>Extend Kalman Filter</i>)
KF	–	Filtro de Kalman (<i>Kalman filter</i>)
OI	–	Interpolação Ótima (<i>Optimal Interpolation</i>)
PDF	–	Função densidade de Probabilidade (probability density function)
PMC	–	Perceptron de Múltiplas Camadas
PNT	–	Previsão Numérica de Tempo
PSAS	–	Physical-Space Statistical Analysis System
SC	–	Correções Sucessivas (<i>Successive Corrections</i>)
4D-Var	–	Four-dimensional variational data assimilation
RNA	–	Redes Neurais Artificiais
NASA	–	National Aeronautics and Space Administration
SIS	–	Amostragem e Reamostragem por Importância <i>sampling importance resample</i>
MV	–	Método Variacional
FP	–	Filtro de Partícula

LISTA DE SÍMBOLOS

a	– constante responsável pela inclinação da tangente hiperbólica
b_k	– Viés
\mathbf{B}	– Matriz de covariância do “ <i>background</i> ”
d	– Operador derivada; vetor de saída desejada (alvo) das RNA
\mathbf{d}	– inovação ou incremento observacional
$E\{.\}$	– Valor esperado
ET	– Erro de treinamento
ETA	– Modelo Regional de Previsão Numérica do Tempo
EV	– Erro de validação
\mathbf{F}	– Matriz que representa o sistema dinâmico
\mathbf{H}	– Matriz que representa o sistema de observação
\mathbf{I}	– Matriz Identidade
$J(.)$	– Função Custo
\mathbf{K}	– Matriz de ganho de Kalman
l	– Índice referente a camada oculta das RNA
L	– Índice referente a camada de saída das RNA
$L(t_0, t)$	– Propagador do modelo tangente linear
\mathbf{M}	– Matriz do modelo tangente linear
o	– Vetor obtido na saída da RNA
p	– probabilidade; função de distribuição
\mathbf{P}^a	– Matriz de covariância do erro de análise
\mathbf{Q}	– Matriz de covariância do erro de modelagem
\mathbf{R}	– Matriz de covariância do erro de observação
R_c	– Número de Rayleigh crítico
\mathbf{u}	– Velocidade de vento zonal
\mathbf{W}	– Matriz de peso
x	– Intensidade do movimento convectivo
\mathbf{x}^a	– Vetor de estado de análise
y	– Diferença de temperatura entre as correntes ascendentes e descendentes
\mathbf{y}	– Vetor de observação
y_k	– Sinal de saída do neurônio
z	– Distorção do perfil vertical de temperatura em relação a um perfil linear
σ	– desvio padrão; Número de Prandtl
σ^2	– variância
\mathbf{v}	– Velocidade de vento meridional
α	– cosseno diretor
δ	– Função delta de Dirac; operador variacional; gradiente local
∂	– operador diferencial parcial
λ	– Multiplicador de Lagrange
θ	– Vetor de pesos
$\varphi(.)$	– Função de ativação
$(\hat{.})$	– Refere-se a estimativa

1 INTRODUÇÃO

Ultimamente, métodos de assimilação de dados têm sido desenvolvidos e usados em muitas áreas de pesquisa, como por exemplo, em clima espacial ([HARTER et al., 2008](#)). Porém, a primeira aplicação desses métodos foi em meteorologia, tornando-se uma componente fundamental para a previsão numérica do tempo e clima ([KALNAY, 2003](#)).

As previsões do tempo são baseadas em modelos físico-matemáticos da atmosfera, que são resolvidos numericamente com o auxílio do computador. A primeira experiência real para se prever a dinâmica do tempo foi desenvolvida pelo meteorologista e matemático inglês L. F. Richardson, resolvendo as equações que governam os movimentos atmosféricos ([RICHARDSON, 1992](#)). Apesar do insucesso do trabalho de Richardson, sua tentativa foi de grande valia para o desenvolvimento desta área da ciência.

Segundo [Daley \(1993\)](#), 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 ([HARTER, 1999](#)).

A melhoria da Previsão Numérica do Tempo (PNT) tem conseqüências positivas em vários setores da sociedade, por exemplo, agricultura, indústria da pesca, do turismo, da moda e em atividades de entretenimento e lazer. Esses fatores tornam relevante as investigações dessa área da ciência.

A inovação tecnológica, que permitiu o surgimento de supercomputadores, a evolução dos modelos numéricos de previsão e da rede de observação, proporcionaram maior confiabilidade na previsão numérica do tempo. Porém, fenômenos não-lineares e caóticos representados por modelos matemáticos têm uma relação intrínseca com a condição inicial (CI), o que significa que, de acordo com a CI utilizada para executar os modelos de previsão, depois de alguns passos de tempo ocorrerá o desagregamento das dinâmicas, devido a pequenas incertezas das condições iniciais. Em outras palavras, pequenas incertezas no estado inicial do sistema conduzem rapidamente a incapacidade de previsão do estado futuro ([GREBOGI et al., 1987](#)). Portanto, a melhor representação da condição inicial produzirá a melhor previsão.

Assim, o problema consiste em: como determinar a melhor condição inicial? Surge então

uma nova ciência, a *Assimilação de Dados*, que utiliza dados de observações para melhor determinar a CI (KALNAY, 2003). Assimilação de dados pode ser definida como a ciência de se ter uma adequada *combinação* de dados de um modelo matemático de previsão com dados de observação para se determinar o dado de análise (ou CI) (CAMPOS VELHO et al., 2007a; HARTER et al., 2008).

É importante ressaltar que, mesmo considerando por hipótese que os modelos determinísticos e as observações fossem perfeitos, a natureza caótica da atmosfera limitaria a previsibilidade para alguns dias, em torno de duas semanas (KALNAY, 2003). A natureza caótica relacionada a meteorologia foi primeiramente observado nos trabalhos de Lorenz (1963), Lorenz (1965). Lorenz observou que a solução de sistemas de equações semelhantes às que governam os movimentos atmosféricos apresentam dependência sensível com relação as condições iniciais fornecidas no início da integração. Desde então, a previsibilidade numérica da atmosfera de forma determinística, baseada em modelos dinâmicos, tem sido amplamente discutida (MENDONÇA; BONATTI, 2002).

A evolução histórica dos métodos de assimilação de dados meteorológicos passa pelo Ajuste de Funções, Correções Sucessivas, Análise de Correções, Interpolação Ótima, Métodos Variacionais, Filtro de Kalman e Métodos de Monte Carlo, (DALEY, 1993; KALNAY, 2003). Todos esses métodos são diferentes formas de combinar uma previsão de “*background*” (estimativa *a priori* do modelo de previsão) com observações, resultando no dado de análise (CI).

Mesmo com os avanços dos métodos de assimilação nas últimas décadas, existe, porém, 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 têm grau de liberdade da ordem de 10^8 a 10^9 . Por exemplo, um modelo com resolução de 1 grau (latitude e longitude) e 20 níveis verticais, tem $360 \times 180 \times 20 = 1,3 \times 10^{11}$ 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 em um total de 5 milhões de variáveis a serem iniciadas. 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 de que o grau de liberdade (KALNAY, 2003).

No começo da PNT, o custo computacional da análise era negligenciado em relação ao custo da 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 (HARTER, 2004). 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 ao alto custo computacional das técnicas de assimilação de dados, há uma intensa pesquisa voltada à satisfazer o binômio: melhorar qualitativamente a análise proveniente da assimilação, com redução do esforço computacional requerido.

No Brasil, mais precisamente no CPTEC, em colaboração com o *Data Assimilation Office-Nacional Aeronautic and Space Administration* (DAO-NASA), foi implementado um sistema de assimilação de dados global e regional para a América do Sul, o GPSAS e RPSAS, respectivamente. O sistema, denominado de *Regional Physical-space Statistical Analysis System* (RPSAS), associado ao modelo regional *ETA*, consiste em uma variação do método de Interpolação Ótima, em que um filtro é aplicado no espaço físico (pontos de observação) ao invés de trabalhar no espaço do modelo. Para uma descrição mais detalhada a respeito o sistema de assimilação de dados do CPTEC consultar Cintra (2004).

A técnica de Redes Neurais Artificiais (RNA) tem registrado um número crescente de aplicações na área de meteorologia. Em Gardner e Dorling (1998) há uma revisão destas aplicações. Hsieh e Tang (1998) sugerem a técnica de RNA como uma possível metodologia para assimilação. No entanto, não é no mesmo sentido da definição de assimilação de dados da literatura. A RNA de Hsieh e Tang (1998) é usada para substituir uma equação desconhecida.

Os autores Hsieh e Tang (1998) chamam a atenção para os seguintes aspectos próprios desse método: instabilidade não-linear, grande dimensão da aplicação e dificuldade em interpretar os resultados das RNA. Estas são consideradas as dificuldades em se aplicar RNA em meteorologia e oceanografia. No entanto, no mesmo trabalho são feitas sugestões de como contornar esses problemas. As sugestões são: controlar a instabilidade não-linear com médias de previsões por conjunto (“ensemble”), aplicar a análise de componentes principais para reduzir a dimensão do problema e interpretar o sinal da saída dos neurônios das camadas não-lineares por meio de análise espectral. Em Tang e Hsieh (2001), é aplicado um modelo híbrido dinâmico-neural, assim chamado pelos autores por utilizarem a rede Perceptron de Múltiplas Camadas (PMC) para substituir uma das equações do sistema de Lorenz, no contexto de assimilação de dados, com o Método Variacional em quatro 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, nos casos de não-linearidade fraca, a metodologia apresenta resultados satisfatórios. Porém, para o 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.

O Laboratório Associado de Computação e Matemática Aplicada (LAC) do INPE foi pioneiro na aplicação de RNA para assimilação de dados (NOWOSAD et al., 2000), em uma abordagem independente e sem conhecimento dos trabalhos de outros pesquisadores. De fato, não haviam implementações do método na literatura. Nowosad et al. (2000) apresentou o uso de um Filtro de Kalman Estendido Adaptativo e Redes Neurais Perceptrons de Múltiplas Camadas, essas técnicas foram aplicadas para três modelos de sistemas dinâmicos não-lineares: sistema de tempo discreto caótico de Hénon, sistema de Lorenz em regime caótico e modelo computacional DYNAMO¹. A RNA utilizada para fazer assimilação de dados foi treinada para “emular” o filtro de Kalman, com o objetivo de reduzir o custo computacional. Nowosad et al. (2000) concluiu que para sistemas de alta dimensão as redes treinadas podem ser computacionalmente mais rápidas que os filtro de Kalman. O algoritmo foi paralelizado em CAMPOS VELHO et al. (2002), havendo uma redução significativa do tempo de processamento durante o treinamento do Perceptron de Multiplas Camadas (PMC).

Mais tarde, o trabalho de Harter (2004) apresentou o desempenho das RNA Funções de Base Radial (FBR), Elman (RN-E), Jordan (RN-J) e PMC, avaliando a eficiência destas em “emular” o Filtro de Kalman. As aplicações foram feitas nos sistemas não-lineares caóticos de Lorenz e no modelo DYNAMO, sendo que o PMC também foi aplicado no modelo de iteração não-linear entre as ondas de Langmuir, Whistler e Alfvén (Modelo de 3 ondas)². Este foi um dos resultados inéditos do trabalho de Harter (2004), a aplicação de assimilação de dados com RNA em clima espacial. Ressalta-se que as RNA desenvolvidas por Harter (2004) foram implementadas uma técnica conhecida como validação cruzada (*cross-validation*), o que permitiu que se conhecesse totalmente a superfície de erros de treinamento e validação, proporcionando a obtenção do melhor conjunto de pesos para o problema investigado. O ganho com a diminuição do espaço de busca, foi um resultado importante da pesquisa de Harter (2004), pois em problemas de minimização de gradiente, em aplicações de grande dimensão, diminuir o espaço de busca pode ser o fator que pode tornar a aplicação operacionalmente viável.

Dando continuidade a essas investigações, esta pesquisa de mestrado tem como objetivo avaliar o desempenho da rede neural PMC treinada para emular: O Filtro de Kalman, Filtro de Partículas e o Método Variacional, ambas as técnicas aplicadas ao sistema de Lorenz em regime caótico.

¹Simulador meteorológico baseado na equação de Água Rasa 1D (LYNCH, 1989).

²Aplicação em assimilação de Dados em Clima espacial, veja (HARTER, 2004) pag. 93

Assim, o texto desta dissertação está estruturado da seguinte forma: no Capítulo 2 revisa-se algumas técnicas empíricas para assimilação de dados, tais como: Correções Sucessivas, Interpolação Ótima e Análise de Correções. No Capítulo 3 faz-se uma breve introdução sobre a teoria de estimação, base para o desenvolvimento do Filtro de Kalman e Filtro de Partículas, em seguida descreve-se essas técnicas. No Capítulo 4 apresenta-se o Método Variacional, antes faz-se uma breve revisão do Cálculo Variacional. No Capítulo 5 fala-se sobre Redes Neurais Artificiais, apresentando o modelo neuronal e descrevendo a arquitetura de rede utilizada. O Capítulo 6 é reservado para os resultados. Por fim, no Capítulo 7 apresenta-se as conclusões e sugestões de trabalhos.

2 TÉCNICAS TRADICIONAIS EM ASSIMILAÇÃO DE DADOS

A qualidade da Previsão do Tempo é resultado de uma boa formulação dos modelos físico-matemáticos e dos dados observados rotineiramente. Os desafios relacionados a qualidade da previsão, consiste na não linearidade do sistema, a sensibilidade à variação de suas condições iniciais, formulação dos processos físicos e métodos numéricos (DALEY, 1993; KALNAY, 2003). Na Figura 6.2, tem-se a ilustração dessa sensibilidade.

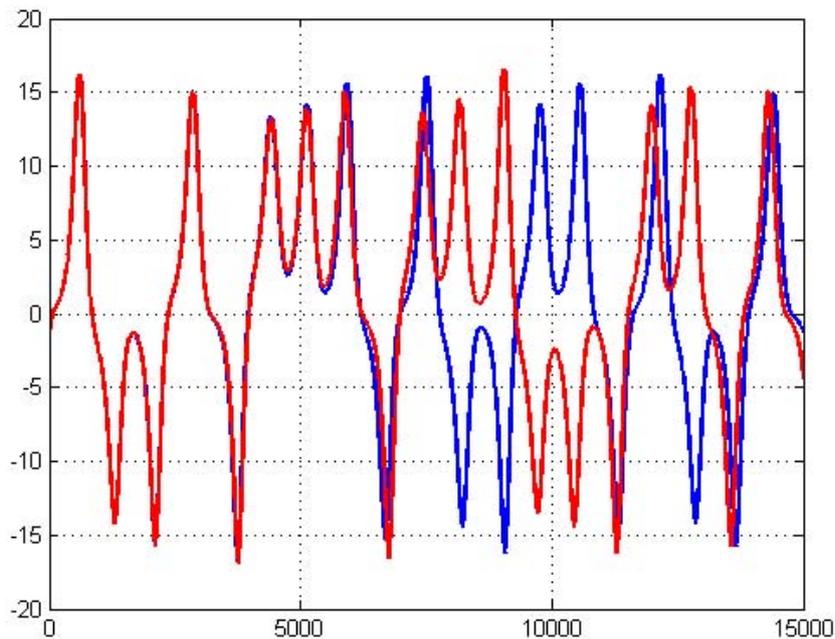


Figura 2.1 - Ilustração da sensibilidade das condições iniciais. Para condições iniciais ligeiramente diferentes, o modelo produz trajetórias completamente diferente.

A previsão numérica do tempo (PNT) é um problema de valor inicial: “*Dado uma estimativa do estado atual da atmosfera (condições iniciais), superfície apropriada e condições de contorno laterais, o modelo simula (prever) a sua evolução*”, (KALNAY, 2003). Portanto, quanto mais exato for a estimativa da condição inicial, melhor será a qualidade da previsão. Para isso, necessita-se da utilização de ferramentas de assimilação de dados para estimar boas condições iniciais para inicializar os modelos numéricos de previsão. Essa aproximação consiste na combinação de duas fontes de dados, que são: dados de observações e dados do modelo físico-matemático de previsão. Assim, a determinação da condição inicial é feita a partir de uma combinação estatística dessas duas fontes de informações. Essa aproximação é conhecida como *assimilação de dados* (KALNAY, 2003). A

análise é o melhor ajuste tanto às observações quanto ao conhecimento prévio do estado da atmosfera, mantido o vínculo físico e dinâmico do sistema (NOWOSAD, 2001).

A Figura 2.2 ilustra a aplicação da técnica de assimilação de dados. Nota-se que, à medida que são inseridas observações no sistema, a dinâmica da estimativa (curva vermelha) está próxima da referência de verdade (curva azul), uma vez interrompendo a inserção das observações, ocorre o desacoplamento das dinâmicas. Retomando-se a inserção de observações a dinâmica da estimativa ajusta-se novamente a dinâmica da referência da verdade. Essa aplicação foi feita no sistema de Lorenz em regime caótico, que será descrito no Capítulo 6.

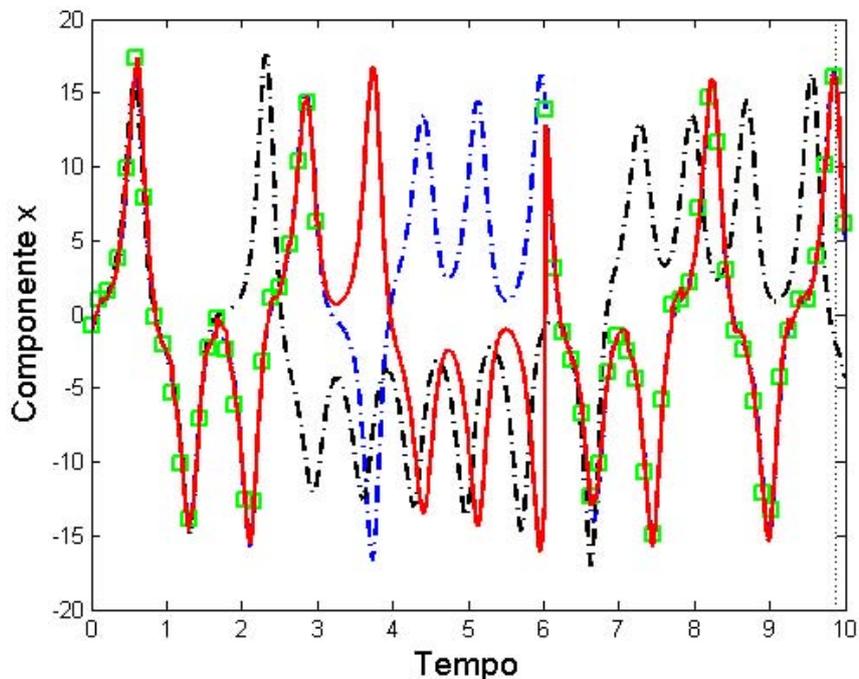


Figura 2.2 - Série temporal da componente x do sistema de Lorenz. Curva azul: referência de verdade; curva preta: estimativa "a priori"; quadrados verdes: observações; curva vermelha: estimativa.

Talagrand (1997) definiu que a assimilação de observações meteorológicas ou oceanográficas podem ser descritas como um processo por meio do qual todas as informações disponíveis são utilizadas para estimar com exatidão o estado do fluido atmosférico ou oceânico. As informações disponíveis consistem basicamente de observações adequadas e de leis físicas que governam a evolução do fluido. As observações são as informações meteorológicas coletadas de diversos instrumentos, como, por exemplo, satélites e radares, sendo que uma grande vantagem da utilização de satélites para coletas de dados é que

estes cobrem áreas de baixa densidade de informações meteorológicas como, por exemplo, oceanos e florestas.

A base dos algoritmos de assimilação de dados pode ser expressa como uma correção ou seqüência de correções dos estados do modelo. Este processo consiste de duas fases, que são: a fase de previsão e a fase de análise ou incremento, representados matematicamente por:

i. Fase de previsão:

$$\mathbf{x}_n^p = \mathbf{F}(\mathbf{x}_{n-1}^a) \quad (2.1)$$

ii. Fase de análise:

$$\mathbf{x}_n^a = \mathbf{x}_n^p + \mathbf{W} \underbrace{(\mathbf{y}_n^o - \mathbf{H}(\mathbf{x}_n^p))}_{\mathbf{d}} \quad (2.2)$$

em que o vetor de estados previsto \mathbf{x}_n^p é definido como $\mathbf{x}_n^p = [x_n \ y_n \ z_n]^T$. Os sobrescritos p e a , representam os passos de previsão e análise, respectivamente; \mathbf{F} representa o modelo de previsão, que é o sistema de Lorenz. Na equação 2.2 o termo \mathbf{d} expressa o incremento do estado do modelo \mathbf{x}_n^p com novas informações de um conjunto de observações \mathbf{y}_n^o ; \mathbf{W} é a matriz de ponderação, \mathbf{y}_n^o o vetor de observações e \mathbf{H} é o operador linear que representa o sistema de observação.

Os métodos de assimilação de dados: Correções Sucessivas, Análise de correções, Interpolação Ótima e o Método Variacional são baseados na equação da análise 2.2 e diferenciam-se quanto a sua origem, seu procedimento de resolução e pela forma de combinar a estimativa inicial (“*background*”) com as observações para produzir a análise (KALNAY, 2003). Lorenc (1986) utilizou argumentos bayesianos para derivar a equação da análise e mostrou que essas técnicas estão relacionadas.

Como dito anteriormente as técnicas aqui apresentadas são aplicadas ao sistema de Lorenz, que será descrito no Capítulo 6. Este tem sido utilizado na literatura por muitos pesquisadores para testar técnicas de assimilação de dados, por exemplo, por Miller et al. (1994), devido apresentar sensibilidade a variação de suas condições iniciais, característica encontrada com freqüência na atmosfera.

A seguir são apresentados algumas técnicas tradicionais em assimilação de dados. No processo de assimilação as observações foram inseridas a cada 12 passos de tempo para todas as técnicas abordadas neste capítulo.

2.1 Correções Sucessivas

O método de Correções Sucessivas (SC) foi o primeiro método de análise usado em assimilação de dados em quatro dimensões (BERGTHORSON; DÖÖ, 1955). A técnica é uma aproximação empírica, sendo que seu algoritmo é iterativo e pode ser escrito como:

$$\mathbf{x}^{n+1}(i) = \mathbf{x}^n(i) + \mathbf{W}[\mathbf{y}(i) - \mathbf{H}(\mathbf{x}^n(i))] \quad (2.3)$$

em que $\mathbf{x}^0(i) = \mathbf{x}^b(i)$, com $\mathbf{x}^b(i)$ é o estado “background”, ou seja, na primeira iteração a previsão é igual ao estado “a priori”. Depois da primeira estimativa as iterações seguintes são obtidas por “sucessivas correções” conforme a equação 2.3, n representa o índice de correção, \mathbf{W} é a matriz de ponderação e \mathbf{y} é o vetor de observações. O algoritmo é parado depois de n correções, o dado de análise é dado por $\mathbf{x}^a(i) = \mathbf{x}_n(i)$. A matriz de ponderação é dada por $\mathbf{W} = 0.5\mathbf{I}$ (LAWLESS, 2004).

Na Figura 2.3 apresenta-se o resultado da assimilação de dados para a componente x do sistema de Lorenz obtido com o método empírico de Correções Sucessivas e o gráfico do erro de assimilação e previsão para esta variável é mostrado na Figura 2.4. Tem-se que, o gráfico azul é a referência de verdade, que consideramos a integração do sistema de Lorenz, admitindo-se que o modelo é perfeito; o gráfico preto é uma estimativa inicial da previsão, a qual denominamos de “background”; os quadrados verdes são as observações sintéticas, que são geradas a partir da integração do sistema de Lorenz adicionando-se um ruído gaussiano com variância 0.5; e por fim temos a análise, que é obtida da combinação entre o “background” e as observações.

O método de assimilação de dados apresentado nessa seção é considerado como uma técnica empírica, a seguir apresentam-se técnicas baseadas na teoria de estimação estatística.

2.2 Interpolação Ótima

No método de Interpolação Ótima (OI) a correção devido à disponibilidade de observações \mathbf{y} é feita estatisticamente. O cálculo da análise para essa técnica é dado por:

$$\mathbf{x}^a(i) = \mathbf{x}^b(i) + \underbrace{\mathbf{W}[\mathbf{y}^o(i) - \mathbf{H}(\mathbf{x}^b(i))]}_d \quad (2.4)$$

em que:

$$\mathbf{W} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad (2.5)$$

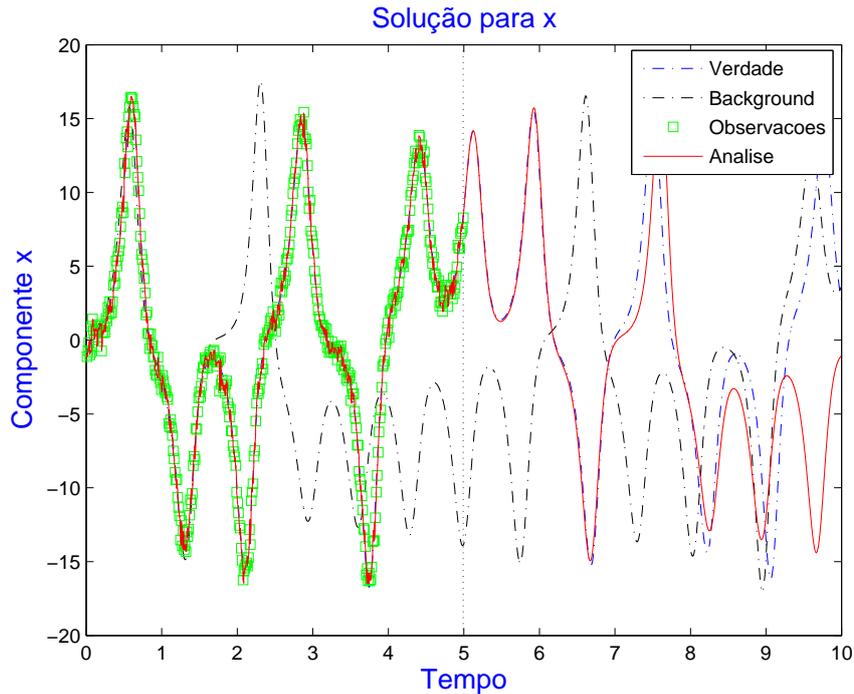


Figura 2.3 - Solução para a componente x . A janela de assimilação de intervalo $[0; 5]$, observações inseridas a cada 12 passos de tempo. Do intervalo $]5; 10]$ tem-se a previsão que foi gerada com a análise em $t = 5$, análise obtida com o método de correções sucessivas.

A equação 2.4 mostra que a análise é obtida por uma soma à estimativa inicial do produto da matriz de peso ótimo (ou matriz de ganho) dado pela equação 2.5 e a inovação (\mathbf{d}). A matriz de peso \mathbf{W} é denominada também de matriz de ganho \mathbf{K} , a qual aparece no algoritmo do Filtro de Kalman. O peso ótimo é calculado a partir das matrizes de covariâncias do erro de observação \mathbf{R} e da matriz de covariância do erro “background” \mathbf{B} . Os erros das observações consistem em “erros de representatividade” e erros de medida do próprio instrumento de coleta de dados atmosféricos, por exemplo, satélites e radiossondas. Se todas as hipóteses estatísticas são exatas, ou seja, se as matrizes de covariâncias \mathbf{B} e \mathbf{R} são conhecidas exatamente, a equação 2.4 fornece a análise ótima. No entanto, na prática as estatísticas são apenas aproximações e portanto diz-se que a equação 2.4 fornece uma “interpolação estatística”, não necessariamente uma “interpolação ótima”. As Figuras a seguir mostram os resultados obtidos com o método OI.

Nas Figuras 2.5 e 2.7 apresenta-se o resultado em assimilação de dados obtido com o método OI. As observações foram inseridas a cada 12 passos de tempo, a variância do erro de observação é 0,5. Na Figura 2.5 o tempo máximo de assimilação é $t = 5$ a partir daí tem-se a previsão, a Figura 2.6 mostra o erro de assimilação e previsão. Já na Figura 2.7 temos duas janelas de assimilação, uma do tempo $t = 0$ até $t = 5$ e outra de $t = 6$

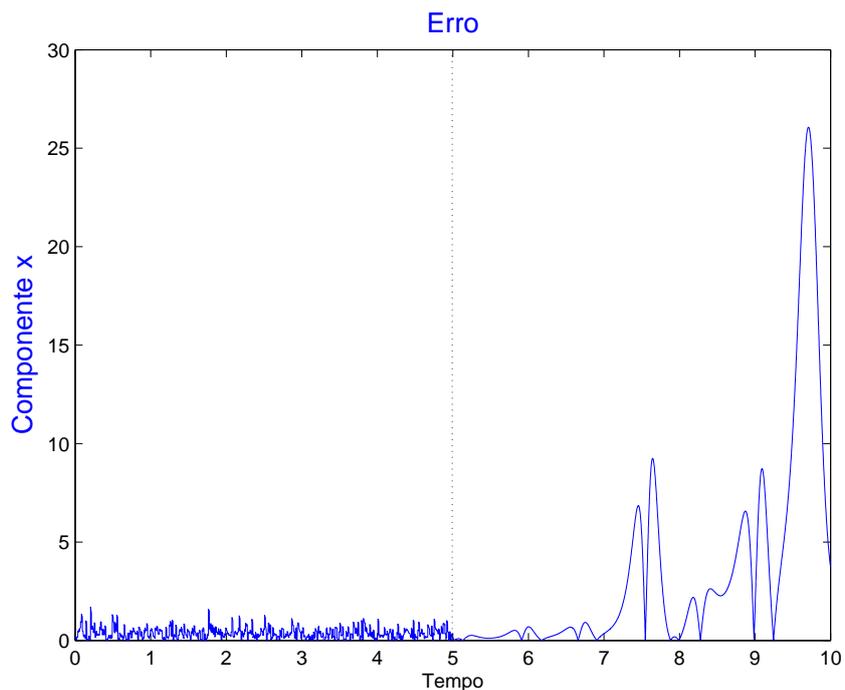


Figura 2.4 - Erro de assimilação com o método de correções sucessivas do intervalo $[0; 5]$ e o erro de previsão do intervalo $]5; 10]$.

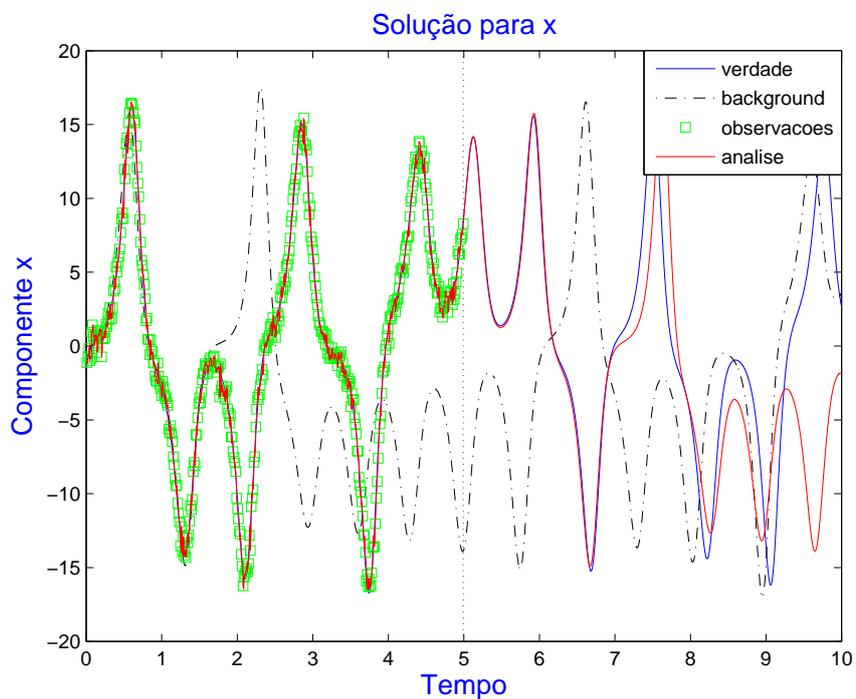


Figura 2.5 - A janela de assimilação de intervalo $[0; 5]$ com o método OI, observações inseridas a cada 12 passos de tempo. Do intervalo $]5; 10]$ tem-se a previsão que foi gerada com a análise em $t = 5$

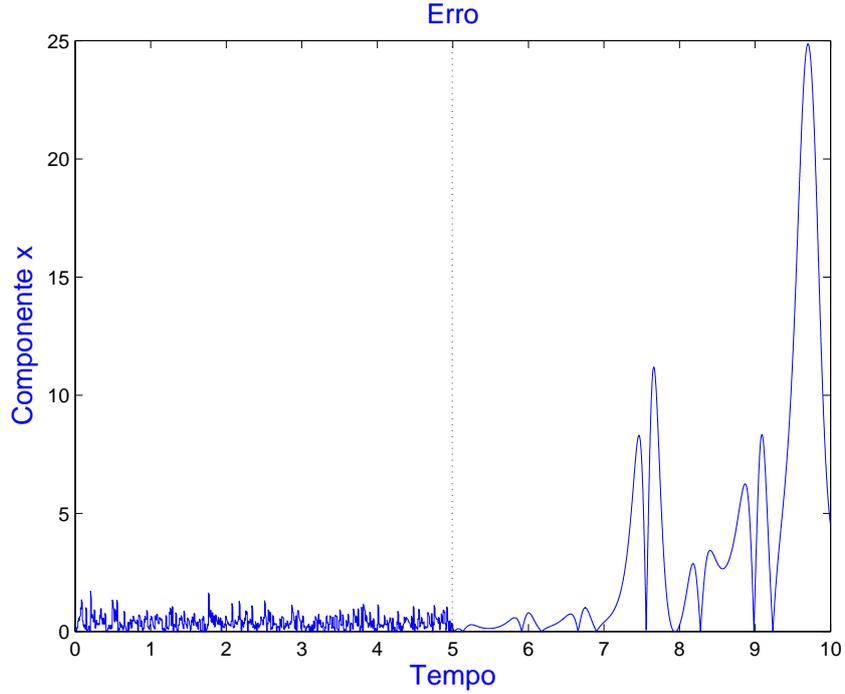


Figura 2.6 - O erro de assimilação do intervalo $[0; 5]$, com o método de Interpolação Ótima, e o erro de previsão do intervalo $]5; 10]$.

ao tempo $t = 10$, nota-se que a medida em que se retoma a inserção de observações no modelo, a análise (gráfico vermelho), que não mais acompanhava a referência de verdade (gráfico azul), aproxima-se novamente, diminuindo o erro (veja o gráfico 2.8).

2.3 Análise de Correções

O algoritmo da análise de correções, o qual está descrito em (LORENC et al., 1991), é uma versão modificada do método de correções sucessivas descrito na seção 2.1.

$$\mathbf{x}^{j+1}(i) = \mathbf{x}^j(i) + \mathbf{KQ}[\mathbf{y}^j(i) - \mathbf{H}(\mathbf{x}^j(i))] \quad (2.6)$$

$$\mathbf{y}^{j+1}(i) = \mathbf{y}^j(i) - \mathbf{Q}[\mathbf{y}^j(i) - \mathbf{H}(\mathbf{x}^j(i))] \quad (2.7)$$

em que:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T\mathbf{R}^{-1} \quad (2.8)$$

$$\mathbf{Q} = (\mathbf{H}\mathbf{K} + \mathbf{I})^{-1} \quad (2.9)$$

O algoritmo é parado depois de k correções no qual a análise é dada por:

$$\mathbf{x}^a(i) = \mathbf{x}^k(i) \quad (2.10)$$

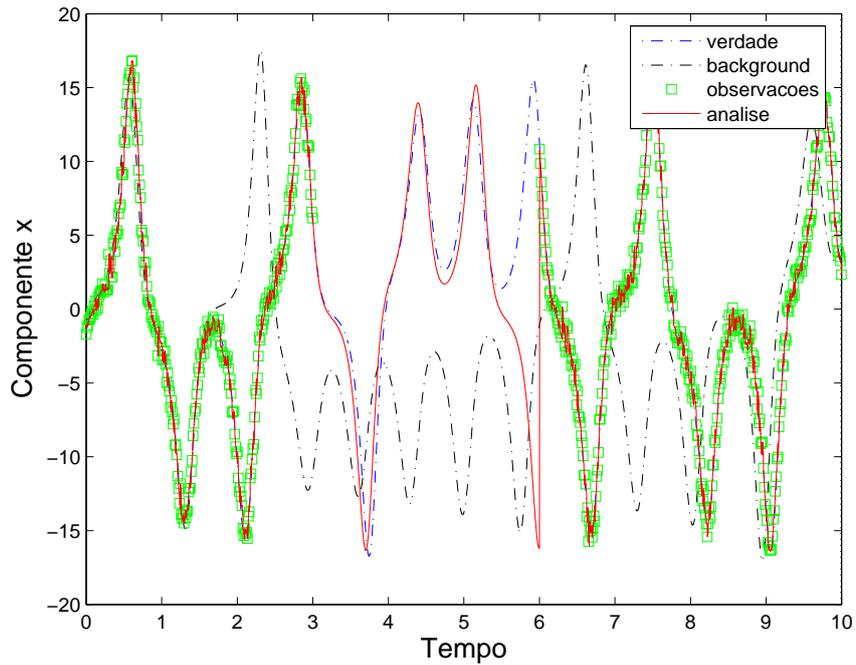


Figura 2.7 - Nessa figura tem-se duas janelas de assimilação com o método OI; uma de intervalo $[0; 5]$ e outra $[6; 10]$. Nota-se que o desacoplamento da análise (gráfico vermelho) se dá no tempo $t = 5, 4$, mas a medida em que volta-se a inserir observações no modelo no tempo $t = 6$ a dinâmica da análise volta a se ajustar ao gráfico azul (verdade).

As figuras 2.9 e 2.10 mostram o resultado obtido com esse método.

As técnicas de assimilação de dados revisadas neste capítulo foram uma das primeiras metodologias usadas para fazer previsão numérica do tempo. A diferença entre elas, consiste basicamente, no modo como é determinada a matriz de ponderação \mathbf{W} . Novas metodologias têm sido investigadas, como: Filtro de Kalman, Filtro de Partículas, Método Variacional e Redes Neurais Artificiais, que são tratados nos capítulos a seguir.

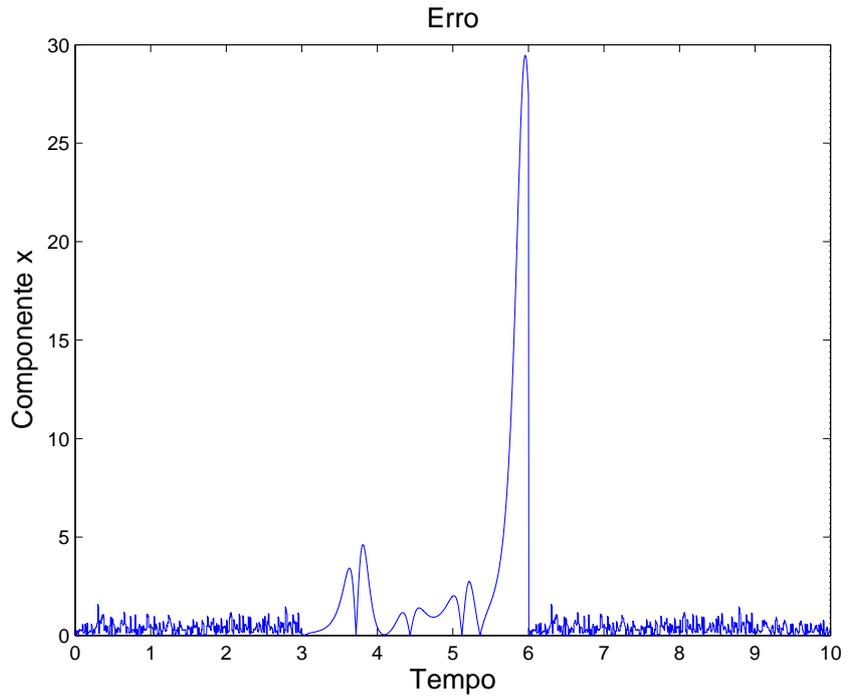


Figura 2.8 - O erro de assimilação nos intervalos $[0; 5]$ e $]6; 10]$, método OI.

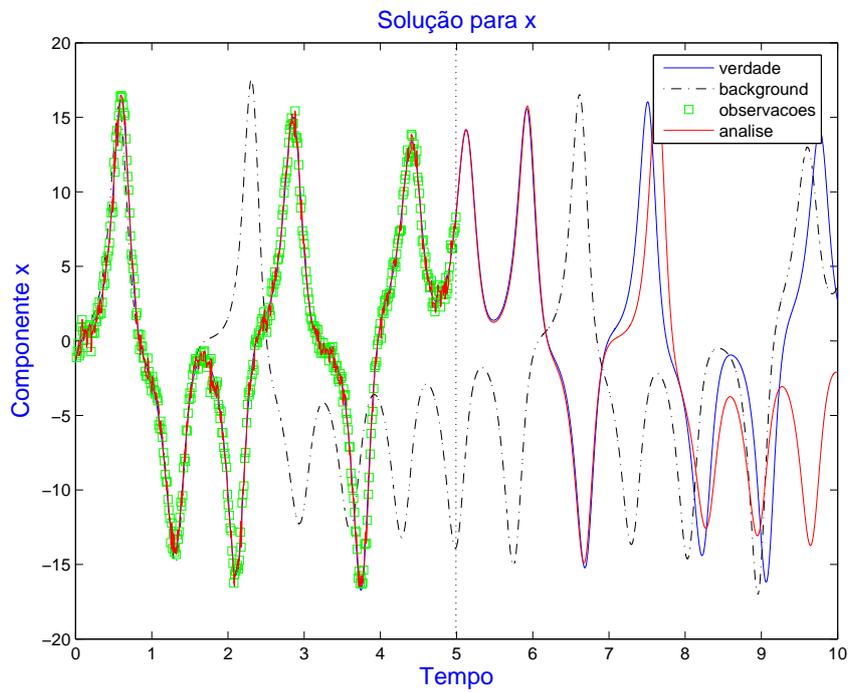


Figura 2.9 - A janela de assimilação de intervalo $[0; 5]$, com o método de Análise de Correções, observações inseridas a cada 12 passos de tempo. Do intervalo $]5; 10]$ tem-se a previsão que foi gerada com a análise em $t = 5$.

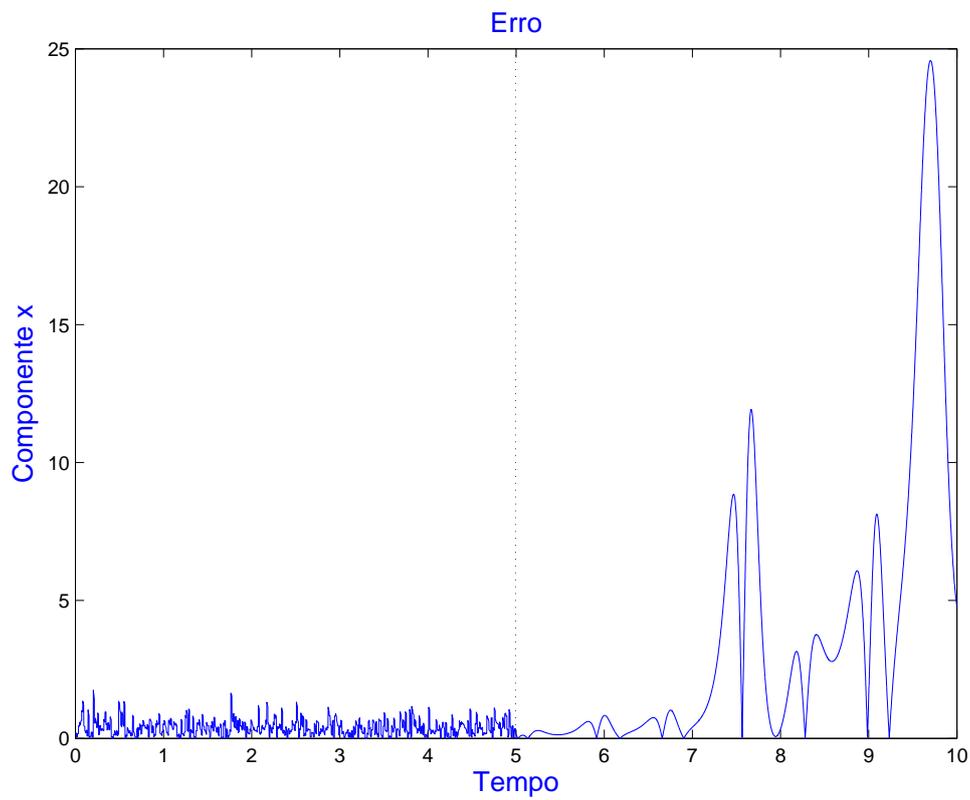


Figura 2.10 - O erro de assimilação do intervalo $[0; 5]$ com o método de Análise de Correções, e o erro de previsão do intervalo $]5; 10]$.

3 FILTRO DE KALMAN E FILTRO DE PARTÍCULAS

O Filtro de Kalman e o Método Sequencial de Monte Carlo, também conhecido como Filtro de Partículas, são métodos baseados na filtragem Bayesiana, que é uma técnica probabilística para a fusão de dados. A técnica combina a formulação matemática de um sistema com observações desse sistema. Ressaltando que esta dissertação trata de Assimilação de Dados, que segundo Harter (2004) é um problema de estimação, reserva-se este capítulo ao Filtro de Kalman e ao Filtro de Partículas. Assim, faz-se uma introdução sobre a teoria de estimação baseada em modelos e por conseguinte apresentam-se o Filtro de Kalman e o Filtro de Partículas.

3.1 Modelos de Sistemas Dinâmicos

A teoria de estimação aqui discutida é baseada na representação matemática de um *sistema*, denominado modelo. Segundo Monteiro (2006) um *sistema* pode ser definido como um conjunto de objetos agrupados por alguma interação ou interdependência, de modo que existam relações de causa e efeito nos fenômenos que ocorrem com os elementos desse conjunto. Um *sistema* é dito ser *dinâmico* quando algumas grandezas que caracterizam seus objetos constituintes variam no tempo.

Portanto, a necessidade de modelos apropriados é fundamental. Considera-se um modelo apropriado aquele que satisfaz o objetivo para o qual foi proposto. Em outras palavras, o modelo deve descrever as propriedades essenciais do sistema em estudo, mas também precisa ser simples o bastante para que seja possível desenvolver um algoritmo de estimação eficiente. Assim, um modelo confiável é essencial para se obter boas estimativas.

Alguns exemplos de modelos são: modelo meteorológico DYNAMO 1D, modelo 3-ondas descritos em Harter (2004) e o modelo de Lorenz, que é o problema teste usado neste trabalho e será descrito na Seção 6.1. Uma formulação geral de um modelo é dada por:

$$F[\dot{z}(t), z(t), \tilde{u}(t), \theta, t] = 0 \quad (3.1)$$

em que o ponto denota diferenciação no tempo, ou seja, $\dot{z} = dz/dt$, z o vetor da variável de estado, \tilde{u} o sinal externo, θ o vetor de parâmetro invariante no tempo e t o tempo. Por fim, a dinâmica do modelo é descrita por uma função F , possivelmente não-linear.

A maioria dos fenômenos físicos são contínuos. Então, faz-se uma breve discussão introdutória baseada em modelos de tempo contínuo. No entanto, modelos de tempo discreto podem ser derivados de modelos de tempo contínuo (SCHÖN, 2006). Na equação 3.1 há dois tipos importantes de sinais externos \tilde{u} , que devem ser tratados separadamente. O

primeiro tipo é constituído por um sinal de entrada conhecido, denotado por u . Exemplos típicos incluem sinais de controle ou perturbações medidas. O segundo tipo é uma entrada não conhecida, denotada por w . Estes sinais são tipicamente usados para modelar perturbações não conhecidas, que são descritas usando-se processos estocásticos.

Em 3.1, a função F pode ser uma equação diferencial algébrica. Se esta equação contiver variáveis externas descritas por processos estocásticos, será denominada de *equação diferencial algébrica estocástica*. Os modelos sempre contêm elementos de incertezas, implicando na necessidade em se ter habilidade para lidar com equações diferenciais algébricas estocásticas.

Além do modelo para descrever o comportamento do sistema, precisa-se também de um modelo para descrever como as medidas ruidosas estão relacionadas às variáveis do sistema, ou seja, precisa-se de um modelo de medida. Desde que não se pode realizar medidas infinitesimais, as observações são obtidas em tempo discreto de acordo com:

$$H[y(t_k), z(t_k), u(t_k), e(t_k), \theta, t_k] = 0 \quad (3.2)$$

em que $y \in \mathbb{R}^{n_y}$ denota a medida, $e \in \mathbb{R}^{n_e}$ o ruído da medida, t_k o tempo discreto, θ o vetor de parâmetro de tempo invariante e H denota uma possível função não-linear que descreve como a medida é obtida. A equação da medida definida em 3.2 é implícita (SCHÖN, 2006). Uma equação de medida explícita mais específica, que é o tipo mais comum, é dada por:

$$y(t_k) = h[z(t_k), u(t_k), e(t_k), \theta, t_k] \quad (3.3)$$

No entanto, há aplicações que requerem a sua representação na forma de equações de medidas implícitas.

A representação de um modelo por uma equação diferencial estocástica algébrica é dada por:

$$F[\dot{z}(t), z(t), u(t), w(t), \theta, t] = 0 \quad (3.4a)$$

$$H[y(t_k), z(t_k), u(t_k), e(t_k), \theta, t_k] = 0 \quad (3.4b)$$

em que $w(t)$ e $e(t_k)$ são processos estocásticos. Para uma definição matematicamente mais completa da teoria de equações diferenciais estocásticas e do cálculo de Itô, consultar o livro de Jazwinski (1970).

Um caso importante do modelo dado pelas equações 3.4a-3.4b aparece quando $\dot{z}(t)$ pode

ser explicitamente resolvido por:

$$\dot{z}(t) = f[z(t), u(t), w(t), \theta, t] \quad (3.5)$$

O modelo resultante é então governado por uma equação diferencial ordinária (EDO), ao invés de uma equação diferencial algébrica. Este modelo é comumente citado como o modelo de espaço de estado em tempo contínuo na comunidade de controle ou sistema dinâmico para matemática física e computação.

3.1.1 Introdução ao estudo de Espaços de Estados de Modelos

A discussão nesta seção está fundamentada na teoria de probabilidades. O objetivo é fornecer uma transição do modelo da equação diferencial algébrica para o modelo de espaço de estados. Consideram-se apenas modelos em tempo discreto.

O modelo do sistema é um modelo dinâmico que descreve a evolução de suas variáveis de estado no decorrer do tempo. Uma propriedade fundamental relacionada ao modelo do sistema é a **propriedade de Markov**, definida a seguir.

Um processo estocástico em tempo discreto $\{x_t\}$ é dito possuir uma **propriedade de Markov** se

$$p(x_{t+1}|x_1, \dots, x_t) = p(x_{t+1}|x_t) \quad (3.6)$$

ou seja, a realização do processo no tempo t contém todas as informações a respeito do passado que é necessária para calcular o comportamento futuro do processo. Então, se a realização presente do processo é conhecida, o futuro é independente do passado. Esta propriedade é às vezes referida como o *princípio da casualidade generalizada*: o futuro pode ser predito do conhecimento do presente (JAZWINSKI, 1970). O modelo do sistema pode então ser descrito como

$$x_{t+1} \sim p_\theta(x_{t+1}|x_1, \dots, x_t) = p_\theta(x_{t+1}|x_t) \quad (3.7)$$

em que na equação 3.7 utilizou-se a propriedade de Markov. A notação $p_\theta(x)$ é usada para descrever uma família de funções de densidade de probabilidade parametrizada por θ . A função densidade de probabilidade $p_\theta(x_{t+1}|x_t)$ descreve a evolução das variáveis de estado no tempo. Em geral, esta pode ser não-gaussiana e incluir não linearidades. Assume-se que o estado inicial pertence a uma função densidade de probabilidade $p_\theta(x_0)$, comumente denominada de probabilidade *a priori*. Além disso, o sistema do modelo pode ser parametrizado por um vetor de parâmetro estacionário θ , como indicado em 3.7. Se o vetor θ não é conhecido, ele deve ser estimado antes do modelo ser usado para seu

objetivo proposto. A tarefa de encontrar este vetor de parâmetros baseado nas medidas disponíveis é conhecido como o problema de *identificação de sistema*. Neste trabalho, a atenção não está voltada para o problema de identificação de sistemas, que é tratado em [Aguirre \(2000\)](#) e [Schön \(2006\)](#). O objetivo aqui é o de determinar as condições iniciais necessárias para executar os modelos numéricos de previsão, o que consiste no problema denominado como *Assimilação de Dados*.

Seja $\{x_t\}$ um processo de Markov não observável diretamente. A informação a respeito desse processo é indiretamente obtida das medidas y_t (observações) de acordo com o *modelo de medida*

$$y_t \sim p_\theta(y_t|x_t) \quad (3.8)$$

O processo de observação $\{y_t\}$ é assumido ser condicionalmente independente do processo das variáveis de estado $\{x_t\}$, isto é

$$p_\theta(y_t|x_1, \dots, x_N) = p_\theta(y_t|x_t), \quad \forall t, \quad 1 \leq t \leq N \quad (3.9)$$

Além disso, assume-se que as observações são mutuamente independentes do tempo, isto é

$$\begin{aligned} p_\theta(y_t|x_1, \dots, x_N) &= \prod_{i=1}^N p_\theta(y_i|x_t, \dots, x_N) \\ &= \prod_{i=t}^N p_\theta(y_i|x_i), \quad \forall t, \quad 1 \leq t \leq N \end{aligned} \quad (3.10)$$

em que a equação 3.9 é usada para obter a igualdade anterior. A discussão acima é resumida pelo **Modelo oculto de Markov** que segundo [Doucet et al. \(2000\)](#), é definido por:

$$x_{t+1} \sim p_\theta(x_{t+1}|x_t) \quad (3.11a)$$

$$y_t \sim p_\theta(y_t|x_t) \quad (3.11b)$$

em que θ é usado para indicar um vetor de parâmetro estacionário. Uma classe restritiva de um modelo é assumir expressões explícitas tanto para o modelo do sistema, quanto para o modelo de medida, resultando no chamado modelo de espaço de estados.

3.1.2 Espaço de Estado de Modelos

De acordo com [Monteiro \(2006\)](#) o espaço de estados, ou espaço de fases, é um espaço m -dimensional, cujos eixos coordenados são o eixo- x_1 , eixo- x_2 , \dots , eixo- x_m . Então, a variável de estado será representada como um ponto de coordenadas $x_1(t), x_2(t), \dots, x_m(t)$ nesse

Há teoremas disponíveis descrevendo como obter relações similares quando o ruído não é aditivo. Para mais detalhes a respeito este tópico, veja [Jazwinski \(1970\)](#).

A hipótese de que as observações são mutuamente independentes no tempo [3.10](#) traduz-se na independência mútua do ruído da medida e_t no tempo,

$$\begin{aligned} p_{\theta}(y_t, \dots, y_N | x_t, \dots, x_N) &= \prod_{i=t}^N p_{\theta}(y_i | x_i) \\ &= \prod_{i=t}^N p_{e_i}(y_i - h(x_i, \theta, i)) \end{aligned} \quad (3.16)$$

Além disso, usando a probabilidade condicional e a propriedade de Markov, tem-se:

$$\begin{aligned} p_{\theta}(x_t, \dots, x_N) &= \prod_{i=t}^{N-1} p_{\theta}(x_{i+1} | x_i) \\ &= \prod_{i=t}^{N-1} p_{w_i}(x_{i+1} - f(x_i, \theta, i)) \end{aligned} \quad (3.17)$$

Conseqüentemente, o ruído do processo w_t também será mutuamente independente no tempo. A discussão anterior explica na realidade como a hipótese traduz o uso do ruído branco, definido no [Apêndice A](#), no modelo em tempo discreto com o ruído aditivo.

Para sistemas gaussianos e lineares, é suficiente que o processo seja não correlacionado de acordo com:

$$E(w_t - Ew_t)(w_s - Ew_s) = 0, \text{ com } t \neq s \quad (3.18)$$

desde que para este caso apenas os dois primeiros momentos estatísticos, isto é, a média e a variância, importam. Porém, para o caso de sistemas não-lineares e não-gaussianos, momentos de altas ordens também devem ser calculados, motivando a necessidade de independência. A definição de ruído branco implica que todas as entidades do processo $\{w_t\}$ são mutuamente independentes. Então, não há informação a respeito das realizações futuras do processo do ruído branco presente nas realizações do passado, implicando que ruído branco é totalmente imprevisível. Quando todas as informações sistêmicas são conhecidas, estas são em geral incorporadas às equações do modelo. O fato é que ruído branco é totalmente aleatório, sem correlação temporal, implicando que este fornece um bom modelo para estes efeitos ([SCHÖN, 2006](#)).

3.1.2.2 Espaço de Estado de Modelos Lineares

O caso especial mais importante de modelos em tempo discreto com ruído aditivo é provavelmente o modelo de espaço de estados lineares, isto é, f e h são funções lineares sujeitas à ruídos gaussianos. A razão para isto é o trabalho fundamental de Kalman (1960) sobre previsibilidade e controle quadrático linear, baseado neste modelo. O modelo de espaço de estados lineares com ruído gaussiano é dado por:

$$x_{t+1} = A_t x_t + B_t u_t + w_t \quad (3.19a)$$

$$y_t = C_t x_t + D_t u_t + e_t \quad (3.19b)$$

em que $w_t \sim \mathcal{N}(0, Q_t)$, $e_t \sim \mathcal{N}(0, R_t)$ e $E\{w_t e_t^T\} = 0$.

Uma propriedade importante do modelo linear 3.19a e 3.19b é que todas as funções de densidade envolvidas são gaussianas. Isto se deve ao fato de que uma transformação linear de uma variável aleatória gaussiana resultará em uma nova variável aleatória gaussiana (SCHÖN, 2006). Além disso, uma função densidade gaussiana é completamente parametrizada por dois parâmetros, o primeiro e o segundo momento estatístico, ou seja, a média e a variância.

3.2 Estimação de Estado Não-Linear

Nesta seção, trata-se da teoria de estimação recursiva de estados não-lineares. O problema de estimação de estados está voltado para uma estrutura probabilística, como visto na seção anterior. Mais especificamente, a aproximação é muito influenciada pela visão Bayesiana da estimação. Isto implica que a solução completa para o problema de estimação é fornecida pela função densidade de probabilidade $p(x_t|Y_s)$, sendo que $\{Y_s\}_{s=1}^t$ representa um conjunto de medidas até o instante t . Esta função de densidade contém toda a informação disponível a respeito a variável de estado (SCHÖN, 2006). Dependendo da relação entre t e s em $p(x_t|Y_s)$, três problemas de estimação diferentes são obtidos

- O problema de filtragem, $t = s$.
- O problema de previsão, $t > s$.
- O problema de interpolação, $t < s$.

Os itens acima indicam a estimação das variáveis de estado no presente, futuro e passado, respectivamente. Quando a representação para $p(x_t|Y_s)$ é obtida, esta pode ser usada para estimar o valor esperado de alguma função g das variáveis de estado, $I(g(x_t))$ de acordo

com:

$$I(g(x_t)) = E_{p(x_t|Y_s)}\{g(x_t)\} = \int_{\mathbf{R}^{n_x}} g(x_t)p(x_t|Y_s)dx_t \quad (3.20)$$

3.2.1 Solução Conceitual

Aqui concentra-se no problema de determinar as funções densidade de probabilidade pertinente ao problema de estimação. A discussão é bem geral, utilizando modelos em tempo discreto com ruído aditivo, isto é

$$x_{t+1} \sim p(x_{t+1}|x_t) \quad (3.21a)$$

$$y_t \sim p(y_t|x_t) \quad (3.21b)$$

As ferramentas necessárias são: o Teorema de Bayes e a propriedade de Markov discutida na [Subseção 3.1.1](#). Considerando duas variáveis estocásticas x e y , usando o teorema de Bayes para funções densidades de probabilidade, tem-se que:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y, x)}{p(y)} \quad (3.22)$$

Considere a densidade de filtragem,

$$p(x_t|Y_t) = p(x_t|y_t, Y_{t-1}) = \frac{p(y_t|x_t, Y_{t-1})p(x_t|Y_{t-1})}{p(y_t|Y_{t-1})} \quad (3.23)$$

em que o denominador de normalização $p(y_t|Y_{t-1})$ pode ser calculado de acordo com

$$\begin{aligned} p(y_t|Y_{t-1}) &= \int_{\mathbf{R}^{n_x}} p(y_t, x_t|Y_{t-1})dx_t \\ &= \int_{\mathbf{R}^{n_x}} p(y_t|x_t, Y_{t-1})p(x_t|Y_{t-1})dx_t \\ &= \int_{\mathbf{R}^{n_x}} p(y_t|x_t)p(x_t|Y_{t-1})dx_t \end{aligned} \quad (3.24)$$

Além disso, derivando a expressão para um passo de predição da densidade $p(x_{t+1}|Y_t)$, integra-se a seguinte equação em relação a x_t ,

$$p(x_{t+1}|Y_t) = p(x_{t+1}|x_t, Y_t)p(x_t|Y_t) = p(x_{t+1}|x_t)p(x_t|Y_t) \quad (3.25)$$

resultando na seguinte expressão:

$$p(x_{t+1}|Y_t) = \int_{\mathbf{R}^{n_x}} p(x_{t+1}|x_t)p(x_t|Y_t)dx_t \quad (3.26)$$

A equação 3.26 é denominada de equação de Chapman-Kolmogorov (JAZWINSKI, 1970). A atualização no tempo é feita de acordo com a equação 3.23, e a recursão é iniciada por $p(x_0|Y_{t-1}) = p(x_0)$ denominada de função densidade de probabilidade *a priori*. Para uma descrição mais detalhada, veja Jazwinski (1970) e Schön (2006).

Há entretanto, um problema com esta solução. Quando se trata de integrais de várias dimensões, a solução analítica existe apenas em poucos casos. A maior parte das soluções analíticas descritas ocorre quando o modelo dinâmico é linear e os ruídos são gaussianos. Isso tem sido intensivamente objeto de estudos nas últimas décadas (ARULAMPALAM et al., 2002; DOUCET, 1998; GORDON et al., 1993). Uma vez que, todas as funções de densidades envolvidas no problema são gaussianas e executam somente operações lineares, a função densidade das variáveis de estado do sistema também terão sua distribuição gaussiana e a solução ótima é fornecida pelo filtro de Kalman (KALMAN, 1960).

3.2.2 Estimativa Pontual

A tarefa de encontrar uma estimativa pontual pode, em termos abstratos, ser pensado como o problema de encontrar uma transformação m_t , que faça uso da informação das medidas e do sinal de entrada conhecido para produzir estimativas dos estados de interesse.

$$m_t : U_s \times Y_s \rightarrow \mathbf{R}^{n_x} \quad (3.27)$$

Todas as informações disponíveis nas medidas devem ser processadas e inferidas na função densidade $p(x_t|Y_s)$. Normalmente, na prática não se precisa conhecer a função densidade de probabilidade. Ao invés disso, precisa-se conhecer como os valores das variáveis de estado evoluem no tempo. Também precisa-se avaliar a qualidade desses valores. É razoável afirmar que uma estimativa é inútil se não for possível avaliar quão boa ela é. Desde que uma estrutura probabilística é empregada, esta permite o uso de ferramentas disponíveis na teoria de probabilidade e estatística para se ter acesso a qualidade da estimativa (SCHÖN, 2006).

Esta seção está voltada para um dos mapeamentos mais comuns presentes na literatura representado pela equação 3.27. A maioria das estimativas são de fato baseadas na aproximação da função densidade de probabilidade $p(x_t|Y_s)$, mas a estimativa também pode ser baseada sobre considerações determinísticas (ANDERSON; MOORE, 1979; JAZWINSKI, 1970).

Do ponto de vista probabilístico, uma estimativa pontual atraente é fornecida pela escolha de valores que minimizam a variância do erro de estimação, denominada como a estimativa

de mínima variância (MV):

$$\hat{x}^{MV} = \arg \min_{\hat{x}} E\{\|x - \hat{x}\|^2|y\} \quad (3.28)$$

sendo $\|x\|^2 = x^T x$. É possível derivar uma expressão explícita para esta estimativa. Tem-se que:

$$\min E\{\|\hat{x} - x\|^2|y\} \iff \frac{\partial E}{\partial \hat{x}} = 0 \quad (3.29)$$

Como,

$$\begin{aligned} E\{\|\hat{x} - x\|^2|y\} &= E\{(x - \hat{x})^T(x - \hat{x})|y\} \\ &= E\{x^T x|y\} - 2\hat{x}^T E\{x|y\} + \hat{x}^T \hat{x} \\ &= \|\hat{x} - E\{x|y\}\|^2 + E\{\|x\|^2|y\} - \|E\{x|y\}\|^2 \end{aligned} \quad (3.30)$$

Então,

$$\frac{\partial E}{\partial \hat{x}} = \frac{\partial}{\partial \hat{x}} \left[\|\hat{x} - E\{x|y\}\|^2 + \underbrace{E\{\|x\|^2|y\}}_a - \underbrace{\|E\{x|y\}\|^2}_b \right] \quad (3.31)$$

Os termos a e b da equação 3.31 são independentes de \hat{x} . Então,

$$\frac{\partial}{\partial \hat{x}} \|\hat{x} - E\{x|y\}\|^2 = 0 \quad (3.32)$$

Assim,

$$\hat{x}^{MV} = E\{x|y\} = \int xp(x|y)dx \quad (3.33)$$

O cálculo acima explica o nome, erro mínimo quadrático médio (MMSE) do inglês *minimum mean square error*, que é comumente usado como um nome alternativo para a estimativa 3.33.

Outra estimativa pontual é a estimativa de máxima *a posteriori* (MAP), dada por:

$$\hat{x}^{MAP} = \arg \max_x p(x|y) = \arg \max_x p(y|x)p(x) \quad (3.34)$$

Na segunda igualdade da equação 3.34 é empregado a regra de Bayes, considerando que a minimização é executada em relação a x .

3.2.3 Sistemas Não-Lineares

Geralmente, os problemas encontrados na prática são de natureza não-linear. Isto implica na necessidade de resolver problemas de estimação no contexto de sistemas não-lineares. Há uma grande quantidade de representações disponíveis para esses sistemas. Porém, uma

representação comum é através de um modelo em tempo discreto com ruído aditivo, dado pelo sistema de equações 3.14a-3.15b, repetido a seguir por conveniência

$$x_{t+1} = f(x_t, t) + w_t \quad (3.35a)$$

$$y_t = h(x_t, t) + e_t \quad (3.35b)$$

com $w_t \sim \mathcal{N}(0, Q_t)$ e $e_t \sim \mathcal{N}(0, R_t)$. Frequentemente, não existe solução analítica para o problema de estimação recursiva não-linear. Então, é necessário obter soluções aproximadas para o problema. Nas seções a seguir, apresentam-se duas classes de aproximações, a aproximação local e global.

3.2.3.1 Aproximação Local

A idéia empregada no método local é aproximar o modelo não-linear por um modelo linear e gaussiano. Este modelo é apenas válido localmente, podendo ser aplicado o Filtro de Kalman (KF). A aproximação local é obtida pela linearização do modelo não-linear 3.35a e 3.35b por aplicação da expansão em série de Taylor de primeira ordem em torno da estimativa atual, isto é

$$f(x_t, t) \approx f(\hat{x}_{t|t}, t) + \left. \frac{\partial f(x, t)}{\partial x} \right|_{x=\hat{x}_{t|t}} (x_t - \hat{x}_{t|t}) \quad (3.36a)$$

$$h(x_t, t) \approx h(\hat{x}_{t|t-1}, t) + \left. \frac{\partial h(x, t)}{\partial x} \right|_{x=\hat{x}_{t|t-1}} (x_t - \hat{x}_{t|t-1}) \quad (3.36b)$$

Usando esta aproximação em 3.35a e 3.35b tem-se que:

$$x_{t+1} = f(\hat{x}_{t|t}, t) - F_t \hat{x}_{t|t} + F_t x_t + w_t \quad (3.37a)$$

$$y_t = h(\hat{x}_{t|t-1}, t) - H_t \hat{x}_{t|t-1} + H_t x_t + e_t \quad (3.37b)$$

em que $F_t = \left. \frac{\partial f(x, t)}{\partial x} \right|_{x=\hat{x}_{t|t}}$ e $H_t = \left. \frac{\partial h(x, t)}{\partial x} \right|_{x=\hat{x}_{t|t}}$.

A aproximação do modelo dado nas equações 3.37a e 3.37b é linear e sujeito a ruídos gaussianos em x_t o que significa que o FK pode ser aplicado. O resultado é o Filtro de Kalman Estendido (EKF), descrito no algoritmo apresentado na tabela 3.1. Considerando o modelo linear em tempo discreto com ruído aditivo, uma aproximação subótima para a função densidade de probabilidade $p(x_t|Y_t)$, obtida pela linearização, é recursivamente dada de acordo com o algoritmo apresentado na tabela 3.1 (ANDERSON; MOORE, 1979) e (JAZWINSKI, 1970).

Tabela 3.1 - Algoritmo: Filtro de Kalman Estendido

Algoritmo: Filtro de Kalman Estendido
$\hat{p}(x_t Y_t) = \mathcal{N}(x \hat{x}_{t t}, P_{t t})$
$\hat{p}(x_{t+1} Y_t) = \mathcal{N}(x \hat{x}_{t+1 t}, P_{t+1 t})$
em que:
$\hat{x}_{t+1 t} = f(\hat{x}_{t t})$
$P_{t+1 t} = F_t P_{t t} F_t^T + Q_t$
$K_t = P_{t t-1} H_t^T (H_t P_{t t-1} H_t^T + R_t)^{-1}$
$\hat{x}_{t t} = \hat{x}_{t t-1} + K_t (y_t - h(\hat{x}_{t t-1}, t))$
$P_t = P_{t t-1} - K_t H_t P_{t t-1}$

3.2.3.2 Aproximação Global

A solução para o problema de estimação recursiva foi apresentado na [Subseção 3.2.1](#). Uma das primeiras aproximações baseadas em um conjunto de parâmetros é a aproximação por somas de gaussianas ([SORENSEN; ALSPACH, 1971](#)). A densidade é aproximada usando uma soma de gaussianas de acordo com:

$$p(x_{t+1}|Y_t) \approx \sum_{i=1}^N q_t^{(i)} \mathcal{N}(x|\hat{x}_{t|t}^{(i)}, P_{t|t}^{(i)}), \quad \sum_{i=1}^N q_t^{(i)} = 1, \quad q_t^{(i)} \geq 0, \quad \forall i \quad (3.38)$$

Uma outra aproximação é fornecida pelo filtro *point-mass* originalmente sugerido por [Bucy e Senne \(1971\)](#), o que como o próprio nome sugere, aproxima a densidade de filtragem por um conjunto de pontos sobre uma grade,

$$p(x_{t+1}|Y_t) \approx \sum_{i=1}^N q_t^{(i)} \delta(x_t - x_t^{(i)}), \quad \sum_{i=1}^N q_t^{(i)} = 1, \quad q_t^{(i)} \geq 0, \quad \forall i \quad (3.39)$$

A idéia tem sido refinada nos últimos anos usando-se, por exemplo, aproximações constantes lineares por partes e interpolação *splines* ([SCHÖN, 2006](#)). Uma aproximação, que pode ser interpretada como uma extensão do filtro *point-mass* é o fornecido pelo método Sequencial de Monte Carlo, referido na literatura como Filtro de Partículas ([GORDON et al., 1993](#)). Este é explanado na [Seção 3.3](#).

3.2.4 Filtragem e Predição

O caso especial obtido assumindo um modelo gaussiano e linear [3.19a](#) e [3.19b](#) permite uma solução explícita para as expressões apresentadas na [Subseção 3.2.1](#). A filtragem e um passo de predição é dado pelo KF, introduzido por [Kalman \(1960\)](#), [Kalman e Bucy \(1961\)](#). O KF é freqüentemente usado em controle e problemas de estimação. Depois de suas aplicações na área espacial ([JAZWINSKI, 1970](#)), ele tem sido aplicado em diversos

ramos da ciência, por exemplo, meteorologia e oceanografia (DALEY, 1993; KALNAY, 2003). A seguir apresenta-se o Filtro de Kalman.

Introduz-se a seguinte notação, $\hat{x}_{t|s}$, que representa a estimativa do estado x no tempo t usando a informação disponível, ou seja, as medidas até o instante s . Em outras palavras, $\hat{x}_{t|s} = E\{x_t|Y_s\}$.

Filtro de Kalman

Considere as equações 3.19a e 3.19b, assume-se que o estado inicial é distribuído como $x_0 \sim \mathcal{N}(\bar{x}_0, \bar{P}_0)$. Então, a estimativa para a função densidade de filtragem e um passo avançado de predição da função densidade de probabilidade são ambas normal, de acordo com:

$$\hat{p}(x_t|Y_t) = \mathcal{N}(x|\hat{x}_{t|t}, P_{t|t}) \quad (3.40a)$$

$$\hat{p}(x_{t+1}|Y_t) = \mathcal{N}(x|\hat{x}_{t+1|t}, P_{t+1|t}) \quad (3.40b)$$

em que:

$$\hat{x}_{t+1|t} = f(\hat{x}_{t|t}) \quad (3.41a)$$

$$P_{t+1|t} = F_t P_{t|t} F_t^T + Q_t \quad (3.41b)$$

$$K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1} \quad (3.41c)$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t (y_t - h(\hat{x}_{t|t-1}, t)) \quad (3.41d)$$

$$P_t = P_{t|t-1} - K_t H_t P_{t|t-1} \quad (3.41e)$$

com valores iniciais $\hat{x}_0 = \bar{x}_0$ e $P_0 = \bar{P}_0$.

Existem diferentes formas para a prova deste resultado. Em Anderson e Moore (1979) encontram-se demonstrações alternativas. Em Harter (2004) o Filtro de Kalman é obtido da solução recursiva do problema de mínimos quadrados ponderados, baseado no trabalho de Sorenson (1970).

A atualização da medida é dada pelas as equações 3.41d e 3.41e. Nestas equações é que a informação da medida atual y_t é incorporada dentro da estimativa. A equação 3.41d mostra que a estimativa do estado é ajustada como uma média ponderada da estimativa anterior e a nova informação disponível em y_t . As incertezas são reduzidas em 3.41e como uma consequência direta do fato de que uma nova informação foi adicionada. Além disso, a atualização no tempo corresponde à predição, implicando em aumento de incerteza 3.41b. Devido ao fato de que o processo de ruído w_t , por definição, não é previsível, a evolução da variável de estado é obtida pelo uso da parte determinística do modelo dinâmico, como

em 3.41a.

3.3 Método Seqüencial de Monte Carlo (Filtro de Partículas)

O método *Seqüencial de Monte Carlo* ou *Filtro de Partículas* trata com o problema de estimar recursivamente a função densidade de probabilidade $p(x_t|Y_s)$. De acordo com o ponto de vista Bayesiano $p(x_t|Y_s)$, contém todas as informações estatísticas disponíveis a respeito a variável de estado x_t baseado nas informações contidas nas medidas Y_s . Esta função densidade de probabilidade pode ser usada para formar várias estimativas das variáveis de estados, de acordo com:

$$I(g(x_t)) = E\{g(x_t)|Y_s\} = \int_{\mathbf{R}^{n_x}} g(x_t)p(x_t|Y_s)dx_t \quad (3.42)$$

A idéia fundamental, base do método seqüencial de Monte Carlo, é representar a função densidade de probabilidade por um conjunto de amostras com seus pesos associados. Este conjunto de amostras é denominado de partículas, então o nome Filtro de Partículas. A função densidade de probabilidade $p(x_t|Y_s)$ é aproximada por uma função de densidade empírica (nota-se a similaridade com a equação 3.39):

$$p(x_t|Y_s) \approx \sum_{i=1}^M \tilde{q}_t^{(i)} \delta(x_t - x_{t|s}^{(i)}), \quad \sum_{i=1}^M \tilde{q}_t^{(i)} = 1, \quad \tilde{q}_t^{(i)} \leq 0, \quad \forall i \quad (3.43)$$

em que $\delta(\cdot)$ é a função delta de Dirac¹ e $\tilde{q}_t^{(i)}$ denota os pesos associados a cada partícula $x_{t|s}^{(i)}$. Para se obter esta aproximação, exige-se a habilidade em gerar números aleatórios de distribuições complexas. A aproximação 3.43 pode também ser obtida usando-se a idéia de integração estocástica (SCHÖN, 2006).

Na Subseção 3.3.1 apresenta-se uma suposição não realística de que se pode gerar amostras de uma função de densidade alvo (“target”). O objetivo é ilustrar a idéia da amostragem, e motivar à próxima seção. Apresenta-se a solução de amostragem e reamostragem por importância, a qual é usada para derivar o Filtro de Partículas.

3.3.1 Amostragem Perfeita

O objetivo é calcular estimativas para a equação 3.42 baseado na hipótese de que se tem acesso a M amostras independente e identicamente distribuídas (i.i.d), $\{x^{(i)}\}_{i=1}^M$, de uma

¹A função Delta de Dirac, heurísticamente, pode ser representada:

$$F(x^{(i)}) = \int_{-\infty}^{+\infty} F(x)\delta(x - x^{(i)})dx, \quad \text{em que } \delta(x - x^{(i)}) = \begin{cases} 0 & \text{se } x \neq x^{(i)}, \\ \infty & \text{se } x = x^{(i)}. \end{cases} \quad (3.44)$$

função densidade alvo (*target density*) $t(x)$. Do ponto de vista prático, esta hipótese é não realística (DOUCET, 1998). Apesar disso, ilustra a idéia fundamental, base do método Sequêncial de Monte Carlo. Usando amostras $\{x^{(i)}\}_{i=1}^M$ uma estimativa empírica da função densidade $t(x)$ pode ser estabelecida de acordo com:

$$\hat{t}_M(x) = \sum_{i=1}^M \frac{1}{M} \delta(x - x^{(i)}) \quad (3.45)$$

Usando esta densidade empírica, uma estimativa de $I(g(x))$ é obtida de

$$\begin{aligned} \hat{I}(g(x)) = \hat{E}\{g(x)|Y\} &= \int g(x) \hat{t}(x) dx \\ \hat{E}\{g(x)|Y\} &= \int g(x) \sum_{i=1}^M \frac{1}{M} \delta(x - x^{(i)}) \\ \hat{E}\{g(x)|Y\} &= \frac{1}{M} \sum_{i=1}^M \int g(x) \delta(x - x^{(i)}) dx \\ \hat{E}\{g(x)|Y\} &= \frac{1}{M} \sum_{i=1}^M g(x^{(i)}) \end{aligned} \quad (3.46)$$

em que $\delta(x - x^{(i)})$ indica a função delta de Dirac em $x^{(i)}$, que vale zero para todo $x^{(i)} \neq x$.

Esta estimativa é sem viés, e de acordo com a lei forte dos grandes números² (SCHÖN, 2003), tem-se que:

$$\lim_{M \rightarrow \infty} \hat{I}(g(x)) \xrightarrow{a.s.} I(g(x)) \quad (3.47)$$

em que $\xrightarrow{a.s.}$ indica convergência “quase certa” (*almost sure*). Assume-se que a variância $\sigma^2 = I(g^2(x)) - I^2(g(x)) < \infty$ e o teorema do limite central possa ser aplicado, resultando em:

$$\lim_{M \rightarrow \infty} \sqrt{M} (\hat{I}(g(x)) - I(g(x))) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad (3.48)$$

em que \xrightarrow{d} indica a convergência da distribuição segundo Doucet et al. (2001). Assim, usando um número grande de amostras $\{x^{(i)}\}_{i=1}^M$, pode-se estimar uma quantidade $I(g(x))$ de acordo com a equação 3.46.

A hipótese fundamental discutida aqui é a possibilidade de se obter amostras independentes e identicamente distribuídas de uma distribuição $t(x)$. No entanto, na prática esta hipótese é raramente válida (SCHÖN, 2003). Para utilizar as idéias discutidas acima, precisa-se

²A lei forte dos grandes números é um teorema em probabilidades que descreve a longo prazo a estabilidade de uma variável aleatória (PAPOULIS, 1984).

de habilidade para gerar números aleatórios de distribuições de densidades complexas. Tem-se feito muitas pesquisas em relação a este problema, e diferentes métodos tem sido usados para tratá-lo (ARULAMPALAM et al., 2002; GORDON et al., 1993).

Uma solução clássica para este tipo de problema consiste em utilizar o método de *amostragem por importância* (*importance sampling*), o qual é descrito na [Subseção 3.3.2](#)

3.3.2 Geração de Números Aleatórios

Com a necessidade de representar funções densidades de probabilidades por um conjunto de amostras aleatórias, exige-se a habilidade em gerar números aleatórios. Porém, desde que não se pode gerar amostras diretamente da função densidade $t(x)$, a idéia é empregar uma densidade alternativa, da qual seja possível extrair amostras, denominada de densidade de amostragem $s(x)$. A única restrição imposta sobre $s(x)$ é que se $\forall x \in \mathbb{R}^{n_x}$, $t(x) > 0 \Rightarrow s(x) > 0$. Quando uma amostra $\bar{x} \sim s(x)$ é extraída, a probabilidade de que esta foi de fato gerada pela função densidade alvo pode ser calculada. Esse valor calculado pode então ser usado para se decidir se \bar{x} deve ser considerado ou não uma amostra de $t(x)$. Essa probabilidade é denominada de probabilidade de aceitação, expressa por $q(\bar{x})$ e definida pela seguinte relação:

$$t(\bar{x}) \propto q(\bar{x})s(\bar{x}) \tag{3.49}$$

Os vários tipos de métodos de Monte Carlo existentes na literatura dependem de como essa probabilidade de aceitação $q(\bar{x})$ é calculada. Os métodos mais comuns são brevemente descritos a seguir. Para mais detalhes a respeito a geração de números aleatórios, consultar [Rubinstein \(1981\)](#).

3.3.2.1 Importância da Amostragem e Reamostragem

O algoritmo SIR (*Sampling Importance Resampling*) é uma extensão da amostragem por importância. Assim, começa-se a exposição sobre o algoritmo SIR com uma explicação do algoritmo de *Amostragem por Importância*. Na discussão desta técnica, a densidade de amostragem $s(x)$ é conhecida como *densidade proposta* (*proposal density*), *função de importância* (*importance function*) ou *densidade amostral por importância* (*importance sampling density*) (ARULAMPALAM et al., 2002; DOUCET, 1998).

A amostragem por importância consiste em se introduzir uma função densidade de probabilidade alternativa, que seja conhecida e fácil de se extrair amostras para a estimação de $I(g(x))$, em oposição à distribuição de probabilidade de $t(x)$, que é considerada difícil de se extrair amostras, (veja ilustração na figura 3.1).

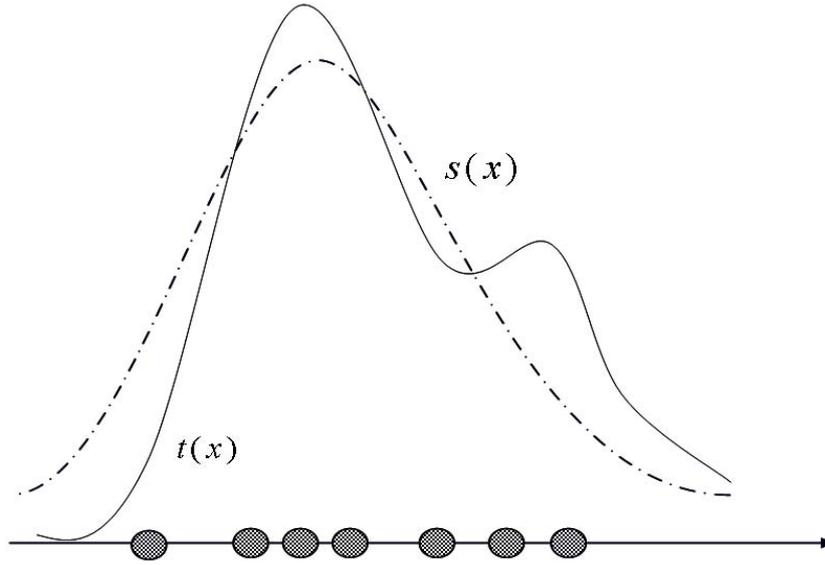


Figura 3.1 - Ilustração da amostragem por importância. $t(x)$ é a PDF verdadeira, representada pela linha cheia e $s(x)$ é a distribuição proposta (ou de importância) representada pela linha pontilhada.

Na amostragem por importância, a integral na forma da equação 3.42 pode ser escrita como,

$$I(g(x)) = E\{g(x)|Y\} = \int_{\mathbf{R}^{n_x}} g(x) \frac{t(x)}{s(x)} s(x) dx_t \quad (3.50)$$

em que $s(x)$ é a função de importância.

O método de Monte Carlo via amostragem por importância objetiva utilizar um número M de amostras independentes extraídas de $s(x)$ para obter uma soma ponderada para aproximar 3.50, sendo que $q(x) = \frac{t(x)}{s(x)}$ é chamado de *peso de importância*, ou *taxa de importância*. Baseado na discussão feita na seção 3.3.1 e de acordo com Rubinstein (1981) a estimativa para $I(g(x))$ pela geração de $M \gg 1$ amostras $\{x^{(i)}\}_{i=1}^M$ de $s(x)$ é

$$\hat{I}(g(x)) = \frac{1}{M} \sum_{i=1}^M q(x^{(i)}) g(x^{(i)}) \quad (3.51)$$

em que:

$$q(x^{(i)}) = \frac{t(x^{(i)})}{s(x^{(i)})} \quad i = 1, \dots, M \quad (3.52)$$

$q(x^{(i)})$ é chamado de peso de importância. Na maioria das aplicações para estimação de variáveis de estados com o procedimento de amostragem por importância, o fator de normalização na densidade alvo não é conhecido. A normalização para os pesos é dada por:

$$\tilde{q}(x^{(i)}) = \frac{q(x^{(i)})}{\sum_{i=1}^M q(x^{(i)})} \quad (3.53)$$

em que $q(x^{(i)})$ é definido em 3.52. Esta normalização para M finito introduzirá uma polarização na estimativa. Porém, da lei forte dos grandes números a estimativa é assintoticamente não polarizada. Assim, a função densidade alvo pode ser aproximada de acordo com:

$$\hat{t}_M(x) = \sum_{i=1}^M \tilde{q}(x^{(i)})\delta(x - x^{(i)}) \quad (3.54)$$

Os pesos de importância contém informações a respeito de quão provável a respectiva amostra associada é gerada a partir da função densidade alvo. Portanto, o peso de importância pode ser utilizado como uma probabilidade de aceitação, que permite gerar aproximadamente amostras independentes $\{\tilde{x}^{(i)}\}_{i=1}^M$ da função densidade alvo. A aproximação $\hat{t}(x)$ dado em 3.54 é definida usando-se um número finito de amostras $\{x^{(i)}\}_{i=1}^M$. Logo, o processo de geração de amostras da função densidade alvo é limitado para estas amostras. Mais especificamente, isto é realizado pela reamostragem entre as amostras, de acordo com:

$$Pr(\tilde{x}^{(i)} = x^j) = \tilde{q}(x^{(j)}), \quad i = 1, \dots, M \quad (3.55)$$

O passo de reamostragem é fundamental para garantir a convergência do método. Isto foi primeiramente realizado por [Gordon et al. \(1993\)](#) e será descrito em [Subseção 3.3.3](#).

3.3.2.2 Aceitação e Rejeição de amostras

O problema inerente ao algoritmo SIR é que as amostras produzidas são apenas aproximadamente distribuídas de acordo com a densidade alvo. Na amostragem por aceitação e rejeição, as amostras produzidas serão exatamente distribuídas de acordo com a densidade alvo. No entanto, este algoritmo sofre de outras desvantagens.

Se existir uma constante $L > 0$ tal que:

$$t(x) \leq Ls(x), \quad \forall x, \quad (3.56)$$

então, pode-se usar o algoritmo a seguir para gerar M amostras da densidade alvo.

Tabela 3.2 - Aceitação e Rejeição de Amostras

Algoritmo: Aceitação e Rejeição de Amostras
1. Gera-se um número aleatório, $\tilde{x} \sim s(x)$ e calcula-se $q(\tilde{x}) = \frac{t(\tilde{x})}{Ls(\tilde{x})}$
2. Aceita-se \tilde{x} como uma amostra de $t(x)$ com probabilidade $q(\tilde{x})$, isto é, $Pr(x^{(i)} = \tilde{x}) = q(\tilde{x})$ se \tilde{x} não for aceito volta para o passo 1.
3. Repete o passo 1. e 2. para $i = 1, \dots, M$

O algoritmo mostrado na tabela 3.4 é um método eficiente de amostragem no sentido de que as amostras extraídas de $t(x)$ são mutuamente independentes e exatas. De qualquer forma, como mencionado acima, o algoritmo sofre de algumas limitações. A primeira de todas, é que se deve encontrar um limite superior L , o que pode ser difícil. Além disso, uma vez que esse limite superior tenha sido encontrado, pode ser provado que $Pr(\tilde{x} \text{ aceitável}) = \frac{1}{L}$, que tipicamente é um número muito pequeno.

3.3.3 Filtro de Partículas

Considere o problema de filtragem, em que a densidade alvo é dada pela densidade de filtragem, $t(x_t) = p(x_t|Y_t)$. Para usar a idéia esboçada nas seções anteriores, faz-se necessário escolher uma função densidade de amostragem apropriada $s(x_t)$ e uma probabilidade de aceitação correspondente. Utilizando o teorema de Bayes e a propriedade de Markov, tem-se:

$$p(x_t|Y_t) = p(x_t|y_t, Y_{t-1}) = \frac{p(y_t|x_t)p(x_t|Y_{t-1})}{p(y_t|Y_{t-1})} \propto p(y_t|x_t)p(x_t|Y_{t-1}) \quad (3.57)$$

que sugere as seguintes escolhas

$$\underbrace{p(x_t|Y_t)}_{t(x_t)} \propto \underbrace{p(y_t|x_t)}_{q(x_t)} \underbrace{p(x_t|Y_{t-1})}_{s(x_t)} \quad (3.58)$$

A equação 3.58 é semelhante a equação 3.49, então, pode-se empregar os algoritmos discutidos na Subseção 3.3.2 para se obter amostras da densidade alvo. O filtro de partículas é tipicamente originado do sistema da amostragem por importância. Para outras derivações do filtro veja (DOUCET, 1998; ARULAMPALAM et al., 2002; SCHÖN, 2003). Considerando que o filtro é derivado da técnica de amostragem e reamostragem por importância, tem-se que a probabilidade de aceitação $\{\tilde{q}^{(i)}\}_{i=1}^M$ é calculada de acordo com:

$$\tilde{q}^{(i)} = \frac{q(x_{t|t-1}^{(i)})}{\sum_{j=1}^M q(x_{t|t-1}^{(j)})} = \frac{p(y_t|x_{t|t-1}^{(i)})}{\sum_{j=1}^M p(y_t|x_{t|t-1}^{(j)})} \quad (3.59)$$

em que $x_{t|t-1}^{(i)} \sim p(x_t|Y_{t-1})$. As partículas preditas $\{x_{t|t-1}^{(i)}\}_{i=1}^M$ são geradas do modelo dinâmico, e as partículas filtradas geradas do instante de tempo anterior $\{x_{t-1|t-1}^{(i)}\}_{i=1}^M$. Os detalhes podem ser compreendido do cálculo a seguir, que é resultado do uso da

atualização do tempo 3.23.

$$\begin{aligned}
s(x_t) &= p(x_t|Y_{t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|Y_{t-1})dx_{t-1} \\
&\approx \int p(x_t|x_{t-1}) \sum_{i=1}^M \frac{1}{M} \delta(x_{t-1} - x_{t-1|t-1}^{(i)}) dx_{t-1} \\
&= \sum_{i=1}^M \frac{1}{M} \int p(x_t|x_{t-1}) \delta(x_{t-1} - x_{t-1|t-1}^{(i)}) dx_{t-1} \\
&= \sum_{i=1}^M \frac{1}{M} p(x_t|x_{t-1}^{(i)})
\end{aligned} \tag{3.60}$$

Portanto, as partículas preditas são obtidas pela passagem das partículas filtradas através do sistema dinâmico.

De acordo com a equação 3.59, a probabilidade de aceitação $\tilde{q}_t^{(i)}$ depende da função de probabilidade $p(y_t|x_{t|t-1})$. Isso faz sentido, pois a probabilidade revela quão provável a medida obtida está relacionada ao estado presente. Quanto melhor uma certa partícula explicar a medida recebida, é alta a probabilidade de que esta partícula tenha sido extraída da função densidade de probabilidade real.

Um novo conjunto de partículas $\{x_{t|t}^{(i)}\}_{i=1}^M$ que aproxima $p(x_t|Y_t)$ é gerada pela reamostragem em substituição entre as partículas preditas, pertencendo a densidade de reamostragem

$$Pr\{x_{t|t}^{(i)} = x_{t|t-1}^{(j)}\} = \tilde{q}_t^{(j)}, \quad i = 1, \dots, M \tag{3.61}$$

Se este procedimento é recursivamente repetido no tempo, a seguinte aproximação

$$p(x_t|Y_t) \approx \sum_{i=1}^M \frac{1}{M} \delta(x_t - x_{t|t}^{(i)}) \tag{3.62}$$

é obtida e tem-se de fato derivado o Filtro de Partículas introduzido por (GORDON et al., 1993). O filtro de partículas é inicializado pela extração de amostras da função densidade de probabilidade *a priori* $p_{x_0}(x_0)$. Na atualização da medida, a nova medida é usada para designar uma probabilidade, representada pelo peso de importância normalizado, para cada partícula. Esta probabilidade é calculada usando a função de probabilidade que descreve o quão provável é obter uma medida dado a informação disponível na partícula. O peso de importância normalizado e a partícula correspondente constituem uma aproximação da densidade de filtragem. O passo de reamostragem retornará partículas que são igualmente prováveis. A atualização no tempo é uma forma de predizer novas partículas de acordo com o sistema do modelo. A seguir, descreve-se o algoritmo para o método,

apresentado na tabela 3.3.

Tabela 3.3 - Algoritmo do Filtro de Partículas

Algoritmo: Filtro de Partículas

1. **Inicialização:** Para $i = 1, \dots, M$ inicializa-se as partículas

$$\{x_{0|t-1}^{(i)}\}_{i=1}^M \sim p_{x_0}(x_0)$$

$$x_0 \sim \mathcal{N}(0, 5)$$
 , as partículas iniciais são extraídas de:

$$w_t^{(i)} \sim p_{w_t}(w_t)$$
 o ruído do processo
 com $w_t \sim \mathcal{N}(0, 1)$
 2. **Atualiza a medida:** Para $i = 1, \dots, M$
 Calcula o peso de importância $\{q_t^{(i)}\}_{i=1}^M$ de acordo com:

$$q_t^{(i)} = p(y_t | x_{t|t-1}^{(i)}),$$
 em que:

$$q_t^{(i)} = p(y_t | x_{t|t-1}^{(i)}) = p_{et}(z_t^i)$$
 com:

$$z_t^i = y_t - h(x_t, t)$$

$$p_{et} = \frac{1}{\sigma_z \sqrt{2\pi}} e^{-\frac{(z-\bar{z})^2}{2\sigma_z^2}}$$
 com

$$\sigma_z^2 = \frac{1}{M-1} \sum_i z_t^{(i)} \text{ e } \bar{z} = \frac{1}{M} \sum_i z_t^{(i)}$$
 normaliza $\tilde{q}_t^{(i)} = \frac{q_t^{(i)}}{\sum_{j=1}^M q_t^{(j)}}$
 3. **Reamostragem:** Para $i = 1, \dots, M$ (Análise)
 extrair M partículas, com substituição, de acordo com:

$$Pr\{x_{t|t}^{(i)} = x_{t|t-1}^{(j)}\} = \tilde{q}_t^{(j)}, \quad i = 1, \dots, M$$
 esse passo é explicado na [Subsubseção 3.3.3.1](#).
 4. **Próximo passo de integração no tempo:** Para $i = 1, \dots, M$
 prediz novas partículas de acordo com

$$x_{t+1|t}^{(i)} = f(x_{t|t}^{(i)}, t) + w_{t|t}^{(i)}$$
 em que f é dado pela equação 3.15a.
 5. **Calcula-se a estimativa**, que é a média das novas partículas

$$\hat{x} = E\{x_{t+1|t}^{(i)}\}$$
 6. Itera para o passo 2.
-

3.3.3.1 Algoritmo de Reamostragem

O passo de reamostragem consiste da extração de um novo conjunto de partículas $\{x_{t|t}^{(i)}\}_{i=1}^M$ com substituição das partículas anteriores $\{x_{t|t-1}^{(i)}\}_{i=1}^M$, de modo que a probabilidade de extrair $\{x_{t|t}^{(i)}\}_{i=1}^M$ é dada por $\tilde{q}_t^{(i)}$, de acordo com

$$Pr\{x_{t|t}^{(i)} = x_{t|t-1}^{(j)}\} = \tilde{q}_t^{(j)}, \quad i = 1, \dots, M \quad (3.63)$$

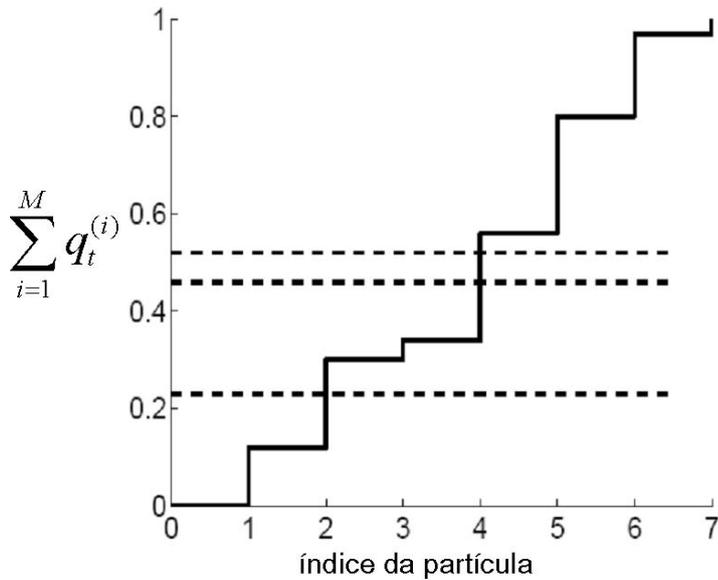


Figura 3.2 - Ilustração do passo de reamostragem do filtro de partículas.

Um modo de fazer isso é usar o algoritmo de reamostragem aleatória simples ilustrado na figura 3.2. A idéia é selecionar o novo conjunto de partículas pela comparação de um conjunto ordenado de números aleatórios distribuídos uniformemente $\mathcal{U}(0, 1)$ com a soma acumulativa dos pesos de importância normalizados. O passo de reamostragem pode de fato ser realizado do acordo com esta idéia. No entanto, existem vários algoritmos de reamostragem disponíveis na literatura. A eficiência do método é determinada pela qualidade da reamostragem e a complexidade computacional. A qualidade da reamostragem é importante para a qualidade global da estimativa.

O algoritmo de reamostragem sistemática apresentado na tabela 3.4 é considerado um dos mais apropriados, segundo Arulampalam et al. (2002).

Tabela 3.4 - Amostragem Sistemática

Algoritmo: Amostragem Sistemática
1. Gera-se M números ordenados de acordo com
$u_k = \frac{(k-1) + \tilde{u}}{M}, \quad \tilde{u} \sim U(0, 1)$
2. As partículas reamostradas são obtidas pela produção
n_i cópias da partícula $x^{(i)}$, em que
$n_i = \text{o número de } u_k \in \left(\sum_{s=1}^{i-1} \tilde{q}_t^{(s)}, \sum_{s=1}^i \tilde{q}_t^{(s)} \right]$

Mesmo com pesquisas extensivas sobre as técnicas de reamostragem, não se pode fugir do fato de que a reamostragem introduz uma dependência entre as diferentes partículas. Isto se deve ao fato de que as partículas associadas a valores de probabilidades maiores são selecionadas muitas vezes, pois a reamostragem é feita a partir de uma função densidade de probabilidade discreta, ao invés de uma contínua. Este problema é comumente citado na literatura como o problema de empobrecimento de amostras. Uma solução foi proposta por [Gordon et al. \(1993\)](#), que consiste em adicionar um ruído aleatório nas partículas para que elas possam diferir umas das outras.

3.3.3.2 Um exemplo para aplicação do Filtro de Partículas e Filtro de Kalman

O objetivo desta seção é apresentar o desempenho do Filtro de Kalman e do Filtro de Partículas. A aplicação é feita no sistema de equações dado a seguir:

$$x_{t+1} = \frac{x_t}{2} + \frac{25x_t}{1+x_t^2} + 8 \cos(1, 2t) + w_t \quad (3.64a)$$

$$y_t = \frac{x_t^2}{20} + e_t \quad (3.64b)$$

O algoritmo apresentado na tabela 3.3 será aplicado para a estimação das variáveis de estado do sistema. A estimativa do Filtro de Kalman é fornecida com o algoritmo dado na tabela 3.1. A equação 3.64a representa o sistema do modelo e a equação 3.64b representa o sistema de medida, sendo $x_0 \sim N(0, 5)$, w_t e e_t são ruídos gaussianos, com $w_t \sim N(0, 10)$ e $e_t \sim N(0, 1)$. Este problema é não-linear, variante no tempo, com ruído aditivo, o qual tem sido utilizado em muitos artigos como um problema teste ([GORDON et al., 1993](#); [ARULAMPALAM et al., 2002](#)).

A implementação foi feita no Matlab. Foram usados os seguintes parâmetros: a variância do erro de observação $R = 1$, do erro de modelagem $B = 10$. Para o filtro de partículas utilizou-se 1000 partículas. As condições iniciais (denominadas partículas) foram geradas usando a função *randn* do *matlab*, a qual gera números aleatórios normalmente distribuídos, ilustrado na figura 3.3.

No passo 2 do algoritmo apresentado na tabela 3.3 o peso de importância $q_t^{(i)}$ é calculado usando a função de probabilidade de acordo com:

$$q_t^{(i)} = p(y_t|x_t) = p_{et}(y_t - h(x_t, t)) \quad (3.65)$$

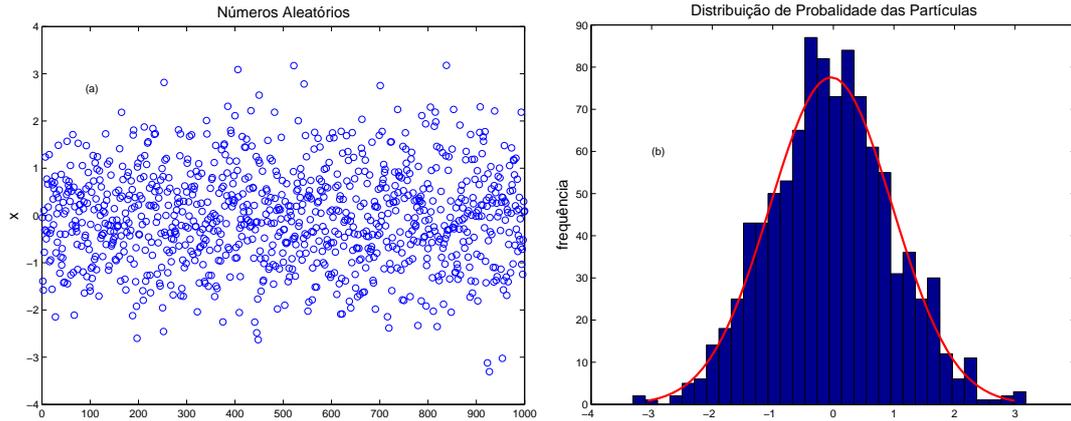


Figura 3.3 - (a): partículas iniciais (números aleatórios); (b): distribuição de probabilidade das partículas iniciais.

em que y_t é dado pela equação 3.64b e $h = \frac{x^2}{20}$ e

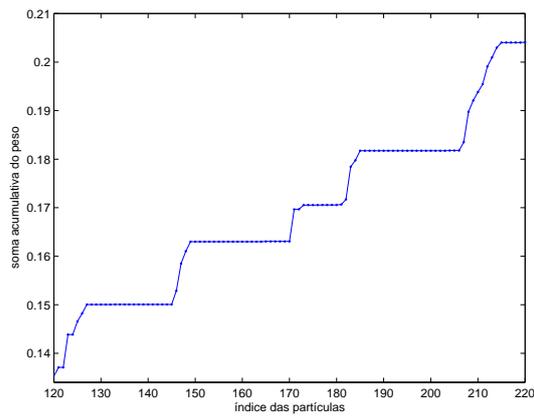
$$p_{et} = \frac{1}{2\pi^{\frac{1}{2}}}\exp\left(-\frac{x^2}{2}\right) \quad (3.66)$$

Uma vez calculado $q_t^{(i)}$ normaliza-o conforme a equação 3.53. O passo de reamostragem é feito conforme o algoritmo apresentado na tabela 3.4. A idéia é selecionar as novas partículas pela comparação de um conjunto ordenado de números aleatórios uniformemente distribuídos entre $U(0, 1)$ com a soma acumulativa dos pesos de importância normalizados. Na figura 3.4 ilustra-se o passo de reamostragem. Pode-se ver que a partícula de índice 130, por exemplo, foi escolhida uma vez, já as partículas de índice 140, 160 e 190 não foram escolhidas, enquanto que a partícula de índice 210 foi escolhida 18 vezes. Este é um dos problemas desta técnica, pois, haverá partículas que serão selecionadas várias vezes. Este problema é denominado de *empobrecimento de amostras*. Uma alternativa para a solução deste problema, como dito anteriormente, é adicionar um ruído aleatório nas partículas.

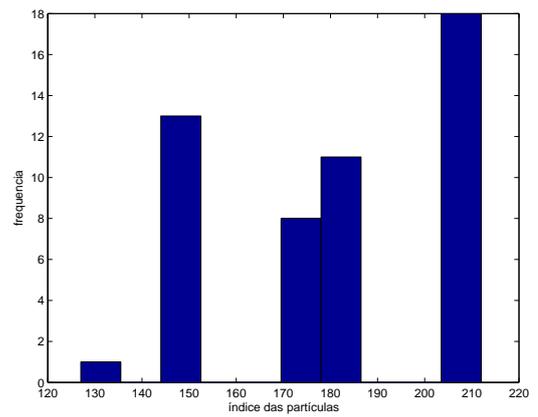
Na figura 3.5(a) e 3.5(b) têm-se a aplicação do Filtro de Kalman e do Filtro de Partículas, respectivamente. Na figura 3.6(a) e 3.6(b) apresenta-se o erro dos métodos. Nota-se que o FP obteve o menor erro, sendo que a média do erro para o KF foi de 8,1966 e para o FP de 2,9690.

Segundo (GORDON et al., 1993) as únicas exigências para a aplicação desta técnica são que:

- $p(x_0)$ esteja disponível para amostragem.
- a probabilidade $p(y_t|x_t)$ é uma função de forma conhecida.

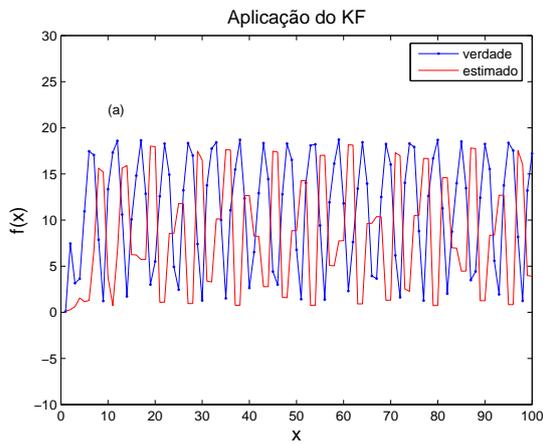


(a) Gráfico da soma acumulativa do peso.

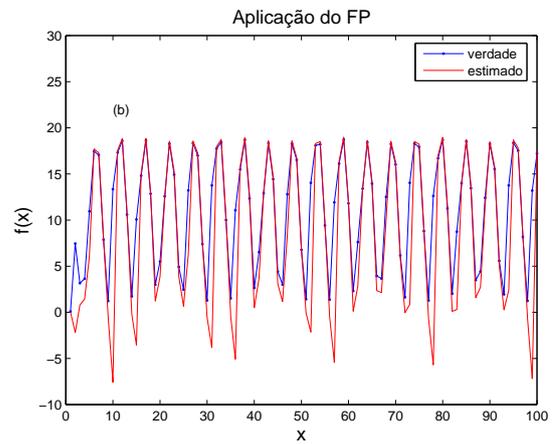


(b) Histograma do peso acumulativo.

Figura 3.4 - Ilustração do passo de reamostragem do filtro de partículas.



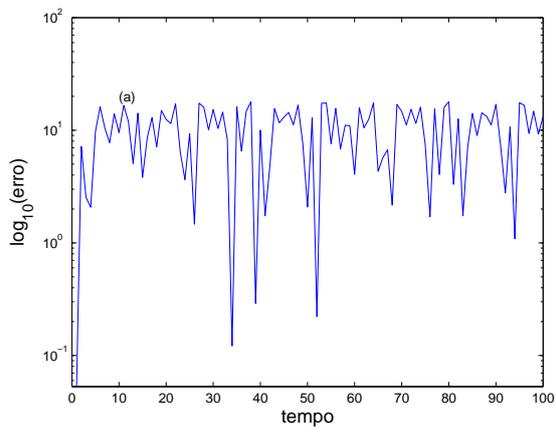
(a) aplicação do Filtro de Kalman



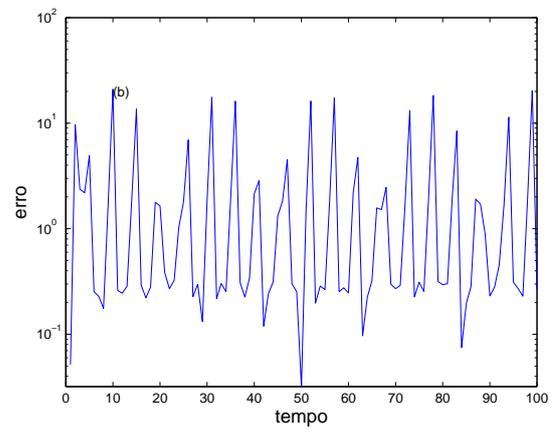
(b) aplicação do Filtro de Partículas

Figura 3.5 - curva azul: verdade; curva vermelha: estimado.

- $p(w_t)$ está disponível para amostragem.



(a) erro do Filtro de Kalman



(b) aplicação do Filtro de Partículas

Figura 3.6 - Gráficos dos erros em escala semilogarítmica. Logarítmica no eixo y e linear no eixo x .

4 MÉTODO VARIACIONAL

A assimilação de dados variacional também é um método de estimação de parâmetros para modelos matemáticos de previsão. A estimativa é feita por otimização do ajuste entre a solução do modelo de previsão e o conjunto de observações. Neste capítulo, apresenta-se o Método Variacional para estimação de condições iniciais de modelos numéricos de previsão. Primeiramente, será feita uma breve introdução sobre cálculo variacional.

4.1 Procedimento Variacional

O cálculo variacional envolve a determinação de pontos estacionários (extremos) de expressões integrais, que são conhecidas como funcionais. Pode-se pensar em um funcional como uma transformação $\varphi[\cdot]$ em que a entrada é uma função e a saída é um número. Isto é,

$$f(x) \xrightarrow{\varphi[f(x)]} y \text{ com } y \in \mathbb{R} \quad (4.1)$$

Uma função $f(x, y)$, sendo x e y variáveis independentes, tem um valor estacionário no ponto (x_0, y_0) se, em uma vizinhança infinitesimal em torno do ponto, a taxa de variação da função em cada direção possível deste ponto é zero.

O conceito de valor estacionário pode ser examinado pelo uso de um operador variacional δ , primeiramente introduzido por Lagrange (DALEY, 1993). Este operador é em muitas formas similar ao operador diferencial ordinário d , mas há uma sutil diferença. O operador diferencial d refere-se ao deslocamento infinitesimal real, enquanto o operador variacional δ refere-se ao deslocamento infinitesimal virtual. A diferença entre deslocamento real e virtual pode ser compreendida através de um exemplo simples. Considere uma bola que está em repouso no ponto mais baixo do interior de uma cavidade parabólica. A bola não se moverá (veja figura 4.1). No entanto, se deseja descobrir como a energia potencial varia quando a bola é deslocada de sua posição de repouso (equilíbrio). Para isso, pode ser feito um deslocamento exploratório na vizinhança de sua posição de repouso. O deslocamento exploratório é virtual, e é denominado de variação da posição.

Matematicamente, o operador δ é usado para avaliar a vizinhança do ponto (x_0, y_0) . Suponha um pequeno deslocamento virtual a partir do ponto (x_0, y_0) , dado por $(\delta x, \delta y)$. A variação da função δf pode ser expressa como:

$$\delta f = \frac{\partial f}{\partial x} \delta x + \frac{\partial f}{\partial y} \delta y \quad (4.2)$$

δf é chamada de primeira variação da função f . As variações δx e δy são escritas em

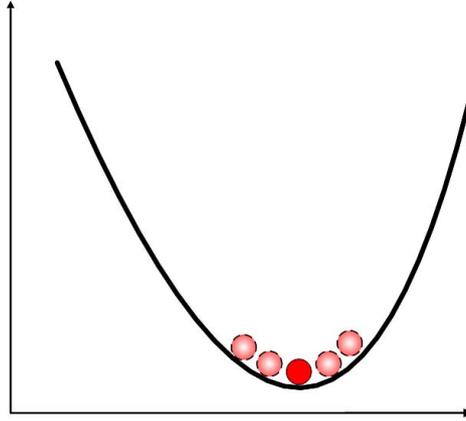


Figura 4.1 - Ilustração de uma bola (círculo vermelho) em estado de equilíbrio. Os círculos pontilhados representam o deslocamento virtual.

termos de seus cosenos diretores:

$$\delta x = \varepsilon \alpha_x \text{ e } \delta y = \varepsilon \alpha_y \quad (4.3)$$

em que α_x e α_y são cosenos diretores da direção virtual, ε é um parâmetro pequeno que tende a zero. A taxa de variação da função na direção especificada é dada por:

$$\frac{\partial f}{\varepsilon} = \frac{\partial f}{\partial x} \alpha_x + \frac{\partial f}{\partial y} \alpha_y \quad (4.4)$$

Para o ponto (x_0, y_0) ser um valor estacionário, $\partial f/\varepsilon$ deve ser zero para qualquer deslocamento virtual, independentemente da direção. Conseqüentemente,

$$\frac{\partial f}{\partial x} = 0 \text{ e } \frac{\partial f}{\partial y} = 0 \quad (4.5)$$

A condição que toda derivada parcial da função f seja nula em um ponto é uma condição necessária e suficiente para a função f ter um valor estacionário no ponto.

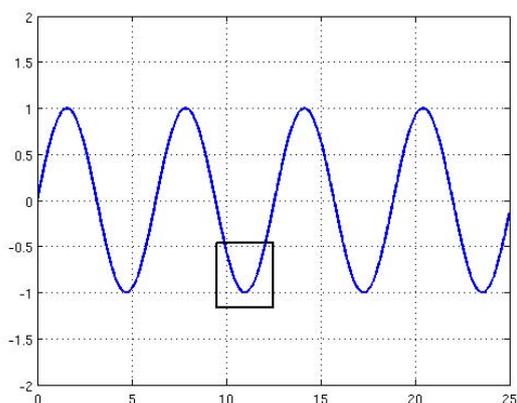
A segunda derivada pode ser usada para determinar se o valor estacionário é um ponto de máximo, ou de mínimo, ou nem um deles (cela). Assim, em uma dimensão o ponto estacionário $x = x_0$ é um ponto de máximo ou mínimo se:

$$\left. \frac{d^2 f}{dx^2} \right|_{x=x_0} \quad (4.6)$$

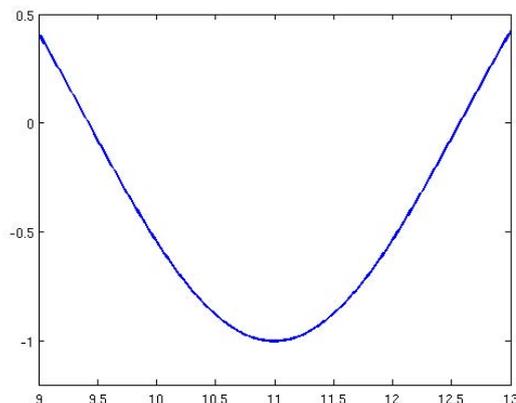
é menor ou maior que zero, respectivamente.

Para se determinar o valor estacionário ou extremo de uma função, é necessário especificar

o domínio em que ele será encontrado. A função $\sin(x)$, por exemplo, tem um número infinito de valores estacionários e extremos, conforme ilustrado na figura 4.2(a). No entanto, para o domínio $x_a < x < x_b$, onde $x_b = x_a + \pi$, haverá apenas um valor estacionário de $\sin(x)$, conforme figura 4.2(b).



(a) Função com vários máximos e mínimos.



(b) Função seno com um mínimo para o intervalo $9 < x < 13$

Figura 4.2 - Gráfico da função seno.

Em muitos problemas do cálculo variacional ocorrem restrições. Se deseja encontrar o valor estacionário da função $f(x, y)$ sujeita a condição auxiliar, $g(x, y) = 0$, há casos em que é possível reescrever essa condição como $y = h(x)$. Desse modo, determinar o ponto estacionário de $f(x, y)$ é equivalente a encontrar o ponto estacionário de $f(x, h(x))$. Assim, reduziu-se o problema bidimensional com restrições a um problema unidimensional sem restrições.

No entanto, nem sempre é possível descrever o problema com restrições da forma descrita acima. Então, utiliza-se o método de multiplicadores de Lagrange para esses casos.

Toma-se a variação da função $f(x, y)$ com restrição $g(x, y) = 0$, ou seja,

$$\delta f = \frac{\partial f}{\partial x} \delta x + \frac{\partial f}{\partial y} \delta y \quad (4.7)$$

Formando um novo funcional $f_1 = f + \lambda g$, onde λ é uma função ainda não determinada. A primeira variação da nova função f_1 é

$$\delta f_1 = \delta(f + \lambda g) = \delta f + \lambda \delta g + g \delta \lambda = \delta f \quad (4.8)$$

pois o operador δ obedece as mesmas regras de multiplicação e adição que o operador diferencial d . A função $g = 0$, $\delta g = 0$, δf_1 é uma função de x, y e λ , e pode ser escrita como:

$$\begin{aligned}\delta f_1 &= \frac{\delta f_1}{\delta x} \delta x + \frac{\delta f_1}{\delta y} \delta y + \frac{\delta f_1}{\delta \lambda} \delta \lambda \\ \delta f_1 &= \left(\frac{\delta f}{\delta x} + \lambda \frac{\delta g}{\delta x} \right) \delta x + \left(\frac{\delta f}{\delta y} + \lambda \frac{\delta g}{\delta y} \right) \delta y\end{aligned}\quad (4.9)$$

com $\frac{\partial f_1}{\partial \lambda} = g = 0$.

O ponto estacionário de f ocorre quando $\delta f = 0$. De 4.8, um ponto estacionário de f_1 é também um ponto estacionário de f . Então a condição para um ponto estacionário de f_1 , sujeito a restrição $g(x, y) = 0$ é

$$\frac{\delta f}{\delta x} + \lambda \frac{\delta g}{\delta x} = 0, \quad \frac{\delta f}{\delta y} + \lambda \frac{\delta g}{\delta y} = 0 \quad (4.10)$$

As duas equações de 4.10 mais a restrição podem então ser resolvidas para encontrar o valor estacionário. O multiplicador de Lagrange λ pode ser interpretado como uma medida da sensibilidade do valor da função f em função da restrição.

Em N dimensões, (x_1, \dots, x_N) , um ponto estacionário da função $f(x_1, \dots, x_N)$ sujeito a M restrições $g_1(x_1, \dots, x_N) = 0, \dots, g_M(x_1, \dots, x_N) = 0$ pode ser encontrado da mesma forma descrita acima. O resultado equivalente à equação 4.10 neste caso é dado por:

$$\frac{\partial}{\partial x_n} \left(f + \sum_{m=1}^M \lambda_m g_m \right) = 0, \quad 1 \leq n \leq N \quad (4.11)$$

4.2 Valor estacionário de uma Integral Definida

I é dito ser um funcional da função $u(x)$ em um intervalo (x_a, x_b) quando este depende de todos os valores $u(x)$, com $x_a \leq x \leq x_b$ (DALEY, 1993). Exemplos de funcionais são:

$$I(u(x)) = \int_{x_a}^{x_b} F(u(x)) dx$$

$$I(u(x)) = \int_{x_a}^{x_b} F(x, u, u', u'', \dots, u^n) dx$$

onde: $u' = \frac{du}{dx}, \dots, u^n = \frac{d^n u}{dx^n}$.

O domínio de um funcional é um conjunto de funções admissíveis, ao invés de uma região

do espaço cartesiano. Por exemplo, o domínio de um funcional pode ser o conjunto de todas as funções positivas (DALEY, 1993).

O problema fundamental do cálculo variacional é: *Dado um domínio de funções admissíveis, determinar a função $u(x)$ de um funcional $I(u(x))$, que é um valor estacionário desse funcional.* Um problema simples do cálculo variacional é determinar a função $u(x)$ que minimiza o funcional:

$$I(u(x)) = \int_{x_a}^{x_b} F(x, u, u') dx$$

sujeito às condições de contorno $x_a = \alpha$ e $x_b = \beta$. Aqui, α e β são fixos e assume-se que u é contínua e possui derivada de segunda ordem. Esse problema foi resolvido por Euler, porém uma solução elegante foi dada por Lagrange (DALEY, 1993). Suponha que $u(x)$ minimiza o funcional. Considere esta função levemente modificada, isto é

$$\tilde{u}(x) = u(x) + \varepsilon\eta(x) \tag{4.13}$$

onde $\eta(x)$ é uma função diferenciável e arbitrária, com $\eta(x_a) = \eta(x_b) = 0$ e ε é um valor pequeno que tende a zero.

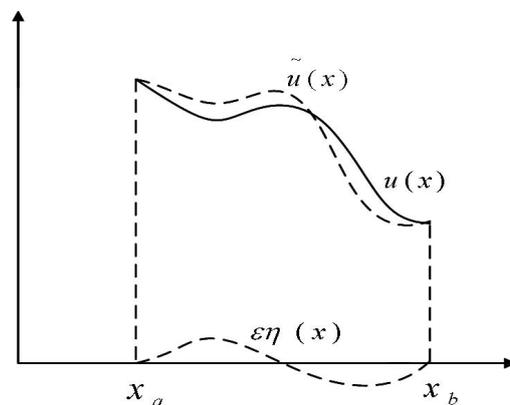


Figura 4.3 - Ilustração de $u(x)$, $\tilde{u}(x)$ e $\varepsilon\eta(x)$ usado na derivação das equações de Euler-Lagrange.

Fonte: adaptada de Daley (1993, 247)

A figura 4.3 descreve esquematicamente as funções $u(x)$, $\tilde{u}(x)$ e $\varepsilon\eta(x)$. A diferença entre $u(x)$ e $\tilde{u}(x)$ é chamada de variação da função u :

$$\delta u = \tilde{u}(x) - u(x) = \varepsilon\eta(x) \tag{4.14}$$

A diferença entre o operador variacional δ e o operador diferencial d está agora mais

claro, pois δu e du são ambas mudanças diferenciais da função u . Entretanto, du se refere a mudança infinitesimal de $u(x)$ causada pela variação infinitesimal do argumento dx , enquanto δu é uma variação infinitesimal de $u(x)$ que produz uma nova função $u(x) + \varepsilon \eta(x)$. No processo de variação, u se altera, enquanto a variável x não se altera. No entanto, se as condições de contorno estão predeterminadas, então esses valores não podem variar:

$$\delta x = 0, \quad \delta u|_{x_a} = 0, \quad \delta u|_{x_b} = 0 \quad (4.15)$$

O operador δ comuta com o operador de derivada normal e integral:

$$\frac{d}{dx}(\delta u) = \frac{d}{dx}(\varepsilon \eta(x)) = \varepsilon \frac{d\eta}{dx} = \frac{\tilde{u}}{dx} - \frac{du}{dx} = \delta \left(\frac{du}{dx} \right) \quad (4.16)$$

Analogamente,

$$\delta \int_{x_a}^{x_b} F(x, u, u') dx = \int_{x_a}^{x_b} \delta F(x, u, u') dx \quad (4.17)$$

Considerando a variação do integrando $F(x, u, u')$

$$\begin{aligned} \delta F(x, u, u') &= F(x, u + \delta u, u' + \delta u') - F(x, u, u') \\ &= F(x, u + \varepsilon \eta, u' + \varepsilon \eta') - F(x, u, u') \end{aligned} \quad (4.18)$$

Expandindo F em série de Taylor em torno de u e u' e desprezando os termos de ordem superior, pois ε é pequeno, tem-se:

$$\delta F(x, u, u') = \varepsilon \left(\frac{\partial F}{\partial u} \eta + \frac{\partial F}{\partial u'} \eta' \right) \quad (4.19)$$

Considerando a variação da integral I

$$\delta I = \delta \int_{x_a}^{x_b} F dx = \int_{x_a}^{x_b} \delta F dx = \varepsilon \int_{x_a}^{x_b} \left(\frac{\partial F}{\partial u} \eta + \frac{\partial F}{\partial u'} \eta' \right) dx \quad (4.20)$$

Integrando o segundo termo por partes:

$$\delta I = \varepsilon \int_{x_a}^{x_b} \eta \left(\frac{\partial F}{\partial u} - \frac{d}{dx} \frac{\partial F}{\partial u'} \right) dx + \varepsilon \eta(x) \frac{\partial F}{\partial u'} \Big|_{x_a}^{x_b} \quad (4.21)$$

δI é chamado de primeira variação do funcional I . Para determinar o ponto estacionário é análogo a 4.1, $\frac{\delta I}{\varepsilon} = 0$. Visto que, $\eta(x) = 0$ para $x = x_a$ e $x = x_b$ e $\eta(x)$ é arbitrário, a integral 4.21 deve se anular para qualquer função $\eta(x)$. Isso ocorre se

$$\frac{\partial F}{\partial u} - \frac{d}{dx} \frac{\partial F}{\partial u'} = 0 \text{ com } x_a \leq x \leq x_b \quad (4.22)$$

A condição 4.22 é necessária e suficiente para anular δI e é chamada de equação de *Euler-Lagrange*. Tomar $\delta I = 0$ é uma condição necessária, mas não suficiente para um extremo de I . A função $u(x)$ que satisfaz a equação de Euler-Lagrange é denominada de *função estacionária*. Para uma abordagem generalizada do problema consultar (DALEY, 1993).

4.3 Análise Variacional

Na prática, o funcional a ser minimizado pode ser substituído por uma forma discreta, representado, de modo geral, pela equação 4.23. Essa equação contém o termo que mede a distância entre a estimativa *a priori* e o somatório no tempo da função custo para cada incremento observacional calculado com respeito ao modelo integrado no tempo da observação (KALNAY, 2003).

O objetivo desta técnica é encontrar condições iniciais $x(t_0)$, de um modelo tal que minimize alguma quantidade escalar J , conhecida como uma função custo ou função de penalidade (*penalty*). O funcional J é definido como (BANNISTER, 2001; LAWLESS, 2004; KALNAY, 2003):

$$\begin{aligned}
 J(\mathbf{x}(t_0)) = & \underbrace{\frac{1}{2}(\mathbf{x}(t_0) - \mathbf{x}^b(t_0))^T \mathbf{B}^{-1}(\mathbf{x}(t_0) - \mathbf{x}^b(t_0))}_{J_b} \\
 & + \underbrace{\frac{1}{2} \sum_{i=0}^N (\mathbf{y}_i^o - \mathbf{H}(\mathbf{x}_i))^T \mathbf{R}^{-1}(\mathbf{y}_i^o - \mathbf{H}(\mathbf{x}_i))}_{J_o}
 \end{aligned} \tag{4.23}$$

em que J_b é o termo de modelagem e J_o é o termo de observação. \mathbf{B} é a matriz de covariância do erro de modelagem, \mathbf{R} é a matriz de covariância do erro de observação, \mathbf{y}^o é o vetor de observações e \mathbf{H} é o operador não-linear que representa o sistema de observação.

O vetor de estado \mathbf{x} e \mathbf{x}^b existe no espaço do modelo e consiste de N elementos. O significado de cada elemento depende do tipo de modelo e como este representa o campo meteorológico. No modelo de grade, os vetores conterão informações pertinentes aos valores de cada campo (u, v, θ, p e q , que representam vento zonal, vento meridional, temperatura, pressão e vapor de água, respectivamente) em cada posição da grade.

A variação da função custo 4.23 quando, a variável de controle $\mathbf{x}(t_0)$ é alterada por uma pequena perturbação $\delta \mathbf{x}(t_0)$ é dada por:

$$\delta J = J[\mathbf{x}(t_0) + \delta \mathbf{x}(t_0)] - J[\mathbf{x}(t_0)] \approx \left[\frac{\partial J}{\partial \mathbf{x}(t_0)} \right]^T \cdot \delta \mathbf{x}(t_0) \tag{4.24}$$

em que o gradiente da função custo $\left[\frac{\partial J}{\partial \mathbf{x}(t_0)}\right]_j = \frac{\partial J}{\partial \mathbf{x}_j(t_0)}$ é um vetor coluna. Para obter o mínimo do funcional J por um método iterativo, é necessário integrar o modelo adjunto \mathbf{M}_i^T , que é o transposto do modelo tangente linear (\mathbf{M}_i), descrito na seção 4.4.

O gradiente do primeiro termo J_b do funcional 4.23 com respeito a $\mathbf{x}(t_0)$ é dado por:

$$\frac{\partial J_b}{\partial \mathbf{x}(t_0)} = \mathbf{B}_0^{-1}[\mathbf{x}(t_0) - \mathbf{x}^b(t_0)] \quad (4.25)$$

O gradiente do segundo termo J_o de 4.23 é mais complicado, devido a $\mathbf{x}_i = M_i[\mathbf{x}(t_0)]$, em que M representa o modelo de previsão. Se se introduz uma perturbação ao estado inicial, então $\delta \mathbf{x}_i = \mathbf{L}(t_0, t_i)\delta \mathbf{x}_0$, assim que:

$$\begin{aligned} \frac{\partial(H(\mathbf{x}_i) - \mathbf{y}_i^o)}{\partial \mathbf{x}(t_0)} &= \frac{\partial H}{\partial \mathbf{x}_i} \frac{\partial M}{\partial \mathbf{x}_0} \\ &= \mathbf{H}_i \mathbf{L}(t_0, t_i) \\ &= \mathbf{H}_i \prod_{j=i-1}^0 \mathbf{L}(t_j, t_{j+1}) \end{aligned} \quad (4.26)$$

Como indicado na equação 4.26, as matrizes \mathbf{H}_i e \mathbf{L}_i são Jacobianas linearizadas $\frac{\partial H}{\partial \mathbf{x}_i}$, $\frac{\partial M}{\partial \mathbf{x}_0}$. \mathbf{L} é a matriz que propaga uma perturbação inicial ao tempo final de integração (modelo tangente linear) (LORENZ, 1965).

Portanto, o gradiente do termo de observação é dado por:

$$\left[\frac{\partial J_o}{\partial \mathbf{x}(t_0)}\right] = \sum_{i=0}^N \mathbf{L}(t_i, t_0)^T \mathbf{H}_i^T \mathbf{R}_i^{-1} [\mathbf{H}(\mathbf{x}_i) - \mathbf{y}_i^o] \quad (4.27)$$

A equação 4.27 mostra que cada iteração de minimização desse funcional requer o cálculo do gradiente, ou seja, o cálculo do incremento $[H(\mathbf{x}_i) - \mathbf{y}_i^o]$ no tempo de observação t_i durante a integração avançada, multiplicando-o por $\mathbf{H}_i^T \mathbf{R}_i^{-1}$ e integrando esses incrementos ponderados do tempo final ao tempo inicial usando o modelo adjunto (transposto do modelo tangente linear). Desde que a integração adjunta é comum em vários intervalos de tempo, o somatório 4.27 pode ser rearranjado mais convenientemente. Assume-se, por exemplo, que o intervalo de assimilação é de 00 h a 12 h, e que há observação a cada 3 horas (conforme figura 4.4). Calcula-se durante a integração avançada os incrementos de observação negativos, de acordo com:

$$\bar{\mathbf{d}}_i = \mathbf{H}_i^T \mathbf{R}_i^{-1} [\mathbf{H}(\mathbf{x}_i) - \mathbf{y}_i^o] = -\mathbf{H}_i^T \mathbf{R}_i^{-1} [\mathbf{y}_i^o - \mathbf{H}(\mathbf{x}_i)] \quad (4.28)$$

O modelo adjunto $\mathbf{L}^T(t_i, t_{i-1}) = \mathbf{L}_{i-1}^T$ aplicado sobre o modelo avançado é de t_i a t_{i-1} . Então, escreve-se 4.27 de acordo com o exemplo ilustrado na figura 4.4 como:

$$\frac{\partial J_o}{\partial \mathbf{x}_o} = \bar{\mathbf{d}}_0 + \mathbf{L}_0^T \{ \bar{\mathbf{d}}_1 + \mathbf{L}_1^T [\bar{\mathbf{d}}_2 + \mathbf{L}_2^T (\bar{\mathbf{d}}_3 + \mathbf{L}_3^T \bar{\mathbf{d}}_4)] \} \quad (4.29)$$

Das equações 4.25 e 4.27 ou 4.29 obtém-se então o gradiente da função custo e o algoritmo de minimização modifica apropriadamente a variável de controle $\mathbf{x}(t_0)$. Depois desta mudança, uma nova integração avançada e o novo incremento observacional é calculado e o processo é repetido.

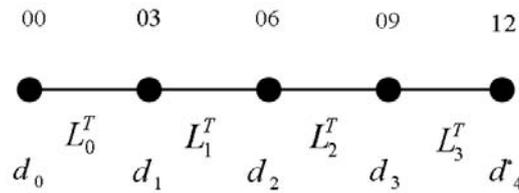


Figura 4.4 - Esquema do cálculo do gradiente da função custo relativo à observação para o período de 12 horas, observações a cada 3 horas e o modelo adjunto é integrado para trás dentro de cada intervalo.

Fonte: adaptada de Kalnay (2003, 183)

4.4 Modelo Tangente Linear e Modelo Adjunto

Considere um modelo não-linear. Uma vez que esse modelo tenha sido discretizado no espaço, por exemplo, usando diferenças finitas, o modelo pode ser escrito como um conjunto de n equações diferenciais acopladas:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}), \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{e} \quad \mathbf{F} = \begin{bmatrix} F_1 \\ \vdots \\ F_n \end{bmatrix} \quad (4.30)$$

Esse é o modelo na forma diferencial. Um modelo atmosférico consiste de um sistema de equações de diferença, que, por exemplo, usando o método de Crank-Nicholson podem ser descritas da forma:

$$\mathbf{x}^{n+1} = \mathbf{x}^n + \Delta t \mathbf{F} \left(\frac{\mathbf{x}^n + \mathbf{x}^{n+1}}{2} \right) \quad (4.31)$$

Uma solução numérica para 4.30 começando do tempo inicial t_0 pode ser obtida pela integração numérica do modelo usando 4.31 entre o tempo inicial e final. Isso fornece uma

solução do modelo não-linear que depende apenas das condições iniciais:

$$\mathbf{x}(t) = M[\mathbf{x}(t_0)] \quad (4.32)$$

em que M é a integração no tempo do método numérico da condição inicial ao tempo t . Uma pequena perturbação $\mathbf{y}(t)$ pode ser adicionada na integração do modelo $\mathbf{x}(t)$:

$$\begin{aligned} M[\mathbf{x}(t_0) + \mathbf{y}(t_0)] &= M[\mathbf{x}(t_0)] + \frac{\partial M}{\partial \mathbf{x}} \mathbf{y}(t_0) + O[\mathbf{y}(t_0)^2] \\ &= \mathbf{x}(t) + \mathbf{y}(t) + O[\mathbf{y}(t_0)^2] \end{aligned} \quad (4.33)$$

Em algum dado tempo, a evolução linear da perturbação $\mathbf{y}(t)$ será dada por:

$$\frac{d\mathbf{y}}{dt} = \mathbf{J}\mathbf{y} \quad (4.34)$$

em que $J = \frac{\partial \mathbf{F}}{\partial \mathbf{x}}$ é o Jacobiano de \mathbf{F} .

Esse sistema de equação diferencial ordinária é um modelo tangente linear na forma diferencial. A solução entre t_0 e t pode ser obtida por integração de 4.34 no tempo usando o mesmo método de diferença finita usado para o modelo não-linear 4.30, então

$$\mathbf{y}(t) = \mathbf{L}(t_0, t)\mathbf{y}(t_0) \quad (4.35)$$

Aqui $\mathbf{L}(t_0, t) = \frac{\partial M}{\partial \mathbf{x}}$ é uma matriz conhecida $m \times m$ como a matriz do modelo tangente linear, que propaga uma perturbação inicial do tempo t_0 até a tempo final t . Lorenz (1965) introduziu o conceito de modelo tangente linear de um modelo atmosférico, que o obteve diretamente de 4.33, desconsiderando os termos quadráticos de alta ordem na perturbação \mathbf{y} :

$$M[\mathbf{x}(t_0)] + \mathbf{L}(t_0, t)\mathbf{y}(t_0) = \mathbf{x}(t) + \mathbf{y}(t) \approx M[\mathbf{x}(t_0) + \mathbf{y}(t_0)] \quad (4.36)$$

Acrescentando um pequena perturbação de tamanho ε no vetor $\mathbf{y}_i(t_0) = \varepsilon \mathbf{e}_i$ e aplicando em 4.36 e subtraindo de 4.32 obtém-se a matriz que define o modelo tangente linear:

$$L(t_0, t)[\varepsilon e_1, \dots, e_n] = \varepsilon \mathbf{L}(t_0, t) = [\mathbf{y}_1(t), \dots, \mathbf{y}_n(t)] \quad (4.37)$$

A norma Euclidiana de um vetor é o produto interno do vetor com ele mesmo:

$$\|\mathbf{y}\|^2 = \mathbf{y}^T \mathbf{y} = \langle \mathbf{y}, \mathbf{y} \rangle \quad (4.38)$$

A norma Euclidiana de $\mathbf{y}(t)$ está relacionada à perturbação inicial por:

$$\|\mathbf{y}(t)\|^2 = (\mathbf{L}\mathbf{y}(t_0))^T \mathbf{L}\mathbf{y}(t_0) = \langle \mathbf{L}\mathbf{y}(t_0), \mathbf{L}\mathbf{y}(t_0) \rangle = \langle \mathbf{L}^T \mathbf{L}\mathbf{y}(t_0), \mathbf{y}(t_0) \rangle \quad (4.39)$$

O *adjunto* de um operador \mathbf{K} é definido pela propriedade $\langle \mathbf{x}, \mathbf{K}\mathbf{y} \rangle \equiv \langle \mathbf{K}^T \mathbf{x}, \mathbf{y} \rangle$. Neste caso, o *adjunto* do modelo tangente linear $\mathbf{L}(t_0, t)$ é simplesmente o *transposto* do modelo tangente linear.

Agora assumindo que se separa o intervalo de tempo (t_0, t) em dois intervalos de tempo sucessivos, por exemplo, se $t_0 < t_1 < t$, tem-se:

$$\mathbf{L}(t_0, t) = \mathbf{L}(t_1, t)\mathbf{L}(t_0, t_1) \quad (4.40)$$

Desde que o adjunto do modelo tangente linear é o transposto dele, a propriedade da transposta do produto também é válida, isto é:

$$\mathbf{L}^T(t_0, t) = \mathbf{L}^T(t_0, t_1)\mathbf{L}^T(t_1, t) \quad (4.41)$$

A equação 4.40 mostra que o modelo tangente linear pode ser calculado como um produto de matrizes do modelo tangente linear correspondendo a integrações curtas. A equação 4.41 mostra que o adjunto do modelo pode ser separado dentro de um único passo de tempo, mas ele é calculado para trás no tempo, começando do último passo de tempo t e terminando com o primeiro passo de tempo em t_0 . Para mais detalhes, consultar [Kalnay \(2003\)](#).

5 REDES NEURAIS

Algumas questões a respeito o cérebro humano são fatos desafiadores para a ciência. Questões do tipo: Como o cérebro processa informações? Como ele é organizado? Quais são os mecanismos envolvidos no funcionamento cerebral? Segundo [Arbib \(1987\)](#), citado por [Haykin \(2004\)](#), o sistema nervoso humano pode ser visto como um sistema de três estágios, como exemplificado no diagrama da figura 5.1. No entanto, o estudo do cérebro tornou-se mais fácil através do trabalho pioneiro dos neurologistas espanhóis Raymon e Cajal no ano de 1911, que introduziram a idéia dos *neurônios* como constituintes estruturais do cérebro, ou seja, os neurônios são unidades básicas de processamento do cérebro. Analogamente, modelos simplificados dos neurônios biológicos, os denominados *neurônios artificiais* constituem as unidades básicas de processamento das redes neurais artificiais (RNA).

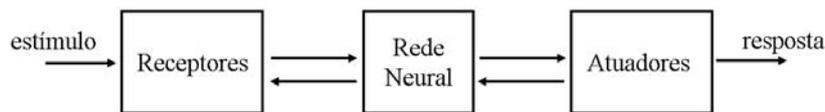


Figura 5.1 - Representação do diagrama em blocos do sistema nervoso.

O cérebro é especialista em desempenhar funções como reconhecimento de padrões, controle motor, percepção, inferência, intuições e etc. Entretanto, o cérebro também é impreciso, realiza generalizações incorretas e, acima de tudo, é geralmente incapaz de explicar suas próprias ações ([ZUBEN; CASTRO, 2000](#)). De acordo, com [Haykin \(2004\)](#) o cérebro é um computador (sistema de processamento de informações) altamente complexo, não linear e paralelo. Então, uma rede neural artificial é uma máquina que é projetada para modelar a maneira como o cérebro realiza uma tarefa particular ou uma função de interesse. O objetivo aqui está voltado para a classe das redes neurais artificiais que realizam computação útil através de um processo de *aprendizagem*. Portanto, uma definição para RN de acordo com [Haykin \(2004\)](#) é:

Uma RNA é um processador maciçamente paralelamente distribuído, constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso. Ela se assemelha ao cérebro em dois aspectos:

1. *O conhecimento é adquirido pela rede a partir de seu ambiente por meio de um processo de aprendizagem.*
2. *Forças de conexão entre os neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.*

O procedimento utilizado para realizar o processo de aprendizagem é denominado de algoritmo de aprendizagem, cuja função é modificar os pesos sinápticos da rede de uma forma ordenada para alcançar um objetivo de projeto desejado. Uma rede neural artificial é um arranjo de unidades caracterizado por:

- um conjunto de unidades de processamento (neurônios).
- um conjunto de pesos, que são as conexões entre as unidades de processamento, onde o conhecimento da rede é armazenado.
- altamente paralelo, controle distribuído.

5.1 Neurônio Biológico

Aqui descreve-se um neurônio biológico, ressaltando suas partes principais, com o objetivo de mostrar uma analogia entre a funcionalidade de um neurônio biológico com um neurônio artificial.

De acordo com (KOVACS, 1996) o neurônio ou célula nervosa ilustrado na figura 5.2 aparece de diferentes formas e tamanhos no cérebro humano, o qual é composto basicamente por dendritos, axônios, sinapses e núcleo ou soma, o qual contém os cromossomos. Os axônios tem a função de transmitir informações para outros neurônios, já os dendritos de receber informações em forma de sinais, que são pulsos elétricos conhecidos como impulsos *nervosos* ou *potenciais de ação*; as sinapses, grosso modo, pode-se dizer que são regiões eletroquimicamente ativas, compreendidas entre duas membranas celulares: a membrana *pré-sináptica*, por onde chega um estímulo proveniente de uma outra célula, e a membrana *pós-sináptica*, que é a membrana do dendrito. Nesta região intersináptica, o estímulo nervoso que chega à sinapse é transferido para a membrana dendrital através de substâncias conhecidas como *neurotransmissores*. O resultado dessa transferência é uma alteração no potencial elétrico da membrana *pós-sináptica*. Dependendo do tipo de neurotransmissor, a conexão sináptica será excitatória ou inibitória, sendo que uma conexão excitatória provoca uma alteração no potencial da membrana que contribui para a formação de um impulso nervoso no axônio de saída, enquanto uma conexão inibitória age no sentido oposto.

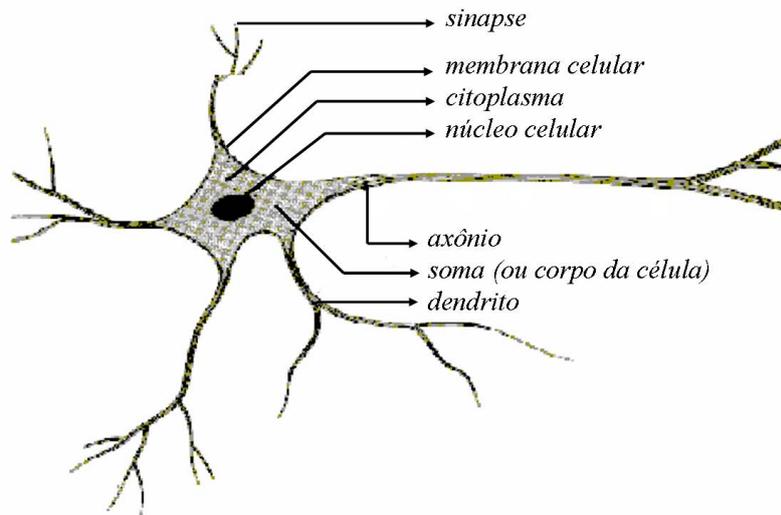


Figura 5.2 - Representação de um neurônio biológico.

5.2 Neurônio Artificial

Segundo Haykin (2004) um neurônio artificial é uma unidade de processamento de informações que é fundamental para a operação de uma rede neural. Pode-se identificar três elementos básicos do modelo neural artificial.

- Um conjunto de *sinapses*, cada uma das quais é caracterizada por um peso correspondente. Especificamente, um sinal z_i na entrada da sinapse i conectada ao neurônio k é multiplicado pelo peso sináptico θ_{ik} ;
- Um *somador* para somar os sinais de entrada, ponderados pelas sinapses respectivas de cada neurônio;
- Uma *função de ativação* para limitar a amplitude de saída do neurônio. Normalmente, a faixa de amplitude da saída de um neurônio é o intervalo $[0, 1]$ ou $[-1, 1]$.

No modelo de um neurônio artificial, exemplificado na figura 5.3, está incluído um limiar μ_k , que tem o efeito de acrescentar um grau de liberdade a cada neurônio. A unidade de processamento em uma rede neural é uma combinação linear com vários pesos de entrada, seguido por uma função de ativação. O k -ésimo neurônio pode ser descrito pelo o

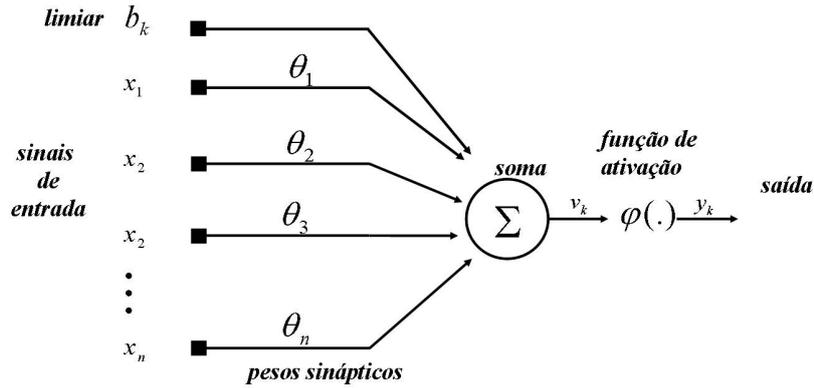


Figura 5.3 - Representação de um neurônio artificial.

acoplamento de duas equações:

$$u_k = \sum_{j=1}^m \theta_{kj} \mathbf{x}_j \quad (5.1)$$

$$\mathbf{y}_k = \varphi(u_k + b_k) \quad (5.2)$$

em que $\mathbf{x}_1, \dots, \mathbf{x}_m$ 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 5.2, conforme

$$v_k = u_k + b_k = \sum_{j=1}^m \theta_{kj} \mathbf{x}_j + b_k \quad (5.3)$$

5.2.1 Funções de ativação

A função $\varphi(\cdot)$ representa a função de ativação, a qual restringe a amplitude do sinal na saída de um neurônio. O papel da função de ativação é simular características não lineares de um neurônio biológico (ABELEM, 1994). A seguir, identificam-se três tipos básicos de função de ativação:

1. *função degrau*: é uma função utilizada para valores binários. O primeiro neurônio usando um dispositivo binário foi introduzido por McCulloch e Pitts (1943), no qual a saída é pulso ou não pulso, sendo que suas entradas têm 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 inibitório-

rios. Se esse resultado for maior ou igual a certo limiar, então a saída do neurônio é pulso, e caso contrário é não pulso. Matematicamente a função é representada por:

$$\varphi(v) = \begin{cases} 1 & \text{se } v > 0 \\ 0 & \text{se } v \leq 0 \end{cases} \quad (5.4)$$

2. *função tangente hiperbólica*: é uma função sigmóide que varia entre $[-1, 1]$, a qual foi utilizada na implementação da camada oculta PMC, é dada por:

$$\varphi(v_j) = \tanh\left(\frac{av_j}{2}\right); \text{ para } a = 1 \quad (5.5)$$

3. *função exponencial*

$$\varphi(v_j) = \exp\left(-\frac{(v_j - \mu)^2}{2\sigma^2}\right); \text{ com } \sigma = 1, \mu = 0 \quad (5.6)$$

O valores nas entradas devem ser normalizados 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.

5.3 Funcionamento das Redes Neurais

Uma rede neural artificial, segundo [Fausett \(1994\)](#), é caracterizada por:

1. o padrão de conexão entre os neurônios, o que determina a arquitetura da rede.
2. o método de determinação dos pesos sobre as conexões (chamado de treinamento ou algoritmo de aprendizagem).
3. a função de ativação

Os neurônios podem ser dispostos de tal maneira a formar uma ou mais camadas. Neste texto, as entradas da rede são denominadas de camada de entrada, a primeira camada sem contato com o meio externo é chamada de camada escondida e os valores na saída da rede são chamados de camada de saída.

Nas redes onde o sinal de entrada projetam-se diretamente sobre a camada de saída, a rede é denominada de *feed-forward* (alimentação para frente). Quando há uma realimentação entre uma das camadas da rede, ela é chamada de rede recorrente.

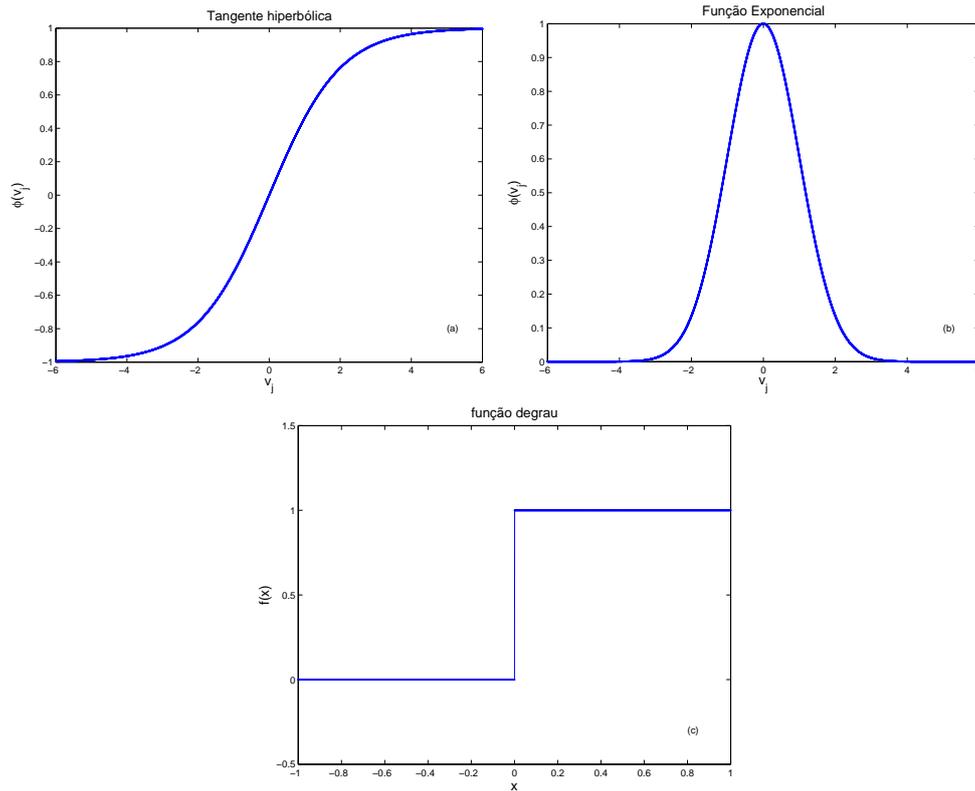


Figura 5.4 - Exemplos de funções de ativação mais utilizadas: (a) tangente hiperbólica; (b) exponencial; (c) função degrau.

O funcionamento de uma rede neural é realizado através de duas fases: a fase de aprendizagem, também conhecido como treinamento, e a fase de ativação. No processo de aprendizagem um conjunto de exemplos são apresentados à rede, a qual extrai as características necessárias para representar a informação fornecida. Essas características são armazenadas nos pesos sinápticos, os quais são utilizados posteriormente no processo de ativação da rede, gerando assim respostas para o problema. Neste trabalho a resposta fornecida pela rede é o sinal estimado.

Um fato importante na fase de treinamento das RNA é adotar um critério de parada. Normalmente, a rede é treinada até que um número máximo de “*épocas*” de treinamento, definido *a priori*, seja atingido. Defini-se uma *época de treinamento* como sendo a apresentação de todo um conjunto (ou padrões) de treinamento da rede. O critério de parada em função do erro objetivo pode ser aplicado com paradigma de aprendizagem supervisionada, o qual defini-se mais adiante. Uma definição de aprendizagem, adaptada de Mendel e McLaren (1970), no contexto de redes neurais disponível em (HAYKIN, 2004) é:

Aprendizagem é um processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimulação pelo ambiente no qual a rede está inserida. O

tipo de aprendizagem é determinado pela maneira pela qual a modificação dos parâmetros da rede ocorre.

5.4 Regras de Aprendizagem

Algoritmo de aprendizagem é um conjunto de procedimentos bem definidos para adaptação dos parâmetros de uma rede neural para que a mesma possa aprender uma determinada função. Os tipos de aprendizagem são: supervisionados ou não-supervisionados. Na aprendizagem supervisionada, tanto a entrada quanto a saída da rede são conhecidas, sendo que a rede opera no sentido de encontrar pesos que minimizem a diferença entre a entrada e a saída desejada em um sentido estatístico, geralmente o erro médio quadrático. Já na aprendizagem não-supervisionada somente os padrões de entrada estão disponíveis na rede. 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 usualmente aplicadas para o processo de aprendizagem supervisionado são: correção de erros e a regra delta, o qual foi generalizado para o treinamento do perceptron de múltiplas camadas, conhecido como algoritmo de retropropagação do erro. Na seção a seguir, descreve-se o algoritmo de retropropagação segundo (HAYKIN, 2004).

5.5 Algoritmo Backpropagation ou de Retropropagação

1. **Iniciar os pesos:** assumindo que nenhuma informação prévia está 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. **Propagação:** sendo um exemplo do conjunto de treinamento representado por $(x(n), d(n))$, com um vetor de entrada $x(n)$ aplicado a camada de entrada de nós sensoriais e o vetor resposta desejada $d(n)$ apresentado a 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 dado por:

$$v_j^{(l)}(n) = \sum_{i=0}^{m_o} \theta_{ji}^{(l)}(n) y_j^{(l-1)}(n) \quad (5.7)$$

em que $y_j^{(l-1)}(n)$ é o sinal de saída do neurônio i na camada anterior $l - 1$, 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 $y_{j0}^{(l)}(n) = b_{j0}^{(l)}(n)$ é o vies aplicado ao neurônio j na camada l . O sinal de saída do neurônio j na camada l é:

$$y_j^l = \varphi(v_j(n)) \quad (5.8)$$

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 na camada de saída (i.e., $l = L$ em que L é denominado a profundidade da rede), se faz $y_j^{(L)} = o_j(n)$.

Calcula-se o sinal do erro

$$e_j^{(n)} = d_j^{(n)} - o_j^{(n)} \quad (5.9)$$

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

4. **Retropagação:** Calcula-se os gradientes locais da rede definidos por:

$$\delta_j^{(l)}(n) = \begin{cases} e_j^{(L)}(n) \varphi_j^{(\prime)}(v_l^{(L)}(n)) & \text{para o neurônio } j \text{ na camada de saída } L, \\ \varphi_j^{(\prime)}(v_l^{(L)}(n)) \sum_k \delta_k^{(l+1)}(n) \theta_{kj}^{(l+1)}(n) & \text{para o neurônio } j \text{ na camada oculta } l. \end{cases} \quad (5.10)$$

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

$$\theta_{ji}^{(n+1)} = \theta_{ji}^{(n)} + \alpha[\theta_{ji}^{(n)} - \theta_{ji}^{(n-1)}] + \eta \delta_j^{(l)}(n) y_i^{(l-1)}(n) \quad (5.11)$$

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

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

5.6 Perceptron de Múltiplas Camadas

Em uma rede multicamadas são identificadas a camada de entrada, uma ou mais camadas ocultas e a camada de saída de nós computacionais, ilustrada na figura 5.5. A camada de

entrada é formada por padrões que serão apresentados à rede durante a fase de treinamento (também chamados de conjuntos sensoriais ou nós de fonte). O sinal de entrada se propaga para frente através da rede, camada por camada. Segundo Haykin (2004), a PMC tem três características distintas:

- o modelo de cada neurônio da rede inclui uma função de ativação não linear, sendo que esta função de ativação deve ser diferenciável em todos os pontos;
- a rede contém uma ou mais camadas de neurônios ocultos. Estes neurônios ocultos capacitam a rede a aprender tarefas complexas extraindo progressivamente as características mais significativas dos padrões de entrada;
- a rede exibe alto grau de conectividade, determinada por sua sinapses.

Os valores na entrada devem ser normalizados 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.

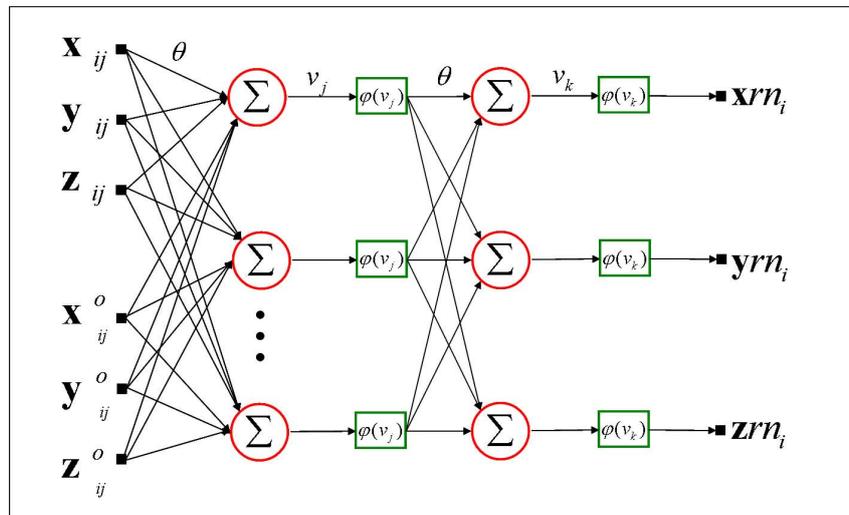


Figura 5.5 - Ilustração da Rede Perceptron de Múltiplas Camadas

6 RESULTADOS NUMÉRICOS

Neste capítulo, são apresentados os resultados obtidos com as técnicas de assimilação de dados abordadas neste trabalho. Essas técnicas foram aplicadas ao sistema dinâmico de Lorenz, que tem sido utilizado por muitos pesquisadores como um problema teste, por ser um sistema que possui reconhecidas similaridades com o fluido atmosférico. Apresenta-se o sistema de Lorenz e em seguida os resultados com as respectivas técnicas de assimilação.

6.1 Sistema de Lorenz

Lorenz (1963) percebeu o fato de que a atmosfera, como qualquer sistema dinâmico com instabilidades, tem um limite finito de previsibilidade, que ele estimou ser em torno de duas semanas, mesmo se o modelo de previsão seja perfeito e as condições iniciais sejam conhecidas quase que perfeitamente. A partir de seus experimentos numéricos, identificou que apesar do sistema ser determinístico, era impossível prever com precisão o tempo no futuro baseado em informações no presente a partir de medidas tomadas com limite finito de precisão. Então, Lorenz revelou o teorema fundamental da previsibilidade (KALNAY, 2003): *Sistemas instáveis têm um limite finito de previsibilidade, enquanto, os sistemas estáveis são infinitamente previsíveis, desde que eles sejam estacionários ou periódicos.*

O movimento de uma camada de fluido, como o da atmosfera terrestre, que é mais quente na base do que no topo, é um importante problema em meteorologia e em outras aplicações da dinâmica dos fluidos. Se a diferença de temperatura vertical ΔT for pequena, há uma variação linear de temperatura com a altura, mas não há movimento convectivo da camada fluida. No entanto, se ΔT for suficientemente grande, o ar mais quente se eleva, deslocando o ar mais frio que está por cima e provoca um movimento convectivo permanente. Se a diferença de temperatura aumentar ainda mais, o escoamento convectivo permanente é perturbado e se instala um movimento turbulento, mais complicado (BOYCE; DIPRIMA, 1998).

O matemático e meteorologista Edward N. Lorenz em 1963, investigou esse problema e propôs um sistema de equações para modelar esse fenômeno. Esse sistema é relativamente simples, mas resultam em um padrão de complexidade infinita: o Atrator de Lorenz (LORENZ, 1963), ilustrado na Figura 6.1.

O fluido aquecido formado na parte mais baixa tende a elevar-se e possuir um movimento cilíndrico, criando correntes de convecção. Se o calor é suficientemente intenso, a convecção acontece de uma maneira turbulenta e irregular. Esse fenômeno também ocorre na atmosfera terrestre e, baseando-se na propriedade de sensibilidade das soluções das equações de Lorenz (veja Figura 6.2), em que uma pequena alteração em relação às condi-

ções iniciais pode gerar grandes e drásticas mudanças no comportamento do sistema. Esse fenômeno ilustra a impossibilidade de se fazer previsão do tempo a longo prazo. O modelo de Lorenz é governado por um sistema acoplado de equações diferenciais ordinárias de primeira ordem, não-linear de três variáveis. Essas equações foram derivadas das equações de Saltzman (1962),

$$\begin{aligned}\frac{dx}{d\tau} &= \sigma(y - x) \\ \frac{dy}{d\tau} &= \rho x - y - xz \\ \frac{dz}{d\tau} &= xy - \beta z\end{aligned}\tag{6.1}$$

em que σ , ρ , β são parâmetros do modelo, neste caso correspondem aos seguintes valores 10, 28 e $8/3$, respectivamente. A solução numérica deste modelo resulta no atrator ilustrado pela Figura 6.1.

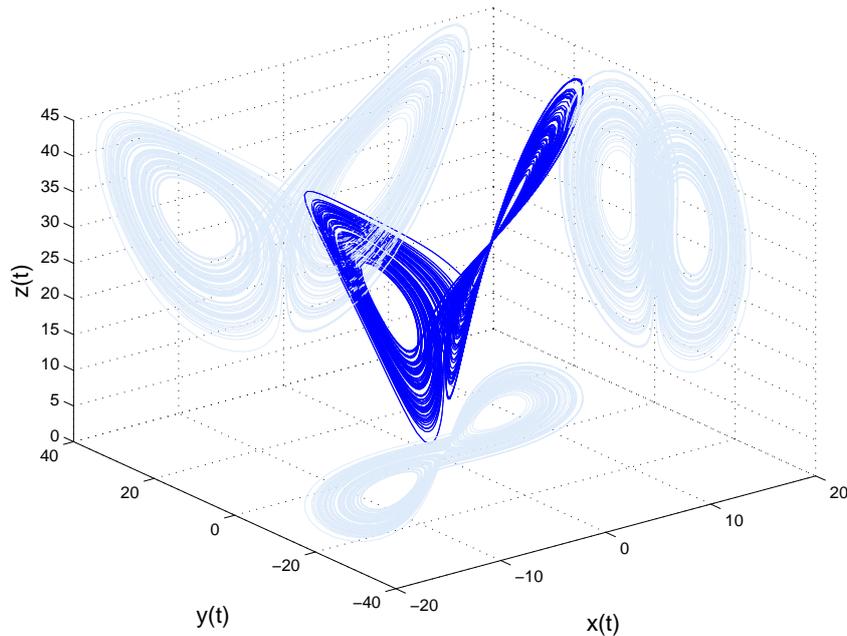


Figura 6.1 - Atrator de Lorenz caótico com condições iniciais $(-1, 5; 1, 5; 20)$ e sua projeções nos eixos xy , xz , e yz .

Nas equações 6.2 as componentes x , y e z são dependentes do tempo para $\tau = \pi^2 H^{-2}(1 + a^2)\kappa t$ sendo que H é a altura (ou profundidade) da camada, a é o número de onda, κ é a condutividade térmica e t é o tempo. Os parâmetros σ , ρ , e β correspondem à $\sigma = \kappa^{-1}\nu$ é o número de Prandtl; $\rho = R_c^{-1}R_a$ é o número de Rayleigh com $R_c = \pi^4 a^{-2}(1 + a^2)^3$ e $R_a = g\alpha H^3 \delta T \nu^{-1} \kappa^{-1}$ sendo que as constantes g , α e ν representam,

respectivamente, a aceleração da gravidade, o coeficiente de expansão térmica e viscosidade térmica, ressaltando que o valor mínimo para R_c é $\frac{27\pi^4}{4}$ quando $a^2 = \frac{1}{2}$; e por fim $\beta = 4(1 + a^2)^{-1}$.

As componentes x , y e z para esse sistema correspondem a: x é proporcional à intensidade do movimento convectivo, y é proporcional a diferença de temperatura entre as correntes ascendentes e descendentes e z é proporcional ao perfil de temperatura vertical. Os sinais similares de x e y significa que o fluido quente está em movimento ascendente e o fluido frio está em movimento descendente. A componente z é proporcional a perturbação do perfil de temperatura vertical, sendo que valores positivos indicam forte gradiente próximo à fronteira (LORENZ, 1963).

O sistema de Lorenz foi discretizado com o método de Runge Kutta de 2a. ordem com passo numérico $\Delta t = 10^{-3}$, o qual é dado pelo conjunto de equações a seguir:

$$\begin{aligned}
 x_{n+1} &= x_n + \frac{\sigma \Delta t}{2} [2(y_n - x_n) - \sigma \Delta t (y_n - x_n) + \Delta t (\rho x_n - y_n - x_n z_n)] \\
 y_{n+1} &= y_n + \frac{\Delta t}{2} [\rho x - y - xz + \rho(x + \sigma \Delta t (y - x)) - y - \Delta t (\rho x - y - xz) \\
 &\quad - (x + \sigma \Delta t (y - x))(z + \Delta t (xy - \beta z))] \\
 z_{n+1} &= z_n + \frac{\Delta t}{2} (xy - \beta z + (x + \sigma \Delta t (y - x))(y + \Delta t (\rho x - y - xz)) - \\
 &\quad \beta(z + \Delta t (xy - \beta z)))
 \end{aligned} \tag{6.2}$$

Na Figura 6.2, tem-se as séries temporais das componentes x e z . O gráfico azul foi gerado

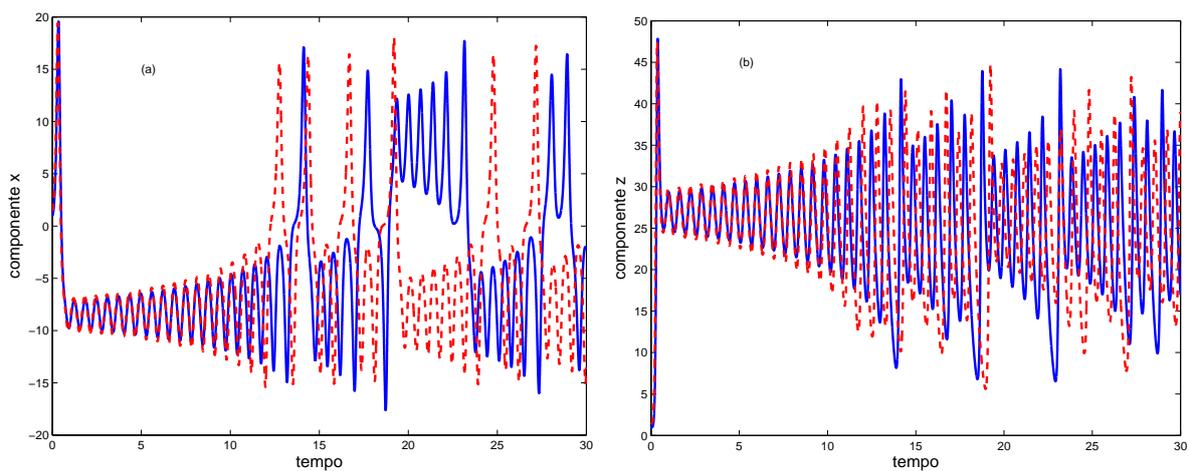


Figura 6.2 - Figuras exemplificando a sensibilidade das condições iniciais do modelo de Lorenz; figura (a): série temporal da componente x , curva contínua $x(0) = 1$ e a curva tracejada $x(0) = 1,05$; figura (b): série temporal da componente z , curva contínua $z(0) = 1$ e a curva tracejada $z(0) = 1,05$.

com a condição inicial $x(0) = 1$, enquanto o gráfico vermelho foi gerado adicionando-se uma pequena perturbação $\delta x = 0,05$ na condição inicial. Nota-se que com o passar do tempo, ocorre o desacoplamento das dinâmicas das séries para ambas as componentes do sistemas, aqui ilustrado para a componente x e z .

A seguir, são apresentados os desempenhos das técnicas de assimilação: Filtro de Kalman, Filtro de Partículas, Método Variacional e Redes Neurais Artificiais.

6.2 Resultados com a Rede Neural emulando o Filtro de Kalman

Para a implementação do Filtro de Kalman foi usado as seguintes matrizes: $\mathbf{Q}_t = 0,1\mathbf{I}$; $\mathbf{R}_t = 0,5\mathbf{I}$; $\mathbf{H} = \mathbf{I}$; matriz de covariância do erro de modelagem, matriz de covariância do erro de observação e o operador que representa o sistema de observação, respectivamente. A matriz de covariância do erro de previsão é inicializada por:

$$\mathbf{P}_0^f = \begin{cases} 10(x_0^f)_i^2 & \text{para } i = j \\ 0 & \text{para } i \neq j \end{cases} \quad (6.3)$$

As observações assimiladas neste trabalho foram geradas artificialmente. Os dados foram gerados a partir da integração do sistema de Lorenz, adicionando-se um ruído aleatório de variância 0.5. Assim, a curva de referência para o método de assimilação é a curva obtida a partir da integração do sistema de Lorenz sem ruído, o que significa que quanto mais próximo as estimativas obtidas com os métodos de assimilação estiverem da dinâmica do sistema, melhor será a estimativa obtida por eles.

Os métodos foram testados variando-se a frequência com que as observações são inseridas no sistema. Notou-se que quanto mais observações são inseridas, melhor é a qualidade da estimativa (veja tabela 6.2).

Para o método proposto de assimilação de dados, que emprega Redes Neurais Artificiais, a determinação do dado de análise (condição inicial) pode ser descrita pela seguinte equação (HARTER, 2004)

$$\mathbf{x}_n^a = F_{RN}(\mathbf{x}_n^p, \mathbf{y}_n^o) \quad (6.4)$$

onde F_{RN} representa a RNA descrita no Capítulo 5. Conforme visto naquele capítulo a rede Perceptron de Múltiplas Camadas é uma rede que exige o treinamento supervisionado, isto é, que seja apresentado a ela a saída desejada, ou o denominado “alvo” da rede. Logo, a rede foi treinada com os seguintes “alvos”: as estimativas do Filtro de Kalman, resultados apresentados nesta seção; Filtro de Partículas (Seção 6.3) e Método Variacional (Seção 6.4). A Figura 6.3 mostra os conjuntos de padrões que são necessários para o treinamento da rede. A fase de treinamento da rede consiste em determinar os melhores

pesos que mapeia os dados de entrada aos dados da saída desejada (“alvo” da rede).

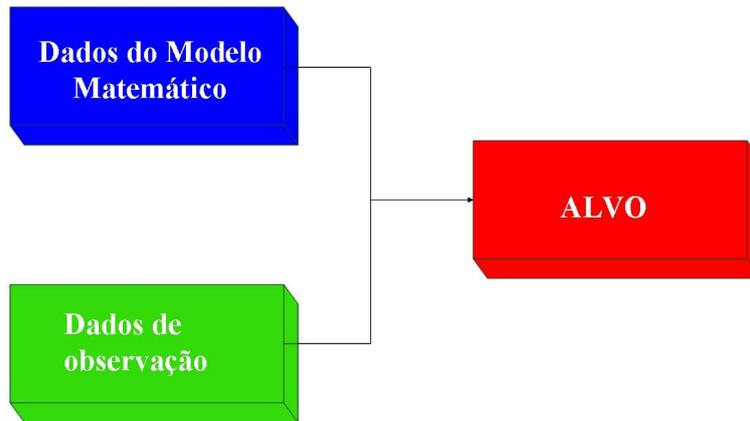


Figura 6.3 - Diagrama esquemático ilustrando a fase de treinamento da RNA. São apresentados os padrões de treinamento: dados do modelo matemático, dados de observação e a saída desejada, ou o “alvo” da rede, que é a estimativa obtida pelo FK, FP e o Método Variacional.

A rede é treinada com 2000 exemplos e 333 dados são usados na validação cruzada. Esta quantidade de padrões foi determinada empiricamente por Harter (2004) e Nowosad (2001), sendo considerada uma quantidade de padrões necessários e suficientes para que a rede possa “aprender”. O erro de treinamento é dado por:

$$ET = \frac{1}{2000} \sum_{k=1}^{2000} \frac{1}{2} \sqrt{(x_k^{alvo} - x_k^{rede})^2} \quad (6.5)$$

Os 333 dados para a validação cruzada foram considerados, segundo Harter (2004), uma quantidade suficiente para analisar graficamente se a rede está conseguindo generalizar, isto é, obter a melhor estimativa com os dados não pertencentes ao conjunto de treinamento. O erro para a validação cruzada é dado por:

$$EV = \frac{1}{333} \sum_{k=1}^{333} \sqrt{(x_k^{alvo} - x_k^{rede})^2} \quad (6.6)$$

A arquitetura da rede utilizada foi composta por uma camada de entrada com 3 neurônios, uma camada oculta com 2 neurônios e uma camada de saída com 3 neurônios, conforme ilustrado na Figura 6.4. A função de ativação implementada nos neurônios foi a função tangente hiperbólica, $\varphi(x) = \tanh(x)$. O algoritmo de treinamento utilizado foi o “back-propagation” ou algoritmo de retropropagação do erro, descrito no Capítulo 5. Os parâmetros

usados na implementação da rede estão apresentados na [Tabela 6.1](#), com constantes de momento α_h na camada escondida e α_β ($\beta = x, y, z$) na camada de saída. De modo análogo, para a taxa de aprendizagem tem-se, η_h na camada oculta e η_β ($\beta = x, y, z$) na camada de saída. A taxa de aprendizagem e constante de momentos são parâmetros da

Tabela 6.1 - Parâmetros da RNA

neurônios	α_h	α_x	α_y	α_z	η_h	η_x	η_y	η_z
2	0,6	0,6	0,6	0,6	0,001	0,001	0,001	0,001

rede, que devem ser ajustados para se determinar pesos precisos, que são calculados de acordo com a equação [5.11](#), repetida a seguir por conveniência.

$$\theta_{ji}^{(n+1)} = \theta_{ji}^{(n)} + \alpha[\theta_{ji}^{(n)} - \theta_{ji}^{(n-1)}] + \eta\delta_j^{(l)}(n)y_i^{(l-1)}(n) \quad (6.7)$$

De acordo com [Haykin \(2004\)](#), quanto menor for a taxa de aprendizagem η , menor serão as variações dos pesos sinápticos (θ) da rede, de uma iteração para a outra, o que pode acarretar a uma taxa de aprendizagem lenta. Por outro lado, se a taxa de aprendizagem for um valor grande, acelerando o processo, ocasiona grandes modificações nos valores dos pesos, tornando a rede instável. Um método simples de aumentar a taxa de aprendizagem, evitando o perigo de instabilidade é usar a constante de momento α , que geralmente varia entre 0,5 e 0,9. Para o problema apresentado neste trabalho, os valores contidos na [tabela 6.1](#) proporcionaram resultados satisfatórios.

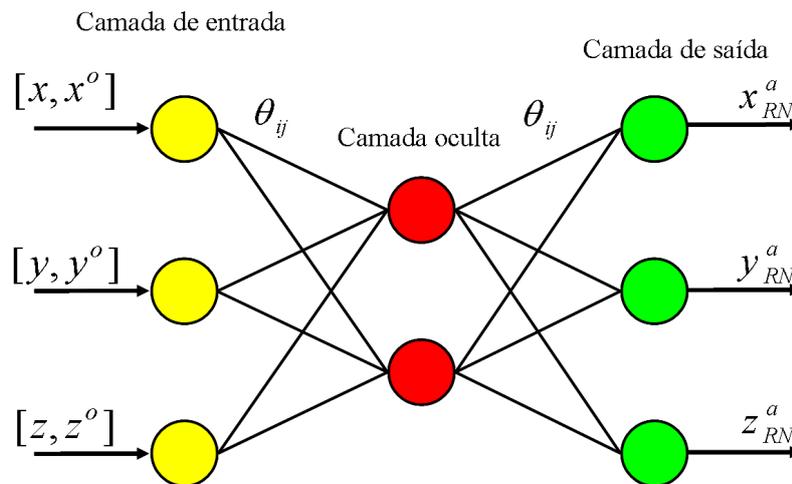


Figura 6.4 - Figura esquemática ilustrando o modelo da rede neural com uma camada oculta, contendo 2 neurônios. Para a rede *feed-forward*, a informação é propagada dos neurônios da camada de entrada para os neurônios na camada de saída.

O algoritmo do Filtro de Kalman no contexto de assimilação de dados é mostrado no diagrama esquemático apresentando na Figura 6.5. O modelo de Lorenz dado pelo sistema de equações 6.2, linearizado através de funções de série de Taylor de primeira ordem, resulta na seguinte matriz:

$$\mathbf{F} = \begin{bmatrix} 1 - \sigma\Delta t & \sigma\Delta t & 0 \\ \Delta t(\rho - z) & 1 - \Delta t & -x\Delta t \\ y\Delta t & x\Delta t & 1 - \beta\Delta t \end{bmatrix} \quad (6.8)$$

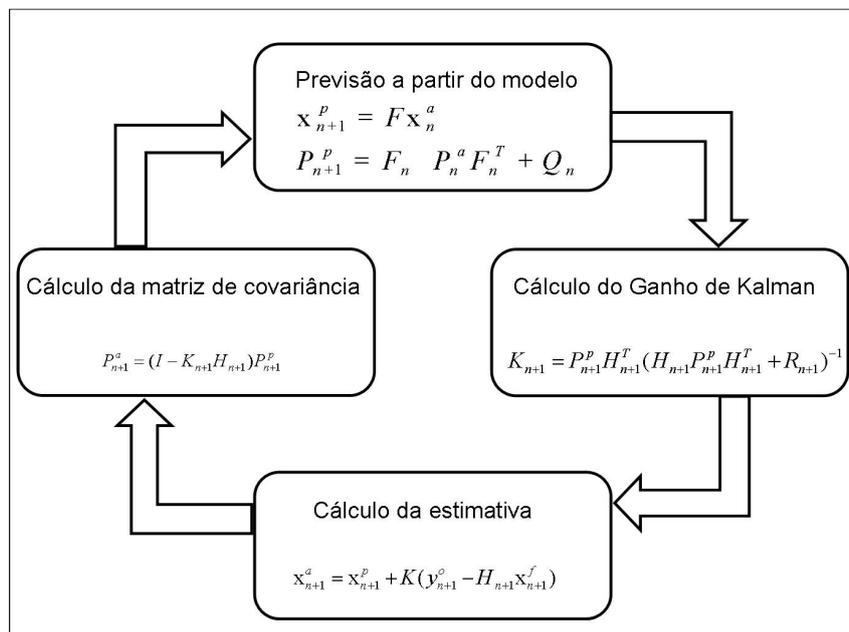


Figura 6.5 - Diagrama esquemático do Filtro de Kalman Linear.

Fonte: (TODLING, 1999).

Uma vez feita a linearização do modelo, o Filtro de Kalman apresentado na Figura 6.5 pode ser aplicado. Na Figura 6.6 tem-se as estimativas das séries temporais x , y e z , resultado da assimilação realizada com o FK e com o Perceptron de Múltiplas Camadas. As observações foram inseridas a cada 50 passos de tempo; a curva tracejada (vermelha) é a estimada com o FK e a curva cheia (azul) é a referência de verdade. Na coluna à direita, tem-se o resultado da assimilação com RNA “emulando” o FK, representada pela curva tracejada (verde), nota-se que a rede obteve um boa estimativa.

O FK é uma técnica eficiente ao realizar assimilação de dados. Uma vantagem desta técnica em relação as apresentadas no Capítulo 2 é que, além de estimar as matrizes de

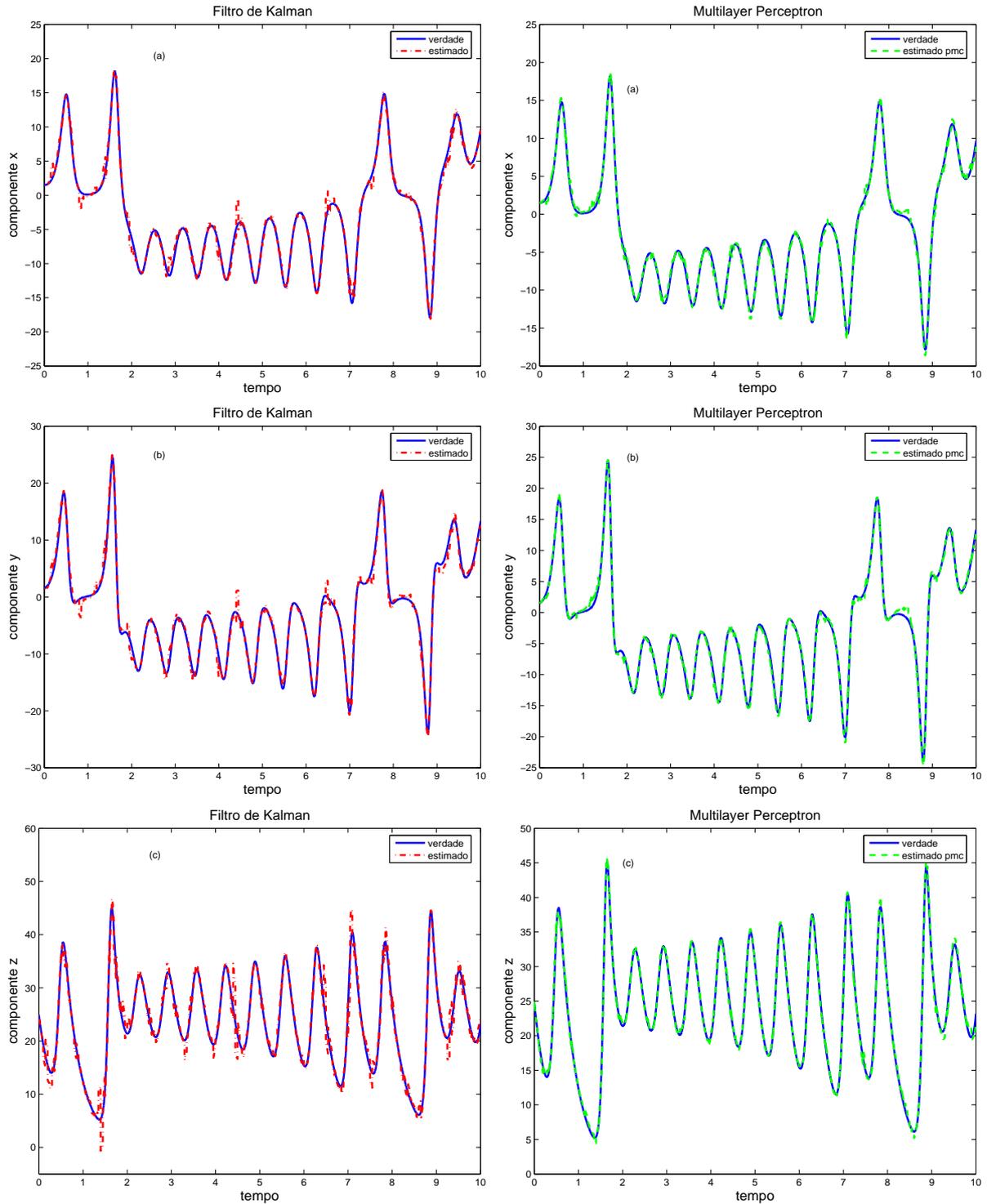
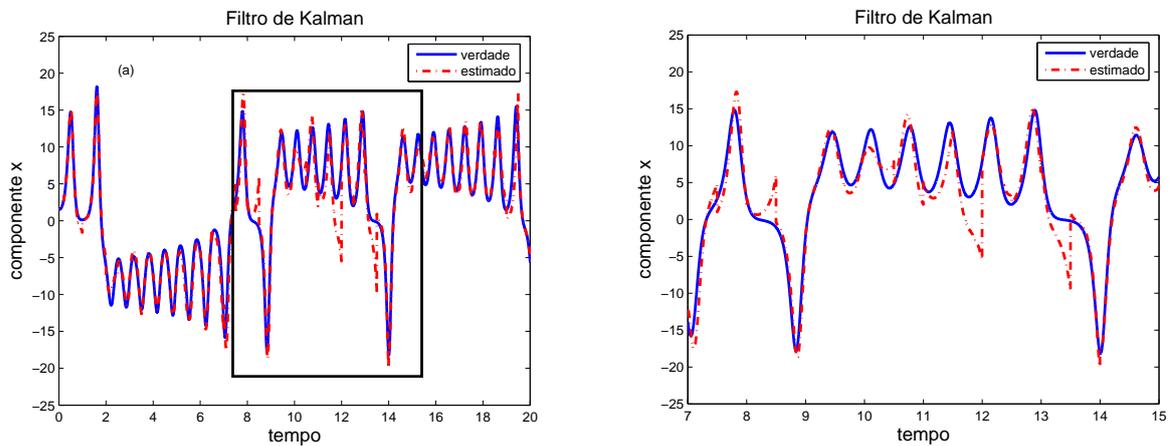


Figura 6.6 - Assimilação de dados com o Filtro de Kalman e com RNA Perceptron de Múltiplas Camadas. As observações foram inseridas a cada 50 passos de tempo. Gráficos à direita: curva tracejada (vermelha) estimado pelo FK, curva cheia (azul) referência de verdade. Analogamente, gráficos à esquerda, curva tracejada (verde) estimado pela RNA, curva cheia (azul) referência de verdade.

covariância do erro de propagação e de análise, ele as atualiza a cada passo de tempo, fornecendo assim, uma melhor estimativa. No entanto, o seu algoritmo sofre desvantagem

em relação a complexidade computacional, devido às várias operações com matrizes, como por exemplo, multiplicação e inversa, que são realizadas a cada iteração, tornando-o um método computacionalmente caro.

As Figuras 6.7 a 6.10 mostram a assimilação realizada a cada 500 passos de tempo. Em 6.7 e 6.8 apresenta a evolução temporal da componente x , estimativas obtidas com o Filtro de Kalman e a RNA Perceptron, respectivamente. Analogamente, em 6.9 e 6.10 mostra a evolução temporal da componente z . A estimativa obtida com o Filtro de Kalman, neste caso, pode-se ver que sua trajetória (curva tracejada vermelha) na região demarcada pelo retângulo e ampliada na Figura 6.7(a), tem-se que a dinâmica contém um certo desacoplamento da curva de referência. Isso ocorre para ambas as componentes do sistema de Lorenz, aqui é ilustrado para a componente x em 6.7 e componente z em 6.9.



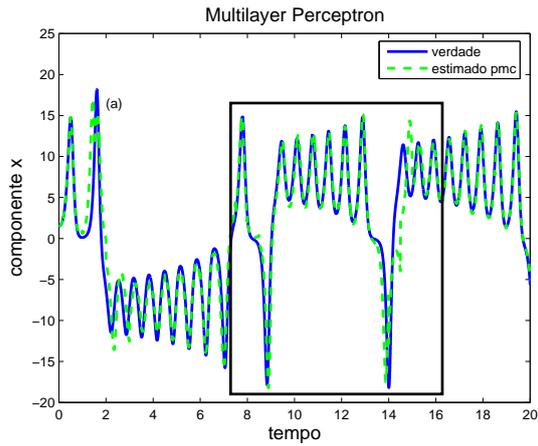
(a) Evolução temporal da componente x .

(b) Ampliação da região do retângulo.

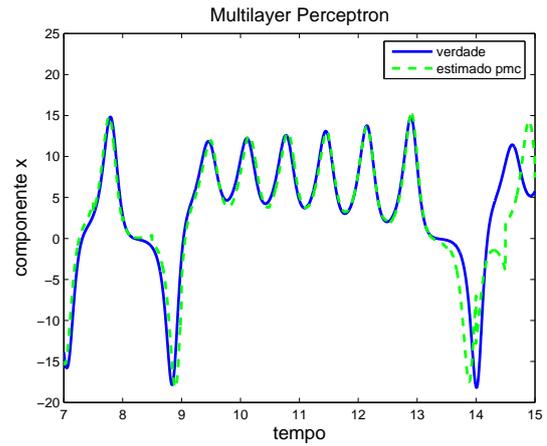
Figura 6.7 - Assimilação de dados a cada 500 passos de tempo, estimativa obtida com o FK: curva tracejada (vermelha) estimado, curva cheia (azul) referência de verdade.

No entanto, a assimilação com a RNA Perceptron de Múltiplas Camadas mostrado nas Figuras 6.8 e 6.10, nota-se claramente que a sua estimativa, neste caso, foi mais eficiente que a obtida pelo Kalman, conforme Figuras 6.8(b) e 6.10(b) para as componentes x e z , respectivamente, onde ocorre o afastamento da dinâmica apenas no instante $t \approx 14$.

Foram feitos experimentos variando a frequência com que as observações são inseridas no sistema, de acordo com a Tabela 6.2. Esta contém os resultados dos erros individuais para cada componente do sistema de Lorenz e a média global, que é ilustrada na Figura 6.11, o gráfico vermelho é a média do erro obtido com Kalman e o gráfico azul é o obtido com a RNA, nota-se que a RNA obteve menor erro. Para observações inseridas a cada 250

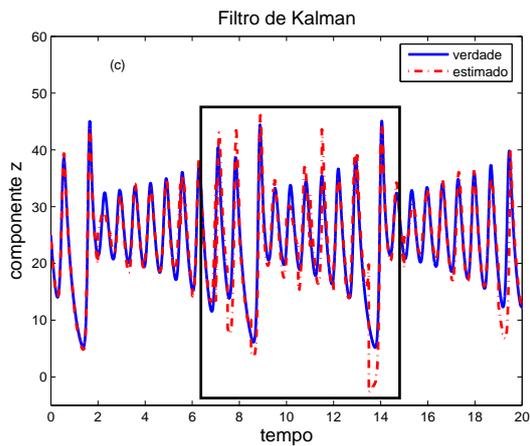


(a) Evolução temporal da componente x .

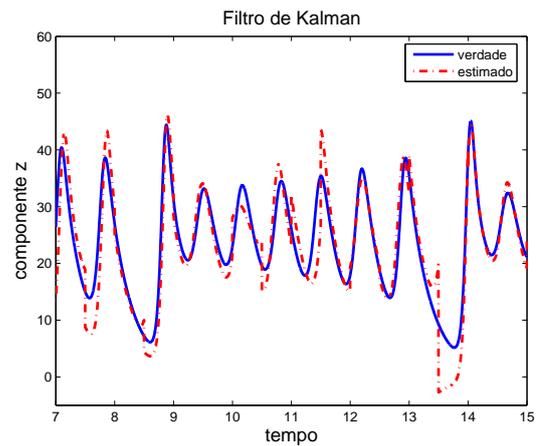


(b) Ampliação da região do retângulo.

Figura 6.8 - Assimilação de dados a cada 500 passos de tempo, estimativa obtida com o Perceptron de Múltiplas Camadas: curva tracejada (verde) estimado, curva cheia (azul) referência de verdade.



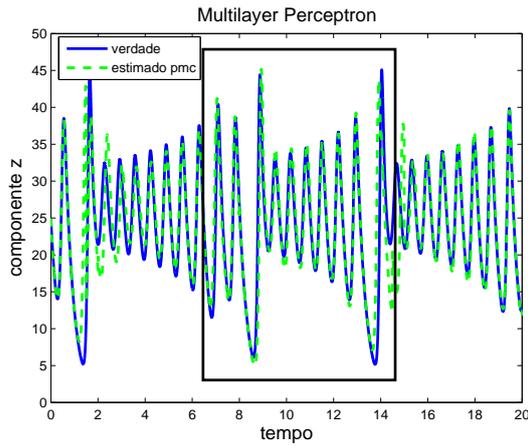
(a) Evolução temporal da componente z .



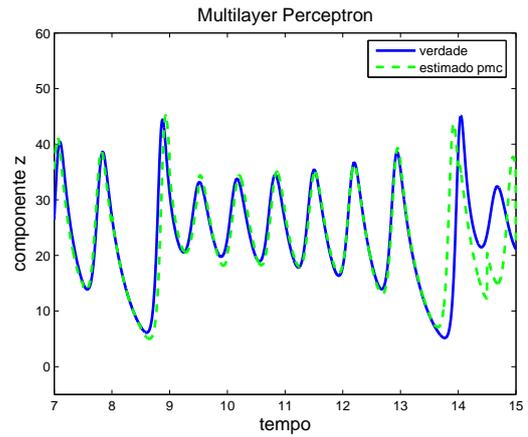
(b) Ampliação da região do retângulo.

Figura 6.9 - Assimilação de dados a cada 500 passos de tempo, estimativa obtida com o Perceptron de Múltiplas Camadas: curva tracejada (verde) estimado, curva cheia (azul) referência de verdade.

passos de tempo o erro foi compatível com o resultado do Filtro de Kalman.



(a) Evolução temporal da componente z .



(b) Ampliação da região do retângulo.

Figura 6.10 - Assimilação de dados a cada 500 passos de tempo, estimativa obtida com o Perceptron de Múltiplas Camadas: curva tracejada (verde) estimado, curva cheia (azul) referência de verdade.

Tabela 6.2 - Erros de Assimilação: Filtro de Kalman e RNA Perceptron.

freq	FK				RNA PMC			
	x	y	z	$(x + y + z)/3$	x	y	z	$(x + y + z)/3$
12	0,3391	0,3862	0,9757	0,5670	0,3781	0,3423	0,3850	0,3685
50	0,4437	0,5802	1,2510	0,7583	0,3851	0,4060	0,3941	0,3951
100	0,4758	0,7821	1,2682	0,8420	0,3820	0,5070	0,5297	0,4729
250	0,8102	1,3174	1,4309	1,1862	0,9171	1,3602	1,2934	1,1902
500	1,9877	3,0699	2,9388	2,6655	2,1839	2,9151	2,8290	2,5745

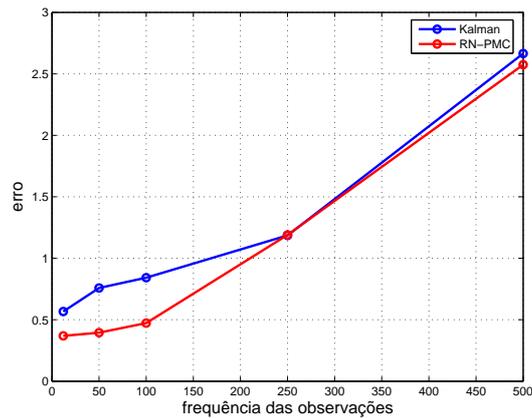


Figura 6.11 - Média global do erro $(x + y + z)/3$ de assimilação para as técnicas do FP e RNA Perceptron de Múltiplas camadas (PMC), eixo x é a frequência das observações e o eixo y é o erro.

6.3 Resultados com a Rede Neural emulando o Filtro de Partículas

Nesta seção são apresentados os resultados da assimilação de dados com o Filtro de Partículas, que foi descrito no [Capítulo 3](#), e a RNA “emulando” esta técnica. O Filtro de Partículas pode ser pensado como uma técnica que se inicializa com um conjunto de condições iniciais, denominadas partículas, e o método selecionará aquelas que produzem trajetórias mais próximas da trajetória de referência. Em oposição ao Filtro de Kalman, o FP trabalha com o problema na sua forma não-linear, fornecendo uma solução aproximada para o problema real. Na implementação do Filtro, foram usadas 1000 partículas (amostras) aleatórias, que foram amostradas diretamente do espaço de estados do sistema de Lorenz. Considerando-se que o ruído nas observações é ruído branco com variância $\sigma = 0,5$, tem-se que a distribuição de probabilidade da observação dado a variável de estado do sistema é dado por:

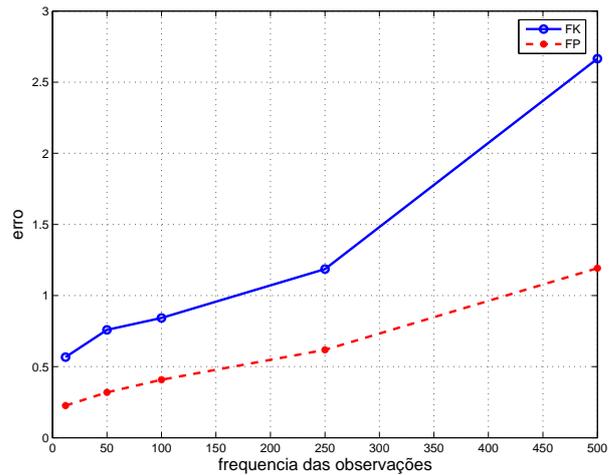
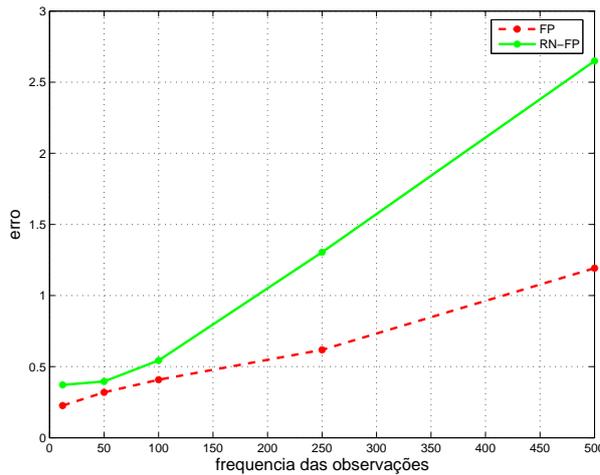
$$p(y|x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(y - \mathbf{H}\mathbf{x})^2}{2\sigma^2} \right] \quad (6.9)$$

Para a assimilação da RNA “emulando” o FP foi utilizado a mesma arquitetura da rede usada para o Filtro de Kalman, ilustrado na [Figura 6.4](#) com parâmetros apresentados em [6.1](#). Para o treinamento foram usados 2000 padrões tanto para os dados do modelo, observações e “alvo” (estimado pelo FP) da rede. Na [Tabela 6.3](#) tem-se os resultados do erro para assimilação variando-se a frequência da inserção de observações, para o Filtro de Partículas e para a RNA Perceptron. Na [Figura 6.12\(b\)](#) nota-se que a assimilação com o FP possui erro menor em relação ao FK. A desvantagem em se trabalhar com o

Tabela 6.3 - Erros de Assimilação: Filtro de Partículas e RNA Perceptron.

freq	FP				RNA PMC			
	x	y	z	$(x + y + z)/3$	x	y	z	$(x + y + z)/3$
12	0,1872	0,2543	0,2378	0,2264	0,3453	0,3609	0,4088	0,3717
50	0,2542	0,3710	0,3314	0,3189	0,3609	0,4504	0,3753	0,3955
100	0,3201	0,4647	0,4391	0,4080	0,4574	0,6244	0,5484	0,5434
250	0,4987	0,7316	0,6232	0,6178	1,0392	1,5009	1,3734	1,3045
500	0,9587	1,4057	1,2134	1,1926	2,1043	2,8108	3,0328	2,6493

Filtro de Partículas é sua complexidade computacional. O passo de reamostragem desta técnica, que o torna um método custoso computacionalmente, pois a cada iteração do método são selecionadas novas partículas para serem propagadas através do sistema do modelo, sendo que a estimativa é dada pelo cálculo da média dessas partículas. Em [Chorin](#)



(a) curva verde: erro da RNA Perceptron; curva vermelha: erro do FP (b) curva vermelha: erro do FP; curva azul: erro do FK

Figura 6.12 - Média global do erro $(x + y + z)/3$ de assimilação; eixo x é a frequência das observações e o eixo y é o erro.

e Krause (2006) é proposto uma estratégia para a redução de dimensão na implementação do Filtro. A estratégia utilizada é sugerida por alguns trabalhos como predição ótima.

Na Figura 6.13 são apresentados os resultados da assimilação com este método e para RNA “emulando” esta técnica. A assimilação foi realizada a cada 500 passos de tempo, na Tabela 6.3, apresenta-se os erros para cada componente do sistema de Lorenz, com assimilação realizada a 12, 50, 100, 250 e 500 passos de tempo, em todos os casos a assimilação com o FP obteve erro menor. A RNA Perceptron emulando o Filtro de Partículas é eficiente em fazer assimilação de dados (FURTADO et al., 2008). No entanto, a abordagem para assimilação de dados com redes neurais possui vantagens no seguinte aspecto: o algoritmo é altamente paralelizável e permite implementação em hardware.

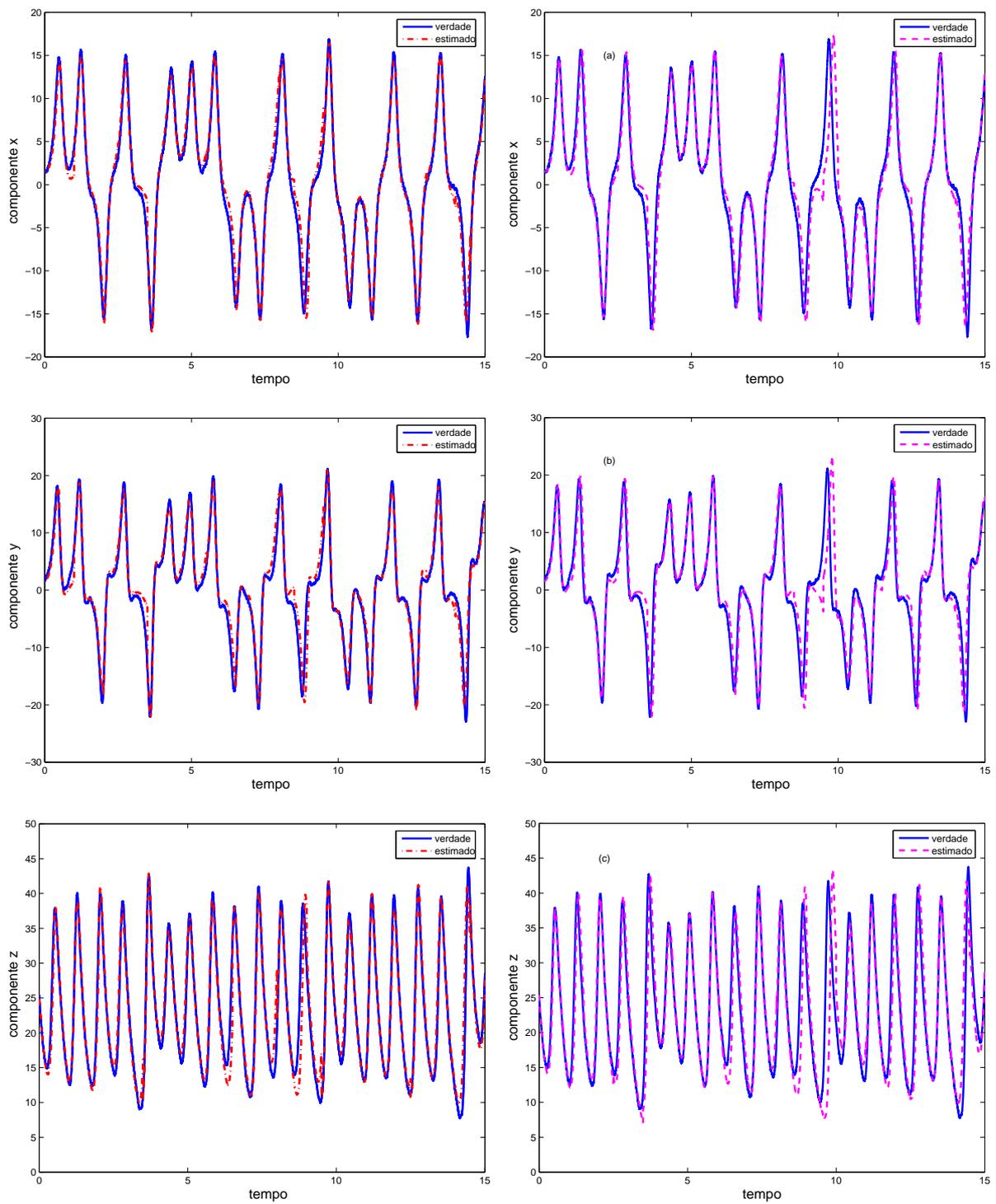


Figura 6.13 - Assimilação de dados com o Filtro de Partículas e com RNA Perceptron de Múltiplas Camadas. As observações foram inseridas a cada 500 passos de tempo. Gráficos à direita: curva tracejada (vermelha) estimado pelo FP, curva cheia (azul) referência de verdade. Analogamente, gráficos à esquerda, curva tracejada (magenta) estimado pela RNA, curva cheia (azul) referência de verdade.

6.4 Resultados com a Rede Neural emulando o Método Variacional

Nesta seção apresenta-se os resultados com o Método Variacional descrito no [Capítulo 4](#). Para a RNA foi usado a mesma arquitetura ilustrada pela [Figura 6.4](#) com parâmetros apresentados na [Tabela 6.1](#). A aplicação da técnica variacional utiliza o funcional [4.23](#), denominado de função custo, que contém o termo de observação J_o e o termo de modelagem J_b . Neste trabalho utilizou-se apenas o termo de observação J_o , isto é, o funcional da forma:

$$J(\mathbf{x}(t_0)) = \underbrace{\frac{1}{2} \sum_{i=0}^N (\mathbf{y}_i^o - \mathbf{H}(\mathbf{x}_i))^T \mathbf{R}^{-1} (\mathbf{y}_i^o - \mathbf{H}(\mathbf{x}_i))}_{J_o} \quad (6.10)$$

em que \mathbf{R} é a matriz de covariância do erro de observação; \mathbf{y} o vetor de observações; $\mathbf{H}=\mathbf{I}$ e \mathbf{x} é o vetor da variável de estado do sistema dinâmico. O método determina o melhor ajuste entre os dados do modelo e dados de observação no período de assimilação. As [Figuras 6.14](#) e [6.15](#), mostram os resultados da assimilação com esta técnica para as componentes x e z , respectivamente. As observações foram inseridas a cada 12 passos de tempo. Nas [Figuras 6.16](#) tem-se o gráfico do erro em escala logarítmica e a curvas de convergência da função custo e da norma do seu gradiente. Em [6.17](#) mostra-se o resultado com a assimilação realizada a cada 500 passos de tempo, na [Figura 6.18](#) tem-se o erro e o gráfico da convergência.

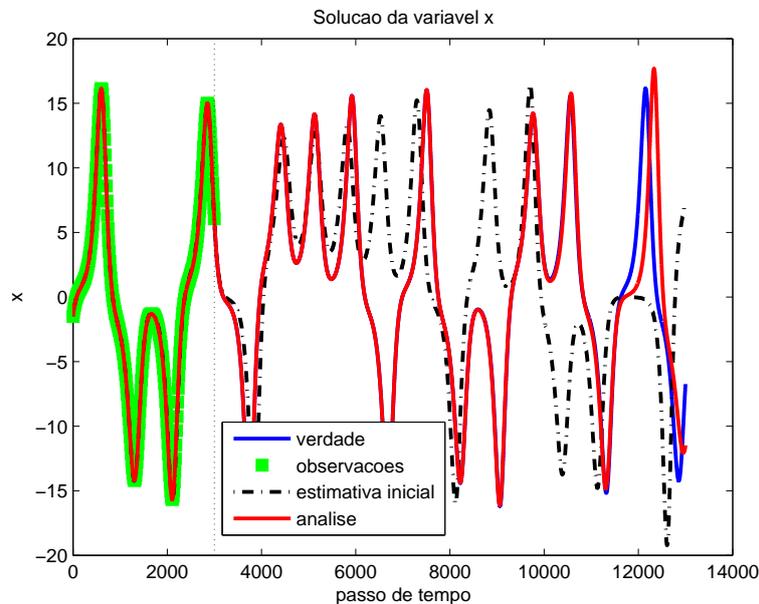


Figura 6.14 - Série temporal da componente x . Assimilação de dados a cada 12 passos de tempo, estimativa obtida com o método variacional: curva vermelha estimado, curva azul referência de verdade.

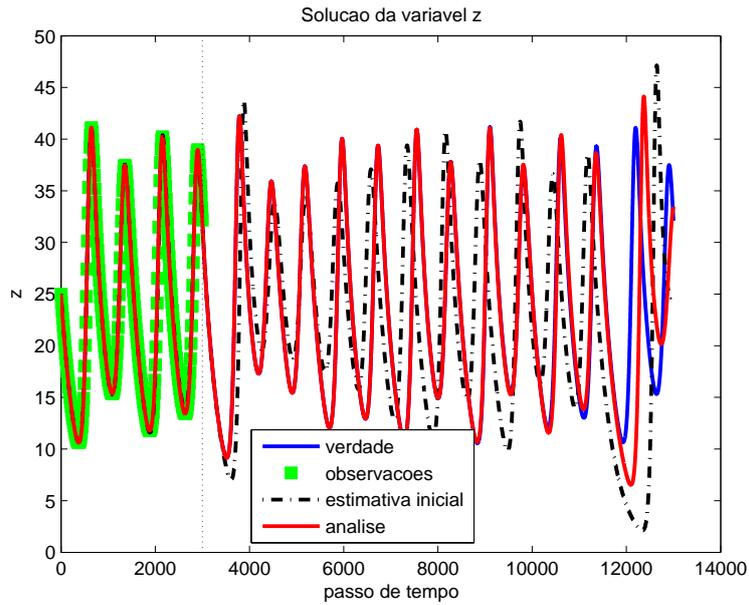
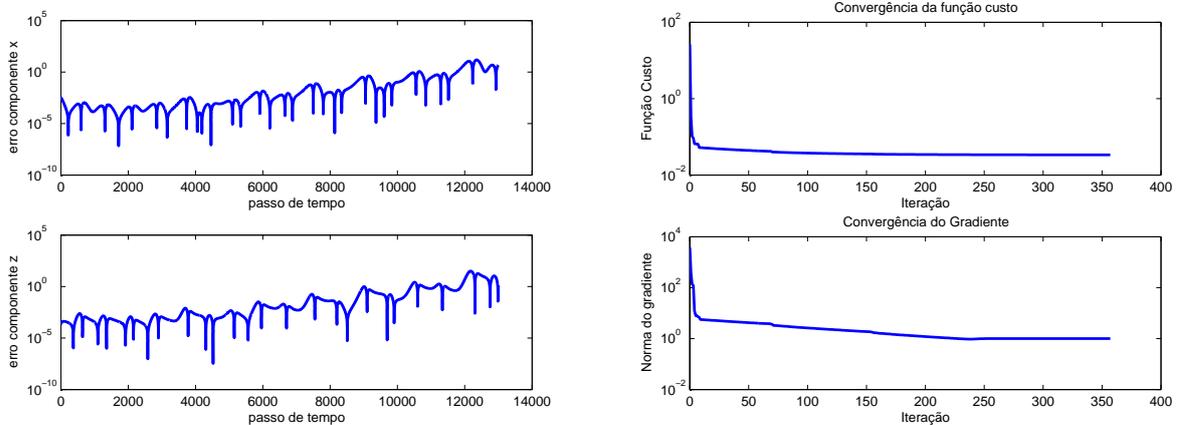


Figura 6.15 - Série temporal da componente z . Assimilação de dados a cada 12 passos de tempo, estimativa obtida com o método variacional; curva vermelha: estimado; curva azul: referência de verdade.



(a) Gráfico do erro em escala logarítmica.

(b) Convergência da função custo e da norma do gradiente.

Figura 6.16 - (a) Gráfico do erro das componentes x e y do modelo de Lorenz, para assimilação realizada a cada 12 passos de tempo; (b) Gráficos de convergência da função custo e da norma do gradiente.

Nas Figuras 6.14, 6.15 e 6.17 a janela de assimilação corresponde a 3 mil passos de tempo, a partir daí, tem-se a previsão. Nota-se que para a janela de assimilação com observações inseridas a cada 12 passos de tempo (Figuras 6.14 e 6.15), se obteve uma boa previsão em

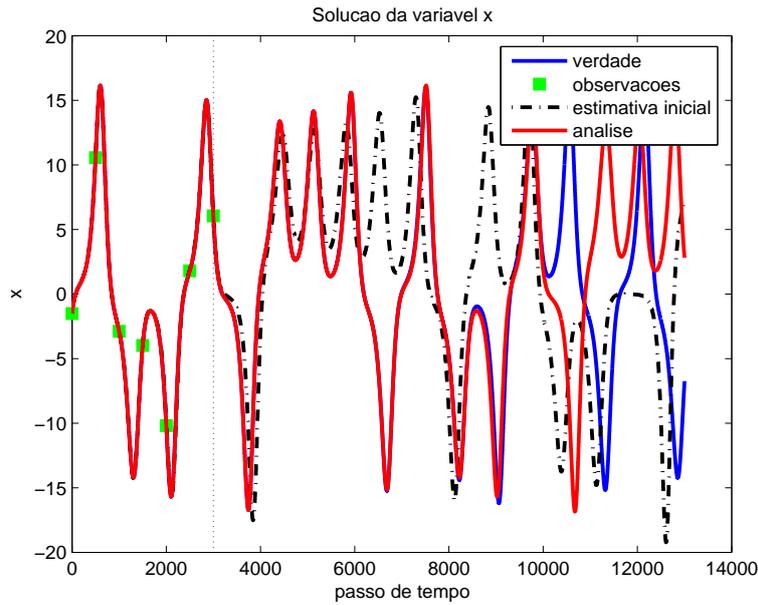
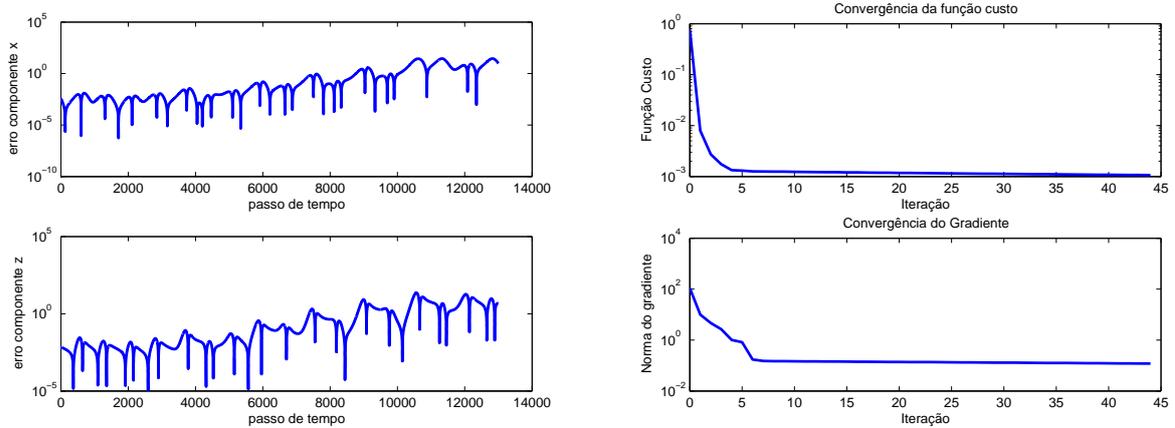


Figura 6.17 - Série temporal da componente x . Assimilação de dados a cada 500 passos de tempo, estimativa obtida com o método variacional; curva vermelha: estimado; curva azul: referência de verdade.



(a) Gráfico do erro em escala logarítmica.

(b) Convergência da função custo e da norma do gradiente.

Figura 6.18 - (a) Gráfico do erro das componentes x e y do modelo de Lorenz, para assimilação realizada a cada 500 passos de tempo; (b) Gráficos de convergência da função custo e da norma do gradiente.

torno de 12 mil passos de tempo, enquanto para a janela de assimilação com observações a cada 500 passos de tempo, em torno de 10 mil passos de tempo. Isso indica que quanto mais observações melhor será a qualidade da previsão. Em 6.19 apresenta-se os resultados da assimilação realizada a cada 12 passos de tempo para componente x do modelo de Lorenz com o método variacional e com a rede neural. Nota-se que a rede foi eficiente ao

“emular” o método Variacional, como pode-se ver por meio do gráfico do erro. Na Figura 6.20 é a diferença absoluta entre a verdade e o estimado e na Figura 6.21 tem-se o erro em escala semilogarítmica. Na Figura 6.22 pode-se visualizar o desempenho da assimilação

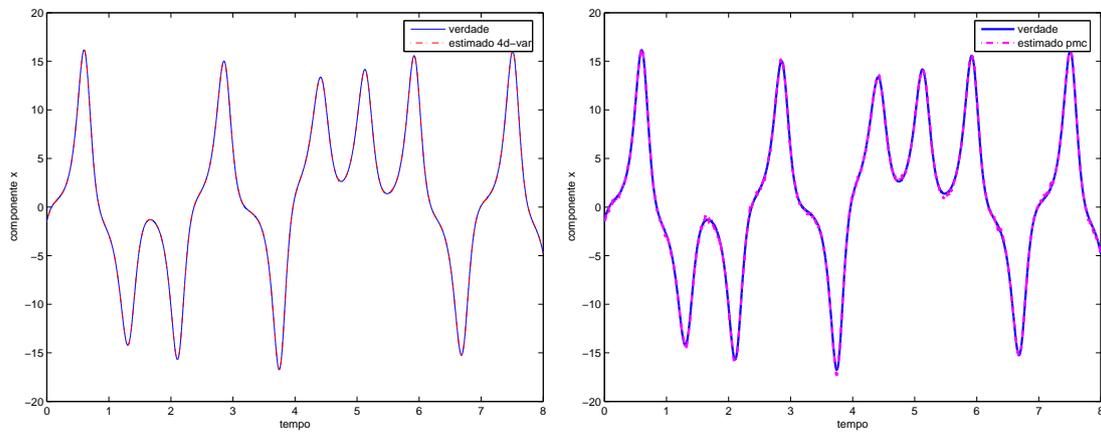


Figura 6.19 - Série temporal da componente x . Assimilação de dados a cada 12 passos de tempo. fig. à direita: estimativa do método variacional; fig. à esquerda: estimativa da RN.

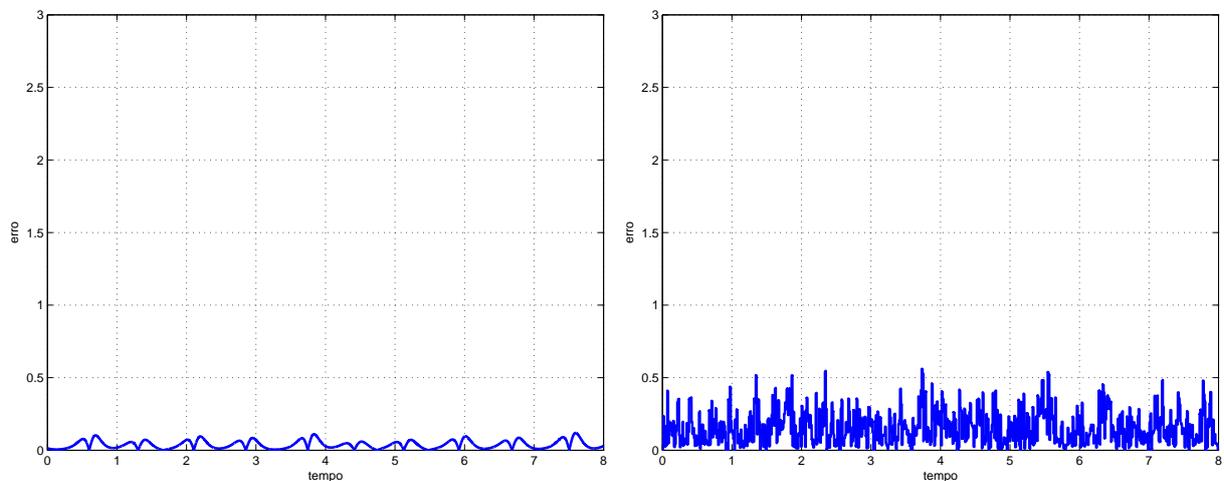


Figura 6.20 - Gráfico do erro da componentes x do modelo de Lorenz, para assimilação realizada a cada 12 passos de tempo; fig. à direita: variacional; fig. à esquerda: RN.

de dados, por meio do gráfico do erro, para as metodologias abordadas neste trabalho. As observações foram inseridas a cada 12 passos de tempo, e cálculo do erro foi feito pela diferença absoluta entre a verdade e o estimado, nota-se que a rede neural foneceu resultados satisfatórios ao “emular” ambas as metodologias, do Filtro de Kalman, Filtro de Partículas e Método Variacional.

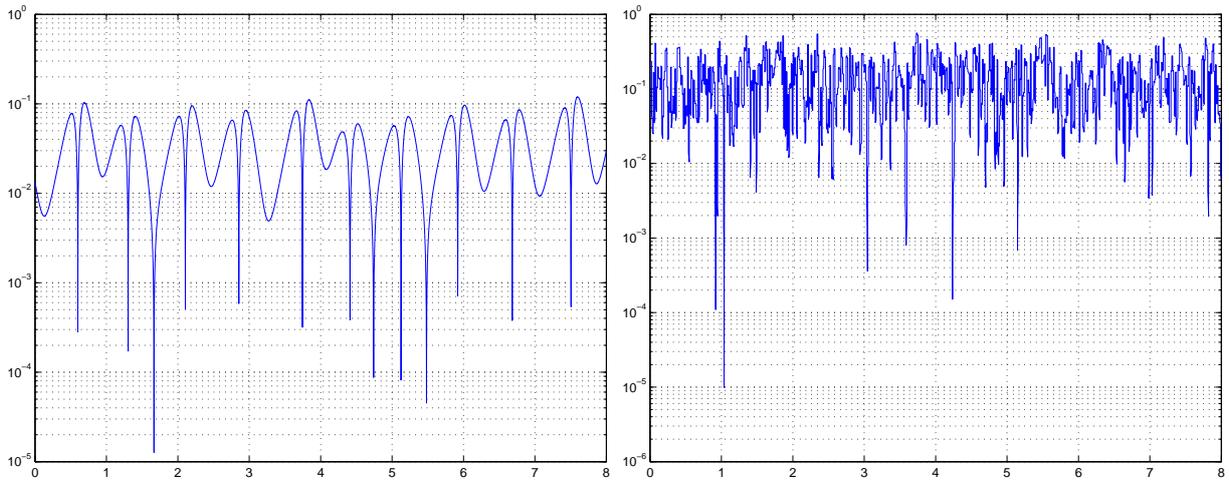


Figura 6.21 - Erros em escala logaritmica; fig. à direita: variacional; fig. à esquerda: RN.

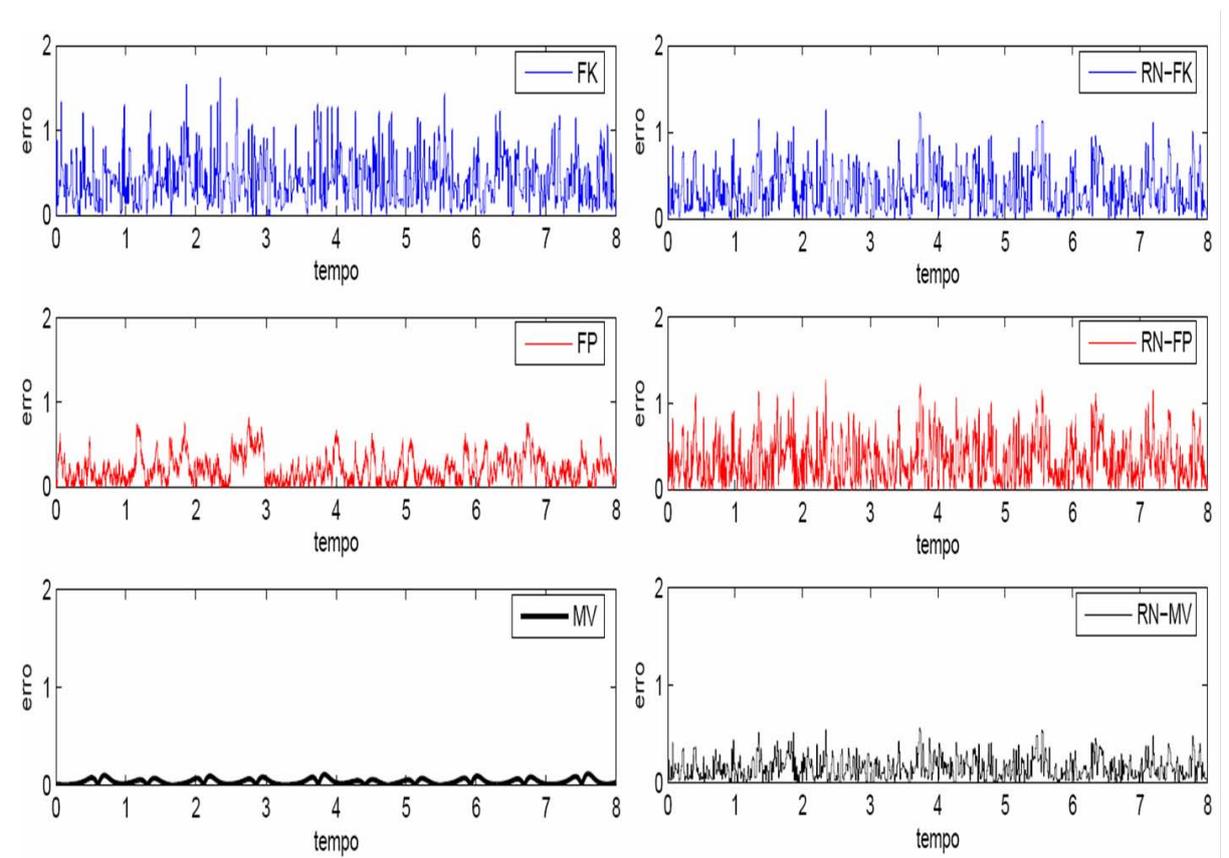


Figura 6.22 - Gráfico dos erros, com observações a cada 12 passos de tempo.

7 CONCLUSÕES

Neste trabalho avalia-se a eficiência da rede Perceptron de Múltiplas camadas em “emular” o Filtro de Kalman, Filtro de Partículas e o Método Variacional, em um contexto de assimilação de dados em dinâmica não linear. O sistema de Lorenz em regime caótico foi utilizado como o problema teste.

O diferencial deste trabalho em relação as pesquisas de [Nowosad \(2001\)](#) e [Harter \(2004\)](#) foi treinar a rede com conjuntos de padrões (exemplos) providos de outros métodos de assimilação, como o Filtro de Partícula ([GORDON et al., 1993](#)) e o Método Variacional ([KALNAY, 2003](#)).

Um resultado significativo, obtido nesta pesquisa, foi que a rede neural PMC, “emulando” o Filtro de Partículas, é eficiente em fazer assimilação com observações inseridas a cada 500 passos de tempo, enquanto o Filtro de Kalman perde a capacidade de estimação para essa frequência de observações.

Uma vantagem em utilizar RNA é que essa apresenta uma menor complexidade computacional em relação ao Filtro de Partículas, a abordagem Variacional e o Filtro Kalman. Outros pontos importantes a serem mencionados são que as RNAs possuem a característica de serem um algoritmo intrinsecamente paralelo e serem propícias a implementação em *hardware*, ou seja, neuro-computadores ([CAMPOS VELHO et al., 2007b](#)). Isso poderá ser uma solução ao desafio da Assimilação de Dados em um 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.

Para o treinamento da RNA 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. Essa técnica explorada no treinamento das redes representa um avanço em relação a metodologia de treinamento da RNA em Assimilação de Dados ([HARTER, 2004](#)).

Os resultados aqui apresentados mostram que a Rede Neural PMC é uma metodologia eficiente para se fazer assimilação de dados.

Como sugestões para trabalhos futuros tem-se o uso da técnica de Redes Neurais para assimilação de dados no Modelo de Água Rasa em duas dimensões ([TODLING; GHIL, 1994](#); [GHIL; TODLING, 1996](#)) (já em desenvolvimento) e a investigação da aplicação da rede recorrente de Hopfield (RH) ([HOPFIELD, 1982](#)). A resposta na RH é obtida quando se alcança um estado estacionário de um sistema de equações diferenciais (cada equação diferencial

representa o comportamento dinâmico de um neurônio). Isso caracteriza o treinamento não-supervisionado. Dessa forma, a análise é obtida como a solução de estado estacionário de um sistema dinâmico a cada ciclo de assimilação. Esse tipo de rede dispensa a aplicação de qualquer outra técnica de Assimilação de Dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABELEM, A. J. G. **Redes neurais artificiais na previsão de séries temporais**. Dissertação (Mestrado em Engenharia Elétrica) — PUC-RJ, Rio de Janeiro, 1994. 84
- AGUIRRE, L. A. **Introdução à identificação de sistemas: técnicas lineares e não-lineares aplicadas a sistemas reais**. Belo Horizonte: Editora UFMG, 2000. 46
- ANDERSON, B. D. O.; MOORE, J. B. **Optimal filtering**. Prentice-Hall: Englewood Cliffs, NJ, 1979. 51, 53, 55
- ARBIB, M. A. **Brains, machines, and mathematics**. New York: Springer Verlag, 1987. 81
- ARULAMPALAM, M. S.; MASKELL, S.; GORDON, N.; CLAPP, T. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. **IEEE Transactions on Signal Processing**, v. 50, n. 2, p. 174–188, 2002. 51, 58, 61, 64, 65
- BANNISTER, R. N. **Elementary 4D-Var**. Reading: University of Reading, 2001. Relatório técnico. 75
- BERGTHORSON, P.; DÖÖ, B. Numerical weather map analysis. **Tellus**, p. 329–340, 1955. 36
- BOYCE, W. E.; DIPRIMA, R. C. **Equações diferenciais elementares e problemas de valores de contorno**. Rio de Janeiro: LTC, 1998. 91
- BRAGA, A.; LUDERMIR, T. B.; CARVALHO, A. C. P. L. F. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: Livros Técnicos e Científicos Editora S.A., 2000. 87
- BUCY, R. S.; SENNE, K. D. Digital synthesis of nonlinear filters. **Automatica**, v. 7, p. 287–298, 1971. 54
- CAMPOS VELHO, H. F.; CINTRA, R. S.; FURTADO, H. C. M. **Introdução à assimilação de dados**. 2007. Disponível em: <www.lac.inpe.br/~haroldo/Curso-DataAssimilation/Curso_Assim_Dados-INPE-2.pdf>. Acesso em: 09 Novembro 2007. 28
- CAMPOS VELHO, H. F.; SILVA, J. D. S.; SHIGUEMORI, E. H. Hardware implementation for the atmospheric temperature retrieval from satellite data. In: INVERSE PROBLEMS, DESIGN AND OPTIMIZATION SYMPOSIUM (IPDO), Sunny Isles Beach, Florida. **Proceedings ...** 2007. Disponível em: <sid.inpe.br/mtc-m17@80/2007/11.27.18.12>. 111

CAMPOS VELHO, H. F.; VIJAYKUMAR, N. L.; STEPHANY, S.; PRETO, A. J.; NOWOSAD, A. G. A neural network implementation for data assimilation using mpi, application of high performace computing in engineering. In: BREBIA, C. A.; MELLI, P.; ZANASI, A. . (Ed.). **Application of high performace computing in engineering**. Southampton: WIT Press, 2002. p. Section 5, 211–220. Disponível em: <Disponívelem:<<http://urlib.net/sid.inpe.br/mtc-m17@80/2007/08.14.13.09>>.Acessoem:22abr.2008.>. 30

CHORIN, A. J.; KRAUSE, P. Dimensional reduction for a bayesian filter. **PNAS**, v. 101, n. 42, p. 15013–15017, Oct. 2006. 103

CINTRA, R. **Implementação do sistema estatístico de assimilação de dados em espaço físico para o modelo global do CPTEC**. 182 p. (INPE-14223-TDI/1124). Dissertação (Mestrado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2004. 29

DALEY, R. **Atmospheric data analysis**. Cambridge: Cambridge University Press, 1993. 27, 28, 33, 55, 69, 72, 73, 75

DOUCET, A. **On sequential simulation-based methods for baysian filtering**. Cambridge, UK, 1998. 51, 57, 58, 61

DOUCET, A.; FREITAS, N. de; GORDON, N. **Sequential monte carlo methods in pratic**. New York, Usa: Springer Verlag, 2001. 57

DOUCET, A.; GODSILL, S.; ANDRIEU, C. On sequential monte carlo sampling methods bayesian filtering. **Statistics and Computing**, v. 10, n. 3, p. 197–208, 2000. 46

FAUSETT, L. **Fundamentals of neural networks: architectures, algorithmos, and applications**. New Jersey: Prentie Hall, 1994. 85

FURTADO, H. C. M.; CAMPOS VELHO, H. F.; MACAU, E. E. Data assimilation: Particle filter and artificial neural network. In: **International Conference on Inverse Problems in Engineering: Theory and Practice (ICIPE)**. Paris: [s.n.], 2008. (artigo submetido). 103

GARDNER, M. W.; DORLING, S. R. Artificial neural networks (the multilayer perceptron - a review of applications in the atmospheric sciences). **Atmospheric Environment**, v. 32, p. 2627–2636, 1998. 29

- GHIL, M.; TODLING, R. Tracking atmospheric instabilities with the Kalman Filter part ii: tow-layer results. **American Meteorological Society Mon. Wea. Rev.**, v. 124, p. 2340–2352, 1996. [111](#)
- GORDON, N.; SALMOND, D.; SMITH, A. Novel approach to nonlinear/non-gaussian bayesian state estimation. **IEE Proceedings**, v. 140, n. 2, p. 107–113, Abril 1993. [51](#), [54](#), [58](#), [60](#), [62](#), [65](#), [66](#), [111](#)
- GREBOGI, C.; OTT, E.; YORKE, J. A. Chaos, strange attractors, and fractal basin boundaries in nonlinear dynamics. **Science**, v. 238, p. 632–638, 1987. [27](#)
- HARTER, F. P. **Uso de filtro digital para iniciar um modelo de área limitada**. 1999-06. 111 p. (INPE-7251-TDI/695). Dissertação (Mestrado em Meteorologia) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 1999. Disponível em: <<<http://urlib.net/sid.inpe.br/deise/1999/11.26.13.10>>. Acesso em: 22abr.2008.>. [27](#)
- HARTER, F. P. **Redes neurais recorrentes aplicadas a assimilação de dados em dinâmica não-linear**. 2004-12-16. 138 p. p. (INPE-14222-TDI/1123). Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2004. [28](#), [30](#), [43](#), [55](#), [94](#), [95](#), [111](#)
- HARTER, F. P.; REMPEL, E. L.; CAMPOS VELHO, H. F. de; CHIAN, A. Application of artificial neural networks in auroral data assimilation. **Journal of Atmospheric and Solar - Terrestrial Physics (submetido)**, 2008. [27](#), [28](#)
- HAYKIN, S. **Redes neurais princípios e prática**. Porto Alegre: Editora Bookman, 2004. [81](#), [83](#), [86](#), [87](#), [89](#), [96](#)
- HOFFMAN, J. D. **Numerical methods for engineers and scientists**. Singapore: McGraw-Hill, 1993. [27](#)
- HOPFIELD, J. Neural networks and physical systems with emergent collective computational abilities. **Biophysics, National Academic Science**, v. 79, n. 8, p. 2554–2558, 1982. [111](#)
- HSIEH, W. W.; TANG, B. Applying neural network models to prediction and data analysis in meteorology and oceanography. **American Meteorology Society**, v. 79, n. 9, p. 1855–1870, September 1998. [29](#)
- JAZWINSKI, A. H. **Stochastic processes and filtering theory**. New York and London: Academic Press, 1970. [44](#), [45](#), [48](#), [51](#), [53](#), [54](#), [121](#)

KALMAN, R. E. A new approach to linear filtering and prediction problems. **Journal of Basic Engineering**, n. 82, p. 35–45, 1960. [49](#), [51](#), [54](#)

KALMAN, R. E.; BUCY, R. S. New results in linear filtering and prediction theory. **Journal of Basic Engineering**, n. 833, p. 95–108, 1961. [54](#)

KALNAY, E. **Atmospheric modeling, data assimilation and predictability**. Cambridge: Cambridge University Press, 2003. [27](#), [28](#), [33](#), [35](#), [55](#), [75](#), [77](#), [79](#), [91](#), [111](#)

KOVACS, Z. **Redes neurais artificiais: princípio e aplicações**. São Paulo: Collegium Cognition, 1996. [82](#)

LAWLESS, A. S. **Data assimilation with the Lorenz equations**. University of Reading, UK, 2004. [36](#), [75](#)

LORENC, A. C. Analysis methods for numerical weather prediction. **Quarterly Journal of the Royal Meteorological Society**, v. 112, p. 1177–1194, 1986. [35](#)

LORENC, A. C.; BELL, R. S.; MACPHERSON, B. The meteorological office analysis correction data assimilation scheme. **Quarterly Journal of the Royal Meteorological Society**, v. 117, n. 497, p. 59–89, 1991. [39](#)

LORENZ, E. N. Deterministic nonperiodic flow. **Journal of the Atmospheric Sciences**, v. 20, n. 2, p. 130–141, january 1963. [28](#), [91](#), [93](#)

_____. A study of the predictability of a 28-variable atmospheric model. **Tellus**, v. 17, p. 321–333, 1965. [28](#), [76](#), [78](#)

LYNCH, P. The slow equations. **Quarterly Journal of the Royal Meteorological Society**, v. 115, p. 201–219, jan 1989. Provided by the SAO/NASA Astrophysics Data System. Disponível em: <<http://adsabs.harvard.edu/abs/1989QJRMS.115..201L>>. [30](#)

MCCULLOCH, W.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, p. 115–133, 1943. [84](#)

MENDEL, J. M.; MCLAREN, R. W. Reinforcement learning control and pattern recognition systems. In: _____. New York: Academic Press, 1970. cap. Adaptive, learning and pattern recognition systems: Theory and applications, p. 287–318. [86](#)

MENDONÇA, A. M.; BONATTI, J. P. O sistema de previsao de tempo global por ensemble do cptec. In: **Anais. SBMET**, 2002. p. 3341–3351. Disponível em: <<http://mtc-m15.sid.inpe.br/col/cptec.inpe.br/walmeida/2003/07.31.15.12/doc/Mendonca>>. [28](#)

- 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. [35](#)
- MONTEIRO, L. H. A. **Sistemas dinâmicos**. São Paulo: Livraria da Física, 2006. 625 p. [43](#), [46](#)
- NOWOSAD, A.; VELHO, H. C.; NETO, A. R. Neural network as a new approach for data assimilation. In: **Congresso Brasileiro de Meteorologia**. Rio de Janeiro, (Brasil): Proceedings in CD-ROM (paper code PT00002), 2000. v. 1, p. 3078–3086. [30](#)
- NOWOSAD, A. G. **Novas abordagens de assimilação de dados metereológicos**. 2001-03-14. 120 p. p. (INPE-8481-TDI/779). Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais, 2001. [34](#), [95](#), [111](#)
- PAPOULIS, A. **Probability, random variables, and stochastic processes**. New York and London: McGraw-Hill, 1984. [57](#)
- 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. [27](#)
- RICHARDSON, L. F. **Weather prediction by numerical process**. Cambridge: University Press, 1992. [27](#)
- RUBINSTEIN, E. Y. **Simulation and the Monte Carlo method**. New York: John Wiley e Sons, 1981. [58](#), [59](#)
- SALTZMAN, B. Finite amplitude free convection as an initial value problem. **Journal of Atmospheric Sciences**, v. 19, p. 329–341, jul 1962. [92](#)
- SCHÖN, T. B. **On Computational methods for nonlinear estimation**. 1047. Dissertação (Thesis) — Linköping Studies in Science and Tecnology, 2003. [47](#), [57](#), [61](#)
- SCHÖN, T. B. **Estimation of nonlinear dynamic systems: theory and applications**. 998. Dissertação (Dissertations) — Linköping Studies in Science and Tecnology, 2006. [43](#), [44](#), [46](#), [48](#), [49](#), [51](#), [54](#), [56](#)
- SORENSEN, H. W. Least-squares estimation: from Gauss to Kalman. **IEEE spectrum**, v. 7, p. 63–68, 1970. [55](#)
- SORENSEN, H. W.; ALSPACH, D. L. Recursive bayesian estimation using gaussian sums. **Automatica**, v. 7, p. 465–479, 1971. [54](#)

TALAGRAND, O. Assimilation of observations, an introduction. **J. Met. Soc. Japan Special Issue**, v. 1B, p. 191–209, 1997. 29, 34

TANG, Y.; HSIEH, W. W. Coupling neural networks to incomplete dynamical systems via variational data assimilation. **American Meteorological Society**, v. 129, p. 818–834, March 2001. 29

TODLING, R. **Estimation theory and foundations of atmospheric data assimilation**. 1999. Disponível em: <citeseer.ist.psu.edu/todling99estimation.html>. 97, 119

TODLING, R.; GHIL, M. Tracking atmospheric instabilities with the Kalman Filter part i: Methodology and one-layer results. **American Meteorological Society Mon. Wea. Rev.**, v. 122, p. 183–204, July 1994. 111

ZUBEN, F.; CASTRO, L. N. d. **Base Biológica: Aspectos funcionais e organizacionais**. 2000. Disponível em: <ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia353_1s07/topico2_07.pdf>. Acesso em: 04 abril 2007. 81

A CONCEITOS BÁSICOS EM PROCESSOS ESTOCÁSTICOS

A.1 Probabilidade Condicional

Se A e B são dois eventos e $P(B) \neq 0$, a probabilidade condicional de A dado B é definida como (TODLING, 1999):

$$P(A|B) \equiv P(A \cap B)|P(B) \quad (\text{A.1})$$

Os eventos A e B são estatisticamente independentes se $P(A|B) = P(A)$. Consequentemente $P(A \cap B) = P(A)P(B)$. Analogamente,

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \quad \forall P(A) \neq 0 \quad (\text{A.2})$$

Combinando as duas relações acima tem-se que:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (\text{A.3})$$

O que é denominado como regra ou (teorema) de Bayes para probabilidades

A.2 Média e Covariância

Uma determinada variável aleatória $\mathbf{x} \in \mathfrak{R}^n$, tem seu valor médio ou esperado dado por:

$$E[\mathbf{x}] = \bar{\mathbf{x}} = \int_{\mathfrak{R}^n} \mathbf{x}f(\mathbf{x})dx \quad (\text{A.4})$$

Onde f é uma função densidade de probabilidade de x . Os momentos em torno do valor médio de \mathbf{x} são conhecidos como momentos centrais e podem ser representados por:

$$E[(\mathbf{x} - \bar{\mathbf{x}})^n] \quad (\text{A.5})$$

sendo n a ordem do momento. A variância, como é chamada para o caso unidimensional, já para o caso multidimensional recebe o nome de covariância, é importante na área de processos estocásticos, pois a variância de uma variável aleatória indica em geral quão longe seus valores se encontram do valor esperado. Assim, a matriz de covariância de um vetor $\mathbf{x} \in \mathfrak{R}^n$, têm dimensão $n \times n$ é definido por:

$$cov[\mathbf{x}] = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T] \quad (\text{A.6})$$

A.3 Variáveis Gaussianas

Uma variável aleatória x é chamada de gaussiana ou normal quando sua função densidade de probabilidade, é expressada por:

$$f_x(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \quad (\text{A.7})$$

Sendo que $\sigma = \sqrt{\text{cov}[x]}$ é o desvio padrão, \bar{x} é a média que representam constantes reais. A seguir têm-se um gráfico ilustrando uma gaussiana de média zero, variando o desvio padrão. A função gaussiana é uma boa aproximação para a representação de ruídos, pois

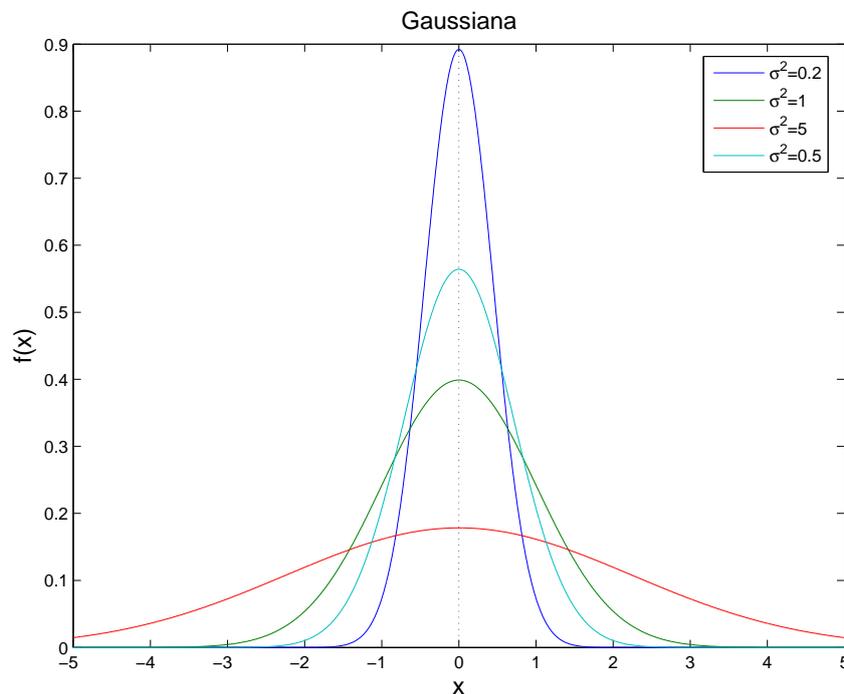


Figura A.1 - Função densidade de probabilidade gaussiana de média zero; ilustra-se distribuições alterando o valor da variância; $\sigma^2 = 0.2$ (azul), $\sigma^2 = 1$ (verde), $\sigma^2 = 5$ (vermelho), $\sigma^2 = 0.5$ (azul esverdeado).

são necessários apenas os momentos de primeira e segunda ordem, isto é, média e variância.

A.4 Processo de Markov

Um processo estocástico de parâmetro contínuo ou discreto $\{x_t, t \in T\}$ é dito um processo de Markov se, para algum conjunto de parâmetro finito $\{t_i : t_i < t_{i+1}\} \in T$ e para cada λ

real (JAZWINSKI, 1970), tem-se:

$$Pr\{x_{t_n}(\omega) \leq \lambda | x_{t_1}, \dots, x_{t_{n-1}}\} = Pr\{x_{t_n}(\omega) \leq \lambda | x_{t_{n-1}}\} \quad (\text{A.8})$$

A propriedade A.8 diz que a lei de probabilidade do processo no futuro, independe do passado, depende apenas do estado presente. Esta propriedade é muitas vezes referida como o princípio da casualidade generalizada, isto é, o futuro pode ser predito do conhecimento presente. Para o processo de parâmetro contínuo, na equação A.8 implica que para $t_1 < t_2$ e todo λ ,

$$Pr\{x_{t_2}(\omega) \leq \lambda | x_\tau, \tau \leq t_1\} = Pr\{x_{t_2}(\omega) \leq \lambda | x_{t_1}\} \quad (\text{A.9})$$

A propriedade A.8 pode ser escrita em termos de função de densidade

$$p(x_{t_n} | x_{t_1}, \dots, x_{t_{n-1}}) = p(x_{t_n} | x_{t_{n-1}}) \quad (\text{A.10})$$

Agora considere a função de densidade conjunta $p(x_{t_n}, x_{t_{n-1}}, \dots, x_{t_1})$ por definição da função densidade de probabilidade condicional pode-se escrevê-la como:

$$p(x_{t_n}, \dots, x_{t_1}) = p(x_{t_n} | x_{t_{n-1}}, \dots, x_{t_1}) p(x_{t_{n-1}}, \dots, x_{t_1}) \quad (\text{A.11})$$

Usando a propriedade A.9,

$$p(x_{t_n}, \dots, x_{t_1}) = p(x_{t_n} | x_{t_{n-1}}) p(x_{t_{n-1}}, \dots, x_{t_1}) \quad (\text{A.12})$$

Continuando esse processo obtém-se finalmente

$$p(x_{t_n}, \dots, x_{t_1}) = p(x_{t_n} | x_{t_{n-1}}) p(x_{t_{n-1}} | x_{t_{n-2}}) \dots p(x_{t_2} | x_{t_1}) p(x_{t_1}) \quad (\text{A.13})$$

Portanto, a lei de probabilidade de um processo de Markov pode ser especificada por descrição de $p(x_t)$ e $p(x_t | x_\tau)$ para todo $t > \tau \in T$. A densidade de probabilidade condicional $p(x_t | x_\tau)$ é chamada de densidade de probabilidade de transição de um processo de Markov.

A.5 Ruído Branco

Um processo estocástico em tempo discreto $\{w_t\}$ é dito ser branco se este é independente no tempo, isto é

$$p(w_t, w_s) = p(w_t) p(w_s), \quad t \neq s. \quad (\text{A.14})$$

Esta definição implica que todas as entidades do processo $\{w_t\}$ são mutualmente independentes.

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.