UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

POLÍTICA DE OPERAÇÃO PREDITIVA ESTABILIZADA VIA TERMO INERCIAL UTILIZANDO "*ANALYTIC SIGNAL*", "*DYNAMIC MODELLING*" E SISTEMAS INTELIGENTES NA PREVISÃO DE VAZÕES AFLUENTES EM SISTEMAS HIDROTÉRMICOS DE POTÊNCIA

Rodrigo Sacchi

Tese apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos exigidos para a obtenção do Título de Doutor em Engenharia Elétrica.

Área de Concentração: **Sistemas Elétricos de Potência**

Orientador: **Prof. Dr. Adriano Alber de França Mendes Carneiro**
Co-orientador: **Prof. Assoc. Ivan Nunes da Silva**

São Carlos, SP
2009

Livros Grátis

http://www.livrosgratis.com.br

Milhares de livros grátis para download.

FOLHA DE JULGAMENTO

Candidato: Engenheiro **RODRIGO SACCHI.**

Tese defendida e julgada em 08/06/2009 perante a Comissão Julgadora:

APROVADO

Prof. Dr. **ADRIANO ALBER DE FRANÇA MENDES CARNEIRO (Orientador)**
(Escola de Engenharia de São Carlos/USP)

APROVADO

Prof. Titular **SECUNDINO SOARES FILHO**
(Universidade Estadual de Campinas/UNICAMP)

APROVADO

Prof. Dr. **ALUIZIO FAUSTO RIBEIRO ARAÚJO**
(Universidade Federal de Pernambuco/UFPE)

APROVADO

Prof. Associado **MARINHO GOMES DE ANDRADE FILHO**
(Instituto de Ciências Matemáticas e de Computação/USP)

APROVADO

Profª. Drª. **ROSANGELA BALLINI**
(Universidade Estadual de Campinas/UNICAMP)

Prof. Titular **GERALDO ROBERTO MARTINS DA COSTA**
Coordenador do Programa de Pós-Graduação em Engenharia Elétrica e
Presidente da Comissão de Pós-Graduação

DEDICATÓRIA

Com muito amor e carinho, dedico este trabalho à
minha esposa Maurícia,
meus pais Selma e Eurydice,
meus irmãos Juliana, Nanci e Luis Carlos
e todos os meus amigos.

AGRADECIMENTOS

Ao professor Adriano Alber de França Mendes Carneiro, pela oportunidade de trabalho, orientação e, principalmente, pela amizade construída desde os tempos de graduação.

Ao professor Ivan Nunes da Silva, pela co-orientação e importantes contribuições a este trabalho de pesquisa.

Aos professores Jose Carlos Principe e William W-G Yeh, pela receptividade, cordialidade e orientação durante o tempo em que estive na UF e UCLA, respectivamente.

Aos professores Aluízio Fausto Ribeiro Araújo, André Carlos Ponce de Leon F. de Carvalho e Mário Oleskovicz, pelas relevantes contribuições ao trabalho de Mestrado.

Aos professores Secundino Soares Filho e Marinho Gomes de Andrade Filho, pelas excelentes contribuições a este trabalho de pesquisa.

À toda minha família, por sempre acreditarem no meu trabalho.

À minha esposa Maurícia, pelo amor, carinho e companheirismo incondicional.

Aos colegas de laboratório (LSEE), pelo agradabilíssimo ambiente de trabalho.

Ao Dr. Roberto Castro, pela oportunidade profissional na CPFL Energia.

Aos novos amigos da CPFL Energia, pela enorme receptividade e profissionalismo.

À Coordenação de Formação de Pessoal de Nível Superior (CAPES), pelas bolsas de Doutorado e Programa de Doutorado com Estágio no Exterior (PDEE).

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo suporte financeiro por meio de Projeto Temático.

Ao Departamento de Engenharia Elétrica da EESC-USP, por toda a infra-estrutura.

"O mais importante da vida
não é a situação em que estamos,
mas a direção para a qual nos movemos".

Oliver Wendell Holmes

RESUMO

SACCHI, R. (2009). **Política de Operação Preditiva Estabilizada via Termo Inercial utilizando "*Analytic Signal*", "*Dynamic Modelling*" e Sistemas Inteligentes na Previsão de Vazões Afluentes em Sistemas Hidrotérmicos de Potência**. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2009.

Este trabalho de pesquisa objetivou a obtenção de uma nova Política de Operação que melhor caracterizasse o comportamento ótimo dos sistemas hidrelétricos de potência, mesmo diante das mais variadas condições hidrológicas. Este trabalho teve duas linhas de investigação. Uma tratou do problema de previsão de vazões afluentes mensais, na busca por abordagens e técnicas que definissem bons modelos de previsão. A outra linha de pesquisa tratou de encontrar uma nova Política de Operação, para o problema de Planejamento da Operação, que fosse capaz de definir uma seqüência de decisões operativas mais estáveis, confiáveis e de menor custo operativo. Na primeira linha de pesquisa, investigou-se três aspectos importantes na definição de um modelo de previsão: técnicas de pré-processamento dos dados, definição automática do espaço de entrada e avaliação do desempenho de alguns modelos de Redes Neurais e Sistemas "*Fuzzy*" como modelos de previsão. Nestes aspectos foram investigadas a utilização da Análise dos Componentes Principais e o tratamento da série temporal de Vazões Afluentes como um sinal discreto, utilizando-se a representação "*Analytic Signal*". Para a definição do espaço de entrada de maneira automática utilizou-se a abordagem da "*Dynamic Modelling*", empregando-se a "*Average Mutual Information*" e "*False Nearest Neighbors*". Para implementação dos modelos de previsão foram estudados e avaliados quatro modelos inteligentes: rede SONARX, rede SONARX-RBF, modelo ANFIS e a rede ESN. Já na outra linha de pesquisa, foi proposta uma Política de Operação que fosse capaz de estabilizar os despachos de Geração Termelétrica e conseqüentemente o Custo Marginal de Operação. A Política de Operação Preditiva Estabilizada via Termo Inercial produziu excelentes resultados operativos, melhorando de forma significativa a performance da Política Preditiva.

Palavras-chave: planejamento da operação, políticas de operação, otimização, sistemas inteligentes, *analytic signal*, *dynamic modelling*.

ABSTRACT

SACCHI, R. (2009). **Predictive Operation Policy Stabilized via Inertial Term using Analytic Signal, Dynamic Modelling and Intelligent Systems on Hydrothermal Power Systems**. Thesis (Ph.D. Thesis) – Engineering School of São Carlos, University of São Paulo, São Carlos, 2009.

This research work aimed at obtaining a new Operation Policy which could better describe the optimal behavior of hydropower systems, even when faced with the most varied hydrological conditions. This research had two lines of investigation. The first one dealt with the monthly water inflow forecasting problem, searching for approaches and techniques which could define efficient forecasting models. Three important aspects to define a forecasting model were investigated: data pre-processing techniques, automatic definition of the embedding and the performance assessment of some Artificial Neural Networks and Fuzzy Systems. Hence, the use of the Principal Components Analysis was investigated and, considering the Water Inflow time series as a discrete signal, the Analytic Signal representation could be used to pre-process the data. Furthermore, the embedding was automatically defined using the Dynamic Modelling approach, by using the Average Mutual Information and the False Nearest Neighbors techniques. The forecasting models were implemented by four intelligent models: SONARX network, SONARX-RBF network, ANFIS model and the ESN network. The other line of investigation came up with a new Operation Policy to solve the Operation Planning problem, defining a more stable, reliable and less costly operative decision sequence. It was proposed an approach to stabilize the Thermoelectric Generation dispatches and, as a result, the Operative Marginal Cost. The Predictive Operation Policy Stabilized via Inertial Term produced excellent operation results, improving the performance of the Predictive Policy.

Keywords: operation planning, operation policies, optimization; intelligent systems, analytic signal, dynamic modelling.

LISTA DE ILUSTRAÇÕES

Figura 2.1 - Sistema hidrotérmico de potência 52
Figura 2.2 - Função de custo com complementação térmica para o parque termelétrico brasileiro, em 2006 53
Figura 2.3 –Fluxograma da Política de Operação Preditiva Estabilizada via Termo Inercial – POLFC-IT 61
Figura 2.4 - Função de CMO em função da complementação térmica 64
Figura 2.5 - Trajetória de CMO médio mensal: de julho de 1999 a março de 2009, submercado SE/CO 65
Figura 3.1 - Série de vazões afluentes médias mensais para a Usina Hidrelétrica de Furnas .. 69
Figura 3.2 - Série Normalizada das vazões afluentes da UHE de Furnas 70
Figura 3.3 – Média e desvio padrão mensal para a série de vazões afluentes da UHE de Furnas 72
Figura 3.4 – Série padronizada da série de vazões afluentes da UHE de Furnas 72
Figura 3.5 - Gráfico dos dados de x_1 e x_2 , variáveis correlacionadas 75
Figura 3.6 - Gráfico dos Componentes Principais (PC): z1 e z2, não correlacionados 75
Figura 4.1 - Série de vazões afluentes médias mensais para a Usina Hidrelétrica de Furnas .. 87
Figura 4.2 – Histograma de vazões afluentes médias mensais para a Usina Hidrelétrica de Furnas 87
Figura 4.3 – "Fourier Power Spectrum" das vazões afluentes médias mensais para a Usina Hidrelétrica de Furnas 89
Figura 4.4 - Logaritmo da série histórica de Vazões Afluentes mensais da Usina Hidrelétrica de Furnas 90
Figura 4.5 - Histograma da distribuição de ocorrências do logaritmo da série de VA mensais: 90
Figura 4.6 - Envelope do logaritmo da série histórica de VA mensais da Usina Hidrelétrica de Furnas 95
Figura 4.7 - Ângulo de Fase do logaritmo da série histórica de VA mensais da UHE de Furnas 95
Figura 4.8 - Projeção do Ângulo de Fase do logaritmo da série histórica de VA mensais da Usina Hidrelétrica de Furnas 96
Figura 5.1 – Processo da "*Non-linear Dynamic Modelling*" 98
Figura 5.2 – "*Average Mutual Information*" (AMI) das Vazões Afluentes a UHE de Furnas 103
Figura 5.3 - "*Average Mutual Information*" (AMI) do Envelope do log(Vazões Afluentes) a UHE de Furnas 104
Figura 5.4 - AMI do Ângulo de Fase do log(Vazões Afluentes) a UHE de Furnas 104
Figura 5.5 – "*False Nearest Neighbors*" FNN das Vazões Afluentes à UHE de Furnas 107
Figura 5.6 - "*False Nearest Neighbors*" FNN da Amplitude do Log(VA) à UHE de Furnas 108
Figura 5.7 -"*False Nearest Neighbors*" FNN do Ângulo de Fase do Log(VA) à UHE de Furnas 108
Figura 6.1 – Topologia dos espaços de entrada e saída, e o mapeamento de características ϕ 117
Figura 6.2 – (a) Vetor de pesos de um dado neurônio *i*. (b) Rede SOM bidimensional 118
Figura 6.3 - Três exemplos da dicotomia de separabilidade-φ para diferentes configurações de cinco pontos em duas dimensões: (a) separabilidade linear; (b) separabilidade quadrática; 134
Figura 6.4 – Arquitetura de uma Rede Neural com Função de Base Radial (RBF) 138

Figura 6.5 – Função de pertinência para um conjunto clássico 146
Figura 6.6 – Função de pertinência para um conjunto *fuzzy* (tipo triangular) 146
Figura 6.7 – Representação de conjuntos clássicos e *fuzzy*..... 147
Figura 6.8 – Sistema de Inferência Fuzzy - FIS..... 150
Figura 6.9 – Os 3 principais tipos de Sistemas de Inferência Fuzzy (FIS)..... 152
Figura 6.10 – modelo FIS de Takagi-Sugeno de 2 entradas e 1 saída 153
Figura 6.11 - arquitetura ANFIS de 2 entradas e 1 saída..... 154
Figura 6.12 - Métodos de particionamento do espaço de entrada: (a) particionamento em *grid*, (b) particionamento em *tree*, (c) particionamento em *scatter*..... 157
Figura 6.13 – Rede Recorrente Simples (SRN) 163
Figura 6.14 - Perceptron Multicamada Recorrente (RMLP) 164
Figura 6.15 - Diagrama de uma "Echo State Nerwork" (ESN) 167
Figura 7.1 – Recuperação das vazões afluentes de treinamento: UH Furnas, série padronizada, ... 174
Figura 7.2 – Recuperação das vazões afluentes de treinamento: UH Furnas, série normalizada, ... 174
Figura 7.3 – Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série padronizada, rede SONARX..... 175
Figura 7.4 – Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série normalizada, rede SONARX..... 175
Figura 7.5 – Recuperação das vazões afluentes de treinamento: UH Furnas, série padronizada, rede SONARX-RBF 176
Figura 7.6 – Recuperação das vazões afluentes de treinamento: UH Furnas, série normalizada, rede SONARX-RBF 176
Figura 7.7 – Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série padronizada, rede SONARX-RBF 177
Figura 7.8 – Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série normalizada, rede SONARX-RBF..... 177
Figura 7.9 – Arquitetura ANFIS com particionamento em *grid*: 16 regras *fuzzy*..... 178
Figura 7.10 – Funções de pertinência do modelo ANFIS, particionamento em *grid*: 2 FP... 178
Figura 7.11 – Recuperação das vazões afluentes de treinamento: UH Furnas, série padronizada, modelo ANFIS – particionamento em *grid* 179
Figura 7.12 – Recuperação das vazões afluentes de treinamento: UH Furnas, série normalizada, modelo ANFIS – particionamento em *grid*..... 179
Figura 7.13 – Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série padronizada, modelo ANFIS – particionamento em *grid* 180
Figura 7.14 - Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série normalizada, modelo ANFIS – particionamento em *grid*..... 180
Figura 7.15 - Arquitetura ANFIS com particionamento em cluster: 3 clusters e 3 regras *fuzzy* 181
Figura 7.16 – Funções de pertinência do modelo ANFIS, particionamento em *cluster*: 3 FP181
Figura 7.17 – Recuperação das vazões afluentes de treinamento: UH Furnas, série padronizada, modelo ANFIS – particionamento em *cluster*..... 182
Figura 7.18 - Recuperação das vazões afluentes de treinamento: UH Furnas, série normalizada, modelo ANFIS – particionamento em *cluster*..... 182
Figura 7.19 - Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série padronizada, modelo ANFIS – particionamento em *cluster*..... 183
Figura 7.20 - Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série normalizada, modelo ANFIS – particionamento em *cluster*..... 183

Figura 7.21 - “Echo States” do modelo ESN treinado com histórico de vazões afluentes mensais 186
Figura 7.22 – Padrões de Treinamento: Vazões Afluentes a Usina Hidrelétrica de Furnas... 186
Figura 7.23 - Padrões de Teste (1972 - 1976): Vazões Afluentes a Usina Hidrelétrica de Furnas 187
Figura 7.24 – Previsão do Envelope do Log(Vazão Afluente) da UHE de Furnas 191
Figura 7.25 – Previsão do Ângulo de Fase do Log(Vazão Afluente) da UHE de Furnas 191
Figura 7.26 – Previsão de Vazão Afluente da UHE de Furnas 192
Figura 7.27 – Previsão de VA da UHE de Furnas: vários-passos-a-frente, 12 meses 193
Figura 7.28 – Previsões anuais de VA à UHE de Furnas 194
Figura 8.1 - Sistema teste, parte da Bacia do Rio Paraná 196
Figura 8.2 - Evolução do custo de operação durante o processo de otimização 197
Figura 8.3 - Trajetória de mercado ao longo do período de otimização 198
Figura 8.4 - Trajetória dos volumes ótimos para o período de Jan/1935 até Dez/1944 199
Figura 8.5 - Trajetória dos volumes ótimos para o período de Jan/1945 até Dez/1954 200
Figura 8.6 - Trajetória dos volumes ótimos para o período de Jan/1955 até Dez/1964 201
Figura 8.7 - Trajetória dos volumes ótimos para o período de Jan/1965 até Dez/1974 202
Figura 8.8 - Trajetória dos volumes ótimos para o período de Jan/1975 até Dez/1984 203
Figura 8.9 - Trajetória dos volumes ótimos para o período de Jan/1985 até Dez/1992 204
Figura 8.10 - Trajetória de Volumes Armazenados da UHE de Furnas: Maio/1953 a Abril/1954 207
Figura 8.11 - – Trajetórias das Gerações Termelétricas e da Demanda: Maio/1953 a Abril/1954 208
Figura 8.12 - Trajetórias das Gerações Hidrelétricas e da Demanda: Maio/1953 a Abril/1954 210
Figura 8.13 - Trajetórias dos Custos Marginais de Operação (CMO): Maio/1953 a Abril/1954 211
Figura 8.14 - Trajetória de Volumes Armazenados da UHE de Furnas: Maio/1953 a Abril/1954 214
Figura 8.15 - Trajetória de Volumes Armazenados da UHE de Marimbondo: Maio/1953 a Abril/1954 214
Figura 8.16 - Trajetória de Volumes Armazenados da UHE de Água Vermelha: Maio/1953 a Abril/1954 215
Figura 8.17 - Trajetórias das Gerações Termelétricas e da Demanda: Maio/1953 a Abril/1954 216
Figura 8.18 - Trajetórias das Gerações Hidrelétricas e da Demanda: Maio/1953 a Abril/1954 217
Figura 8.19 - Trajetórias dos Custos Marginais de Operação (CMO): Maio/1953 a Abril/1954 218

LISTA DE TABELAS

Tabela 3.1 – Média e desvio padrão para a série de vazões afluentes da UHE de Furnas 71
Tabela 6.1 - Algoritmo de aprendizagem híbrida, em 2 passos 156
Tabela 7.1 - Resultados da previsão de vazões afluentes da UH de Furnas: período de 1972 a 1976, histórico normalizado e 3 dados de entrada (t , t-1 , t-2) 171
Tabela 7.2 - Resultados da previsão de vazões afluentes da UH de Furnas: período de 1972 a 1976, histórico padronizado e 4 dados de entrada (PCA) 173
Tabela 7.3 - Erros de Previsão das VA da Usina Hidrelétrica de Furnas 188
Tabela 7.4 – Erros de previsão de Vazões Afluentes da UHE de Furnas 193
Tabela 7.5 – Erros de Previsão Anuais da UHE de Furnas 194
Tabela 8.1 - Tabela comparativa entre as Políticas de Operação: Maio/1953 a Abril/1954 .. 212
Tabela 8.2 - Tabela Custo de Operação com relação ao Ótimo Determinístico: Maio/1953 a Abril/1954 212
Tabela 8.3 - Tabela comparativa entre as Políticas de Operação: Maio/1953 a Abril/1954 .. 219
Tabela 8.4 - Tabela Custo de Operação com relação ao Ótimo Determinístico: Maio/1953 a Abril/1954 219

LISTA DE SIGLAS

AG	Algoritmos Genéticos
AMI	"*Average Mutual Information*"
ANFIS	Sistema de Inferência Neuro-*Fuzzy* Adaptativa
AS	"*Analytic Signal*"
CEPEL	Centro de Pesquisas de Energia Elétrica
CLFC	Controle de Malha Fechada
CMO	Custo Marginal de Operação
DET	Despacho Econômico Térmico
EESC	Escola de Engenharia de São Carlos
EPM	Erro Percentual Médio
ESN	"*Echo State Network*"
FIS	Sistema de Inferência *Fuzzy*
FNN	"*False Nearest Neighbors*"
LSEE	Laboratório de Sistemas de Energia Elétrica
MATQV	Memória Associativa Temporal via Quantização Vetorial
MIMO	Múltiplas Entradas/Múltiplas Saídas (*Multiple-Imput/Multiple-Output*)
MLP	Perceptron de Multicamadas (*MultiLayer Perceptron*)
MRE	Erro Relativo Médio (*Mean Relative Error*)
MSE	Erro Médio Quadrático (*Mean Squared Error*)
NARX	Modelo Auto-Regressivo Não-Linear com Variáveis Exógenas
NFC	Controle "Naive"
OLFC	Controle de Malha Parcialmente Aberta
ONS	Operador Nacional do Sistema Elétrico
PC	Componentes Principais
PCA	Análise dos Componentes Principais
PD	Programação Dinâmica
PDDE	Programação Dinâmica Dual Estocástica
PDE	Programação Dinâmica Estocástica
PNL	Programação Não-Linear
POLFC	Política de Controle Parcialmente Aberto
POLFC-IT	Política de Controle Parcialmente Aberto com Termo Inercial

RBF	Rede com Função de Base Radial (*Radial Basis Function*)
RNA	Rede Neural Artificial
ROO	Regras de Operação Otimizadas
ROR	Regra de Operação de Reservatórios
RPH$_i$	Redução do Potencial Hidrelétrico da Usina *i*
SI	Sistemas Inteligentes
SISO	Entrada-Única/Saída-Única (*Single-Imput/Single-Output*)
SOM	Mapa Auto-Organizável de Kohonen (*Self-Organizing Map*)
SONARX	Modelo NARX Auto-Organizável (*Self-Organizing NARX*)
UHE	Usina Hidrelétrica
USP	Universidade de São Paulo
VA	Vazão Afluente
WTA	Rede competitiva *Winner-Take-All*

LISTA DE SÍMBOLOS

$D(t)$	Demanda média durante o intervalo t;
t	intervalo de planejamento;
MW	Mega Watt
$\overline{MW}$	Mega Watt médio
$GH(t)$	geração hidrelétrica do sistema durante o intervalo t;
$GT(t)$	geração termelétrica complementar durante o intervalo t;
$C(t)$	função de custo com complementação térmica durante o intervalo t;
I	número de intervalos do horizonte de planejamento;
j	taxa de juros relativa a cada intervalo do horizonte de planejamento;
$ghi(t)$	geração média da usina hidrelétrica i, durante o intervalo t, em $\overline{MW}$;
k_i	constante que engloba aceleração da gravidade, densidade da água, rendimento turbina-gerador e fatores de conversão, da usina hidrelétrica i;
$vol_i(t)$	volume do reservatório da usina hidrelétrica i, ao final do intervalo t, em hm^3;
$def_i(t)$	defluência média da usina hidrelétrica i, durante o intervalo t, em m^3/s;
$h_{líq,i}(t)$	altura de queda líquida da usina hidrelétrica i, durante o intervalo t, em m;
$turb_k(t)$	turbinagem média da usina hidrelétrica i, durante o intervalo t, em m^3/s;
$af_{inc,i}(t)$	afluência incremental média à usina hidrelétrica i, durante o intervalo t, em m^3/s;
Ω_k	conjunto das usinas à montante da usina hidrelétrica i;
$vol_{min,i}$	limites mínimo para o volume armazenado na usina hidrelétrica i, em hm^3;
$vol_{máx,i}$	limites máximo para o volume armazenado na usina hidrelétrica i, em hm^3;
$turb_{min,i}$	limites mínimo para a turbinagem da usina hidrelétrica i, em m^3/s;
$turb_{máx,i}$	limites máximo para a turbinagem da usina hidrelétrica i, em m^3/s;
λ	fator de acoplamento entre os reservatórios, $0 \leq \lambda \leq 1$;
$x(t)$	vetor de entrada da rede no instante t;
$w_i(t)$	vetor de pesos associado ao neurônio i;
i^*	neurônio vencedor ;
$h(i^*,i;t)$	função de vizinhança das redes SOM e SONARX;
$\alpha(t)$	é a taxa de aprendizagem de pesos para a fase de treinamento;
$u(t)$	vetor de entrada de um sistema qualquer no instante t;
$y(t)$	vetor de saída de um sistema qualquer no instante t;
$F(\cdot)$	função ou mapeamento não-linear de um dado sistema;
$P(t)$	padrão de entrada das RNAs;
$\hat{y}(t+1)$	saída da rede no instante t

$\overrightarrow{dist}$ distância Euclidiana;

$\| \cdot \|$ Norma Euclidiana[1]

$\varphi_j(r)$ Função de Base Radial da unidade neural j, com relação a distância r;

σ_j é a largura do campo receptivo da função de base radial da unidade j;

μ_j é a posição espacial do centro da função de base radial da unidade j;

η taxa de aprendizagem para a camada de saída da rede RBF;

[1] A Norma utilizada neste trabalho de pesquisa foi a Norma Euclidiana Vetorial: $\| \cdot \|_2$.

SUMÁRIO

Capítulo 1 Introdução 29

1.1 JUSTIFICATIVA (MOTIVAÇÃO) DA PESQUISA 33

1.2 OBJETIVOS DA PESQUISA 34

1.3 ORGANIZAÇÃO GERAL DESTE DOCUMENTO 36

Capítulo 2 Políticas de Operação 39

2.1 POLÍTICA DE CONTROLE EM MALHA ABERTA 42

2.1.1 Controladores Sub-Ótimos 43

2.1.2 Política de Controle Ingênua 44

2.1.3 Política de Controle Preditivo 46

2.2 POLÍTICA DE OPERAÇÃO PREDITIVA ESTABILIZADA VIA TERMO INERCIAL 47

2.2.1 Modelo de Previsão 48

2.2.2 Modelo de Otimização 51

2.2.3 Modelo de Simulação 59

2.2.4 Modelo de Formação de Preço: Custo Marginal de Operação 62

Capítulo 3 Pré-Processamento dos Dados 67

3.1 ANÁLISE DE SÉRIES TEMPORAIS: VAZÕES AFLUENTES 68

3.1.1 Padronização da Série 70

3.2 ANÁLISE DOS COMPONENTES PRINCIPAIS 73

3.2.1 Definições de Componentes Principais 74

3.2.2 Determinação dos Componentes Principais 76

3.2.3 Escolha do Número de Componentes Principais 81

Capítulo 4 Processamento Digital de Sinais em Séries Temporais 85

4.1 ANÁLISE ESTATÍSTICA E ESPECTRAL DA SÉRIE DE VAZÕES AFLUENTES 87

4.2 "ANALYTIC SIGNAL" DA SÉRIE DE VAZÕES AFLUENTES 91

4.2.1 Determinação da Amplitude (Envelope) e Ângulo de Fase 94

Capítulo 5 Princípios da "Dynamic Modelling" 97

5.1 "TIME DELAYS" (τ): "AVERAGE MUTUAL INFORMATION" 100

5.2 "EMBEDDING DIMENSION" (D_E): "FALSE NEAREST NEIGHBORS" 105

Capítulo 6 Sistemas Inteligentes 111

6.1 REDE SONARX 112

6.1.1 A Rede SOM 114

6.1.2 A Rede SOM com Modelos de Sistemas Dinâmicos 121

6.1.3 Análise da Convergência do Modelo SONARX 126

6.1.4 Representação Espaço de Estados da Rede SONARX 129

6.2 REDE SONARX-RBF 130

6.2.1 Teorema de Cover: Separabilidade dos Padrões 132

6.2.2 Problema de Interpolação 134

6.2.3 Modelagem da Rede RBF 138

6.3 FUNDAMENTOS DOS SISTEMAS "NEURO-FUZZY" 143

6.3.1 Fundamentos da Lógica "Fuzzy" 144

6.3.2 Sistemas de Inferência "Fuzzy" 149

6.3.3 Sistema de Inferência "Neuro-Fuzzy" Adaptativo – ANFIS 152

6.4 REDE "ECHO STATE NETWORK" – ESN 161

6.4.1 Redes Neurais Recorrentes 162

6.4.2 Modelagem da Rede ESN 166

Capítulo 7 Resultados da Previsão de Vazões Afluentes 169

7.1 AVALIAÇÕES INICIAIS DA PREVISÃO DE VA UTILIZANDO RNA 170

7.2 PREVISÕES UTILIZANDO ANÁLISE DOS COMPONENTES PRINCIPAIS 172

7.3 PREVISÕES UTILIZANDO A REDE "ECHO STATE NETWORK" 184

7.4 PREVISÕES UTILIZANDO "ANALYTIC SIGNAL" E "DYNAMIC MODELLING" 188

Capítulo 8 Política de Operação POLFC com Termo Inercial 195

8.1 O SISTEMA HIDRELÉTRICO DE TESTE 196

8.1.1 Solução da Otimização Determinística para o Histórico de VA 197

8.2 POLFC-IT APLICADO A UMA USINA HIDRELÉTRICA: UHE DE FURNAS 205

8.3 POLFC-IT APLICADO A SUBSISTEMAS DE USINAS HIDRELÉTRICAS 213

8.3.1 Três UHE em Cascata: Furnas, Marimbondo e Água Vermelha 213

Capítulo 9 Conclusões 221

9.1 CONTRIBUIÇÕES DO TRABALHO 223

9.2 TRABALHOS FUTUROS 224

BIBLIOGRAFIA 225

Capítulo 1

Introdução

No planejamento da operação de sistemas hidrotérmicos de potência busca-se definir uma política de operação que atenda a demanda de energia elétrica do sistema, com confiabilidade e pelo menor custo operativo. Em outras palavras, isto equivale a determinar um cronograma ótimo de geração para cada usina do sistema, a cada intervalo, permitindo que o sistema atenda a demanda de forma confiável, sobretudo, minimizando o custo esperado de operação durante o período de planejamento (Pereira, 1985; Carneiro e Kadowaki, 1996).

Este processo de planejamento é muito complexo, uma vez que as afluências futuras não são conhecidas, caracterizando-o como um problema estocástico. Devido ao grande número de usinas envolvidas no processo de otimização, tem-se um grande número de variáveis, característico de problemas de grande porte. Além disso, as usinas que compõem um sistema de geração são altamente interligadas, tanto eletricamente pelo sistema de transmissão como hidraulicamente pelas inúmeras disposições em cascata, caracterizando-o como interconectado. Como a função de custo a ser minimizada e, algumas equações do modelo matemático são não-lineares, temos também um problema não-linear. Por fim, as decisões tomadas em um intervalo de planejamento dependem das decisões tomadas no passado e determinam a evolução futura do sistema, classificando o problema também como dinâmico (Yeh, 1985; Carneiro, Soares Filho e Bond, 1990).

Devido à complexidade do planejamento e aos diferentes aspectos que devem ser abordados, o mesmo é dividido em horizonte de médio, curto e curtíssimo prazos. Nos

horizontes de médio e curto prazo são tratados os aspectos energéticos da operação, que está relacionado aos estados dos reservatórios, em cada intervalo do horizonte de planejamento. Enquanto que no horizonte de curtíssimo prazo são tratados os aspectos elétricos do sistema como, por exemplo, os limites de transmissão elétrica (Pereira, 1985).

O planejamento de médio prazo, que no Brasil é realizado em um horizonte de cinco anos e com discretização mensal, utilizou a Programação Dinâmica Estocástica (PDE) na definição das Políticas de Operação de seu sistema hidrotérmico de potência, durante as décadas de 1970, 80 e meados de 90. Apesar da PDE possibilitar a representação explicita da estocasticidade do problema, ela apresenta a conhecida "maldição da dimensionalidade" quando preservada a representação individualizada das Usinas Hidrelétricas (UHE) (Bellman, 1962; CEPEL, 1977; Yeh, 1985; Braga Jr. et al., 1991).

Diante deste problema, a representação do sistema hidrelétrico precisou ser feita de maneira agregada, ou seja, através de um reservatório equivalente em energia, ao qual aflui energia, ao invés de água (Arvanitidis e Rosing, 1970a; Arvanitidis e Rosing, 1970b; CEPEL e ELETROBRÁS, 1980). Entretanto, a representação do sistema por reservatório equivalente implica em perda na fidelidade da representação matemática do problema, desconsiderando as características individuais de cada usina hidrelétrica. No caso brasileiro, esta agregação é feita por subsistemas.

Em se tratando desta modelagem, tanto a água armazenada em cada reservatório, como as vazões afluentes devem ser transformadas em energia armazenada e energia afluente, respectivamente. Para tanto, faz-se necessária a utilização de Regras de Operação de Reservatórios (ROR), a qual é de fundamental importância para a definição do Reservatório Equivalente em Energia (Fortunato et al., 1985). Quanto mais próxima da realidade forem estas Regras, mais fiel será o Reservatório Equivalente.

São as Regras de Operação que estabelecem a maneira na qual os reservatórios são operados. A Regra adotada no planejamento de operação do sistema brasileiro é uma Regra de Operação em Paralelo, a qual estabelece que todos os reservatórios em um sistema devem operar com a mesma porcentagem de armazenamento dos reservatórios. Entretanto, vários estudos de operação ótima de reservatórios em geração hidrelétrica mostram que o comportamento de cada reservatório depende de sua localização na cascata (Cruz Jr. e Soares Filho, 1999; da Silva e Finardi, 2001).

Ou seja, diferentes posições relativas dentro das cascatas e diferentes características construtivas: capacidade de armazenamento de água, produtibilidade da usina, altura de queda d'água, etc., são características peculiares de cada usina hidrelétrica. O que praticamente inviabiliza desenvolver ROR gerais para simular suas operações (Wood e Wollenberg, 1984).

O reservatório de cabeceira (montante), por exemplo, deve regularizar a vazão afluente utilizando seu reservatório de acumulação, enquanto que os reservatórios de jusante devem ser mantidos cheios a maior parte do tempo, operando como usinas a fio d'água. Vários estudos têm sugerido que a Regra de Operação em Paralelo subestima a capacidade de geração de sistemas hidrotérmicos.

No final da década de 1990, o Operador Nacional do Sistema Elétrico (ONS) adotou para o planejamento da operação em sistemas hidrotérmicos de potência, um modelo baseado em Programação Dinâmica Dual Estocástica (PDDE) (Pereira e Pinto, 1985; Rotting e Gjelsvik, 1992), o qual permite a representação de múltiplos reservatórios em reservatórios equivalentes, divididos em quatro subsistemas (Norte, Nordeste, Sudeste/Centro-Oeste e Sul), procurando preservar as características das bacias hidrográficas.

O modelo NEWAVE, desenvolvido por (CEPEL, 2000), otimiza e simula a operação do sistema ao longo do período de planejamento, para distintos cenários de seqüências hidrológicas, falhas dos componentes e variações da demanda. Calcula índices de

desempenho, tais como a média dos custos de operação, dos custos marginais, o risco de déficit, os valores médios de energia não suprida, de intercâmbio de energia e de geração hidrelétrica e térmica.

As Políticas de Operação são importantes em várias etapas dentro do Planejamento da Operação, principalmente com o objetivo de estudar a performance do Sistema Hidrotérmico através de simulações operativas. Isto porque os modelos de simulação permitem que a operação do sistema seja simulada sob diversas condições hidrológicas, o que é de suma importância para a avaliação das futuras condições de operação do sistema (Carneiro e Kadowaki, 1996).

No processo de Planejamento da Expansão uma das mais importantes etapas é o dimensionamento e o benefício energético de uma usina hidrelétrica. Este dimensionamento visa especificar cinco parâmetros básicos, os quais determinam a capacidade de produção de energia de um aproveitamento: os volumes mínimo e máximo do reservatório, a potência instalada, a altura nominal e a altura efetiva das turbinas. Durante a definição destes parâmetros, alterações em seus valores devem ser consideradas sempre que os benefícios energéticos incrementais resultantes forem maiores que os respectivos acréscimos de custo. O benefício de um aproveitamento equivale à valorização econômica de seus ganhos energéticos. Estes ganhos correspondem não somente à energia que a usina é capaz de gerar, mas também aos eventuais ganhos que esta usina possa proporcionar às outras usinas do sistema no qual ela será inserida.

Assim, para avaliar os benefícios de uma usina hidrelétrica, deve-se primeiro definir um Sistema de Referência, que corresponde ao sistema existente no qual a usina será incorporada. Deve-se então proceder à simulação da operação deste sistema e determinar seus benefícios energéticos, após a inclusão da nova usina hidrelétrica (Silva Filho et al., 2000).

Desta maneira, fica evidente a grande importância para o adequado aproveitamento do nosso parque hidrelétrico o contínuo aprimoramento e desenvolvimento de Políticas de Operação dentro do Planejamento da Operação, que melhor caracterize o comportamento ótimo dos sistemas hidrelétricos de potência, mesmo diante das mais variadas condições hidrológicas.

1.1 Justificativa (Motivação) da Pesquisa

As Políticas de Operação do setor elétrico merecem estudos detalhados e inovadores, a fim de encontrar melhores decisões operativas e sinalizações de mercado – custo marginal de operação, contribuindo para a diminuição do risco de déficit de energia e dando uma maior segurança ao planejamento da operação. Isto pode ser feito com um amplo investimento no estudo deste problema, levando ao aprimoramento dos modelos existentes e também, o que é muito importante, ao desenvolvimento de modelos alternativos de abordagem ao problema, promovendo o aperfeiçoamento e transparência dos modelos utilizados. Estudos comparativos entre modelos, incluindo nacionais e internacionais, se fazem necessários para uma maior confiabilidade na operação do sistema.

Um fator preponderante neste problema é que a abordagem em uso no Brasil, a Programação Dinâmica Dual Estocástica (PDDE), além do problema relacionado à "maldição da dimensionalidade", ao continuar trabalhando com a representação agregada das usinas hidrelétricas, trabalhar com modelos lineares simples para representar a estocasticidade hidrológica, além de fazer uma aproximação da função de custo futuro ao invés de trabalhar com a própria função. Estes fatores acabam prejudicando o desempenho deste tipo de abordagem.

Dentro desta perspectiva, a Política de Operação Preditiva Estabilizada via Termo Inercial utilizando Sistemas Inteligentes na Previsão de Vazões Afluentes proposta neste trabalho aparece como uma excelente alternativa para o problema do Planejamento da Operação. Isto porque ela combina a previsão de vazões com um otimizador não-linear determinístico a usinas individualizadas, podendo preservar a representação de todas as usinas hidrelétricas do sistema, sem sofrer do problema da "maldição da dimensionalidade".

1.2 Objetivos da Pesquisa

Este trabalho de pesquisa objetiva a obtenção de uma nova Política de Operação que melhor caracterize o comportamento ótimo do sistema. Para tanto, um modelo de Programação Não-Linear (PNL), especialmente desenvolvido para a realização da operação determinística de Sistemas Hidrotérmicos de Potência, utilizado para otimizar a operação de um sistema hidrelétrico real com grandes UHE pertencentes ao Sistema Sudeste Brasileiro e, tendo como função objetivo a minimização do custo com complementação térmica.

Considerando as seqüências de afluências do histórico de vazões pôde-se adotar este tipo de abordagem determinística, e com isso definir quanto cada usina deveria gerar a cada instante para que a operação fosse ótima. Apesar deste tipo de análise ser fictícia, ela é muito importante na realização de uma análise do comportamento dos reservatórios, observando-se como eles relacionam-se entre si.

Os Sistemas Inteligentes (Redes Neurais Artificiais - RNA, Sistemas Fuzzy, Neuro-Fuzzy, etc.) são muito apropriados no mapeamento do comportamento de sistemas dinâmicos complexos (Principe, Wang et al. 1998). Este mapeamento entrada-saída (input/output) pode ser feito do tipo caixa preta (black box), já que a descrição analítica do sistema é desconhecida (Principe and Wang 1995).

No contexto deste trabalho de pesquisa, os dados de entrada foram as vazões afluentes históricas, tendo como variáveis da saída a previsão de futuras vazões afluentes. A previsão de futuras vazões afluentes as UHE é de fundamental importância para a tomada de decisão por parte do operador do sistema de geração.

De posse desta previsão, em cada estágio do período de planejamento, o modelo de otimização proposto determina uma decisão ótima com base na atual previsão das futuras vazões afluentes e no atual estado de armazenamento do sistema, caracterizando assim o comportamento otimizado do sistema de geração.

Este trabalho de pesquisa propõe a Política de Operação Preditiva Estabilizada via Termo Inercial utilizando Sistemas Inteligentes na Previsão de Vazões Afluentes, trazendo avanços tanto na obtenção de modelos de previsão de vazões, com o objetivo de se obter melhores previsões, quanto nos modelos de otimização, para que possam encontrar decisões operativas mais estáveis.

Com a captura do comportamento ótimo operativo, a Política de Operação proposta pôde ser integrada a um simulador da operação de sistema hidrelétrico, onde a cada instante de tempo é necessário que ela decida quanto cada usina deve gerar para que a demanda solicitada seja atendida.

Estas simulações não são determinísticas e, a cada intervalo de simulação, as decisões devem ser tomadas sem nenhum conhecimento sobre o futuro. Com isto, buscou-se que esta Política de Operação fosse capaz de reproduzir decisões bem próximas ao comportamento ótimo para todo o Sistema Hidrelétrico.

1.3 Organização Geral deste Documento

Este documento está dividido em 9 capítulos, tendo-se a seguir uma breve descrição deles, iniciando com o Capítulo 2 que trata de Políticas de Operação, apresentando de forma resumida algumas Políticas de Operação aplicadas no Planejamento da Operação, bem como a apresentação da Política de Controle em Malha Aberta.

Ainda no Capítulo 2 é descrita toda a metodologia proposta da Política de Operação Preditiva Estabilizada via Termo Inercial, onde são descritos seus modelos de Previsão, dois modelos de Otimização, modelo de Simulação e o modelo de Formação de Preço de Curto Prazo. Além de descrever de que maneira tais modelos se integram nesta proposta.

Já o Capítulo 3 trata das técnicas de Pré-Processamento dos Dados, a fim de transformar o conjunto de dados em algo bem mais tratável por parte destes Sistemas Inteligentes (SI). Dentre os principais tratamentos aplicáveis a um conjunto de dados, podemos destacar alguns como: Normalização, Padronização, Análise dos Componentes Principais (PCA).

O Capítulo 4 apresenta uma visão alternativa da série temporal de vazões afluentes. Utilizando alguns conceitos de Processamento Digital de Sinais (PDS), este tipo de série temporal pôde ser vista como um sinal discretizado, sujeito a algumas análises deste campo da ciência como a análise espectral. Além disto, foi possível encontrar uma nova representação deste "sinal discretizado": o "*Analytic Signal*" da série de vazões, que é uma representação fasorial da série temporal.

Já o Capítulo 5 discute os principais conceitos da abordagem da "*Dynamic Modelling*" utilizada na representação e identificação de sistemas dinâmicos não-lineares desconhecidos com base nas medições dos dados por ele gerados, como é o caso da série histórica de vazões afluentes. Este capítulo também apresenta um procedimento automatizado capaz de selecionar

as variáveis importantes - dentre as medições passadas, que devem alimentar o modelo de previsão, sem depender da intuição ou experiência do operador.

O Capítulo 6 apresenta os Sistemas Inteligentes (SI) que foram utilizados neste trabalho de pesquisa. Os SI podem ser aplicados isoladamente ou em conjunto para auxiliar o processo decisório. As técnicas estudadas foram: Redes Neurais Artificiais, Lógica Fuzzy e Sistemas de Inferência Fuzzy. Mais especificamente, a rede SOM de Kohonen, a rede RBF, o modelo ANFIS e a rede ESN, sendo que cada uma destas técnicas possui diferentes graus de habilidade para representar o conhecimento humano.

Os Capítulos 7 e 8 estão relacionados à aplicação prática das propostas do trabalho de pesquisa. No Capítulo 7 é apresentada a aplicação das técnicas propostas e os sistemas inteligentes utilizados na obtenção de previsões de futuras vazões afluentes, previsões de um passo à frente e previsões de vários passos à frente para constituir previsões anuais de vazões.

O Capítulo 8 apresenta o desempenho da Política de Operação Preditiva Estabilizada via Termo Inercial proposta neste trabalho, em comparação com algumas outras Políticas de Operação e a própria solução ótima absoluta. As estratégias de operação foram avaliadas em um Sistema Teste, constituído por Usinas Hidrelétricas pertencentes ao Subsistema Sudeste Brasileiro.

Finalmente, o Capítulo 9, apresenta as conclusões decorrentes das análises e dos resultados apresentados neste trabalho de pesquisa, além de discutir algumas sugestões de continuidade na investigação das propostas e contribuições deste trabalho.

Capítulo 2

Políticas de Operação

Como dito anteriormente, no planejamento da operação de sistemas hidrotérmicos de potência busca-se definir uma política de operação que atenda a demanda de energia elétrica do sistema, com confiabilidade e pelo menor custo operativo.

O planejamento da operação de sistemas hidrotérmicos deve levar em consideração uma ampla gama de atividades, desde a otimização pluri-anual de reservatórios até o despacho horário (meia-hora) sob restrições elétricas. Os diferentes horizontes de decisões de programação representam diferentes tipos de impacto no desempenho do sistema, tais como efeitos de médio prazo (probabilidade de déficit energético futuro, valor esperado da geração termelétrica futura, etc.), efeitos de curto prazo (contratos anuais de garantia de pico e atendimento energético entre agentes, programas de manutenção, etc.) e efeitos de curtíssimo prazo (controle de cheia, restrições do sistema de transmissão, etc.). Toda esta complexidade do problema da programação da operação não pode ser representada por um único modelo, sendo então necessário o desenvolvimento de uma "cadeia" de modelos com diferentes níveis de detalhamento na representação do sistema para cada horizonte de planejamento (Pereira, 1985).

Devido ao avanço e popularização dos computadores, os métodos de otimização se tornaram cada vez mais importantes durante as últimas três décadas no gerenciamento, planejamento e operação de sistemas de reservatórios complexos. Tais complexidades deste problema geralmente exigem que as decisões operativas sejam tomadas por um modelo de

otimização e/ou de simulação. A maioria dos modelos de otimização é baseada em algum tipo de programação matemática. A escolha dos métodos depende da característica do sistema considerado, da disponibilidade de dados, e na especificação dos objetivos e restrições do problema.

A Programação Dinâmica (PD), proposta por (Bellman, 1962), tem sido amplamente utilizada no planejamento e gerenciamento de recursos hídricos, bem como para o planejamento da operação hidrotérmica. A popularidade e sucesso desta técnica pode ser atribuída ao fato das características não-lineares e estocásticas destes problemas serem adequadamente tratadas pela formulação da PD (Yeh, 1985). Além disso, a abordagem da PD é uma política de controle de malha fechada destinada para a obtenção, não de valores numéricos ótimos para as variáveis de decisão, mas de uma regra ótima para a seleção, a cada estágio, da decisão ótima para cada estado possível do sistema (Bertsekas, 1995; Martinez e Soares Filho, 2002).

A utilidade da técnica de PD, entretanto, é limitada pela "maldição da dimensionalidade" quando o esforço computacional cresce exponencialmente com o número de variáveis de estado (Bellman, 1962). E, este é exatamente o caso do problema de operação hidrotérmica, onde as variáveis de estado crescem com o número de usinas hidrelétricas e com o número de discretização do espaço de estado, ou seja, níveis de armazenamento e vazão afluente.

Várias abordagens foram sugeridas para superar o problema da dimensionalidade, inclusive a agregação do sistema hidrelétrico através da representação agregada (Arvanitidis e Rosing, 1970a; Arvanitidis e Rosing, 1970b; Durán et al., 1985; Turgeon e Charbonneau, 1998; Cruz Jr. e Soares Filho, 1999).

Entretanto, estas abordagens usam uma representação agregada do sistema hidrelétrico para reduzir a dimensionalidade do problema, perdendo a representação individual de cada

usina hidrelétrica. Logo, perdendo as informações de características construtivas como por exemplo: capacidade de armazenamento de água, produtividade da usina, restrições operativas, posição relativa dentro da cascata, etc.; das quais são extremamente importante para se obter uma estratégia ótima de operação.

Além disso, tem sido proposta a utilização da Programação Dinâmica Dual Estocástica (PDDE), baseada na aproximação da função de custo futuro da PDE de cada estágio por funções lineares por partes. Estas funções de custo futuro aproximadas são obtidas a partir das soluções duais de cada estágio do problema de planejamento e podem ser interpretadas pelos Cortes de Benders, num algoritmo de decomposição estocástica e multi-estágios (Pereira e Pinto, 1985; Pereira, 1989; Rotting e Gjelsvik, 1992; CEPEL, 2000). No entanto, estas aplicações em sistemas reais complexos continuam exigindo uma modelagem via representação agregada, conseqüentemente com perdas das informações individuais das UHE.

Alternativamente, a Política de Controle Preditivo, também conhecida como Política de Controle de Malha Parcialmente Aberta (*Partial Open Loop Feedback Control* – POLFC, em inglês), também tem sido utilizada através da combinação um modelo de otimização determinística do problema de programação hidrotérmica com as previsões de vazões afluentes geradas por um modelo de previsão (Bertsekas, 1995). Em cada estágio do período de planejamento, a POLFC determina a decisão ótima com base nos valores das vazões afluentes futuras, e esta decisão é implementada até que uma nova previsão esteja disponível. Quando esta nova previsão se torna disponível, uma nova decisão ótima é determinada (Becker e Yeh, 1974; Carneiro e Soares Filho, 1989; Yeh et al., 1992; Martinez e Soares Filho, 2002; Barros et al., 2003).

A grande vantagem da Política POLFC é que ela pode ser aplicada numa representação individualizada do sistema hidrelétrico, preservando a informação de cada característica construtiva. Isto é possível porque esta abordagem utiliza um modelo de

otimização determinístico, tratando o aspecto estocástico do problema separadamente através do modelo de previsão.

Os resultados da comparação de desempenho entre PDE e a POLFC feita por (Martinez e Soares Filho, 2002) mostrou maior geração hidrelétrica média para a POLFC. Apesar deste fato, o custo total de operação ficou maior, por causa da maior flutuação da geração térmica complementar, i.e. um maior desvio padrão.

Com base nisto, o presente trabalho de pesquisa tratou de estabilizar a geração térmica complementar, e conseqüentemente, reduziu o custo total de operação para os mesmos níveis da operação ótima absoluta. A solução mais estável foi obtida usando dois modelos de otimização não-lineares determinísticos, com representação individual de cada UHE, e aplicando o Termo Inercial (TI), responsável por estabilizar a geração termelétrica.

2.1 Política de Controle em Malha Aberta

A política de controle malha aberta surge como uma abordagem alternativa ao controle em malha fechada na solução de problemas envolvendo variáveis aleatórias. A idéia principal desta política é considerar as variáveis aleatórias fixadas em algum valor conhecido previamente.

Para o problema de planejamento da operação, o objetivo da política de controle em malha aberta é encontrar uma seqüência de decisões operativas (vazões turbinadas) das usinas hidrelétricas com base em alguma expectativa das variáveis aleatórias (vazões afluentes), i.e. encontrar uma solução determinística hipotética.

Apesar desta aparente simplicidade da política de controle em malha aberta, ela também utiliza realimentações de estados (níveis de armazenamento) para melhorar o desempenho do sistema.

Assim, para cada estágio de planejamento, informações de estado do sistema são utilizadas como condição inicial na solução do problema determinístico para o restante do horizonte de planejamento, mas apenas a solução ótima do primeiro estágio de otimização é implementada no sistema. No estágio seguinte, novas informações de estado do sistema são utilizadas, a partir do qual o processo é repetido até o final do horizonte.

A grande vantagem da política de controle em malha aberta é a possibilidade da utilização de qualquer tipo de modelagem e algoritmo de otimização para resolver o problema determinístico de cada estágio de planejamento.

2.1.1 Controladores Sub-Ótimos

A solução numérica de muitos problemas de otimização seqüencial da Programação Dinâmica é computacionalmente impraticável ou infactível devido a dimensionalidade do problema.

Por esta razão, os problemas reais requerem a utilização de políticas sub-ótimas que podem ser mais facilmente calculadas e implementadas do que a política ótima, mas com a expectativa de não resultar em piora significativa do valor da função objetivo em relação à solução ótima absoluta.

Para o caso da Programação Dinâmica, uma das possibilidades é simplificar ou modificar o modelo para que a PD possa ser computacionalmente factível. No caso do problema de planejamento da operação a PD tem sido utilizada na versão discreta da Programação Dinâmica Estocástica e na versão linearizada da Programação Dinâmica Dual Estocástica, além das simplificações já mencionadas da representação agregada do sistema de usinas hidrelétricas.

Alternativamente, para a solução de problemas de otimização seqüenciais de planejamento da operação existem algumas abordagens de política de controle sub-ótimas. Dentre elas tem-se a Política de Controle Ingênua e a Política de Controle Preditivo (ou Política de Controle de Malha Parcialmente Aberta).

Considerando que J^*_{MA} é uma solução ótima obtida em malha aberta e J^*_{MF} é uma solução ótima obtida em malha fechada, então:

$$J^*_{MF} \leq J^*_{MA} \tag{2.1}$$

pois a classe de controladores em malha aberta não leva em consideração novas informações do sistema ao longo do *horizonte* e, a diferença $J^*_{MA} - J^*_{MF}$ pode ser chamada de *valor da informação* (Bertsekas, 1976).

O objetivo do uso de informações, pelo uso de realimentações, em políticas de controle sub-ótimas é reduzir o custo ótimo de J^*_{MA} à J^*_{MF}.

Evidente que qualquer solução de controle sub-ótima π que leve em consideração informações ao longo do horizonte, somente será apropriada caso o valor da função objetivo J_π satisfaça:

$$J^*_{MF} \leq J_\pi \leq J^*_{MA} \tag{2.2}$$

caso contrário, tais informações prejudicariam o desempenho da política sub-ótima.

2.1.2 Política de Controle Ingênua

A Política de Controle Ingênua (*Naive Feedback Control* – NFC, em inglês) é um esquema de controle baseado na idéia que tem sido utilizada com considerável sucesso por muitos anos. Também chamada de Controlador do Equivalente Certo, sua concepção e justificativa advêm da teoria de realimentação que empregou as realimentações como um

dispositivo capaz de compensar incertezas e ruídos do sistema dinâmico; negligenciando a incerteza ao fixar as quantidades incertas com algum valor típico (por exemplo, o valor esperado) e projetando um esquema de controle realimentado para o correspondente sistema determinístico (Bertsekas, 1976).

No caso do planejamento da operação, a política NFC pode ser implementada ao considerar as incertezas do problema, as vazões afluentes, em seus valores médios, ou seja, utilizar a Média de Longo Termo (MLT) do histórico de vazões no lugar das futuras vazões afluentes. A solução obtida seria uma seqüência de controle essencialmente determinística.

Desta forma, para cada estágio de planejamento, informações de estado do sistema são utilizadas como condição inicial na solução do problema determinístico para o restante do horizonte de planejamento. A política NFC implementa no sistema apenas a solução ótima do primeiro estágio de otimização. No estágio seguinte, novas informações de estado do sistema são utilizados, a partir do qual o processo é repetido até o final do horizonte.

Vale lembrar que o principio do equivalente certo é válido para muitos problemas de controle estocástico lineares, fornecendo resultado da função objetivo bem próximos do valor ótimo. O que pode não ser verdade para sistema não-lineares (Bertsekas, 1995).

Para o caso brasileiro, (Toscano, 2009) apresentou recentemente alguns resultados mostrando que o modelo ODIN de otimização não-linear determinístico e a usinas individualizadas alimentado pelas vazões afluentes da MLT, em comparação ao modelo NEWAVE, que utiliza a PDDE, e em utilização oficial pelo ONS em seu planejamento da operação, apresentaram performance semelhantes e com ligeira vantagem para a política NFC em relação a PDDE.

A política NFC forneceu o despacho hidrotérmico de menor custo operativo, com geração hidráulica com menores variações, não oscilando bruscamente entre períodos secos e

úmidos. Assim, sua operação demonstrou-se mais econômica, segura, e sustentável a longo prazo.

2.1.3 Política de Controle Preditivo

A Política de Controle Preditivo ou Política de Controle em Malha Parcialmente Aberta (*Partial Open Loop Feedback Control* – POLFC, em inglês) é uma forma de controle sub-ótimo considerada intermediária entre o Controle Ótimo e a Política de Controle em Malha Aberta (*Open Loop Feedback Control* – OLFC, em inglês). A política POLFC utiliza medições do passado para estimar as probabilidades condicionadas de variáveis aleatórias, sendo especialmente recomendada em problemas que apresentem previsões destas variáveis aleatórias durante o processo de controle ao longo do horizonte de planejamento (Bertsekas, 1976).

Logo, a política POLFC combina a otimização determinística da solução em malha aberta com modelos de previsão das variáveis aleatórias do problema.

No caso do planejamento da operação, a política POLFC pode ser implementada com a utilização de qualquer modelo de otimização determinístico, alimentado por qualquer modelo de previsão de suas variáveis aleatórias, as vazões afluentes.

Desta forma, para cada estágio de planejamento, informações de estado do sistema e as previsões das futuras vazões afluentes são utilizadas como condição inicial na solução do problema determinístico para o restante do horizonte de planejamento. A política POLFC implementa no sistema apenas a solução ótima do primeiro estágio de otimização. No estágio seguinte, novas informações de estado do sistema e as novas previsões são utilizadas, a partir do qual o processo é repetido até o final do horizonte.

Os resultados apresentados por (Martinez e Soares Filho, 2002) revelaram que em períodos de estiagem, importantes para o planejamento da operação, a política de controle de malha parcialmente aberta (que combina previsão de vazões afluentes com otimização determinística) se revelou mais eficiente do que a política de controle de malha fechada (que utiliza a PDE), onde a quantidade de água disponível é bem limitada e o modelo estocástico não é capaz de prevê-lo eficientemente, ao contrário da técnica de malha parcialmente aberta que é mais eficiente para prever este tipo de cenário.

Como o que ocorreu de fato na crise energética de 2001, quando o modelo NEWAVE não foi capaz de sinalizar corretamente a crise de fornecimento que estava por vir, nos meses que antecederam a subida abrupta do preço da energia de curto prazo (preço spot) e o início do racionamento de energia, dando início à crise energética. Fato semelhante voltou a ocorrer recentemente, no início de 2008, quando do atraso do período chuvoso, sendo verificado uma grande variação do PLD entre os meses de janeiro a maio de 2008 (Sacchi, Marques e Castro, 2008).

2.2 Política de Operação Preditiva Estabilizada via Termo Inercial

Tanto a Programação Dinâmica Estocástica (PDE) quanto a sua versão Dual (PDDE) têm sido extensamente utilizadas no problema de planejamento da operação hidrelétrica em diversos países. Entretanto, apesar de suas habilidades em lidar com as características não-lineares e estocásticas deste problema, estas abordagens exigem grande esforço computacional devido a "maldição da dimensionalidade", o que requer simplificações de modelagem para lidar com sistemas reais.

Outra abordagem para a programação hidrotérmica de longo prazo é a combinação de um modelo de otimização determinístico com um modelo de previsão de vazões afluentes,

numa Política de Controle de Malha Parcialmente Aberta (POLFC) ou Política de Controle Preditiva. Apesar da Política POLFC não exigir nenhuma simplificação na modelagem e apresentar melhor desempenho comparado com a PDE, esta política produz uma solução instável para o Despacho da Geração Termelétrica (GT), que continua sendo muito custosa.

Desta forma, foi proposto neste trabalho de pesquisa um avanço para a Política POLFC que fosse capaz de estabilizar o despacho de GT e de aproximar tal solução da Solução Ótima. A solução mais estável é obtida usando dois modelos de otimização não-linear determinísticos, com representação individualizadas de cada Usina Hidrelétrica (UHE), e o Termo Inercial (TI), responsável por estabilizar o despacho de GT.

2.2.1 Modelo de Previsão

Análise e previsão de vazões afluentes são extremamente importantes no planejamento da operação de sistemas hidrelétricos de potência. Uma das maiores dificuldades de se obter boas previsões da série de vazões afluentes está em seu comportamento "aparentemente" aleatório em seus períodos úmidos e secos de cada ano.

Uma das preocupações deste trabalho de pesquisa foi o estudo e aplicação de técnicas de Pré-Processamento dos Dados. Estas técnicas são de extrema importância, pois, quando aplicadas podem transformar o conjunto de dados em algo bem mais tratável por parte do previsor. Dentre os principais tratamentos aplicáveis a um conjunto de dados, podemos destacar alguns como: Normalização, Padronização, Análise dos Componentes Principais (PCA), etc.

Os modelos de previsão com base na metodologia Box & Jenkins têm sido amplamente utilizados em problemas de previsão de vazões (Salas, Boes e Smith, 1982;

MecLeod, 1994). Entretanto, estes modelos paramétricos auto-regressivos assumem uma relação linear entre os valores da série (Box, Jenkins e Reinsel, 1994).

Nos últimos anos, Redes Neurais Artificiais (ANN) têm sido sugeridas para a análise de séries temporais devido suas habilidades para tratar de mapeamentos de entrada-saída não-lineares (Weigend, Rumelhart e Huberman, 1991; Haykin, 1998). A rede Perceptrons de MultiCamadas (MLP) com o algoritmo "backpropagation", a mais popular delas, tem sido aplicada na previsão de vazões afluentes com resultados interessantes (Atiya et al., 1999). Em seguida, outros tipos de ANN também foram aplicadas em problemas de previsão de vazões com muito bons resultados (Coulibaly e Anctil, 1999; Figueiredo et al., 2004; Sacchi, Carneiro e Araújo, 2004; Valenca, Ludermir e Valenca, 2005).

Todos estes modelos de previsão têm utilizado o modelo Não-Linear Auto-regressivo (NAR) para as séries de Vazões Afluentes, definido como:

$$VA(t+1) = F[VA(t), VA(t-1), ..., VA(t-n_{VA}+1)] \quad \textbf{(2.3)}$$

onde n_{VA} é a ordem de memória do modelo de previsão.

Neste tipo de problema, o uso de ANN normalmente exige um grande esforço na modelagem do mapeamento, uma vez que elas estão tratando de um problema muito difícil do mundo real. Conseqüentemente, isto implica na tarefa de ajustar um grande número de parâmetros.

Outro grande problema em aplicações de previsão é a definição de um espaço de entrada "ótimo", o qual define dentre as medidas do passado, quais devem ser consideradas como dados de entrada para o previsor. E que tradicionalmente tem sido definida pelo processo de "tentativa-e-erro".

A opção encontrada neste trabalho para se definir o espaço de entrada baseou-se na Teoria do Caos (Hilborn, 2000). Caos vinha sendo visto e representado como uma forma de "ruído", quando analisado por ferramentas de processos dinâmicos lineares, tais como a

análise de Fourier. Mais recentemente, entretanto, Caos tem sido caracterizado como algo que compreende a classe de sinais (ou séries temporais) classificados entre movimentos senoidais regulares (ou quasiperiódicos) e imprevisibilidade (verdadeiro comportamento estocástico). (Kaplan e Glass, 1995; Abarbanel, 1996).

Com base nesta teoria, foi possível explorar a obtenção de uma Modelagem Dinâmica Não-Linear usando Sistemas Inteligentes da série de Vazões Afluentes, apresentada por (Abarbanel, 1996; Haykin e Principe, 1998; Principe, Euliano e Lefebvre, 1999).

E para a definição do espaço de entrada "ótimo" foi utilizado o teorema "*Delay-Embedding*" proposto por (Takens, 1981). Takens mostrou que para um sistema dinâmico e observável genérico, o mapeamento das coordenadas de *atraso* (*delay*, em inglês) de um "*manifold*" *d*-dimensional suave e compacto em um "*manifold*" no R^{2d+1} é um *difeomorfismo*, onde *d* é a dimensão do espaço de fase (estado) do sistema dinâmico.

Ainda de acordo com o teorema de Takens, a estrutura geométrica *d*-dimensional de um sistema desconhecido, o qual governa a série temporal, pode ser desdobrada a partir da série temporal observada no espaço reconstruído *D*-dimensional, onde $D \geq 2d+1$, d é a dimensão do espaço de fase (estado) do sistema, D é o "*embedding*" (espaço de entrada) e, d_E é a "*embedding dimension*" (ou seja, a dimensão do espaço de entrada), i.e. o número mínimo de *atrasos* suficiente para a reconstrução da dinâmica da série temporal.

Assim, o mapeamento preditivo $f : R^{d_E} \rightarrow R$ do modelo Não-Linear Auto-regressivo (NAR) determinístico da série de Vazões Afluentes pode ser expresso por:

$$VA(t+T) = f(VA(t), VA(t-\tau), \ldots, VA(t-(d_E-1)\cdot\tau) \qquad \textbf{(2.4)}$$

onde τ é o "*Time Delay*" e d_E é o "*Embedding Dimension*"; os pontos centrais desta reconstrução.

Além disto, explorou-se as técnicas de Processamento Digital de Sinais para o pré-processamento dos dados, por meio da representação, decomposição e visualização da série mensal de vazões afluentes através "*Analytic Signal*" da série (Oppenheim e Schafer, 1999).

Em processamento de sinais, o "*Analytic Signal*" é definido como uma representação por números complexos do sinal original, sendo que a parte real é composta pelo próprio sinal original, enquanto que a parte imaginária é definida pela Transformada de *Hilbert* do sinal original (Marple Jr., 1999; Oppenheim e Schafer, 1999). Esta representação do sinal é bastante útil para se calcular atributos instantâneos da série temporal, especificamente a Amplitude, Ângulo de Fase e Freqüência instantâneas.

2.2.2 Modelo de Otimização

Como dito anteriormente, foram implementados dois modelos de otimização. Estes modelos de otimização da operação estão preocupados com as decisões de médio e curto prazo de um sistema hidrotérmico de potência, aquelas relacionadas às variações de volume dos reservatórios ao longo dos anos e à quantidade de geração do sistema de usinas. O objetivo é encontrar decisões operativas que minimizem o custo de operação do sistema e que garantam níveis mínimos de confiabilidade de suprimento (Pereira, 1985; Soares Filho, 1987; Soares Filho e Carneiro, 1991).

Todo sistema hidrotérmico de potência, tendo em vista o planejamento energético, pôde ser representado por dois subsistemas: um subsistema hidrelétrico, e um subsistema termelétrico, como ilustrado na Figura 2.1, onde toda a geração é injetada a um sistema de transmissão, ilimitado e sem perdas, que irá atender toda a Demanda média *D(t)*, durante o intervalo *t*, em $\overline{MW}$.

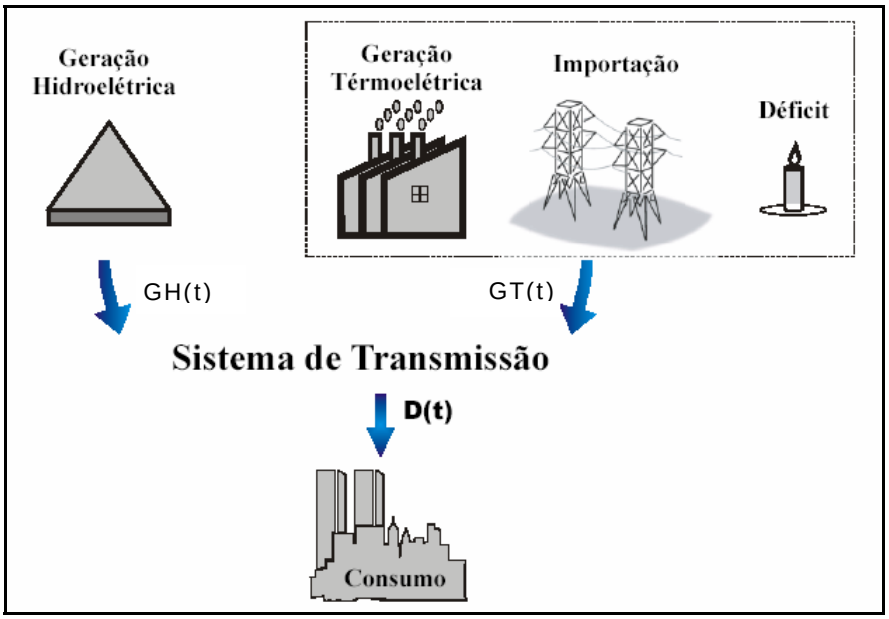


Figura 2.1 - Sistema hidrotérmico de potência

O Subsistema Hidrelétrico é constituído basicamente pelas usinas hidrelétricas, tendo seu custo operativo simplificado e considerado nulo, uma vez que o "combustível" empregado na obtenção de energia é a água. Mesmo em situações onde são cobradas taxas pelo uso da água, o custo de operação das usinas hidrelétricas é relativamente baixo. Ele é responsável por toda a geração das hidrelétricas *GH(t)*, durante o intervalo *t*, em $\overline{MW}$.

Já o Subsistema Termelétrico foi representado através de uma função de custo que engloba o custo do combustível utilizado na operação das usinas termelétricas, o custo de importação de energia de outros sistemas, e pelo custo da falta de suprimento energético, chamado de custo do déficit. Este subsistema é responsável por toda a geração não hidráulica *GT(t)*, chamada de complementação térmica, durante o intervalo *t*, em $\overline{MW}$.

Determinar o comportamento ótimo de um sistema de geração predominantemente hidrelétrico, como o brasileiro, implica em substituir, na medida do possível, a geração de origem não-hidráulica por geração hidráulica. Isto pode ser traduzido por encontrar as trajetórias ótimas de volume armazenado nos reservatórios, o que significa conhecer as vazões turbinadas e vertidas em cada usina, a cada intervalo de tempo (Soares Filho, 1987).

Assim, para cada intervalo mensal *t*, o sistema hidrotérmico tem como custo de operação justamente o custo do sistema não-hidráulico complementar, *C(t)*, o qual depende da demanda *D(t)* e da geração hidráulica *GH(t)*. Quanto maior a geração hidráulica, menor a complementação térmica *GT(t)* correspondente.

A função C(t), ilustrada na Figura 2.2, é o resultado do Despacho Econômico Térmico (DET) (Lyra, Tavares e Soares Filho, 1984) aplicada ao Parque Termelétrico Brasileiro de 2006. Observa-se que primeiramente são utilizadas as fontes com menor custo de geração (usinas nucleares) e conforme a participação destas fontes aumenta, usinas mais caras (usinas a gás natural, carvão, óleo bruto, diesel) vão sendo despachadas até que sejam necessárias importações, e em último caso, cortes de carga cujos custos são representados pelo custo do déficit, de até 5% da demanda.

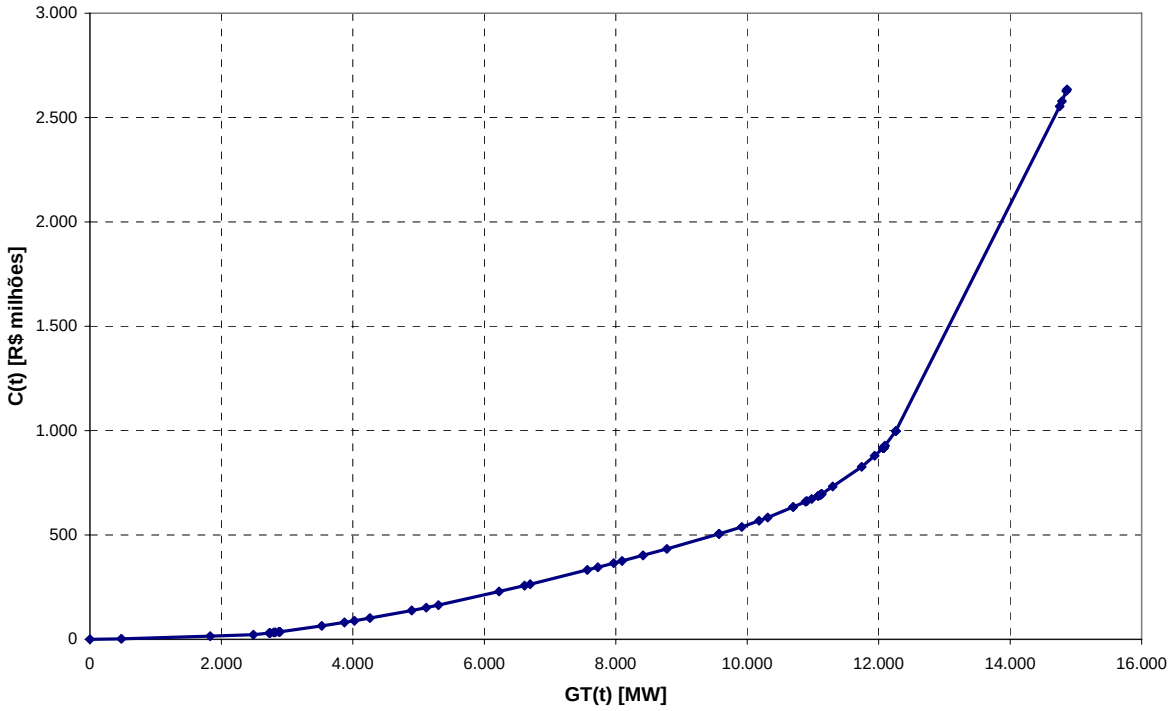


Figura 2.2 - Função de custo com complementação térmica para o parque termelétrico brasileiro, em 2006

Desta maneira, o problema do planejamento ótimo da operação determinística de um sistema hidrotérmico de potência pode ser formulado da seguinte forma:

$$\min \sum_{t=1}^{T} \frac{1}{(1+j)^t} \cdot C(GT(t))$$

$$s.a. \begin{cases} GH(t) + GT(t) = D(t) \\ GH(t) = \sum_{i=1}^{NH} gh_i(t) \\ gh_i(t) = k_i \cdot h_{líq,i}(t) \cdot turb_i(t) \\ h_{líq,i}(t) = h_{mon,i}(vol_i(t)) - h_{jus,i}(def_i(t)) - ph_i(turb_i(t)) \\ def_i(t) = turb_i(t) + vert_i(t) \\ vol_i(t+1) = vol_i(t) + \left(-turb_i(t) - vert_i(t) + af_{inc,i}(t) + \sum_{k \in \Omega_i} turb_k(t) \right) \cdot \Delta t \\ turb_{mín,i}(h_{líq,i}(t)) \le turb_i(t) \le turb_{máx,i}(h_{líq,i}(t)) \\ vol_{mín,i} \le vol_i(t) \le vol_{máx,i} \\ vol_i(0) = Vol_{Inicial,i} \\ vol_i(T) = Vol_{Final,i} \end{cases} \quad \textbf{(2.5)}$$

onde:

T : número de intervalos do horizonte de planejamento;

$C(t)$: custo de operação do parque complementar térmico, em R$;

j : taxa de juros relativa a cada intervalo do horizonte de planejamento;

$D(t)$: demanda média durante o intervalo t, em $\overline{MW}$;

$GT(t)$: geração média das fontes não-hidráulicas durante o intervalo t, em $\overline{MW}$;

$GH(t)$: geração hidráulica total média durante o intervalo t, em $\overline{MW}$;

$gh_i(t)$: geração média da usina hidrelétrica i, durante o intervalo t, em $\overline{MW}$;

k_i : constante que engloba aceleração da gravidade, densidade da água, rendimento médio turbina-gerador e fatores de conversão, da usina hidrelétrica i;

$h_{líq,i}(t)$: altura de queda líquida da usina hidrelétrica i, durante o intervalo t, em m;

$turb_i(t)$: vazão turbinada média da usina hidrelétrica i, durante o intervalo t, em m^3/s;

$h_{mon,i}(t)$: altura de montante da usina hidrelétrica i, durante o intervalo t, em m;

$h_{jus,i}(t)$: altura de jusante da usina hidrelétrica i, durante o intervalo t, em m;

$vol_i(t)$: volume do reservatório da usina hidrelétrica i, ao final do intervalo t, em hm^3;

$def_i(t)$: vazão defluente média da usina hidrelétrica i, durante o intervalo t, em m^3/s;

$vert_i(t)$: vazão vertida média da usina hidrelétrica i, durante o intervalo t, em m^3/s;

$ph_i(t)$: perda hidráulica nas tubulações da da usina hidrelétrica i, durante o intervalo t, em m;

$af_{inc,i}(t)$: vazão afluente incremental média à usina hidrelétrica i, durante o intervalo t, em m^3/s;

Ω_k : conjunto das usinas imediatamente à montante da usina hidrelétrica i;

Δt : intervalo de discretização do horizonte de planejamento, em meses (ou semanas, etc.);

$vol_{min,i}$ e $vol_{máx,i}$: limites mínimo e máximo para o volume armazenado na usina hidrelétrica i, em hm^3;

$turb_{min,i}$ e $turb_{máx,i}$: limites mínimo e máximo para a vazão turbinada da usina hidrelétrica i, em m^3/s;

$Vol_{Inicial,i}$: volume inicial da usina hidrelétrica i, no início do horizonte de planejamento, em hm^3;

$Vol_{Final,i}$: volume final da usina hidrelétrica i, ao final do horizonte de planejamento, em hm^3;

A implementação deste modelo de otimização foi feita através do Software LINGO®, o que facilitou bastante o trabalho. Este software foi escolhido por ter uma interface extremamente amigável e ser capaz de representar de forma simples e direta qualquer modelagem de Programação Não-Linear. Além de possuir um eficiente pacote de algoritmos capaz de resolver este tipo de problema.

Com o objetivo de melhorar a performance da Política POLFC e aproximar a solução da solução ótima determinítica, este trabalho de pesquisa propõe a implementação da Política POLFC Estabilizada via Termo Inercial.

Este Termo Inercial (TI) pode ser entendido como uma representação equivalente do fenômeno físico de momento inercial, onde este TI estaria representando a "massa" de um dado corpo que se move em uma dada direção com velocidade constante, até que este corpo passa a sofrer a ação de uma força perpendicular na tentativa de mudar a trajetória do movimento; neste momento, a massa do corpo tende a impedir (ou minimizar) o conseqüente desvio. Ou seja, quanto maior a massa (TI, para o Problema da Operação), menor a influência

sofrida, o desvio (a variação entre duas Gerações Termelétricas consecutivas: *ΔGT = GT(t)-GT(t-1)*, para o Problema da Operação), pelo corpo.

Desta forma, a implementação do TI se daria da seguinte forma:

$$\Delta GT_{POLFC-IT} = (1-it)\cdot \Delta GT_{POLFC} \qquad \textbf{(2.6)}$$

onde *it* é o Termo Inercial (0 < *it* < 1), ΔGT_{POLFC} seria a variação da Geração Termelétrica resultante da Política POLFC entre dois estágios (meses) consecutivos, $\Delta GT_{POLFC-IT}$ é a variação da Geração Termelétrica minimizada pelo Termo Inercial.

Fazendo o desenvolvimento da Equação 2.6, temos que:

$$\begin{aligned}
&\Delta GT_{POLFC-IT} = (1-it)\cdot \Delta GT_{POLFC} \\
&GT_{POLFC-IT}(t) - GT_{POLFC-IT}(t-1) = (1-it)\cdot (GT_{POLFC}(t) - GT_{POLFC-IT}(t-1)) \\
&GT_{POLFC-IT}(t) - \cancel{GT_{POLFC-IT}(t-1)} = GT_{POLFC}(t) - \cancel{GT_{POLFC-IT}(t-1)} - it\cdot GT_{POLFC}(t) + it\cdot GT_{POLFC-IT}(t-1) \\
&GT_{POLFC-IT}(t) = it\cdot GT_{POLFC-IT}(t-1) + (1-it)\cdot GT_{POLFC}(t)
\end{aligned}$$

Logo, o novo despacho de Geração Termelétrica do estágio (mês) seguinte, *GT(t)*, pode ser expresso da forma:

$$GT_{POLFC-IT}(t) = it\cdot GT_{POLFC-IT}(t-1) + (1-it)\cdot GT_{POLFC}(t) \qquad \textbf{(2.7)}$$

onde *it* é o Termo Inercial (0 < *it* < 1), $GT_{POLFC}(t)$ seria a Geração Termelétrica resultante da Política POLFC para o próximo estágio (mês), $GT_{POLFC-IT}(t-1)$ é a Geração Termelétrica que efetivamente simulada (operada) no último estágio (mês), $GT_{POLFC-IT}(t)$ é a Geração Termelétrica estabilizada via Termo Inercial para o próximo estágio (mês).

Para que fosse possível a utilização desta abordagem com Termo Inercial ao problema com mais de uma UHE, foi necessária a implementação de um segundo modelo de otimização não-linear determinístico, com algumas modificações em relação ao primeiro modelo utilizado – Equações 2.5, para que fosse possível encontrar um despacho ótimo individual de cada UHE, com base na solução do primeiro modelo.

Com isto, os dois modelos passaram a ser utilizados de forma sucessiva, a cada estágio, para a obtenção das:

- Geração Termelétrica total e Geração Hidrelétrica total para o próximo mês *GT(t)* e *GH(t)* , como as variáveis de decisão do primeiro modelo;
- Geração Hidrelétrica de cada UHE para o próximo mês, $gh_i(t)$, passou a ser determinada pelo segundo modelo.

Para a implementação do segundo modelo foram necessárias duas modificações. A primeira modificação foi na Função Objetivo, onde o somatório passou a contar não mais a partir do estágio *t=1*, mas sim do estágio *t=2*. Tal modificação se deve ao fato do Custo Imediato já ter sido definido como conseqüência da solução de despacho da Geração Termelétrica obtida pelo 1º modelo de otimização e, redefinido via Termo Inercial como $GT_{POLFC\text{-}IT}(t)$. Desta maneira, definisse como função objetivo do 2º modelo de otimização, minimizar o Custo Futuro de Operação:

$$\min \sum_{t=2}^{T} \frac{1}{(1+j)^t} \cdot C(GT(t)) \qquad \textbf{(2.8)}$$

A segunda modificação foi a adição de mais uma restrição de igualdade ao conjunto de restrições da Equação 2.5, onde é fixada a Geração Termelétrica do próximo mês pela solução obtida pelo 1º modelo de otimização e, redefinido via Termo Inercial:

$$GT(t=1) = GT_{POLFC-IT}(t) \qquad \textbf{(2.9)}$$

Definidos os dois modelos de otimização, cujos objetivos são de minimizar o Custo de Operação ao longo do horizonte de planejamento, Função Objetivo da Equação 2.5 e Equação 2.8, respectivamente:

$$\min \sum_{t=1}^{T} \frac{1}{(1+j)^t} \cdot C(t) \quad \text{e} \quad \min \sum_{t=2}^{T} \frac{1}{(1+j)^t} \cdot C(t)$$

Desta forma, pode-se analisar as condições de otimalidade de 1a. ordem destes problemas, para o caso de uma UHE com capacidade de armazenamento ilimitada, sem perda de generalidade (Read, 1982), como se segue.

Em se tratando de capacidade de armazenamento irrestrito, as *T* restrições de balanço hídrico podem ser substituídas por uma única restrição para todo o horizonte de planejamento:

$$vol(t+1) = vol(t) - turb(t) - vert(t) + af(t) + \sum_{k \in \Omega_i} turb_k(t) \qquad t=1\ ,\ \dots\ ,\ T$$

$$vol(T) = vol(0) - \sum_{t=1}^{T} turb(t) - \sum_{t=1}^{T} vert(t) + \sum_{t=1}^{T} af(t)$$

A Função Lagrangeana do problema pode ser expressa como:

$$\begin{aligned} L = & \sum_{t=1}^{T} \left(\frac{1}{(1+j)^t} C(GT(t)) + \lambda(t) \cdot (D(t) - GT(t) - GH(t) \right) \\ & + \varphi \cdot \left(vol(T) - vol(0) - \sum_{t=1}^{T} turb(t) - \sum_{t=1}^{T} vert(t) + \sum_{t=1}^{T} af(t) \right) \\ & + \sigma \cdot (vol(T) - Vol_{Final}) \end{aligned} \qquad \textbf{(2.10)}$$

onde $\lambda(t)$, φ e σ são os multiplicadores de Lagrange associados às restrições de atendimento à demanda pelas unidades geradoras, balanço hidráulico de todo o horizonte de planejamento e atendimento à restrição de armazenamento final, respectivamente.

Derivando a Função Lagrangeana *L* pela Geração Termelétrica *GT*, definimos o Custo Marginal de Operação *CMO*:

$$\frac{dL}{dGT(t)} = \frac{dC(GT(t))}{dGT(t)} - \lambda(t) = 0 \quad \rightarrow \quad CMO(t) = \lambda(t) = \frac{dC(GT(t))}{dGT(t)} \qquad \textbf{(2.11)}$$

E derivando a Função Lagrangeana *L* pela Vazão Turbinada *turb*(t), definimos o Valor da Água φ:

$$\frac{dL}{dturb(t)} = -\lambda(t) \cdot \frac{dGH(t)}{dturb(t)} + \varphi = 0 \quad \rightarrow \quad \varphi = \lambda(t) \cdot \frac{dGH(t)}{dturb(t)} \qquad \textbf{(2.12)}$$

Analisando a Equação 2.12, observa-se que a seqüência ótima de decisão iguala o Valor da Água para todo o horizonte de planejamento, ou seja, busca equilibrar uso do recurso hídrico ao longo do horizonte.

Entretanto, de acordo com as Equações 2.11 e 2.12, o CMO correspondente a seqüência ótima de decisão pode apresentar valores distintos ao longo do horizonte de planejamento, sendo constante apenas na hipótese de produtibilidade $\frac{dGH(t)}{dturb(t)}$ constante.

Esta análise em torno da solução ótima para o caso das condições de otimalidade para o caso de uma UHE com capacidade de armazenamento ilimitada se justifica, pois serviu de motivação deste trabalho de pesquisa na busca por um mecanismo que pudesse estabilizar a seqüência de decisões operativas para o caso de Políticas de Controle sub-ótimas.

2.2.3 Modelo de Simulação

Como dito anteriormente, a Política de Controle de Malha Parcialmente Aberta (POLFC) combina um modelo de otimização determinístico do problema de programação hidrotérmica com as previsões de vazões afluentes geradas por um modelo de previsão. Em cada estágio do período de planejamento, a POLFC determina a decisão ótima com base nos valores das vazões afluentes futuras, e esta decisão é implementada até que uma nova previsão esteja disponível. Quando esta nova previsão se torna disponível, uma nova decisão ótima é determinada.

Para o estudo comparativo entre algumas políticas de operação, foi necessário também a implementação de um Simulador de Operação, feita através do Software MatLab®. Este software foi escolhido por também possuir uma interface de programação amigável e permitir a visualização gráfica dos resultados operativos encontrados pelas diferentes Políticas de Operação.

Estes dois modelos: Otimizador – Simulador, são integrados de tal forma que a cada estágio de tempo, a cada mês *t*, é necessário executar uma otimização a partir do estágio atual até o final do horizonte de planejamento, *T*, dada uma seqüência de Vazões Afluentes Previstas. O resultado desta otimização define a decisão sub-ótima, Geração Termelétrica total e a Geração Hidrelétrica total para o próximo mês *GT(t)* e *GH(t)*, a ser implementada no Simulador da Operação, onde a seqüência verificada de Vazões Afluentes será diferente da Prevista. Levando todo o Sistema de Geração para o estágio seguinte, mês seguinte *t*.

O fluxograma da Figura 2.3 ilustra a abordagem proposta neste trabalho de pesquisa, a Política de Operação Preditiva Estabilizada via Termo Inercial. Na figura é possível observar que a cada instante de tempo *t*, o modelo de previsão de Vazões Afluentes fornece uma seqüência de previsões das futuras VA até o final do horizonte de planejamento T, com base no próprio histórico de VA. Estas previsões são fornecidas para os dois modelos de otimização não-linear determinística a usinas individualizadas.

De posse destas previsões e do atual estado do sistema, o primeiro modelo de otimização define uma seqüência de decisões operativas até o final do horizonte de planejamento, mas apenas a informação de Geração Termelétrica total do estágio t $GT_{POLFC}(t)$ é passada adiante.

Sobre a Geração Termelétrica total do estágio t $GT_{POLFC}(t)$ é aplicado o conceito do Termo Inercial afim de estabilizar o despacho termelétrico em relação às decisões de despacho termelétrico tomadas num passado recente, segundo a Equação 2.7, produzindo a Geração Termelétrica total do estágio t $GT_{POLFC-IT}(t)$. Em seguida, no caso da operação de apenas uma única UHE, a tal informação é enviada diretamente ao Simulador. Caso contrário, na hipótese da operação de um sistema de UHE, $GT_{POLFC-IT}(t)$ é enviada ao segundo modelo de otimização.

O segundo modelo de otimização define uma seqüência de decisões operativas a partir do instante *t=2* até o final do horizonte de planejamento T, com base na $GT_{POLFC-IT}(t)$ definida pelo Termo Inercial aplicado à indicação de despacho $GT_{POLFC}(t)$ do modelo de otimização 1. Mas apenas as decisões de Geração Hidrelétrica de cada UHE do primeiro estágio *t* $gh_i(t)$ é enviada ao Simulador.

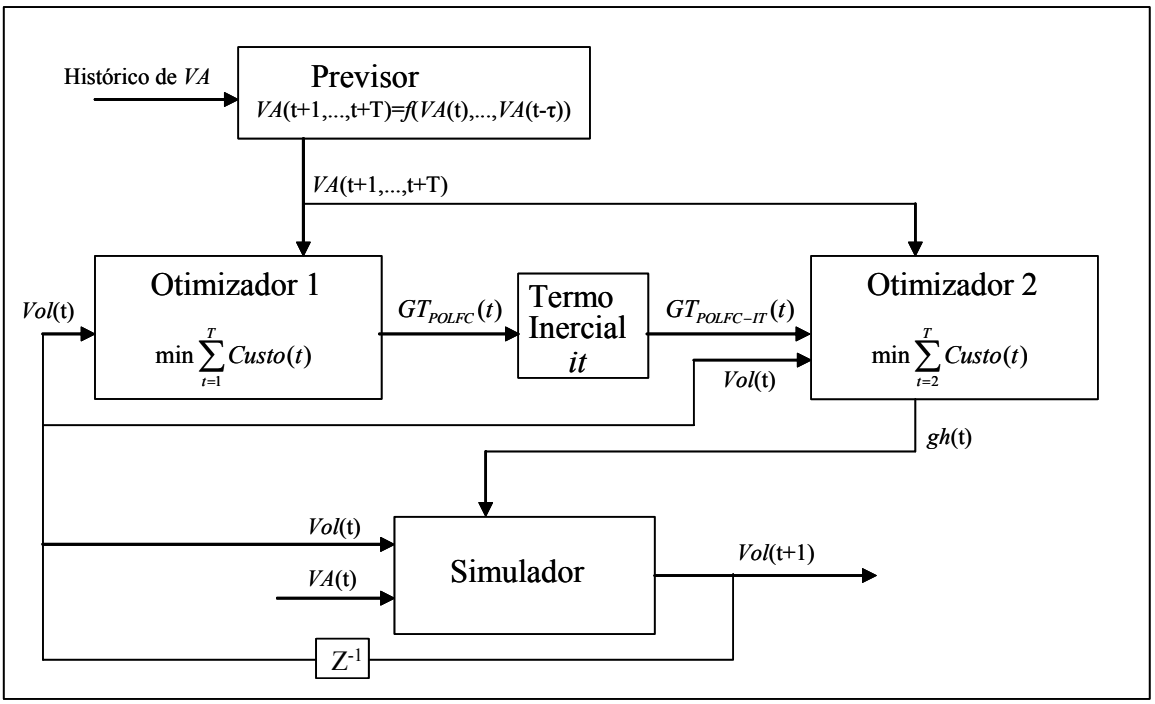


Figura 2.3 –Fluxograma da Política de Operação Preditiva Estabilizada via Termo Inercial – POLFC-IT

Já o modelo de simulação, de posse do atual estado do sistema e das decisões operativas para o primeiro estágio *t*, executa a simulação da operação do sistema diante de uma dada condição hidrológica do estágio *t*, como por exemplo as Vazões Afluentes às UHE do sistema no estágio *t* $VA_i(t)$, tendo como resultado a evolução de todo o sistema para os estados de armazenamento no estágio seguinte *t+1* , $Vol_i(t+1)$.

E finalmente, em se tratando de uma abordagem com realimentações, estes novos estados de armazenamentos $Vol_i(t+1)$ devem realimentar o próprio modelo de simulação

para o estágio seguinte $t+1$, bem como os dois modelos de otimização, para que todo o processo se repita a cada estágio ao longo do horizonte de planejamento.

2.2.4 Modelo de Formação de Preço: Custo Marginal de Operação

O Setor Elétrico Brasileiro passou, depois de décadas, por um processo de reestruturação. Em 1998, como resultado do Projeto de Reestruturação do Setor Elétrico Brasileiro (Projeto RE-SEB), deu-se o início do Novo Modelo do Setor Elétrico, responsável basicamente pela: desverticalização das empresas do setor – em empresas de geração, transmissão, distribuição e comercialização; início do Mercado de Energia Elétrica; e a abertura do setor ao investimento privado em novos empreendimentos (RE-SEB, 1998).

Entretanto, nos anos que se seguiram, um dos pilares fundamentais da reestruturação, a expansão do sistema, não ocorreu no ritmo suficiente para manter o equilíbrio entre Oferta e Demanda de energia elétrica. Em 2001, o setor elétrico sofreu uma grave crise de abastecimento que culminou em um plano de racionamento de energia elétrica. Em 2002, foi instituído o Comitê de Revitalização do Modelo do Setor Elétrico, cujo trabalho resultou na criação, em 2004, do Novo Modelo do Setor Elétrico Brasileiro.

Desta forma, o setor de energia elétrica tem utilizado cada vez mais os indicadores econômicos na determinação das Tarifas de energia para o consumidor final, o Preço de Liquidação da Diferença (PLD), as tarifas do sistema de transmissão, dentre outros. Em especial, os parâmetros Custo Marginal de Operação (CMO) e Valor da Água são utilizados também na decisão operativa do sistema elétrico.

O Brasil optou pela formação do preço da energia elétrica de curto prazo através de modelos computacionais para o despacho otimizado do sistema hidrotérmico de geração.

Neste caso, a metodologia para determinação do PLD é operacionalizada através dos programas NEWAVE / DECOMP.

Apesar de na teoria da Programação Dinâmica Dual Estocástica (PDDE), o CMO ser definido como a derivada da Função de Custo Futuro pelo estado de armazenamento:

$$CMO(t) = -\frac{\partial FCF(t+1)}{\partial EArm(t+1)} \tag{2.13}$$

onde $FCF(t+1)$ é a Função de Custo Futuro do estágio $t+1$ até o final do horizonte de planejamento e $EArm(t+1)$ é o estado de armazenamento do reservatório equivalente no início do estágio $t+1$.

Esta definição advém do desenvolvimento das Equações 2.14 e 2.15, uma vez que a Função de Custo Total é dada pela soma da Função de Custo Imediato FCI(t) do estágio t mais a Função de Custo Futuro $FCF(t+1)$, definida por:

$$CT = FCI(t) + FCF(t+1) \tag{2.14}$$

derivando a Equação 2.14 por $EArm(t+1)$ e igualando a zero, como critério de otimalidade de 1a. ordem, temos:

$$\begin{aligned} &\frac{\partial CT}{\partial Earm(t+1)} = \frac{\partial FCI(t)}{\partial Earm(t+1)} + \frac{\partial FCF(t+1)}{\partial Earm(t+1)} = 0 \\ &\therefore \frac{\partial FCI(t)}{\partial Earm(t+1)} = -\frac{\partial FCF(t+1)}{\partial Earm(t+1)} \end{aligned} \tag{2.15}$$

Diferentemente, neste trabalho de pesquisa propõe-se definir o Custo Marginal de Operação (CMO) como a derivada da função Custo $C(t)$ em relação à Geração Termelétrica $GT(t)$:

$$CMO(t) = \frac{dC(GT(t))}{dGT(t)} \tag{2.16}$$

A Figura 2.4 mostra a função de CMO em função da complementação termelétrica, como conseqüência do DET aplicado ao Parque Termelétrico.

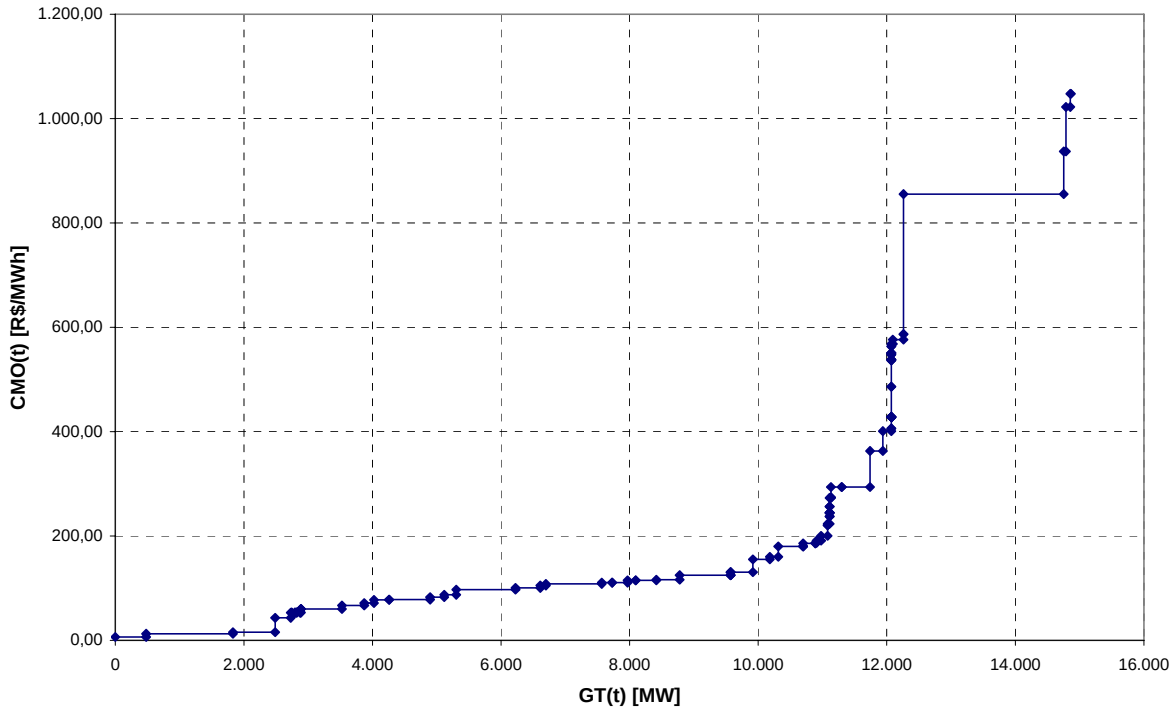


Figura 2.4 - Função de CMO em função da complementação térmica para o parque termelétrico brasileiro, em 2006

Como cada unidade geradora termelétrica define seu Custo Operativo unitário através de um valor constante, i.e. um custo fixo por unidade gerada e independente da quantidade despachada, a função Custo *C(t)* fica sendo definida como uma função linear por partes. Conseqüentemente, a função de CMO fica sendo definida como uma função em degraus e, neste caso apresentando um comportamento exponencial, onde cada patamar está relacionado àquela unidade termelétrica a ser despachada marginalmente.

Com base nas características apresentadas pelas Figuras 2.2 e 2.4, optou-se neste trabalho por uma função de Custo Operativo *C(t)* do tipo polinomial de ordem três, capaz de apresentar esta característica exponencial também para sua derivada de 1ª ordem:

$$C(t) = c \cdot \left[GT(t)\right]^3 \tag{2.17}$$

Já a Figura 2.5 ilustra a grande volatilidade do CMO praticado, no recém criado Mercado de Energia Elétrica Brasileiro, desde julho de 1999 até março de 2009.

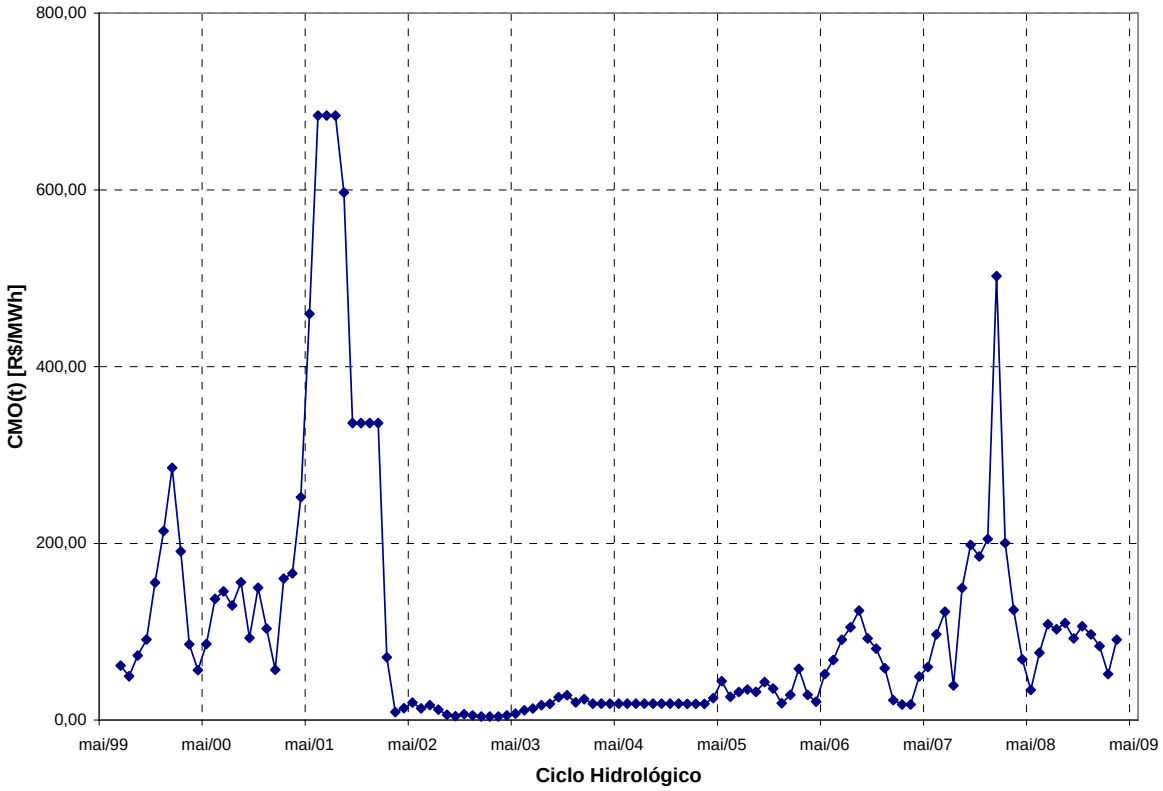


Figura 2.5 - Trajetória de CMO médio mensal: de julho de 1999 a março de 2009, submercado SE/CO

Estes valores de CMO foram resultados obtidos pelo ONS através dos modelos matemático-computacionais NEWAVE e DECOMP, desenvolvidos pelo CEPEL com base na PDDE. Estes são os modelos atualmente vigentes no Setor Elétrico Brasileiro.

Na Figura 2.5 é possível observar claramente que na crise energética de 2001, o modelo NEWAVE não foi capaz de sinalizar com antecedência, através de uma elevação prévia e moderada no CMO, o déficit energético que estava por vir. Com isto, nos meses abril e maio de 2001 houve uma subida abrupta do CMO, o que culminou no racionamento de energia elétrica decretado em junho de 2001.

Este fato vem a ratificar os resultados apresentados por (Martinez e Soares Filho, 2002), onde revelaram que em períodos de estiagem, importantes para o planejamento da operação, a política de controle de malha parcialmente aberta (que combina previsão de vazões afluentes com otimização determinística) se revelou mais eficiente do que a política de controle de malha fechada (que utiliza a PDE), onde a quantidade de água disponível é bem limitada e o modelo estocástico não é capaz de prevê-lo eficientemente, ao contrário da técnica de malha parcialmente aberta que é mais eficiente para prever este tipo de cenário.

Diante disto, foi adotada em 2002 a "Curva de Aversão ao Risco", que representa a evolução ao longo do período dos requisitos mínimos de armazenamento de energia de um subsistema, necessários ao atendimento pleno da carga, sob hipóteses pré-definidas de afluências, intercâmbios inter-regionais e carga, e com toda a geração térmica (inclusive as térmicas emergenciais da Comercializadora Brasileira de Energia Emergencial - CBEE) despachada em sua produção máxima, de forma a se garantir níveis mínimos operativos ao longo do período.

Entretanto, ainda na Figura 2.5 é possível notar uma nova elevação abrupta do CMO no mês de janeiro de 2008, após um grande período em que o CMO se mostrou extremamente baixo, sem qualquer sinalização prévia do que poderia ocorrer com o comportamento do CMO nos primeiros meses de 2008, além de não garantir, por meio de despacho termelétricos preventivos, a segurança do abastecimento energético.

Capítulo 3

Pré-Processamento dos Dados

A aplicação de Sistemas Inteligentes (SI), tais como as Redes Neurais, Sistemas *Fuzzy* e/ou Neuro-*Fuzzy*, utilizados neste trabalho de pesquisa, pode ser feita com maior eficiência se aplicadas técnicas de Pré-Processamento dos Dados, a fim de transformar o conjunto de dados em algo bem mais tratável por parte destes SI.

Dentre os tratamentos convencionais que podem ser aplicados a um conjunto de dados, podemos destacar: Normalização, Padronização, Análise dos Componentes Principais (PCA), etc. , e que serão tratados neste capítulo.

Além destas abordagens, este trabalho de pesquisa investigou a aplicação de um conjunto de técnicas não convencionais para o pré-processamento de séries temporais, tais como: "*Analitic Signal*" – abordagem de Processamento Digital de Sinais, "*Dynamic Modelling*", e que serão discutidas nos próximos dois capítulos.

De volta ao conteúdo deste capítulo, com a finalidade de garantir estabilidade numérica ao SI que esteja sob "treinamento", é usual que o conjunto de dados sofra ao menos o processo de Normalização dos dados. Isto para que todo o conjunto de dados que originalmente possuía diferentes amplitudes de variação para cada uma das dimensões do problema, agora passe a apresentar um Universo de Discurso (*Range*, em inglês) limitado, ou seja, preferencialmente unitário [0,1].

Outro procedimento para o tratamento de um conjunto de dados, que apresente um comportamento não-estacionário, ou seja, variante no tempo, é a Padronização dos dados.

Com isto, é possível transformar uma série temporal não-estacionária, em uma série estacionária, bem mais fácil de ser tratada tanto por ferramentas estatísticas convencionais quanto por não convencionais como os próprios SI.

Neste capítulo serão apresentadas tais ferramentas auxiliares na Análise de Séries Temporais, bem como a Análise dos Componentes Principais, que tem se mostrado extremamente útil e eficaz na redução da dimensão de um conjunto de dados que eventualmente possua alta dimensionalidade e um certo grau de redundância em seus dados.

3.1 Análise de Séries Temporais: Vazões Afluentes

Uma série temporal pode ser descrita, de maneira simplificada, como uma seqüência de valores numéricos que representam a evolução temporal de um processo qualquer.

Formalmente, uma série temporal pode ser definida como um conjunto de observações de uma variável aleatória indexada no tempo, denotada por $[x_t, t \in \Re_+]$, onde $\Re_+$ representa o conjunto dos números reais positivos.

A análise de séries temporais tem como objetivos básicos a modelagem do fenômeno em estudo, a obtenção de conclusões em termos estatísticos e a adequação de um modelo em termos de predição.

Muitos dos métodos estatísticos tradicionais, como os modelos Auto-Regressivos (AR), Periódico Auto-Regressivo (PAR), Auto-Regressivos de Média Móvel (do inglês, *AutoRegressive Moving Average* – ARMA) e Auto-Regressivo Integrado de Média Móvel (do inglês, *AutoRegressive Integrated Moving Average* – ARIMA), para a análise de séries temporais foram aplicados com muito sucesso em diversas séries temporais de problemas reais (Box, Jenkins e Reinsel, 1994). Entretanto, estas técnicas assumem relações lineares

entre as variáveis envolvidas, e como muitas das séries encontradas em aplicações práticas não possuem relações lineares, sua análise e predição tornaram-se menos eficientes.

Neste contexto, nos últimos anos, considerável atenção tem sido dedicada a métodos alternativos para o estudo de séries com padrões não-lineares, destacando-se a utilização de Sistemas Inteligentes (SI) tais como: Redes Neurais Artificiais e Redes Neuro-*Fuzzy* (Ballini, 2000).

Neste trabalho de pesquisa são analisadas as séries de vazões naturais afluentes médias mensais cujos dados foram observados somente para valores inteiros de t. Desta forma, as séries serão denotadas como $[x_1, x_2, x_3, \ldots, x_N]$, sendo x_t a observação da série no instante t = 1,2, ... , N.

A Figura 3.1 mostra, como um exemplo, a série de vazões médias mensais da Usina Hidrelétrica de Furnas, localizada no Rio Grande, compreendendo o período de janeiro de 1931 a dezembro de 2006, totalizando uma série com 912 meses.

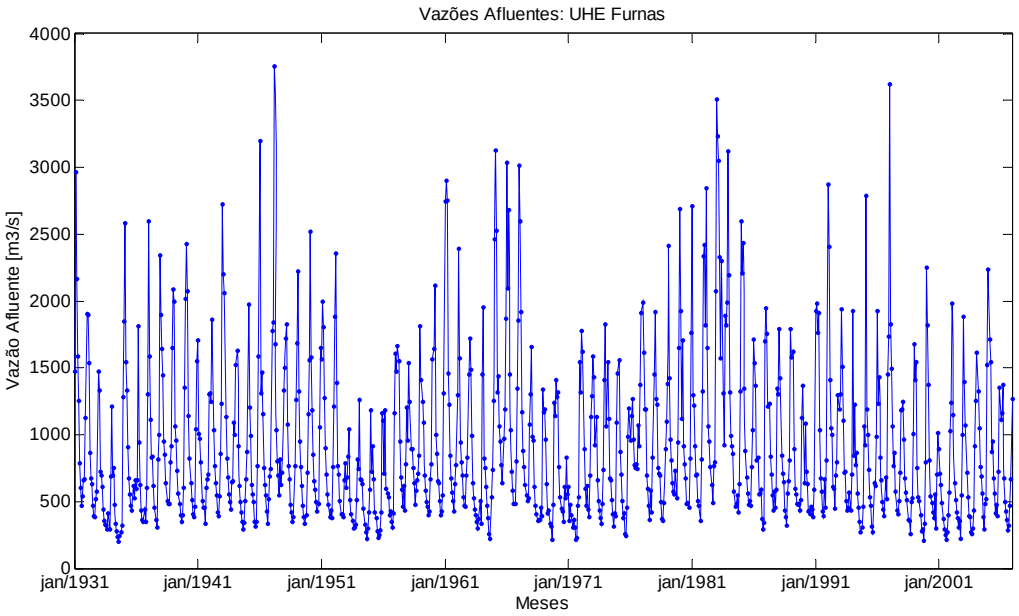


Figura 3.1 - Série de vazões afluentes médias mensais para a Usina Hidrelétrica de Furnas

Num primeiro momento, aplicou-se somente o procedimento de Normalização da série, que consiste em dividir toda a série pelo maior valor registrado, x_{max} , mantendo assim o comportamento sazonal presente na série original, o que pode ser visto na Figura 3.2:

$$z_t = \frac{x_t}{x_{\max}} \tag{3.1}$$

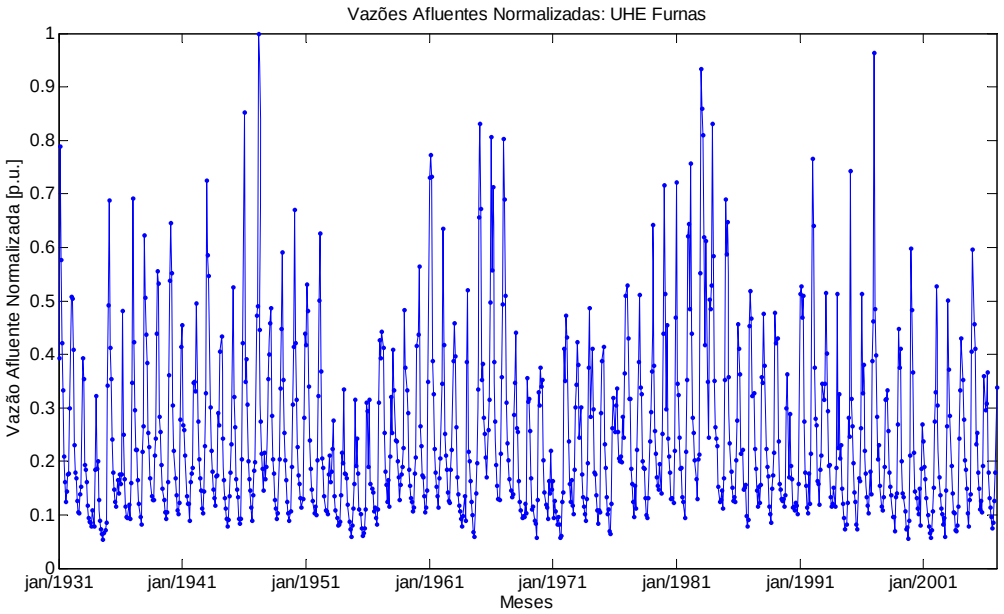


Figura 3.2 - Série Normalizada das vazões afluentes da UHE de Furnas

3.1.1 Padronização da Série

De forma geral, uma série temporal com não-estacionariedade sazonal pode ser representada por dois componentes: um sazonal s_t e outro estacionário $\hat{z}_t$. Ou seja, a vazão x_t pode ser escrita como:

$$x_t = s_t + \hat{z}_t + a_t \tag{3.2}$$

onde a_t é um componente aleatório, com média zero e variância constante, chamado de ruído branco.

Para tornar a série estacionária, é necessário remover a componente sazonal. Para isso realiza-se a seguinte transformação nos dados da série:

$$\hat{z}_{i,m} = \frac{x_{i,m} - \hat{\mu}_m}{\hat{\sigma}_m} \tag{3.3}$$

onde $x_{i,m}$ denota a vazão do ano $i = 1,2,...,n$, no mês $m = 1,2,...,12$.

Sendo que $\hat{\mu}_m$ e $\hat{\sigma}_m$ representam, respectivamente, a média e o desvio padrão de cada mês ao longo de todos os anos do histórico, que podem ser expressos por:

$$\hat{\mu}_m = \frac{1}{n} \cdot \sum_{i=1}^{n} x_{i,m} \tag{3.4}$$

$$\hat{\sigma}_m = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_{i,m} - \hat{\mu}_m)^2} \tag{3.5}$$

A Tabela 3.1 mostra a média e o desvio padrão para cada mês da série de vazões. Esta média histórica mensal da série de vazões é também chamada de Média de Longo Termo (MLT).

Tabela 3.1 – Média e desvio padrão para a série de vazões afluentes da UHE de Furnas

Meses	MLT	Desvio Padrão
Janeiro	1.739,2	677,0
Fevereiro	1.656,1	627,2
Março	1.478,3	593,9
Abril	1.005,5	344,3
Maio	743,3	231,8
Junho	615,5	246,8
Julho	506,7	154,2
Agosto	418,3	121,7
Setembro	439,6	226,8
Outubro	515,1	221,9
Novembro	728,7	309,6
Dezembro	1.240,0	457,9

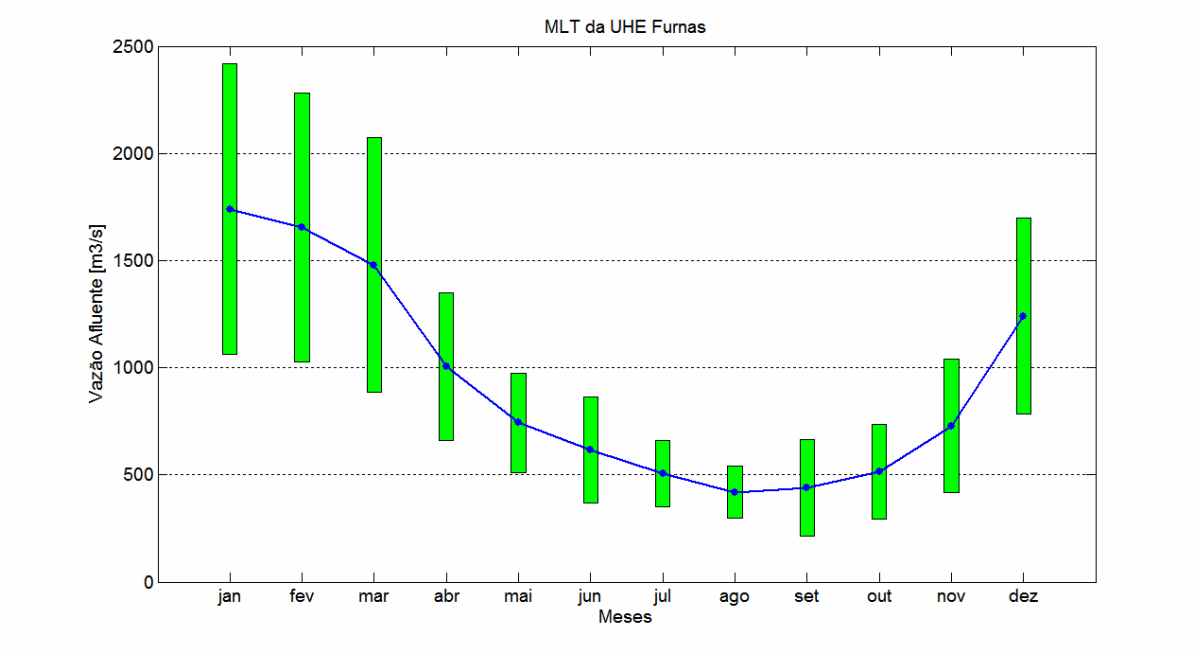


Figura 3.3 – Média e desvio padrão mensal para a série de vazões afluentes da UHE de Furnas

Como pode ser observado na Figura 3.3, a MLT apresenta um comportamento sazonal e bem definido, com seu valor máximo em janeiro e seu valor mínimo em agosto. Além disso, apresenta também um comportamento típico para o desvio padrão ao longo do ano, onde temos um maior desvio padrão justamente em janeiro (de maior afluência média) e um menor desvio padrão em agosto (de menor afluência média).

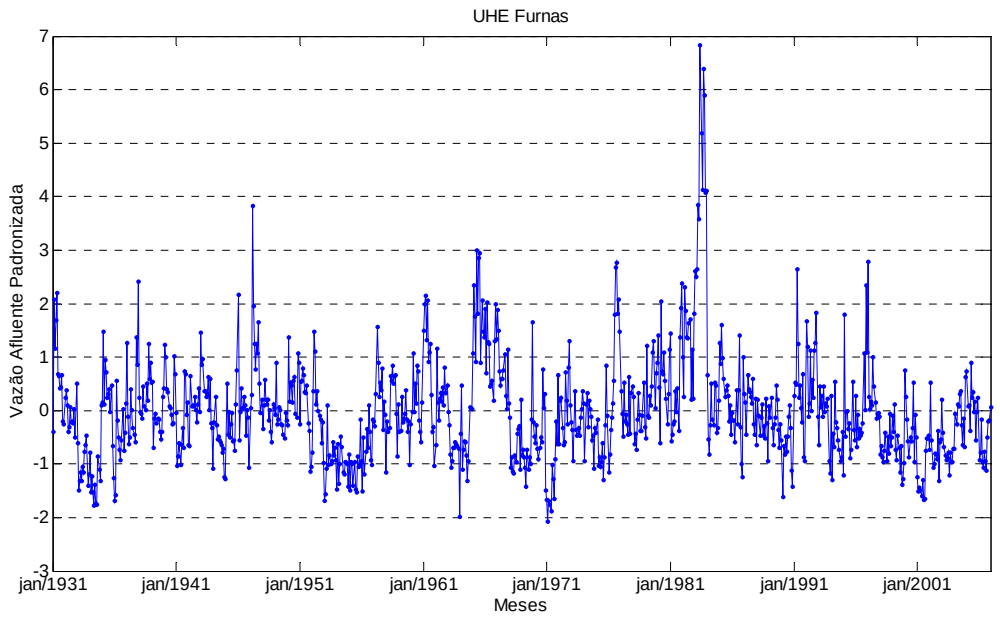


Figura 3.4 – Série padronizada da série de vazões afluentes da UHE de Furnas

A Figura 3.4 mostra a série de vazões afluentes padronizada da UHE de Furnas $\hat{z}_{i,m}$, que possui, aproximadamente, média zero e variância unitária, ou seja:

$$\bar{z} = \frac{1}{N}\frac{1}{12} \cdot \sum_{i=1}^{N}\sum_{m=1}^{12} z_{i,m} \approx 0$$

$$Var[z] = \frac{1}{N}\frac{1}{12} \cdot \sum_{i=1}^{N}\sum_{m=1}^{12} (z_{i,m} - \bar{z})^2 \approx 1.0$$

3.2 Análise dos Componentes Principais

A Análise dos Componentes Principais (*Principal Components Analysis* – PCA, em inglês) é provavelmente a mais antiga e conhecida técnica de análise multivariada. Ela foi inicialmente introduzida por (Pearson, 1901), e posteriormente desenvolvida de forma independente por (Hotelling, 1933). Como vários dos métodos multivariados, ela só passou a ser amplamente utilizada com o advento dos computadores.

A idéia central da Análise dos Componentes Principais (PCA) é reduzir a dimensionalidade de um conjunto de dados que contenha um grande número de variáveis linearmente correlacionadas, enquanto preserva o máximo possível da variação presente no conjunto de dados.

Esta redução de dimensionalidade é obtida pela transformação do conjunto de dados em um novo conjunto de variáveis, os Componentes Principais (PC), os quais são não correlacionados, e ordenados de tal modo que os primeiros possam representar a variação presente em todas as variáveis originais.

O cálculo dos Componentes Principais (PC) se resume à resolução do problema de autovalores e autovetores de uma matriz simétrica semidefinida positiva. Com isto, tem-se que a definição e o cálculo dos PC são bastante diretos, mas esta aparente técnica simples tem uma vasta gama de aplicações, bem como um grande número de diferentes derivações.

3.2.1 Definições de Componentes Principais

Suponha que **x** seja um vetor de p variáveis aleatórias, e que as variâncias das p variáveis aleatórias e a estrutura das covariâncias e correlações entre as p variáveis sejam significativos.

No caso de p grande, isto é, muitas variáveis, normalmente não é conveniente simplesmente observar as p variâncias e todas as ½ $p(p$-1) correlações e covariâncias. Uma alternativa seria observar algumas (<< p) das "novas" variáveis, das quais preservariam a maior parte da informação fornecida por estas variâncias e correlações ou covariâncias. Apesar da PCA não ignorar as covariâncias e correlações, ela dá prioridade às variâncias.

O primeiro passo é observar a função linear $z_1 = \boldsymbol{\alpha}_1' \cdot \mathbf{x}$ dos elementos de **x** que tenham variância máxima, onde $\boldsymbol{\alpha}_1$ é um vetor de p constantes: $\alpha_{11}, \alpha_{12},..., \alpha_{1p}$, e ' denota transposto, tal que:

$$\mathbf{z}_1 = \boldsymbol{\alpha}_1'\mathbf{x} = \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1p}x_p = \sum_{j=1}^{p} \alpha_{1j}x_j \qquad \textbf{(3.6)}$$

Em seguida, observe a função linear $z_1 = \boldsymbol{\alpha}_1' \cdot \mathbf{x}$, não correlacionada com $z_2 = \boldsymbol{\alpha}_2' \cdot \mathbf{x}$, que tenha variância máxima, e assim por diante; até que no k-ésimo estágio a função linear $\boldsymbol{\alpha}_k' \cdot \mathbf{x}$ seja encontrada com a variância máxima e não correlacionada com $z_1 = \boldsymbol{\alpha}_1' \cdot \mathbf{x}$, $z_2 = \boldsymbol{\alpha}_2' \cdot \mathbf{x}$, ... , $z_{k-1} = \boldsymbol{\alpha}_{k-1}' \cdot \mathbf{x}$. Desta forma, a k-ésima "nova" variável, $z_k = \boldsymbol{\alpha}_k' \cdot \mathbf{x}$, é definida como o k-ésimo Componente Principal (PC).

É possível encontrar até p PC, mas se espera que a maior parte da variação em **x** possa ser explicado por m PC, onde $m << p$. Esta redução de dimensionalidade pode ser um tanto complexa, logo, por conveniência, pode-se tomar como um exemplo ilustrativo o caso simples onde p = 2. A vantagem de p = 2 é de que os dados possam ser visualizados num plano, bidimensional, como pode ser visto na Figura 3.5.

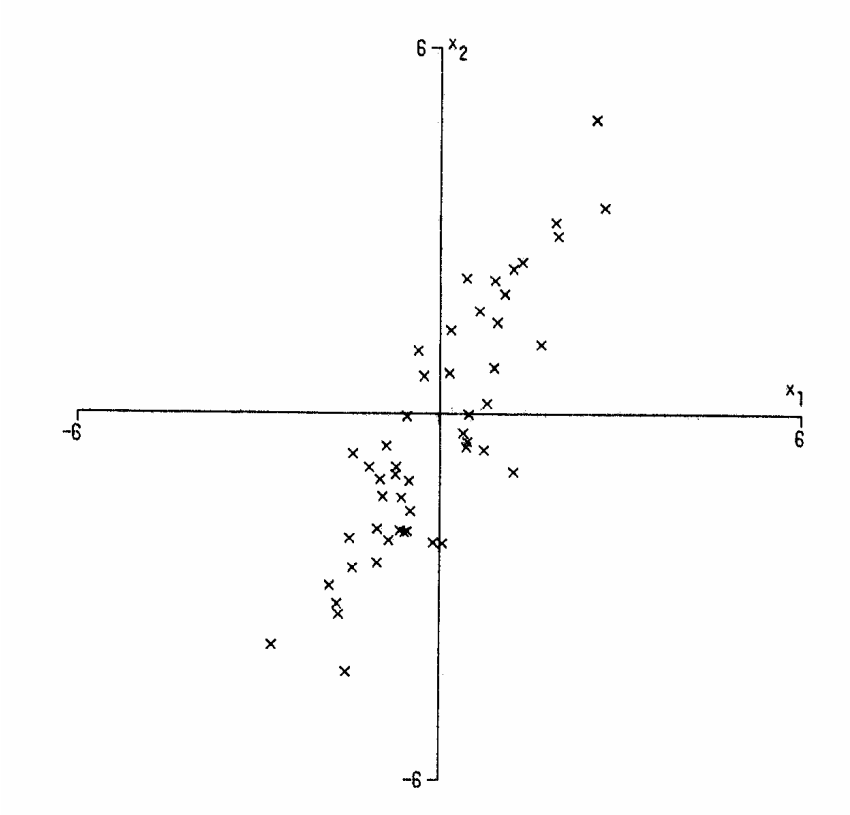


Figura 3.5 - Gráfico dos dados de x_1 e x_2 , variáveis correlacionadas

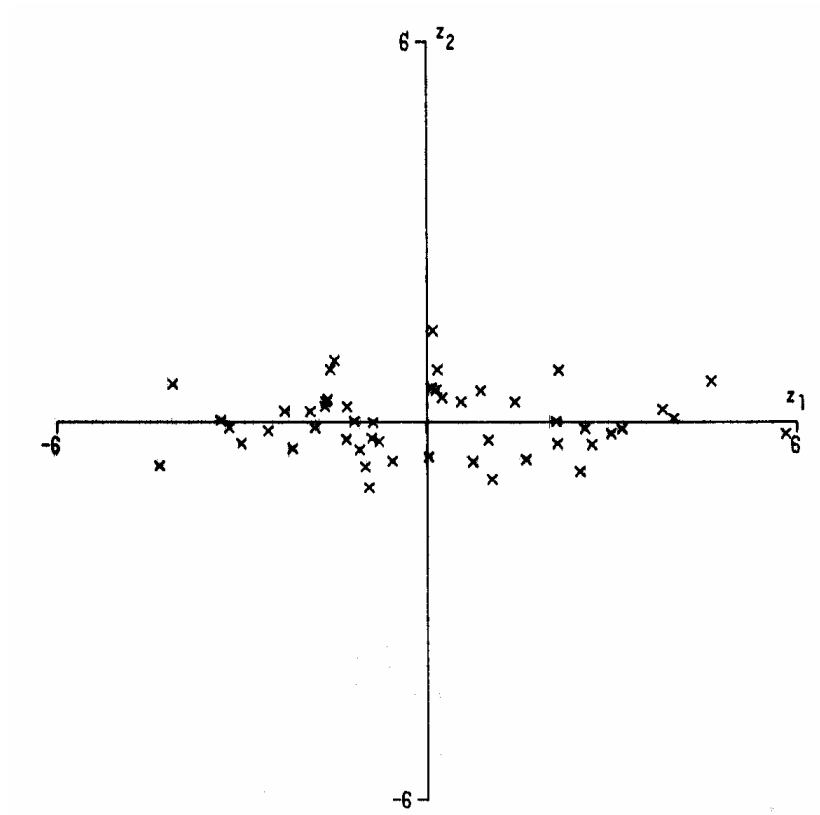


Figura 3.6 - Gráfico dos Componentes Principais (PC): z1 e z2, não correlacionados

A Figura 3.5 mostra os dados correlacionados das variáveis x_1 e x_2. Há uma considerável variação para as duas variáveis, embora uma variação um pouco maior na direção de x_2 do que de x_1.

Se processada a transformação nos Componentes Principais (PC) z_1 e z_2, obtém-se o gráfico da Figura 3.6 , onde se pode notar uma grande variação na direção de z_1 , bem maior que a pequena variação na direção de z_2. Além disto, pode-se observar que os PC são não correlacionados, isto porque, como será visto mais adiante, eles serão definidos como "novas" variáveis ortogonais entre si.

Genericamente, se um conjunto de p variáveis tiverem correlações substanciais entre si, então os primeiros PC irão representar a maior parte da variação das variáveis originais. Por outro lado, os últimos PC irão identificar direções com pouca variação, isto é, eles identificarão relações quase que constantes entre as variáveis originais.

Definidos os PC, é preciso saber como determiná-los. Considerando, por enquanto, o caso onde o vetor das variáveis aleatórias **x** possui uma Matriz de Covariância conhecida **C**, onde seu elementos (i,j) são as covariâncias entre os i-ésimo e o j-ésimo elementos de **x** para $i \neq j$, e são a variância do j-ésimo elemento de **x** para $i = j$.

Desta maneira, para $k = 1, 2, \ldots , p$, o k-ésimo PC é dado por $z_k = \boldsymbol{\alpha_k}'\cdot\mathbf{x}$ onde $\boldsymbol{\alpha_k}$ é um autovetor da matriz de covariância **C** correspondente ao k-ésimo maior autovalor λ_k . Além disto, se $\boldsymbol{\alpha_k}$ for definido como um vetor unitário ($\boldsymbol{\alpha_k}'\cdot\boldsymbol{\alpha_k} = 1$), então $\mathrm{var}(z_k) = \lambda_k$, ou seja, a variância de z_k será o próprio autovalor λ_k .

3.2.2 Determinação dos Componentes Principais

Esta ferramenta utiliza-se de variáveis de entrada que, combinadas linearmente, produzem outras de mesma dimensão não correlacionadas. A falta de correlação entre os PC é

uma importante propriedade porque significa que os índices estão medindo diferentes "dimensões" dos dados (Manly, 1986). Os PC são dados pela Equação 3.7 (Kendall, 1975), assumindo que:

$$z_i = \alpha_{i1}x_1 + \alpha_{i2}x_2 + \cdots + \alpha_{ip}x_p \qquad i = 1, \ldots, p \tag{3.7}$$

$$\mathrm{var}(z_i) = \lambda_i \tag{3.8}$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0 \tag{3.9}$$

sendo:

z_i Componentes Principais;

α Autovetores associados aos correspondentes autovalores;

x Valores originais;

λ Autovalores das PC.

A PCA se resume em encontrar autovalores da matriz de variância e covariância dos dados **C**. Portanto, deve-se encontrar para cada variável o valor médio, os desvios entre os valores originais da variável e o valor médio, o desvio padrão e, por fim, a variância e covariância. Em outras palavras, as variáveis devem ser padronizadas para que tenham média *0* e variância *1* antes de proceder à análise, de maneira a evitar uma grande influência de uma variável nos Componentes Principais.

A matriz de variância e covariância é simétrica e tem a forma:

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix} = \begin{bmatrix} 1 & c'_{12} & \cdots & c'_{1p} \\ c'_{21} & 1 & \cdots & c'_{2p} \\ \vdots & \vdots & & \vdots \\ c'_{p1} & c'_{p2} & \cdots & 1 \end{bmatrix} \tag{3.10}$$

A PCA, como mencionado anteriormente, é realizada sobre a matriz de correlação dos dados. Para encontrar os autovalores da matriz de correlação basta resolver a seguinte equação, conforme (Kendall, 1975), porém utilizando a notação de (Manly, 1986):

$$C \cdot A^T = A^T \cdot \Lambda \tag{3.11}$$

ou ainda:

$$\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}^T = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}^T \cdot \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix} \tag{3.12}$$

sendo:

C Matriz de variância e covariância;

Λ Matriz dos autovalores;

A Matriz dos autovetores associados aos autovalores;

λ_i Autovalor *i* da matriz de variância e covariância C;

$\boldsymbol{\alpha_i}$ Autovetor *i* associado aos autovalor *i*.

Executando a multiplicação de matrizes tem-se p^2 equações. Considerando apenas o conjunto das *p* equações derivadas da primeira coluna do produto das matrizes tem-se:

$$\begin{aligned} c_{11} \cdot a_{11} + c_{12} \cdot a_{12} + \cdots + c_{1p} \cdot a_{1p} &= a_{11} \cdot \lambda_1 \\ c_{21} \cdot a_{11} + c_{22} \cdot a_{12} + \cdots + c_{2p} \cdot a_{1p} &= a_{12} \cdot \lambda_1 \\ \cdots \quad + \cdots + \quad \cdots \quad &= \quad \cdots \\ c_{p1} \cdot a_{11} + c_{p2} \cdot a_{12} + \cdots + c_{pp} \cdot a_{1p} &= a_{1p} \cdot \lambda_1 \end{aligned} \tag{3.13}$$

ou ainda:

$$\begin{aligned} (c_{11} - \lambda_1) \cdot a_{11} + c_{12} \cdot a_{12} + \cdots + c_{1p} \cdot a_{1p} &= 0 \\ c_{21} \cdot a_{11} + (c_{22} - \lambda_1) \cdot a_{12} + \cdots + c_{2p} \cdot a_{1p} &= 0 \\ \cdots \qquad &= \vdots \\ c_{p1} \cdot a_{11} + c_{p2} \cdot a_{12} + \cdots + (c_{pp} - \lambda_1) \cdot a_{1p} &= 0 \end{aligned} \tag{3.14}$$

eliminando-se os α , tem-se:

$$\left|C - \lambda_1 \cdot I\right| = 0 \tag{3.15}$$

$$\begin{vmatrix} c_{11} - \lambda_1 & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} - \lambda_1 & \cdots & c_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} - \lambda_1 \end{vmatrix} = 0 \tag{3.16}$$

generalizando, tem-se p raízes:

$$\left|C - \boldsymbol{\lambda} \cdot I\right| = 0 \tag{3.17}$$

Os λ são os autovalores da matriz de correlação que representam as variâncias dos Componentes Principais (var(z_i) = λ_i). Os valores de α associados aos autovalores são chamados autovetores. Os autovalores obtidos da solução da matriz de correlação são reais não negativos (Kendall, 1975).

Uma outra propriedade importante dos autovalores λ é que, quando somados, correspondem à soma dos elementos da diagonal principal (traço) de **C**, ou seja, a soma das variâncias dos p Componentes Principais é igual à soma das variâncias das variáveis originais, descrito como a seguir por (Manly, 1986):

$$\lambda_1 + \lambda_2 + \cdots + \lambda_p = c_{11} + c_{22} + \cdots + c_{pp} \tag{3.18}$$

Os autovetores são obtidos a partir da Equação 3.11 após a determinação dos autovalores. Da Álgebra Linear tem-se que, se autovalores são distintos, os autovetores associados são únicos. Da proposição anterior, a obtenção da matriz **A** tem como condição necessária que os autovalores sejam distintos e, portanto, a variância do Componente Principal é única. A matriz **A** obtida pela técnica PCA é uma matriz ortogonal, no sentido em que os autovalores de **C**, dispostos nas colunas, satisfazem as condições de ortogonalidade, como a seguir:

$$\boldsymbol{\alpha}_{\mathbf{i}}^{\mathbf{T}} \cdot \boldsymbol{\alpha}_{\mathbf{k}} = \begin{cases} 1, & \mathrm{i} = \mathrm{k} \\ 0, & \mathrm{i} \neq \mathrm{k} \end{cases} \tag{3.19}$$

Isto significa que as colunas da matriz **A** representam vetores linearmente independentes, portanto pode-se escrever:

$$A^T \cdot A = I \tag{3.20}$$

$$A^T = A^{-1} \tag{3.21}$$

sendo:

I Matriz identidade

A partir disto, deduz-se que a inversa da matriz **A** é igual à sua transposta, assim pode-se reescrever a Equação 3.11 como a seguir:

$$A^T \cdot C \cdot A = \Lambda \tag{3.22}$$

$$\boldsymbol{\alpha}_{\mathbf{i}}^{\mathbf{T}} \cdot C \cdot \boldsymbol{\alpha}_{\mathbf{k}} = \begin{cases} \lambda_i, & \mathrm{i} = \mathrm{k} \\ 0, & \mathrm{i} \neq \mathrm{k} \end{cases} \tag{3.23}$$

Para reconstrução do vetor de dados original **x**, a partir dos Componentes Principais z_i associados aos autovalores, em ordem crescente, procede-se como a seguir:

$$z = \begin{bmatrix} z_1 & z_2 & \cdots & z_p \end{bmatrix}^T \tag{3.24}$$

$$z = \left[x^T \cdot A\right]^T \tag{3.25}$$

$$z = A^T \cdot x \tag{3.26}$$

Multiplicando-se ambos os lados da Equação 3.26 pela matriz **A** e simplificando o vetor **x** de dados originais, pode ser reconstituído da seguinte forma:

$$\mathbf{x} = A \cdot \mathbf{z} = \sum_{i=1}^{p} z_i \cdot \boldsymbol{\alpha}_{\mathbf{i}} \tag{3.27}$$

Porém, o que se deseja é uma redução de dimensionalidade dos PC para a representação da variável, assim o vetor de dados aproximado pode ser obtido através do truncamento da Equação 3.27 depois do *m*-ésimo termo, como se segue:

$$\mathbf{x}' = \sum_{i=1}^{m} z_i \cdot \boldsymbol{\alpha}_\mathbf{i} \qquad m < p \tag{3.28}$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq \cdots \geq \lambda_p \tag{3.29}$$

O valor de *m* é definido pelos maiores autovalores $\lambda_1, \lambda_2, \ldots, \lambda_m$ da matriz de correlação **C**, descartando as combinações lineares com as mais baixas variâncias. Quanto mais próximos de zero os autovalores, mais efetiva será a redução de dimensionalidade na preservação das informações contidas nos dados de entrada.

3.2.3 Escolha do Número de Componentes Principais

Esta seção mostra como se pode definir adequadamente um subconjunto de Componentes Principais (PC) que representem a variação total dos dados originais **x**. O principal objetivo em muitas aplicações de PCA é substituir os *p* elementos de **x** por um número *m* bem menor de PC, e que descartem muito pouca informação. É crucial que se saiba o número de PC, $m << p$, sem perdas consideráveis de informação.

Usar *m* PC ao invés das *p* variáveis originais **x** reduz consideravelmente a dimensionalidade do problema, se $m << p$; apesar disto, os valores de todas as *p* variáveis originais continuam sendo utilizadas no cálculo dos PC, uma vez que eles são função de todas as *p* variáveis originais **x**.

Várias regras, em sua maioria *ad hoc*, vêm sendo propostas para determinar valores adequados de *m*, número de PC. Dentre as quais podemos citar:

i) porcentagem cumulada da variação total;
ii) tamanho das variâncias dos PC;
iii) o gráfico "scree" ou o diagrama de log-autovalor (LEV);
iv) número de Componentes com autovalores desiguais;
v) escolha de m por validação cruzada;
vi) correlação parcial.

As três primeiras regras para a escolha de m são consideradas bastante empíricas, isto porque, apesar de várias tentativas de deixá-las com um embasamento formal, elas continuam sendo intuitivamente plausíveis, além de funcionarem bem na prática. A quarta regra já possui um certo embasamento matemático-estatístico, mas é baseada nas suposições da distribuição que não é muito realista e, em alguns casos, seleciona mais PC do que seriam necessários na prática. As duas últimas regras também têm embasamento estatístico, a quinta se baseia na validação cruzada e na Decomposição de Valor Singular (SVD), enquanto que a sexta nas correlações parciais.

Dentre todas estas opções de regra para a definição de m, optou-se pela regra (i): *porcentagem cumulada da variação total*. Esta escolha se deve ao fato de ser um critério bastante obvio e razoavelmente simples de ser implementado. Sendo assim, esta regra é apresentada a seguir.

i) porcentagem cumulada da variação total

Talvez o critério mais óbvio para a escolha de m PC seja selecionar um percentual acumulado da variação total, que representem cerca de 80% ou 90% do total de variação das variáveis originais $\mathbf{x}$. Assim, o número necessário de PC seria o menor valor de m para o qual esta porcentagem fosse atingida.

Os Componentes Principais (PC) são escolhidos sucessivamente para que tenham a maior variância possível, sendo que a variância do j-ésimo PC é λj. Além disso, a soma das

variâncias dos p PC é igual a soma das variâncias das p variáveis originais **x**, como visto na Equação 3.18, e reescrita:

$$\sum_{j=1}^{p} \lambda_j = \sum_{j=1}^{p} c_{jj} \tag{3.30}$$

A "porcentagem da variação" dos primeiros k PC pode ser definida por:

$$t_k = 100 \cdot \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{p} c_{jj}} = 100 \cdot \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{p} \lambda_j} \tag{3.31}$$

Sendo que no caso da matriz de correlação, pode ser reescrito como:

$$t_k = \frac{100}{p} \cdot \sum_{j=1}^{k} \lambda_j \tag{3.32}$$

Para a concepção de uma regra, basta que seja definido um limite mínimo percentual, t^*, algo entre 70% e 90%. Em seguida, que os primeiro m PC representem a maior parte das informações contidas nos dados originais **x**;onde m é definido como o menor inteiro, k, para o qual $t_k > t^*$.

Uma série de tentativas devem ser feitas para encontrar m PC, de tal maneira que se atinja este limite mínimo percentual, $t_k > t^*$, exigido para cada tipo de problema.

Capítulo 4

Processamento Digital de Sinais em Séries Temporais

Muitos dos métodos estatísticos tradicionais para a análise de séries temporais foram aplicados com muito sucesso em diversas séries temporais de problemas reais (Box, Jenkins e Reinsel, 1994). Entretanto, estas técnicas assumem relações lineares entre as variáveis envolvidas, e como muitas das séries encontradas em aplicações práticas não possuem relações lineares, sua análise e predição tornaram-se pouco eficientes.

Neste contexto, considerável atenção tem sido dedicada a métodos alternativos para o estudo de séries com padrões não-lineares, destacando-se a utilização de Sistemas Inteligentes (SI) tais como: Redes Neurais Artificiais e Redes Neuro-Fuzzy e/ou Neuro-Fuzzy (Stokelj, Paravan e Golob, 2002; Figueiredo et al., 2004; Kim, Jeong e Ko, 2006; Sacchi et al., 2007).

A utilização de SI pode ser feita com maior eficiência e robustez se aplicadas técnicas de Pré-Processamento dos Dados; a fim de transformar o conjunto de dados em algo bem mais tratável por parte destes SI.

Para o tratamento de um conjunto de dados que apresente um comportamento não-estacionário, ou seja, variante no tempo, normalmente se adotaria uma abordagem estatística com o intuito de se obter uma nova série estacionária, provavelmente a Padronização da série, como discutido no capítulo anterior.

Entretanto, este tipo de abordagem possui o inconveniente de presumir que a série temporal de Vazões Afluentes pode ser tratada como um processo ciclo-estacionário, sendo representado por dois componentes: um sazonal e outro estacionário.

A Padronização dos dados da série pode acarretar em perdas de informação, comprometendo o modelo dinâmico que se pretende ajustar, tendo como conseqüência o comprometimento do desempenho do previsor para a série temporal das futuras Vazões Afluentes mensais.

Alternativamente, optou-se pela utilização da representação da série mensal de Vazões Afluentes através do "*Analytic Signal*" (AS) da série. Tendo como resultado uma nova série, composta por uma seqüência de números complexos, que preserva todas as informações contidas na série original.

Em Processamento Digital de Sinais, o "*Analytic Signal*" é definido como uma representação complexa do sinal original, sendo que a parte real é composta pelo próprio sinal original, enquanto que a parte imaginária é definida pela Transformada de *Hilbert* do sinal original (Oppenheim e Schafer, 1999). Esta representação do sinal é bastante útil para se calcular atributos instantâneos da série temporal, especificamente a Amplitude, Ângulo de Fase e Freqüência instantâneas.

Através da decomposição da série de VA nestas novas grandezas, é possível identificar e compreender melhor este fenômeno natural e algumas de suas características comportamentais. Este tipo de entendimento é muito útil não só para o sistema de geração hidrelétrico, mas também para todos os gestores de recursos hídricos.

A representação da série de VA pelo "*Analytic Signal*" pode ser utilizada na etapa de Pré-Processamento dos dados para a previsão de futuras VA mensais. As previsões passaram então a serem feitas para duas novas séries: Amplitude e Ângulo de Fase.

4.1 Análise Estatística e Espectral da Série de Vazões Afluentes

Dando continuidade à análise estatística iniciada no Capítulo 3, onde a Figura 3.1 é aqui reproduzida pela Figura 4.1, apresenta o histórico disponível para este trabalho das VA da UH Furnas. Nela pode-se observar a ocorrência das mais variadas condições hidrológicas, com uma aparente distribuição de probabilidade diferente da distribuição Normal.

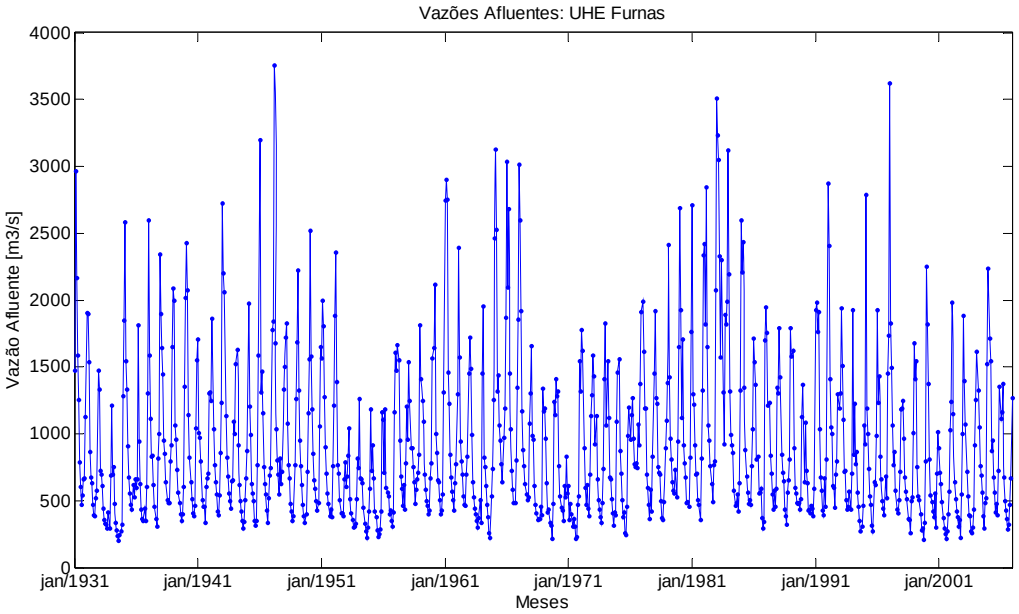


Figura 4.1 - Série de vazões afluentes médias mensais para a Usina Hidrelétrica de Furnas

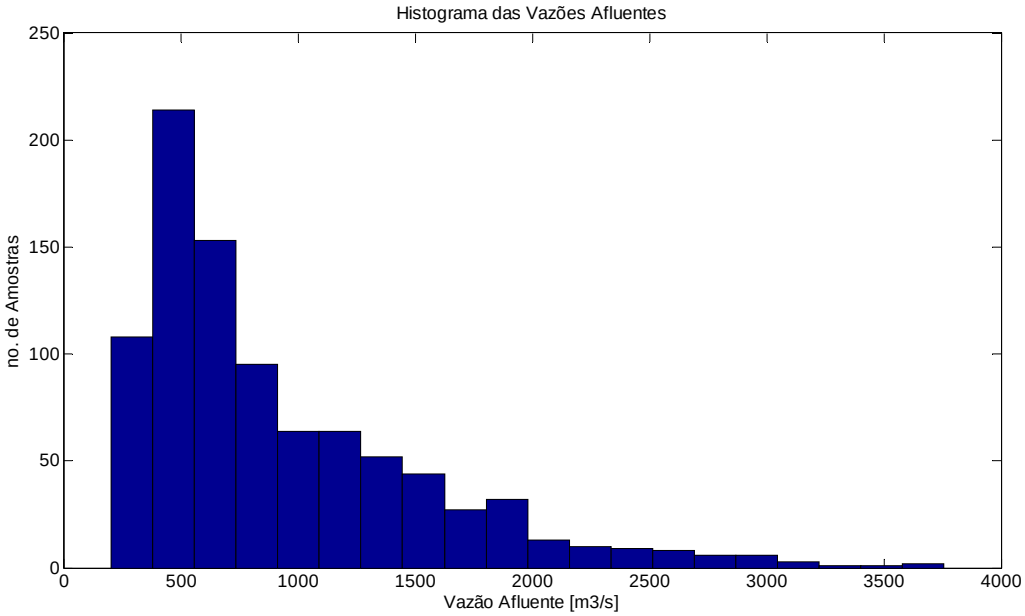


Figura 4.2 – Histograma de vazões afluentes médias mensais para a Usina Hidrelétrica de Furnas

A Figura 4.2, histograma de distribuição de ocorrências de VA, ilustra justamente este comportamento não-uniforme de VA. Pode-se notar que existem muito mais ocorrências de VA baixas e médias, e poucas ou pouquíssimas VA altas e/ou muito altas.

Este comportamento assimétrico é desfavorável para que se possa obter uma boa modelagem do processo estocástico das VA, prejudicando o desempenho do previsor de VA que se queira ajustar.

Com o intuito de identificar características de periodicidade média presente na série histórica de vazões afluentes da UHE Furnas, procedeu-se a avaliação do "*Fourier Power Spectrum*" da série, que pode ser observado na Figura 4.3.

O "*Fourier Power Spectrum*" , segundo (Oppenheim e Schafer, 1999), pode ser calculado a partir da Transformada Discreta de Fourier (DFT), utilizando um dos algoritmos eficientes da Transformada Rápida de Fourier (FFT), para sinais discretos, igualmente espaçados e de finita duração, como é o caso da série temporal de vazões afluentes.

Considerando uma seqüência finita $x[n]$ com N amostras, pode-se calcular a DFT da seguinte forma:

$$X[k] = \sum_{n=1}^{N} x[n] \cdot \omega_N^{(n-1)(k-1)} \tag{4.1}$$

onde:

$$\omega_N = e^{(-2\pi n)/N} \tag{4.2}$$

Calculados os coeficientes da DFT, pode-se calcular o "*Fourier Power Spectrum*" da seguinte forma:

$$P_{xx} = \frac{1}{2 \cdot N} \cdot \log\left(X[k] \cdot X^*[k]\right) \tag{4.3}$$

onde $X^*[k]$ é o complexo conjugado dos coeficientes da DFT.

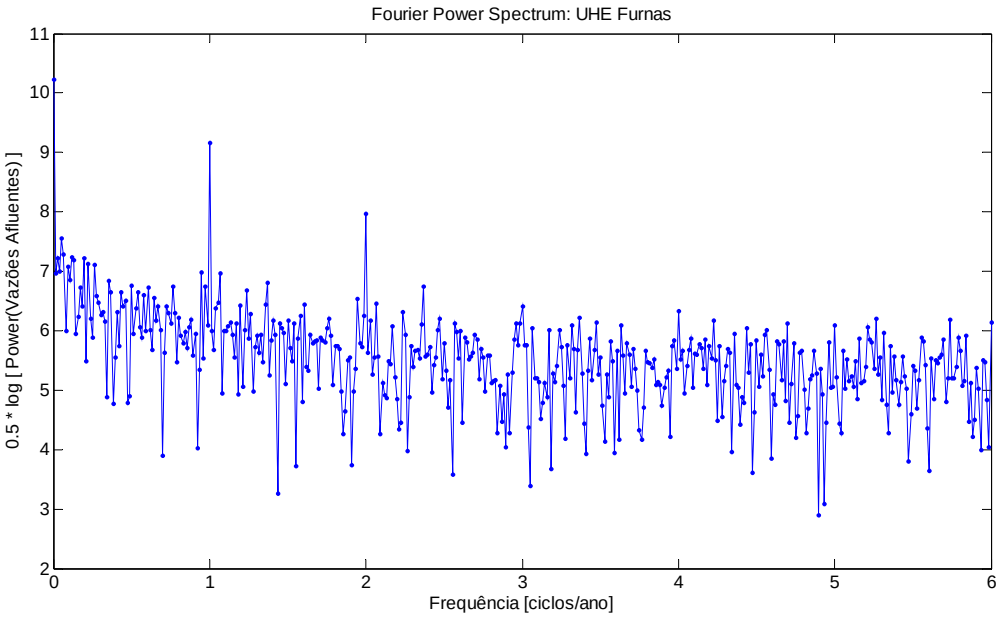


Figura 4.3 – "Fourier Power Spectrum" das vazões afluentes médias mensais para a Usina Hidrelétrica de Furnas

Na Figura 4.3 é possível observar a presença da componente de freqüência de 0 (zero) ciclos/ano, correspondente ao primeiro "pico", o que denota a existência de um valor médio para todo o histórico da série de vazões afluentes.

Também é possível observar um segundo "pico" com freqüência de 1 ciclo/ano, ou seja uma componente com periodicidade média de 12 meses, o que denota a tendência de um comportamento cíclico médio anual.

Além disso, é possível observar um terceiro "pico" com freqüência de 2 ciclos/ano, ou seja uma componente com periodicidade média de 6 meses, o que poderia justificar a forte assimetria do histograma das vazões afluentes da Figura 4.2; mas que evidentemente merece maior investigação.

Finalmente, é possível notar na Figura 4.3 que o histórico de vazões afluente da UHE de Furnas apresenta bastante energia para toda a banda do "*spectrum*" apresentado, característico de sinais não-periódicos ou caóticos (Abarbanel, 1996).

De volta a figura 4.2, nela pode-se notar que a distribuição a série de VA da UHE de Furnas parece possuir uma distribuição de probabilidade semelhante a Log-normal. Desta maneira, neste trabalho, optou-se em trabalhar com o Logaritmo da série temporal de VA, *log(VA)* – Figura 4.4, ao invés da série original de VA.

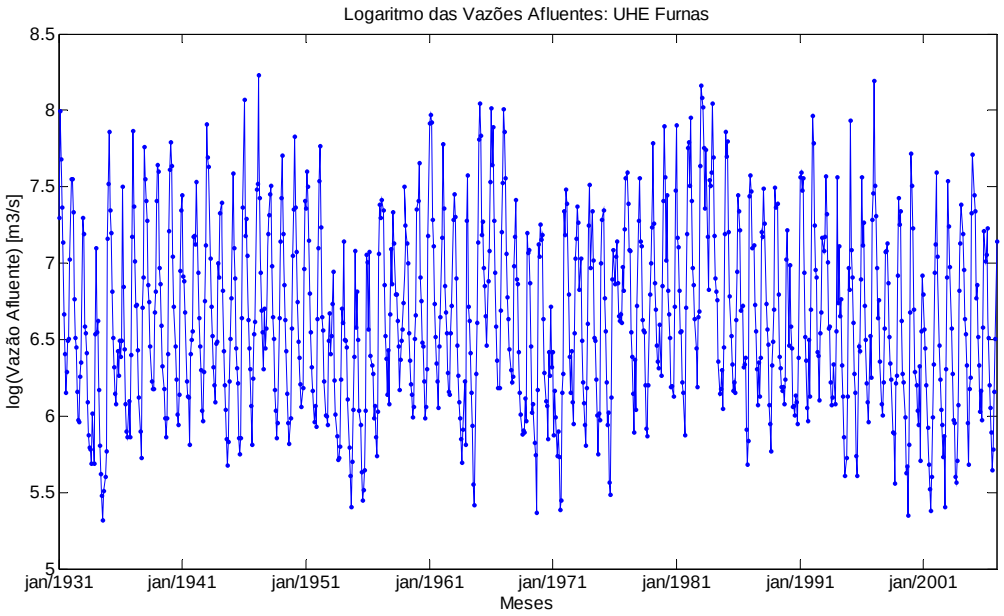


Figura 4.4 - Logaritmo da série histórica de Vazões Afluentes mensais da Usina Hidrelétrica de Furnas

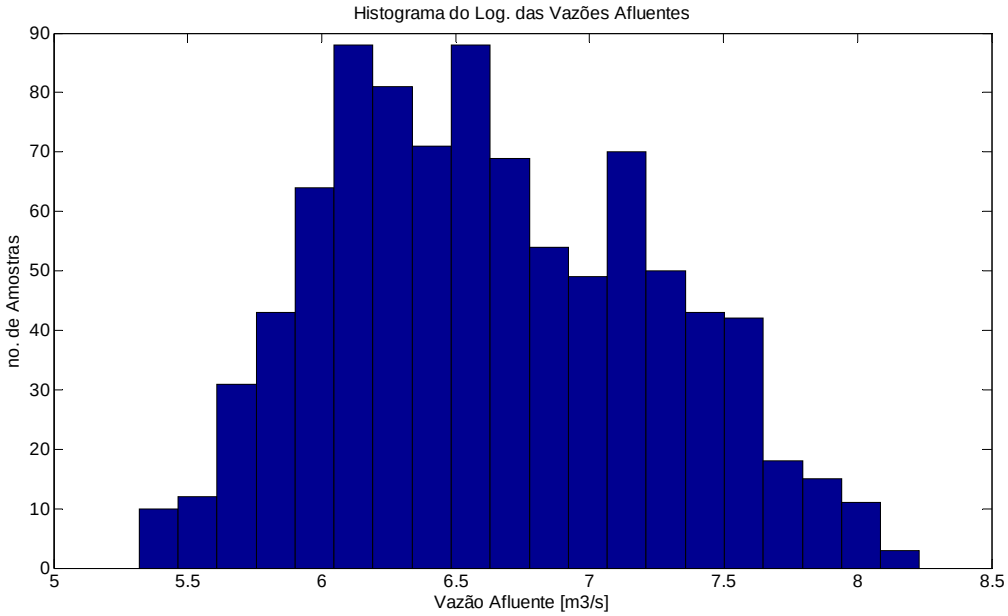


Figura 4.5 - Histograma da distribuição de ocorrências do logaritmo da série de VA mensais: UHE Furnas

Este tipo de transformação é bastante conveniente por tornar a nova série mais próxima de uma distribuição de probabilidade simétrica, o que pode ser notado na Figura 4.5. Além disso, o Logaritmo da série de VA possibilita a representação da série via "*Analytic Signal*" com menor dinâmica, o que é bastante interessante para melhorar a performance do modelo de previsão.

4.2 "Analytic Signal" da Série de Vazões Afluentes

Esta seção apresenta a transformação para a obtenção do "A*nalytic Signal*" de sinais discretos e finitos. Neste trabalho de pesquisa, a série temporal de Vazões Afluentes mensais é vista como um sinal discreto, com discretização mensal, e finito, com registros históricos de 1931 a 2006.

Um sinal contínuo de valores reais possui a propriedade que sua transformada de Fourier (Espectro de Freqüência) é simétrica. Logo, metade do espectro, a porção de freqüências negativas, contém informação redundante em relação à outra metade, porção de freqüências positivas.

O objetivo de (Garbor, 1946) e (Ville, 1948) ao criarem o "*Analytic Signal*" de sinais contínuos era o de remover esta redundância espectral, apagando a porção de freqüências negativas do sinal transformado. O sinal resultante composto por valores complexos, com espectro apenas positivo, preserva toda a informação contida no sinal original composto por valores reais.

Já o trabalho (Marple Jr., 1999) apresenta o algoritmo de transformação para a obtenção do "*Analytic Signal*" para o caso de sinais discretos e finitos, com algumas modificações necessárias. Isto porque existem algumas armadilhas na aplicação das propriedades do caso de sinal contínuo para o caso de sinal discreto.

O uso do "*Analytic Signal*" no lugar do sinal original para muitas aplicações em processamento de sinais tem mostrado mitigar cálculos distorcidos e eliminar termos cruzados indesejáveis entre componentes de freqüências positivas e negativas, o que ocorreria no caso de se utilizar o sinal original (Marple Jr., 1999).

Considere um "*Analytic Signal*" discreto z[n] correspondente a um sinal discreto real x[n] com duração finita. Assume-se que x[n] para $0 \leq n \leq N-1$ é obtido pela discretização do sinal contínuo x(nT) = x[n] em intervalos periódicos de T, e que deve ser definido para prevenir "*aliasing*" (sobreposição de espectros).

O espectro do sinal discreto x[n] pode ser obtido pela Transformada Discreta de Fourier (DTFT) (Marple Jr., 1987):

$$\mathrm{X}(f) = T\sum_{n=0}^{N-1} \mathrm{x}[n] \cdot \exp(-j2\pi fnT) \qquad \textbf{(4.4)}$$

A DFT é calculada através da Transformada Rápida de Fourier (FFT), que calcula a DFT para cada conjunto de freqüências X(f_m) = X[m], no qual $f_m = m/NT$ Hz, e $0 \leq m \leq N-1$.

Existem duas propriedades que devem ser satisfeitas para que z[n] = $\mathrm{z}_r[n] + j\mathrm{z}_i[n]$ seja um "analytic signal" discreto. A primeira é que a parte real resultante deve ser exatamente igual ao sinal discreto original.

$$\mathrm{z}_r[n] = \mathrm{x}[n] \qquad \text{para } 0 \leq \mathrm{n} \leq \mathrm{N}-1 \qquad \textbf{(4.5)}$$

A segunda propriedade exige que os componentes reais e imaginários devam ser ortogonais sobre o intervalo finito.

$$T \cdot \sum_{n=0}^{N-1} \mathrm{z}_r[n] \cdot \mathrm{z}_i[n] = 0 \qquad \textbf{(4.6)}$$

Uma abordagem de filtragem no domínio do tempo para se obter uma aproximação do "*Analytic Signal*" a partir de um sinal discreto foi sugerido por (Oppenheim e Schafer, 1999). Apesar de satisfazer a propriedade de ortogonalidade, Equação 4.6, não preserva os valores do sinal original, desrespeitando a primeira propriedade, Equação 4.5.

Outro procedimento foi proposto por (Reilly, Frazer e Boashash, 1994), satisfazendo a primeira propriedade, pois preserva os valores do sinal original; entretanto, ela não satisfaz a propriedade da ortogonalidade da Equação 4.6.

Alternativamente, (Marple Jr., 1999) propôs uma abordagem no domínio da freqüência, no qual a porção negativa das freqüências de cada período espectral fosse zerada, resultando em um espectro unilateral.

Assumindo que N é par, o procedimento para a obtenção de um "*Analytic Signal*" discreto composto por N-valores complexos a partir de um sinal discreto de N-valores reais, e que satisfaça simultaneamente as duas propriedades das Equações 4.5 e 4.6, é assim descrito:

(i) Calcule os N-valores da DTFT X[m], definidos pela Equação 4.4, usando a FFT das N-amostras reais;

(ii) Faça a transformação do "analytic signal" discreto de N-amostras:

$$Z[m] = \begin{cases} \mathrm{X}[0] & \text{para } m = 0 \\ 2 \cdot \mathrm{X}[m] & \text{para } 1 \le \mathrm{m} \le \dfrac{N}{2} - 1 \\ \mathrm{X}[N/2] & \text{para } m = \dfrac{N}{2} \\ 0 & \text{para } \dfrac{\mathrm{N}}{2} + 1 \le m \le N - 1 \end{cases} \tag{4.7}$$

(iii) Calcule a DTFT inversa dos N-valores, para se obter o "*Analytic Signal*" z[n] do sinal original x[n]:

$$\mathrm{z}[n] = \frac{1}{NT} \sum_{m=0}^{N-1} \mathrm{Z}[m] \cdot \exp(+j2\pi mn / N) \tag{4.8}$$

Como z[n] foi definida com um espectro unilateral, o "*Analytic Signal*" é necessariamente composto por valores complexos, podendo ser representado em termos de seus componentes reais $z_r(n) = Re(z[n])$ e imaginários $z_i(n) = \mathrm{Im}(z[n])$. Além disto, temos que:

$$z_r[n] = x[n] \qquad\qquad z_i[n] = HT\{x[n]\} \tag{4.9}$$

onde *HT* se refere a Transformada de *Hilbert*.

Logo, o "*Analytic Signal*" z[n] pode ser expresso através da representação Fasorial:

$$z[n] = A[n] \cdot e^{j\phi[n]} \tag{4.10}$$

onde A[n] é a Amplitude instantânea do Fasor z[n], o que corresponde ao Envelope do sinal original x[n], e $\phi[n]$ é o Ângulo de Fase instantâneo de z[n].

4.2.1 Determinação da Amplitude (Envelope) e Ângulo de Fase

De posse do "*Analytic Signal*" z[n] em sua representação fasorial, a Figura 4.6 mostra o Envelope, em vermelho, do Logaritmo da série de Vazões Afluentes a UH de Furnas para todo o histórico disponível para este trabalho. Já a Figura 4.7 mostra o segundo componente da representação Fasorial do "*Analytic Signal*", o Ângulo de Fase.

Num primeiro momento, pode-se acreditar que a série de VA seja uma série periódica, uma vez que o Ângulo de Fase da Figura 4.7 evolui praticamente de maneira linear e com uma inclinação média de 0.5237 rad/mês. Desta forma, pode-se dizer que se trata de um sinal com periodicidade média de 11,9978 meses, ou seja, aproximadamente 1 ciclo/ano. Resultado este, em conformidade com o observado anteriormente na Figura 4.3 do item 4.1.

Entretanto, se observada a Figura 4.7 com mais detalhe, é possível notar que

existem algumas oscilações do Ângulo de Fase. A Figura 4.8 apresenta a projeção do Ângulo de Fase em torno do segmento de reta com inclinação de 0.5236 rad/mês, equivalente a um Período de exatos 12 meses/ciclo.

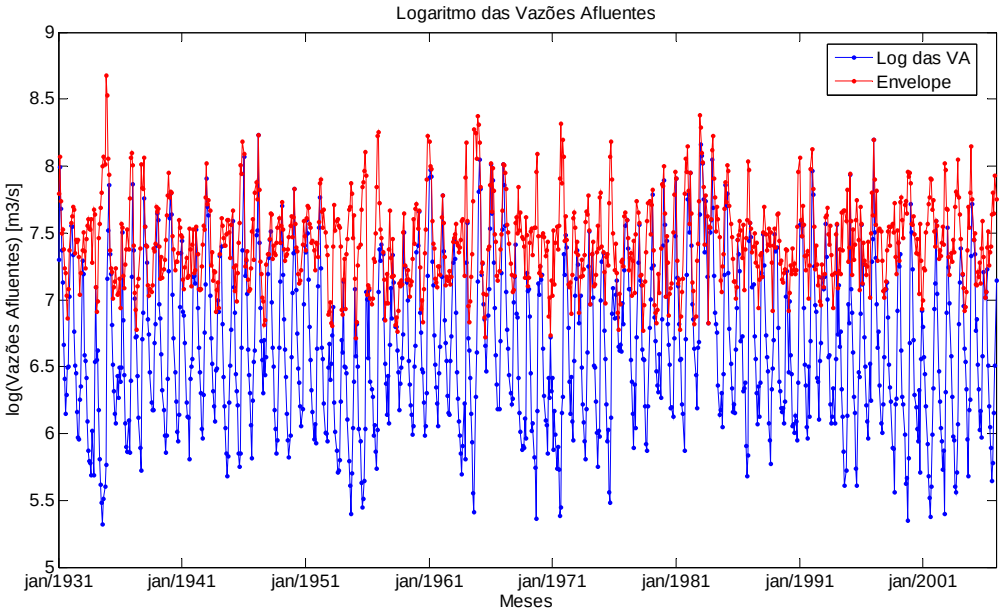


Figura 4.6 - Envelope do logaritmo da série histórica de VA mensais da Usina Hidrelétrica de Furnas

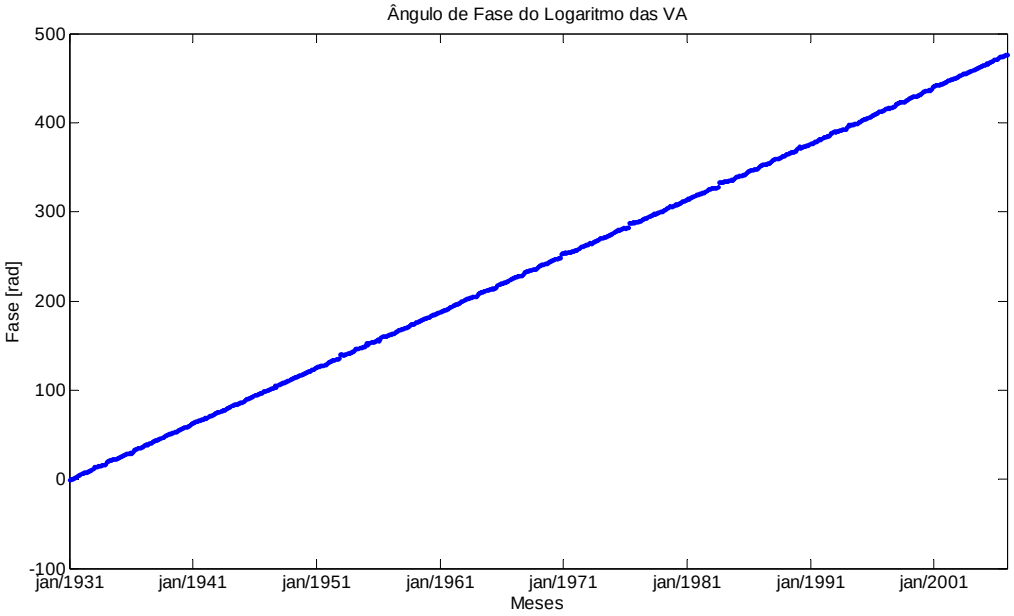


Figura 4.7 - Ângulo de Fase do logaritmo da série histórica de VA mensais da UHE de Furnas

Na Figura 4.8 fica evidente que existem grandes oscilações do Ângulo de Fase,

fazendo com que a série de Vazões Afluentes seja entendida como uma série cíclica, mas não periódica (com periodicidade média de 12 meses).

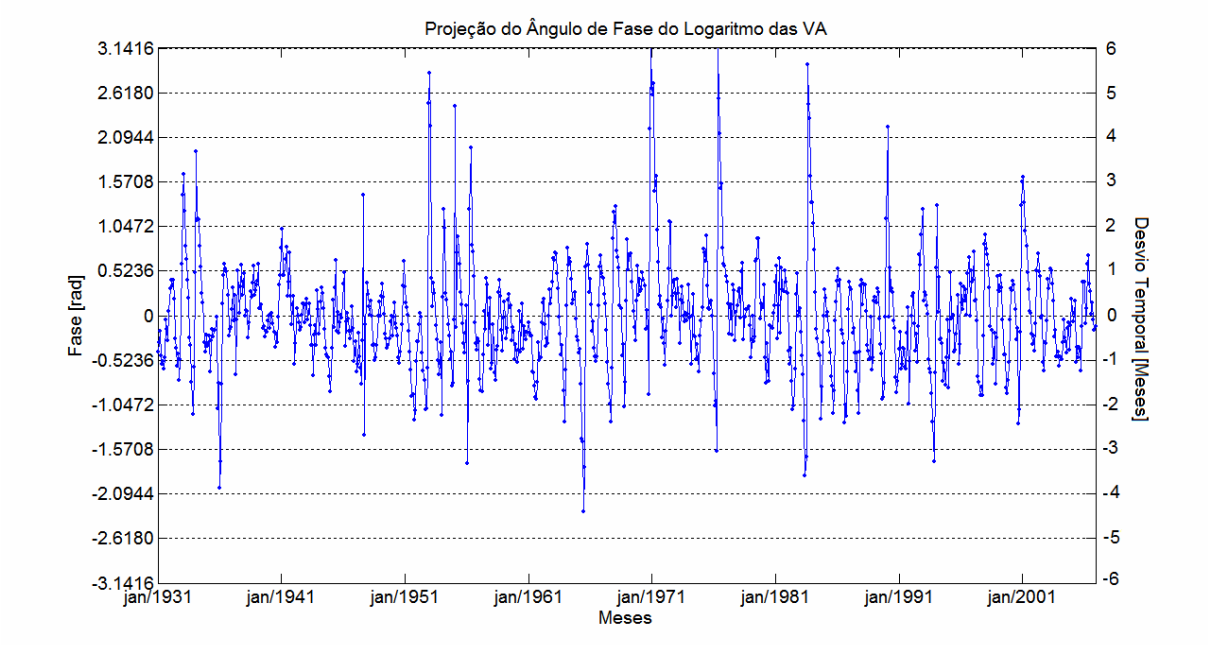


Figura 4.8 - Projeção do Ângulo de Fase do logaritmo da série histórica de VA mensais da Usina Hidrelétrica de Furnas

Além disto, esta série apresenta um desvio temporal bastante grande, podendo chegar a 6 meses de desvio. Em outras palavras, uma determinada quantidade de água que deveria ocorrer num determinado mês, pode vir a acontecer com até 6 meses de atraso ou antecipação. O desvio temporal médio calculado para as VA da UH Furnas foi de aproximadamente 1,26 mês; o que é bastante expressivo, além de ser uma informação bastante valiosa.

Fica evidente que este tipo de comportamento não deve ser desprezado, simplesmente linearizado ou aproximado quando se tem o objetivo de analisar, modelar e prever a série de Vazões Afluentes.

Capítulo 5

Princípios da "Dynamic Modelling"

A modelagem linear tem sido amplamente utilizada no desenvolvimento de modelos que manipulam dados: primeiro em regressão de dados estáticos e posteriormente em filtragem ótima de séries temporais. Do ponto de vista da "*Dynamic Modelling*", os modelos lineares não se mostram interessantes por apresentarem convergência para um ponto de atração, quando estável, ou para um ciclo limite (Principe, 1997).

O objetivo da "*Dynamic Modelling*" é criar um modelo não-linear da série temporal, utilizando apenas os dados disponíveis da série temporal. E, por aprenderem diretamente da série temporal e por serem poderosos mapeadores de funções, as Redes Neurais Artificiais são bastante apropriadas para a "*Dynamic Modelling*" (Sandberg e Xu, 1997).

Conseqüentemente, do ponto de vista metodológico, a "*Dynamic Modelling*" apresenta maior capacidade de modelar fenômenos altamente complexos, em comparação a capacidade limitada dos modelos lineares (Principe, 1997).

Em "*Dynamic Modelling*", um importante passo é a criação de um "*embedding*" da série temporal utilizando o teorema "*Time Delay Embedding*" (Takens, 1981; Takens, 1985; Sauer, Yorke e Casdagli, 1991). A idéia do "*embedding*" é criar um espaço, chamado de Espaço de Reconstrução (*the Reconstruction Space*, em inglês), onde as dinâmicas do sistema não-linear autônomo, que gera a série temporal, podem ser reconstruídas. O teorema de Takens especifica condições que garantem que exista um difeomorfismo entre a trajetória reconstruída e o sistema dinâmico original.

Em (Casdagli, 1989) foi mostrado que a previsão do próximo ponto no Espaço de Reconstrução corresponde ao próximo ponto no domínio do tempo, desde que as trajetórias reconstruídas possam ser usadas na criação de um modelo auto-regressivo do sistema dinâmico (Principe, 1997).

"*Non-linear Dynamic Modelling*" é definida como a identificação do mapeamento $\phi: R^{d_M} \rightarrow R^{d_M}$ que define o sistema dinâmico desconhecido, onde d_M é a dimensão do "*manifold*" original. A Figura 5.1 apresenta os principais passos necessários na modelagem não-linear. Nota-se que o sistema desconhecido é considerado como autônomo, ou seja, não apresenta nenhuma entrada externa, o que significa que a complexidade das séries temporais está somente na modelagem de suas dinâmicas.

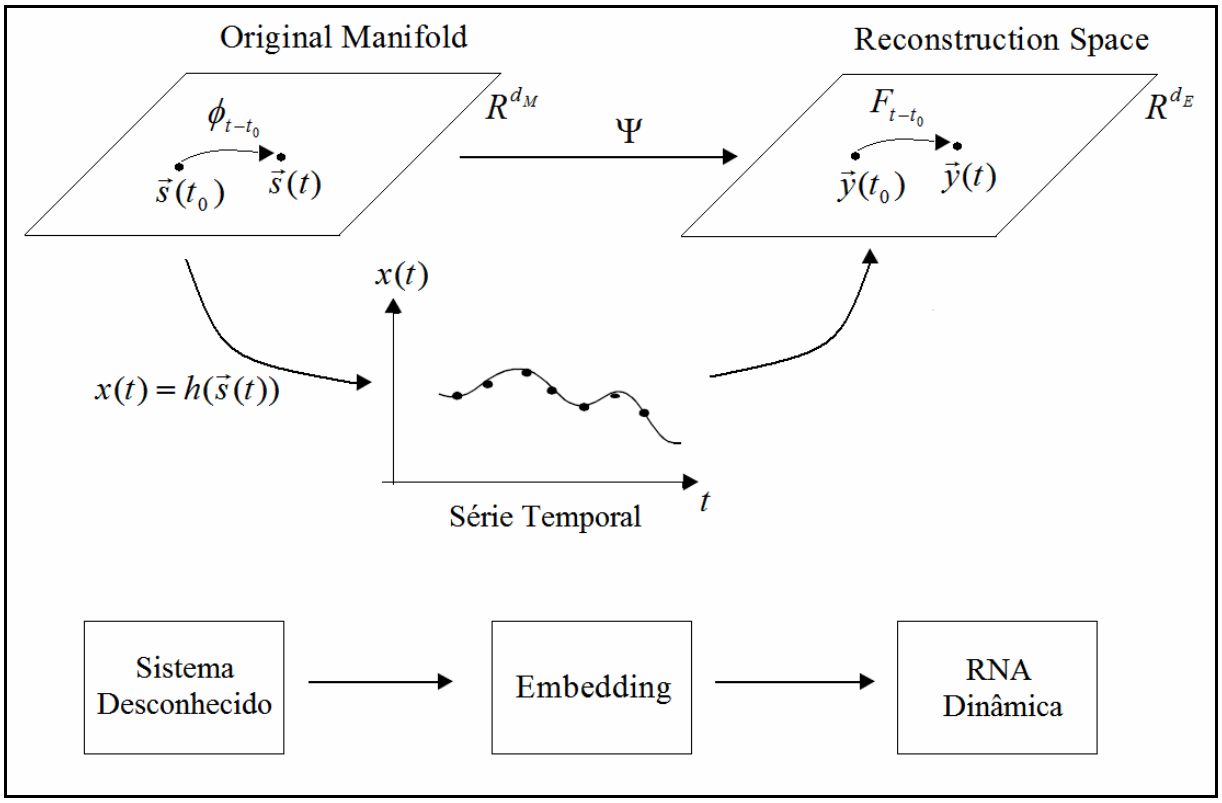


Figura 5.1 – Processo da "*Non-linear Dynamic Modelling*"

Em primeiro lugar, a série temporal deve ser transformada em uma trajetória por meio de uma das técnicas de "*embedding*" (Sauer, Yorke e Casdagli, 1991), dentre as quais a mais

comum é "*Time Delay Embedding*" (Takens, 1985). Caso a dimensão d_E do espaço de reconstrução seja maior que $2 \cdot d_M + 1$, onde d_M é a dimensão do "*manifold*" original, então o mapa Ψ é *difeomorfo* e o fluxo *F* irá preservar a dinâmica do sistema original ϕ.

Segundo, a previsão do próximo ponto no Espaço de Estado identifica o mapeamento não-linear *F* que governa a evolução da trajetória no Espaço de Reconstrução. Ou seja, de acordo com o teorema "*Taken's embedding*", existe um mapa $F : R^{d_E} \rightarrow R^{d_E}$ que transforma o estado reconstruído atual $\vec{y}(t_0)$ no próximo estado $\vec{y}(t)$ que é dado por:

$$F_{t-t_0}(\vec{y}(t_0)) = \Psi \bullet \phi_{t-t_0} \bullet \Psi^{-1}(\vec{y}(t_0)) = \vec{y}(t) \qquad \textbf{(5.1)}$$

onde $\bullet$ é a composição de funções. As relações entre estes mapeamentos e vetores são ilustradas na Figura 5.1.

O mapeamento preditivo é a peça central da modelagem. Estando o mapeamento preditivo uma vez determinado, *F* pode ser obtido por meio de simples operações matriciais. O mapeamento preditivo $f : R^{d_E} \rightarrow R$ pode ser expresso por:

$$x(t+T) = f(\vec{y}(t)) \qquad \textbf{(5.2)}$$

onde $\vec{y}(t) = [x(t), x(t-\tau), \ldots, x(t-(d_E-1)\cdot\tau)]$. A Equação 5.2 define um modelo Não-Linear Auto-regressivo (NAR) <u>determinístico</u> da série.

A existência deste modelo preditivo estabelece a base teórica da "dynamic modelling", no sentido de que abre a possibilidade de se construir, a partir do vetor de série temporal, um modelo que aproxime o mapeamento *f* (Principe, Wang e Kuo, 1998).

Em suma, o problema de "*Dynamic Modelling*" pode ser definido como uma representação apropriada da dinâmica da série pela etapa de "*embedding*", seguido pela identificação do mapeamento *f* (Principe e Wang, 1995).

A etapa de "*embedding*" pode ser implementada por uma linha de *atraso*, também chamada de estrutura de memória, com o tamanho especificado pelo teorema "*Taken's*

embedding" (Haykin e Principe, 1998). Entretanto, que "time lag" τ usar e que dimensão d usar são os pontos centrais desta reconstrução, e que serão discutidos nas seções 5.1 e 5.2, respectivamente.

Após a etapa de "embedding", qualquer sistema adaptativo não-linear pode ser treinado como um previsor para identificar o mapeamento f. Neste trabalho de pesquisa foi utilizada uma abordagem global implementada por Redes Neurais Artificiais (RNA).

5.1 "Time Delays" (τ): "Average Mutual Information"

O teorema de "*embedding*" não faz nenhuma mensão sobre a escolha do "*time delay*" a ser usado na construção dos vetores de dados d_E-dimensionais no espaço de reconstrução. Desta forma, esta seção estabelece um procedimento para a escolha do "*time delay*" τ .

O "*time delay*" não deve ser muito pequeno, uma vez que as coordenadas $x(n)$ e $x(n+\tau)$, que serão usadas no espaço de reconstrução, podem não ser independentes o bastante. Em outras palavras, o sistema a ser modelado não teria o tempo suficiente para explorar suficientemente seu espaço de estado a ponto de produzir nova informação à cerca do espaço de estados. Entretanto, caso τ seja muito grande, a correlação entre os dados $x(n)$ e $x(n+\tau)$ tenderia a ser muito fraca, a ponto de serem considerados dados independentes (ou aleatórios) entre si (Abarbanel, 1996).

Fundamental para a noção de informação entre dados de medições é a idéia de Shannon de Informação Mútua (*Mutual Information*, em inglês) entre dois dados a_i e b_j pertencentes aos conjuntos A e B de dados de medições (Gallager, 1968). A Informação Mútua entre a medição a_i do conjunto de dados $A = \{a_i\}$ e medição b_j do conjunto de dados $B = \{b_j\}$ é a quantidade aprendida pela medição de a_i com relação a medição de b_j . Em bits é dado por:

$$\log_2\left[\frac{P_{AB}(a_i,b_j)}{P_A(a_i)P(b_j)}\right] \tag{5.3}$$

onde $P_{AB}(a,b)$ é a densidade de probabilidade conjunta das medições A e B resultantes em valores a e b. $P_A(a)$ e $P_B(b)$ são as densidades de probabilidade individuais para as medições de A e B. Em sistemas determinísticos estas probabilidades são avaliadas com a construção de histogramas de variações de a_i ou b_j observando suas medições.

Caso a medição de um valor de A resultante em a_i seja completamente independente da medição de um valor de B resultante em b_j, então $P_{AB}(a,b)$ pode ser fatorado: $P_{AB}(a,b) = P_A(a)P_B(b)$ e a quantidade de informação entre as medições, a Informação Mútua, seria zero, como esperado.

A média de todas as medições desta informação estática, chamada de Informação Mútua Média (*Average Mutual Information* – AMI, em inglês) entre os dados de medições A e B é definido por:

$$AMI_{AB} = \sum_{a_i,b_j} P_{AB}(a_i,b_j)\log_2\left[\frac{P_{AB}(a_i,b_j)}{P_A(a_i)P_B(b_j)}\right] \tag{5.4}$$

Esta métrica não está relacionada às regras de evolução lineares ou não-lineares dos dados de medição. Ela é apenas uma idéia teórica que conecta dois conjuntos de dados e estabelece um critério para suas dependências mutuas com base na noção da conexão da informação entre eles. Esta conexão define a noção de que os dados de medição $x(t)$ no tempo t estão conectados aos dados de medição $x(t+\tau)$ no tempo $t+\tau$, por compartilharem informações mutuamente.

Assim, toma-se para o conjunto de dados A os valores observados $x(n)$, e para o conjunto B, os valores de $x(n+\tau)$. A Informação Mútua Média (AMI) entre estas duas

medições, i.e. a quantidade (em bits) aprendida das medições de $x(n)$ por meio das medições de $x(n+\tau)$ é dada por:

$$AMI(\tau) = \sum_{x(n),x(n+\tau)} P(x(n), x(n+\tau)) \log_2 \left[\frac{P(x(n), x(n+\tau))}{P(x(n))P(x(n+\tau))} \right] \quad \textbf{(5.5)}$$

onde $AMI(\tau) \geq 0$. Quando τ se torna grande, o comportamento caótico da série tornam as medidas *x(n)* e $x(n+\tau)$ praticamente independentes , e $AMI(\tau)$ tende a zero.

A função $AMI(\tau)$ pode ser utilizada como um tipo de <u>auto-correlação não-linear</u> para determinar quando os valores de *x(n)* e $x(n+\tau)$ são suficientemente independentes entre eles, a ponto de poderem ser utilizados como coordenadas do vetor de "*time delay*", mas não tão independentes a ponto de não haver qualquer conexão (ou correlação) entre eles (Fraser e Swinney, 1986).

Nesta abordagem a sugestão é que seja definido τ onde ocorrer o primeiro mínimo da Informação Mútua Média $AMI(\tau)$ como o valor a ser utilizado como "*time delay*" da reconstrução do espaço de estados ("*reconstruction space*") (Abarbanel, 1996).

Esta abordagem exige a generalização da noção da AMI de espaços de estados de altas dimensões por meio da substituição de $x(n)$ pelos vetores d_E-dimensionais $\vec{y}(n)$; sendo que tais vetores agregam novas informações incrementais por meio da adição de novos componentes ao vetor.

Na prática, esta abordagem busca um valor de τ que represente e esteja conectado com as informações não-lineares que caracterizem as dinâmicas da série temporal.

A Informação Mútua Média $AMI(\tau)$ foi aplicada à série temporal das Vazões Afluentes da UHE de Furnas para $x(n)$ e $x(n+\tau)$. O resultado é a curva na Figura 5.2, onde é possível observar que o primeiro mínimo ocorre para $\tau = 4$.

Este resultado indica para a utilização do seguinte vetor de "*time delay*":

$$\vec{y}^{VA}(n) = [VA(t), VA(t-4), \ldots, VA(t-(d_E-1)\cdot 4)] \quad \textbf{(5.6)}$$

no espaço de reconstrução da série temporal de VA.

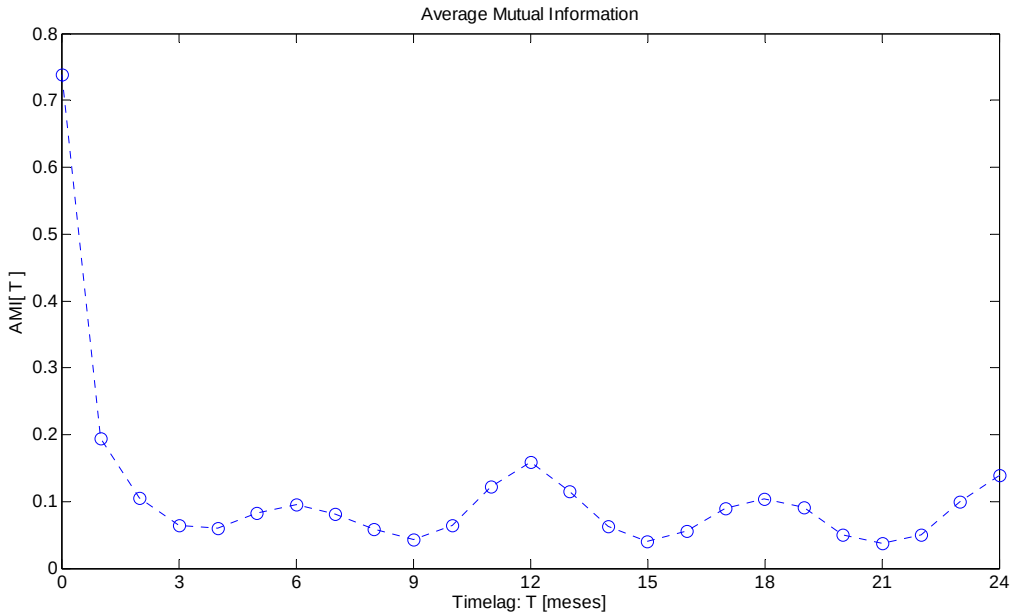


Figura 5.2 – "*Average Mutual Information*" (AMI) das Vazões Afluentes a UHE de Furnas

A seguir as Figuras 5.3 e 5.4 apresentam os resultados da $AMI(\tau)$ para o logaritmo da série temporal de VA decomposta na Amplitude (Envelope) e Ângulo de Fase, respectivamente.

Na Figura 5.3, a curva de $AMI(\tau)$ apresenta o primeiro mínimo em $\tau = 4$. Logo, o vetor de "*time delay*" para a Amplitude A da série temporal é:

$$\vec{y}^{A}(n) = [A(t), A(t-4), \ldots, A(t-(d_E-1)\cdot 4)] \quad \textbf{(5.7)}$$

no espaço de reconstrução da Amplitude da série temporal.

Finalmente, na Figura 5.4, a curva de $AMI(\tau)$ apresenta seu primeiro mínimo em $\tau = 8$. Logo, o vetor de "*time delay*" do Ângulo de Fase ϕ da série temporal é:

$$\vec{y}^{\phi}(n) = [\phi(t), \phi(t-8), \ldots, \phi(t-(d_E-1)\cdot 8)] \quad \textbf{(5.8)}$$

no espaço de reconstrução do Ângulo de Fase da série temporal.

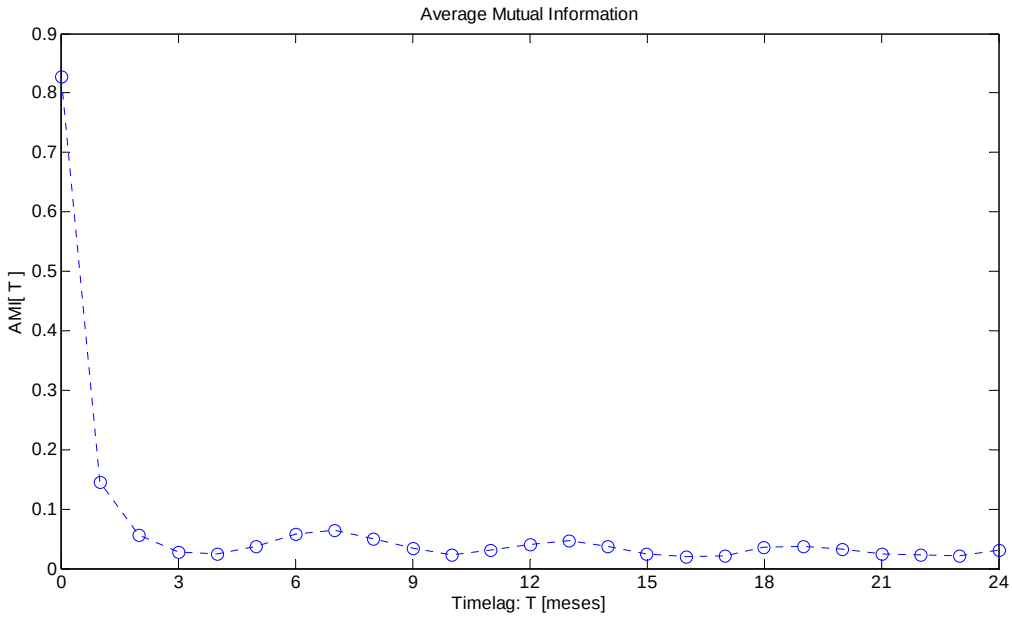


Figura 5.3 - "*Average Mutual Information*" (AMI) do Envelope do log(Vazões Afluentes) a UHE de Furnas

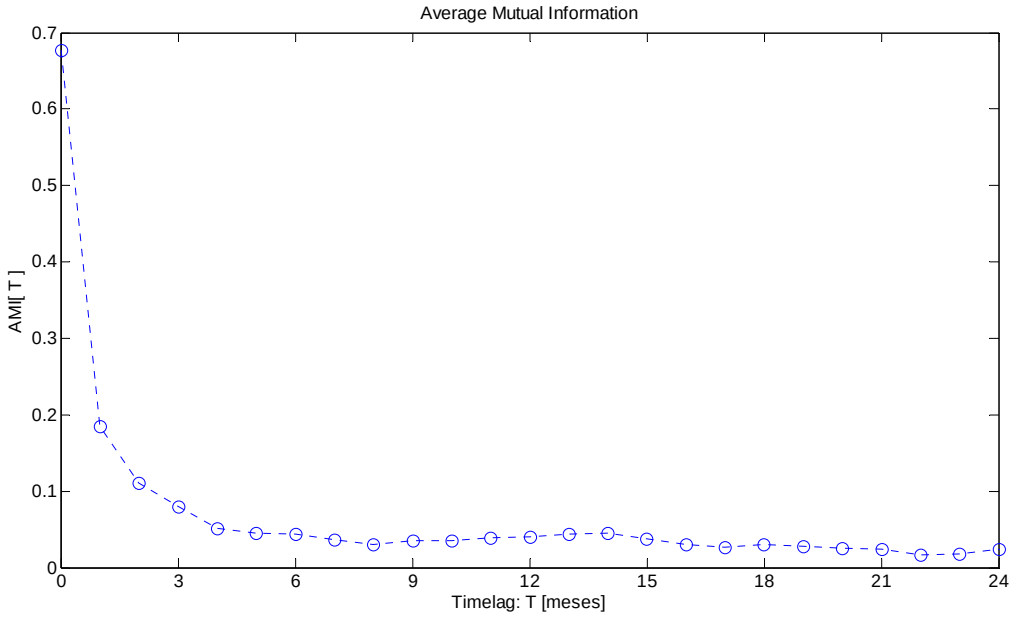


Figura 5.4 - AMI do Ângulo de Fase do log(Vazões Afluentes) a UHE de Furnas

A questão que ainda permanece se refere ao "*Embedding Dimension*" d_E do espaço de reconstrução, o que será discutido na próxima seção.

5.2 "Embedding Dimension" (d_E): "False Nearest Neighbors"

O teorema de "*embedding*" diz que caso a dimensão do atrator definido pelas órbitas do sistema seja d_A, então o atrator certamente pode ser desdobrado em um espaço dimensional inteiro de dimensão d_E , onde $d_E > 2d_A$. Esta não é a dimensão necessária para o desdobramento, mas é suficiente e certamente indica quando parar a adição de componentes ao vetor de "*time delay*" $\vec{y}(n)$.

O objetivo é determinar a dimensão inteira onde o número necessário de coordenadas para desdobrar as órbitas observadas que se sobreponham em projeções do atrator em espaços de baixa dimensão. Logo, a "E*mbedding Dimension*" d_E pode ser entendida como a menor dimensão que desdobre o atrator até que não reste nenhuma destas sobreposições (Abarbanel, 1996).

Supondo-se a reconstrução do espaço de estado na dimensão *d*, que foi definido utilizando o "*time delay*" sugerido pela Informação Mútua Média, cujo vetor de dados seja:

$$\vec{y}(n) = [x(n), x(n-\tau), \ldots, x(n-(d-1)\cdot\tau)] \qquad \textbf{(5.9)}$$

Seu Vizinho Próximo (*Nearest Neighbor*, em inglês) no espaço de estado do vetor $\vec{y}(n)$ é:

$$\vec{y}^{NN}(n) = [x^{NN}(n), x^{NN}(n-\tau), \ldots, x^{NN}(n-(d-1)\cdot\tau)] \qquad \textbf{(5.10)}$$

Caso o vetor $\vec{y}^{NN}(n)$ seja realmente um vizinho de $\vec{y}(n)$, então ele veio para a vizinhança de $\vec{y}(n)$ por origem dinâmica. O vetor $\vec{y}^{NN}(n)$ seria o vetor logo a frente ou logo atrás de $\vec{y}(n)$ ao longo da órbita, caso os intervalos de tempo ao longo da órbita sejam suficientemente pequenos, ou ele chegou na vizinhança de $\vec{y}(n)$ pela evolução ao longo da órbita e ao redor do atrator.

Entretanto, caso o vetor $\vec{y}^{NN}(n)$ seja um falso vizinho de $\vec{y}(n)$ tendo chegado em sua vizinhança pela projeção de uma dimensão superior, uma vez que a atual dimensão d não desdobrou o atrator, então partir para a dimensão superior $d+1$ talvez possa remover este falso vizinho para fora da vizinhança de $\vec{y}(n)$.

Desta forma, é possível definir o "*Embedding Dimension*" d_E ao checar todo vetor de dado $\vec{y}(n)$ e perguntar em qual dimensão todos os falsos vizinhos foram removidos, uma vez que todas as intersecções de órbitas de dimensões reduzidas foram removidas, i.e. o atrator foi desdobrado (Kennel, Brown e Abarbanel, 1992).

Quando a dimensão vai para $d+1$, caso a distância adicional seja maior que a distância de dimensão d entre *vizinhos próximos*, tem-se um falso vizinho. Caso não seja maior, tem-se um verdadeiro vizinho.

A distância entre pontos quando vistos na dimensão $d+1$ relacionada com a distância na dimensão d é:

$$\sqrt{\frac{D_{d+1}^2(n) - D_d^2(n)}{D_d^2(n)}} = \frac{\left|x(n-d\tau) - x^{NN}(n-d\tau)\right|}{D_d(n)} \tag{5.11}$$

onde $D_d(n)$ é a distância Euclidiana entre os vizinhos próximos vistos na dimensão d.

Quando esta métrica é maior que um dado limiar, tem-se um falso vizinho.

Caso se tenha um conjunto limpo de dados de um sistema caótico, o esperado é que a porcentagem de Falsos Vizinhos Próximos (do inglês, *False Nearest Neighbors* – FNN) irá cair de cerca de 100% na dimensão um até quase zero quando d_E for atingido. Depois disto, a porcentagem permanecerá em zero, uma vez que o atrator tenha sido desdobrado, estará desdobrado.

Entretanto, se o sinal estiver contaminado, pode ser que esta contaminação prepondere em relação à própria série a ponto de a dimensão necessária seja para desdobrar a contaminação ao invés da série original. Caso a contaminação seja de alta dimensão, típico de

ruídos, então provavelmente a porcentagem de Falsos Vizinhos Próximos (FNN) não decairá a zero em nenhuma dimensão.

De acordo com o que foi encontrado na seção 5.1 pela Informação Mútua Média (AMI), a série temporal de Vazões Afluentes a UHE de Furnas apresenta um "*time delay*" $\tau = 4$. A Figura 5.5 apresenta a porcentagem de FNN da série temporal de VA em função da dimensão. A porcentagem de FNN foi para zero na "*embedding dimension*" $d_E = 4$, e como esperado permaneceu em zero, definindo um Espaço de Reconstrução $R^{d_E=4}$.

Assim, reescrevendo a Equação 5.6, o vetor de "*time delay*" das VA é definido como:

$$\vec{y}^{VA}(n) = [VA(t), VA(t-4), VA(t-8), VA(t-12)] \quad \textbf{(5.12)}$$

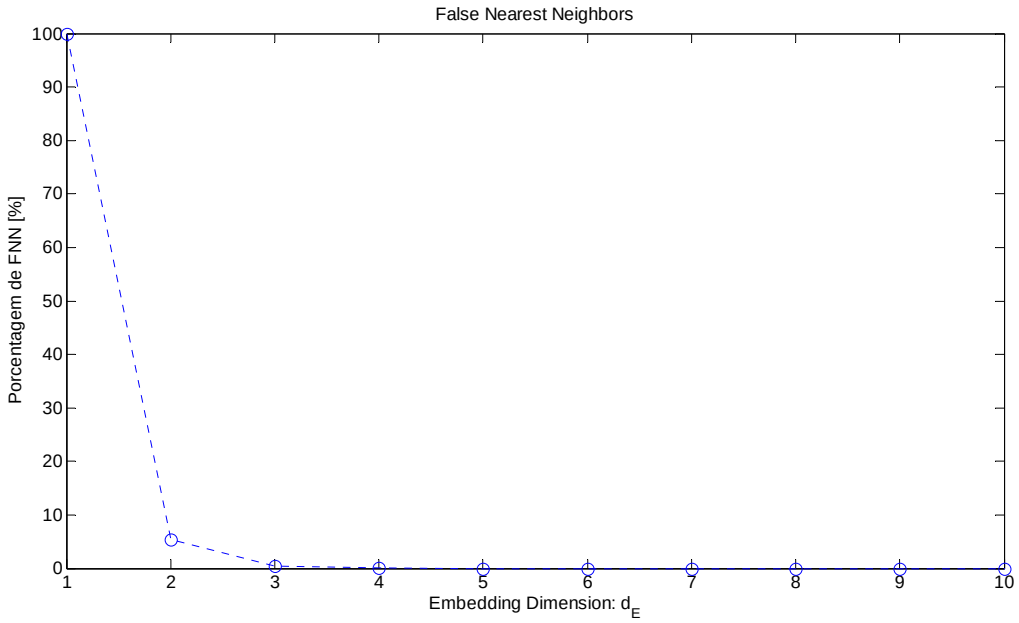


Figura 5.5 – "*False Nearest Neighbors*" FNN das Vazões Afluentes à UHE de Furnas

A Figura 5.6 apresenta a porcentagem de FNN da Amplitude A da série temporal Log(VA) em função da dimensão, utilizando o "*time delay*" $\tau = 4$ definido na seção 5.1. É possível observar que a porcentagem de FNN a zero na "*embedding dimension*" $d_E = 4$. Assim, reescrevendo a Equação 5.7, o vetor de "*time delay*" da Amplitude é definido por:

$$\vec{y}^{A}(n) = [A(t)\,,\, A(t-4)\,,\, A(t-8)\,,\, A(t-12)] \qquad \textbf{(5.13)}$$

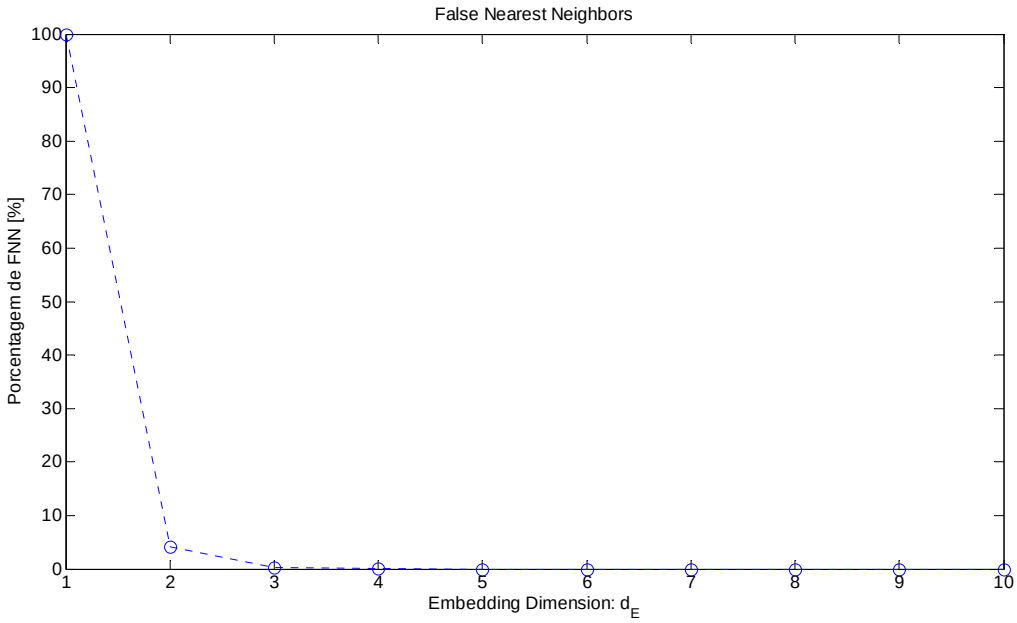


Figura 5.6 - "*False Nearest Neighbors*" FNN da Amplitude do Log(VA) à UHE de Furnas

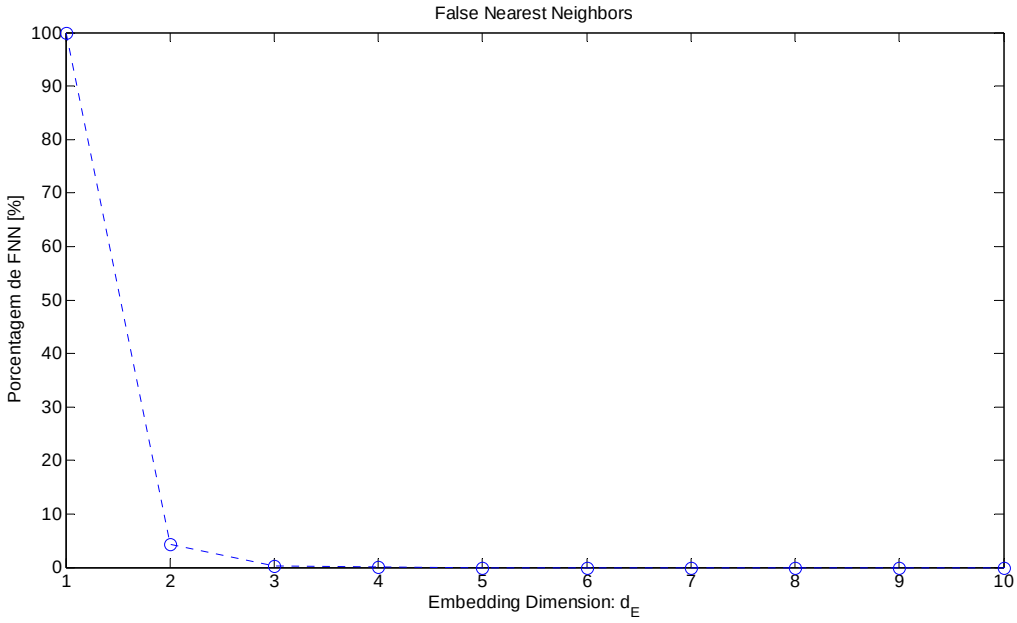


Figura 5.7 -"*False Nearest Neighbors*" FNN do Ângulo de Fase do Log(VA) à UHE de Furnas

E a Figura 5.7 apresenta a porcentagem de FNN do Ângulo de Fase ϕ da série temporal Log(VA) em função da dimensão, utilizando o "*time delay*" $\tau = 8$ definido na seção

5.1. É possível observar que a porcentagem de FNN a zero na "*embedding dimension*" $d_E = 4$. Assim, reescrevendo a Equação 5.8, o vetor de "*time delay*" do Ângulo de Fase é definido por:

$$\vec{y}^{\phi}(n) = [\phi(t), \phi(t-8), \phi(t-16), \phi(t-24)] \quad \textbf{(5.14)}$$

Agora que o Espaço de Reconstrução está totalmente definido para a série temporal original de VA pela Equação 5.12, para a Amplitude do log(VA) pela Equação 5.13 e para o Ângulo de Fase do log(VA) pela Equação 5.13, o modelo de previsão pode ser ajustado para aproximar o mapeamento *f*; objeto dos próximo capítulos.

Capítulo 6

Sistemas Inteligentes

Na era do conhecimento, o nível de capacitação do indivíduo e da empresa é o fator determinante de sua sobrevivência, pois o valor não estará mais no domínio da informação, mas sim em como trabalhar com o conhecimento relacionado a esta informação.

Em razão de os Sistemas Inteligentes (SI) poderem manipular símbolos que representam entidades do mundo real, eles são capazes de trabalhar com conhecimento de forma eficaz. Para entender o que isto significa, é necessário definir o que sejam dados, informação e conhecimento (Rezende, 2005):

O dado é um elemento quantificável sobre um determinado evento. Geralmente, os dados são adquiridos do ambiente operacional, armazenados, selecionados e recuperados de um Banco de Dados. O dado, por si só, não oferece embasamento para o entendimento de uma situação.

A informação é o dado analisado e contextualizado, envolve a interpretação de um conjunto de dados para a definição de um contexto, da referência, ou seja, a base para que se obtenha uma informação a partir de um dado.

Já o conhecimento, refere-se à habilidade de criar um modelo que descreva o objeto e indique as ações a serem implementadas, as decisões a serem tomadas. A compreensão, análise e síntese, necessárias para a tomada de decisões inteligentes, são realizadas a partir do nível do conhecimento.

Uma decisão é o uso explícito de um conhecimento. O conhecimento pode ser representado como uma combinação de estruturas de dados e procedimentos interpretativos que levam a um comportamento conhecido. Este comportamento fornece informações a um sistema que pode, então, planejar e decidir.

A busca da viabilização da transferência desses processos para o computador constitui um campo de pesquisa denominado Inteligência Artificial (IA), com a proposta de desenvolver Sistemas Inteligentes (SI) capazes de simular ou emular o processo de decisão do ser humano.

Os SI podem ser desenvolvidos usando-se algumas técnicas-chave, as quais podem ser aplicadas isoladamente ou em conjunto para auxiliar o processo decisório. As principais técnicas e metodologias usadas pelos SI são: *Aquisição de Conhecimento*, *Aprendizagem de Máquina*, *Redes Neurais Artificiais*, *Lógica Fuzzy* e *Sistemas de Inferência Fuzzy*, *Computação Evolutiva*, *Agentes e Multiagentes* (Rezende, 2005).

Sendo que cada uma destas técnicas possui diferentes graus de habilidade para representar o conhecimento humano.

6.1 Rede SONARX

Nesta seção é apresentada a arquitetura da rede de Mapa Auto-Organizável (*Self-Organizing Map* – SOM, em inglês) de Kohonen com Modelos de Sistemas Dinâmicos, além de toda sua formulação, também serão descritas suas principais características, e inclusive seus campos de aplicação.

Este modelo neural não-supervisionado é capaz de processar padrões espaço-temporais, baseado na rede de Mapa Auto-Organizável (rede SOM). Ele passa a ser uma versão não-supervisionada do modelo NARX (do inglês, *Nonlinear Auto-Regressive model*

with eXogenous input), sendo assim chamado de modelo ou rede NARX Auto-Organizável (SONARX).

Vale ressaltar que a rede SONARX é uma instância particular de uma técnica mais geral que permite utilizar RNA's Temporais competitivas na modelagem entrada-saída de sistemas dinâmicos. Esta técnica recebe o nome de Memória Associativa Temporal via Quantização Vetorial (MATQV) (Barreto e Araújo, 2001a).

Dentre as áreas de aplicação da rede SONARX , destacam-se as seguintes: identificação e controle de sistemas dinâmicos, tendo como exemplo um atuador hidráulico; a aprendizagem de trajetórias robóticas complexas, como controle preditivo; a modelagem e previsão de série temporais, que será objeto de estudo deste trabalho.

Na tentativa de representar de forma reduzida a seqüência de entrada, por meio da técnica de quantização vetorial, utiliza-se então como base do modelo aqui apresentado, a rede SOM. Nesta rede os vetores-protótipos (centróides) representam de forma compacta os elementos da seqüência, através dos vetores de pesos associados aos neurônios da rede SOM.

Uma representação condensada de uma seqüência ou série temporal somente é possível porque a rede SONARX utiliza um processo de aprendizagem estático que precisa de várias apresentações do conjunto de treinamento, ou seja vários ciclos de treinamento, para ajustar os pesos de forma adequada.

A maior limitação da aprendizagem de vários ciclos está na ocorrência de destruição da memória adquirida à medida que novas informações vão sendo apresentadas à rede. Este fenômeno é chamada de Interferência Catastrófica (Grossberg, 1986) apud (Barreto, 2003).

Apesar desta limitação, foi mostrado por (Barreto, 2003), que a rede SONARX tem diversas aplicações, não exigindo alterações substanciais da arquitetura da rede SOM para ser aplicada em vários domínios.

6.1.1 A Rede SOM

A habilidade humana de processar informação foi inicialmente um mistério, sendo Ross Ashby citado como o primeiro a descrever o sistema nervoso como auto-organizado. Em seu livro, (Ashby, 1952) apud (Barreto, 2003) estava preocupado com a noção básica de que o comportamento adaptativo não é inato, mas sim aprendido, e que através da aprendizagem o comportamento de um animal ou sistema muda, ou seja, se adapta.

Haja visto que em (Ashby, 1947) apud (Barreto, 2003) já se pensava em sistemas auto-organizáveis como um caso de sistemas orientados à meta que são capazes, sem ajuda externa explícita, de melhorar sua performance ao longo do tempo, enquanto perseguem seus objetivos e que devem ser avaliados por alguma função de desempenho.

Em outras palavras, o que é usualmente referido como um comportamento auto-organizável é a formação espontânea de estruturas organizadas, como padrões ou comportamentos, a partir de condições iniciais aleatórias.

A questão básica consiste em entender como a ação de vários elementos simples (neurônios), em uma configuração inicial desordenada ou aleatória, dá margem ao surgimento espontâneo (sem regras pré-programadas) de padrões coerentes de comportamento.

A resposta reside essencialmente na seguinte observação feita por (Turing, 1952) apud (Barreto, 2003): “Ordem global pode surgir espontaneamente a partir de interações locais”, observação esta que se aplica tanto ao estudo de cérebro, quanto ao de Redes Neurais Artificiais.

A organização da rede se dá em dois planos diferentes, que interagem entre si, na forma de um laço de realimentação. O primeiro deles é a chamada Ativação, caracterizada pela atividade neural, de cada unidade de processamento, produzida por uma dada rede em resposta a estímulos de entrada. E o segundo é a Conectividade, que seria o conjunto de

conexões ou pesos sinápticos da rede, em modificação sob respostas aos padrões de ativação dos neurônios da rede.

Atualmente há muitas teorias que apóiam a hipótese de auto-organização do sistema nervoso, sendo possível enumerar quatro princípios que parecem estar presentes em cada uma delas:

1- Modificação nos pesos sinápticos tendem a se auto-amplificar;

2- Limitação de recursos leva à competição entre sinapses e, em conseqüência, à seleção das sinapses mais vigorosas, em detrimento de outras menos adequadas;

3- A presença de uma sinapse de valor elevado pode melhorar a adequabilidade (*fitness*) de outras sinapses;

4- Redundância provê conhecimento (informação).

Os quatro princípios sobre aprendizagem não-supervisionada citados anteriormente formam a base neurobiológica subjacente a um importante algoritmo neural: a rede SOM.

Uma representação econômica dos dados com todas as suas inter-relações é um dos problemas centrais da informação, e tal habilidade é também característica da operação do cérebro. Ao processar informação, há uma tendência em comprimí-la, formando representações reduzidas dos fatos mais relevantes, sem perda de conhecimento a cerca de suas inter-relações. Sendo justamente esta a característica de interesse na aplicação da rede Auto-Organizável neste trabalho.

O processamento inteligente da informação, em geral, parece ser a criação de imagens simplificadas do mundo observável em vários níveis de abstração, a partir de uma seqüência de dados de entrada.

Apenas recentemente tem sido possível obter evidências concretas sobre o cérebro formando representações neurais que podem ser consideradas imagens de várias representações sensoriais, projetadas sobre várias áreas corticais. Dá-se o nome genérico de mapas corticais

ou mapas topográficos a essas áreas de projeção. Tais mapas têm importância fundamental para os processos de aprendizagem e memória, pois definem uma representação compacta dos dados sensoriais.

Um conjunto de modelos computacionais, chamados de mapas topológicos auto-organizáveis, vem sendo propostos com o intuito de modelar diversas propriedades de mapas corticais. Estes modelos pertencem à classe das redes neurais artificiais não-supervisionadas. Nesta secção, será apresentada a teoria relativa a um desses mapas topológicos, proposto por (Kohonen, 1989), conhecido como Mapa Auto-Organizável (SOM), cuja teoria apresentada envolve o estudo do algoritmo SOM e determinadas propriedades como convergência e estabilidade. A apresentação da Rede SOM nesta secção objetiva fornecer subsídios necessários à correta utilização do algoritmo.

Sem ter a ambição de modelar fielmente um neurônio real, a rede SOM foi proposta com a finalidade de modelar algumas características essenciais dos mapas corticais cerebrais. Tendo como objetivo principal transformar um sinal de entrada de dimensão qualquer em um arranjo discreto de unidades de processamento. Esta transformação deve ser realizada de forma adaptativa e mantendo relações de similaridade entre os dois espaços.

Topologicamente a rede SOM pode ser tratada da seguinte maneira. Seja **X** um espaço contínuo de dados de entrada, sendo dele selecionado aleatoriamente um conjunto de vetores $\mathbf{x} \in \mathbf{X}$, sujeito a uma organização aleatória de densidade de probabilidade p(**x**). Já a topologia do espaço discreto de saída **A** é definida pelo arranjo geométrico do conjunto de neurônios $\mathbf{i} \in \mathbf{A}$.

Desta maneira pode-se definir uma transformação não-linear $\phi : \mathbf{X} \rightarrow \mathbf{A}$, que leva do espaço de entrada **X** ao espaço de saída **A**. Tal transformação pode ser visualizada na Figura 6.1, sendo chamada de mapeamento de características (*Feature Mapping*, em inglês).

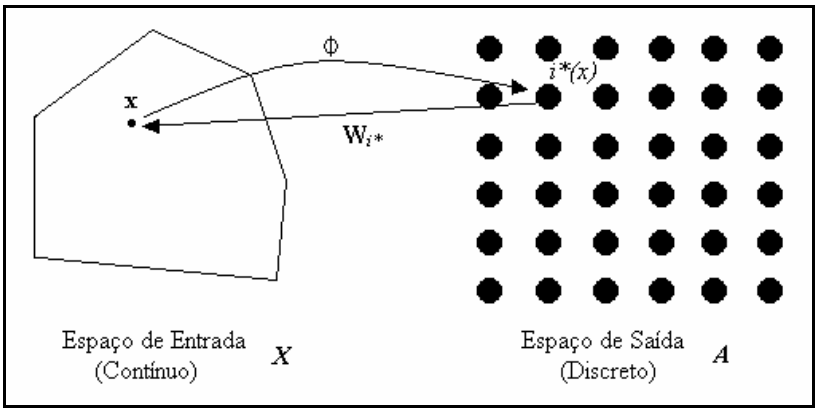


Figura 6.1 – Topologia dos espaços de entrada e saída, e o mapeamento de características ϕ

Estabelecendo assim uma relação neural para tal mapeamento, o espaço de entrada **X** pode representar o conjunto de coordenadas de receptores somato-sensoriais densamente distribuídos sobre toda a superfície do corpo. Já o espaço de saída **A** representa o conjunto de neurônios localizados em uma certa porção do córtex cerebral que responde aos estímulos gerados por tais receptores.

Seja o espaço contínuo de entrada **X** de dimensão n , um vetor de entrada (padrão de entrada) $\mathbf{x} \in \mathbf{X} \subset \Re^n$, pode ser representado por:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{6.1}$$

Da mesma forma, o vetor de pesos associado a cada neurônio da rede tem a mesma dimensão do vetor de entrada. A representação do vetor de pesos de um neurônio i é a seguinte:

$$\mathbf{w}_i = \begin{pmatrix} w_{i1} \\ w_{i2} \\ \vdots \\ w_{in} \end{pmatrix}, \quad i = 1,2,\cdots,M \tag{6.2}$$

onde M é o número total de neurônios da rede.

Cada neurônio i no arranjo está totalmente conectado às n unidades de entrada por meio de termos ponderadores, w_{ij} , $j = 1,2,...,n$, chamados pesos sinápticos, agrupados no vetor $\boldsymbol{w}_i$, como pode ser visto na Figura 6.2(a).

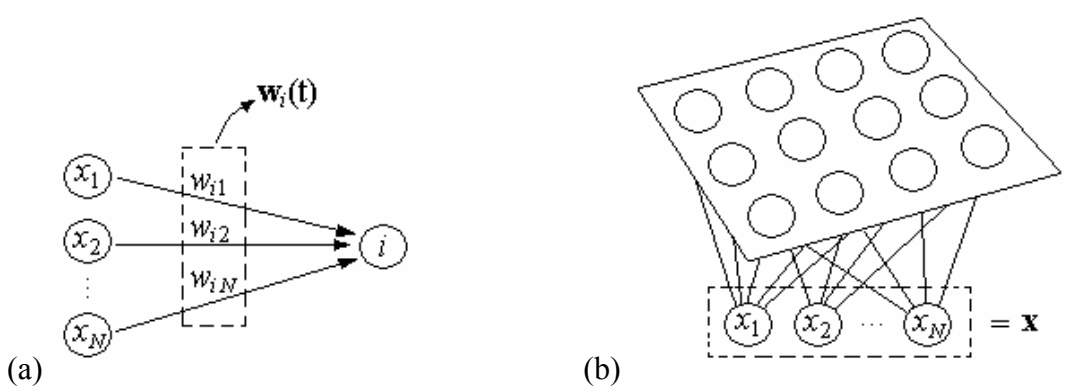


Figura 6.2 – (a) Vetor de pesos de um dado neurônio *i*. (b) Rede SOM bidimensional

Esta rede possui uma estrutura de propagação para frente (*feedforward*, em inglês) com uma única camada de processamento, constituída de M neurônios normalmente dispostos de forma retangular, no caso da topologia bidimensional. Assim a Figura 6.2(b) mostra um diagrama esquemático de um arranjo de neurônios comumente usado como espaço discreto de saída.

Outras configurações devem ser destacadas, tais como de um arranjo unidimensional, onde a camada de processamento é formada por uma única linha de neurônios; e a configuração hexagonal, a qual é bastante adequada para a visualização do mapeamento topográfico ϕ .

Definida a organização da topologia da rede SOM e seus dados de entrada e saída, o algoritmo responsável pela formação do mapa auto-organizável tem como passo inicial a atribuição de valores iniciais (normalmente aleatórios) aos pesos da rede.

Em seguida se dá a Competição, a fim de se determinar dentre todos os neurônios aquele que melhor representa o padrão de entrada, sendo este definido como o neurônio vencedor. Esta etapa é muito importante tanto na fase de treinamento quanto de operação da

rede SOM. A Cooperação representa a interação entre o neurônio vencedor e seus vizinhos, mais intensa a medida que se trata de um neurônio próximo do vencedor. Já o Ajuste dos pesos sinápticos da rede se dá somente durante a fase de treinamento da rede, onde o neurônio vencedor, dado um padrão de entrada, recebe um ajuste mais intenso, enquanto que seus vizinhos são ajustados em menor intensidade.

Para encontrar o neurônio vencedor, o vetor de pesos de todos os neurônios da rede SOM é comparado, um a um, com o vetor de entrada. O neurônio cujo vetor de pesos está mais próximo do vetor de entrada é considerado o vencedor ***i****. A distância euclidiana, definida pela Equação 6.3, pode ser implementada para tal finalidade, como na Equação 6.4.

$$\|\mathbf{x}(t) - \mathbf{w}_i(t)\| = \sqrt{\sum_{j=1}^{n} \left[x_j(t) - w_{ij}(t)\right]^2} \tag{6.3}$$

$$i^*(x(t)) = \arg\min_{i \in A} \|\mathbf{x}(t) - \mathbf{w}_i(t)\| \tag{6.4}$$

onde $i^*(\mathbf{x})$ é o índice que representa o neurônio vencedor para o padrão de entrada $\mathbf{x}(t)$.

O neurônio vencedor $i^*(\mathbf{x})$ determina o centro espacial de um grupo dentre seus vizinhos. Sendo que o neurônio vencedor e sua vizinhança interagem lateralmente de forma cooperativa. A intensidade desta iteração lateral entre o vencedor i^* e um neurônio i qualquer é, em geral, descrita matematicamente na forma de uma função vizinhança $h(i^*,i;t)$. Esta função define o que se chama de vizinhança topológica centrada no neurônio vencedor $i^*(\mathbf{x})$.

A função vizinhança deve satisfazer os seguintes requisitos:

- alcançar seu valor máximo para uma distância lateral nula;
- ser simétrica em relação ao neurônio vencedor;
- sua amplitude decai monotonicamente com o aumento da distância lateral.

Uma escolha comum para $h(i^*,i;t)$ que satisfaz os requisitos acima é a função gaussiana, definida por:

$$h(i^*, i; t) = \exp\left(-\frac{\left\| \mathbf{w}_{i^*} - \mathbf{w}_i \right\|^2}{2\sigma^2(t)} \right) \tag{6.5}$$

onde o parâmetro $\sigma(t)$ define a "largura efetiva" da vizinhança topológica.

A partir deste tipo de função vizinhança, pode-se derivar uma rede competitiva simples, conhecida como rede WTA (do inglês, *Winner-Take-All*), em que apenas o neurônio vencedor tem seu vetor de pesos atualizado a cada iteração do algoritmo.

O ajuste dos pesos sinápticos é o último passo na formação de um mapa auto-organizável. Como a rede SOM é não-supervisionada, o vetor de pesos de um dado neurônio *i* deve ser modificado como função do estímulo de entrada apenas. Durante a adaptação sináptica o objetivo é extrair algum tipo de regularidade presente no espaço de entrada **X**. A partir das suposições feitas por (Hebb, 1949) apud (Barreto, 2003) na tentativa de relacionar estímulos de entrada à uma estrutura neural com suas sinapses neurais, Kohonen propôs uma representação matemática para o ajuste dos pesos para a rede SOM :

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \alpha(t) \cdot h(i^*, i; t) \cdot [\mathbf{x}(t) - \mathbf{w}_i(t)] \tag{6.6}$$

onde $\alpha(t)$ define a taxa de aprendizagem.

A taxa de aprendizagem controla a velocidade com que os pesos sinápticos são modificados, ou seja, adquire novo conhecimento, devendo cair monotonicamente com o treinamento, para garantir tanto a convergência como a estabilidade da rede SOM.

Desta maneira, o mapeamento ϕ resultante da aplicação deste algoritmo SOM, representado pelo conjunto de vetores de pesos sinápticos $\mathbf{w}_i$ que compõem **A**, provê uma aproximação discreta do espaço de entrada **X**. Cada neurônio da rede é responsável por uma determinada região do espaço de entrada que define a sua região de influência ou seu campo receptivo. A região de influência de um dado neurônio pode indicar que todos os padrões de treinamento que ativam tal neurônio compõem um grupo ou categoria. Assim, padrões pertencentes a um mesmo grupo guardam fortes semelhanças entre si, enquanto padrões

pertencentes a grupos diferentes guardam semelhanças menores. Dessa forma, é correto afirmar que uma propriedade intrínseca do algoritmo SOM é a categorização dos padrões de entrada.

Após rotular adequadamente cada grupo com base nos vetores de treinamento, a rede pode ser utilizada para classificar padrões que venham a ser apresentados. O vetor de pesos do neurônio responsável por uma dada categoria é considerado o "protótipo" daquela categoria com o qual o novo padrão de entrada é comparado para determinar se ele pertence àquela categoria. Diante disso, os vetores de pesos dos neurônios da rede podem também receber outras denominações tais como vetores protótipos ou centróides.

Resumidamente, na operação desta rede um dado vetor de entrada $\mathbf{x} \in \mathbf{X}$ é fornecido a rede, o algoritmo SOM primeiro identifica um determinado neurônio vencedor $i^*(\mathbf{x})$ no espaço de saída **A** em concordância com o mapeamento característico ϕ. O vetor de pesos $\mathbf{w}_{i^*}$ associado ao neurônio i^* pode ser visto como um "ponteiro" para o espaço de entrada **X**; ou seja, os elementos do vetor $\mathbf{w}_i$ podem ser vistos como as coordenadas da imagem do neurônio i projetada no espaço de entrada.

6.1.2 A Rede SOM com Modelos de Sistemas Dinâmicos

Conforme apresentado no item 6.1.1, a rede SOM é uma das mais populares arquiteturas de RNAs não-supervisionadas. Sua principal área de aplicação envolve tarefas de reconhecimento de padrões, tais como formação de categorias e agrupamentos (*Clusters*, em inglês), classificação e visualização de dados (Kaski, Kangas e Kohonen, 1998).

Isto se deve, em grande parte, à simplicidade do seu algoritmo de aprendizagem, que pode ser resumido em dois passos principais:

1. Busca pelo neurônio vencedor, $i^*(t)$, para o vetor de entrada $\mathbf{x}(t)$:

$$i^*(t) = \arg\min_{i \in A} \left\| \mathbf{x}(t) - \mathbf{w}_i(t) \right\| \tag{6.7}$$

2. Atualização dos vetores de pesos, $\mathbf{w}_i(t)$, da rede:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \alpha(t) \cdot h(i^*, i; t) \cdot [\mathbf{x}(t) - \mathbf{w}_i(t)] \tag{6.8}$$

onde $\alpha(t)$ é a taxa de aprendizagem e $h(i^*,i;t)$ é uma função vizinhança do tipo gaussiana:

$$h(i^*, i; t) = \exp\left(-\frac{\left\| \mathbf{w}_{i^*}(t) - \mathbf{w}_i(t) \right\|^2}{\sigma^2(t)} \right) \tag{6.9}$$

onde $\sigma(t)$ define a largura da vizinhança, enquanto $\mathbf{w}_{i^*}(t)$ e $\mathbf{w}_i(t)$ são, respectivamente, as posições dos neurônios i^* e i no arranjo geométrico **A**, ao qual eles pertencem. As variáveis $0 < \alpha(t), \sigma(t) < 1$ decaem exponencialmente com o passar do tempo:

$$\alpha(t) = \alpha_0 \left(\frac{\alpha_T}{\alpha_0} \right)^{(t/T)} \quad \text{e} \quad \sigma(t) = \sigma_0 \left(\frac{\sigma_T}{\sigma_0} \right)^{(t/T)} \tag{6.10}$$

onde α_0 e σ_0 são os valores iniciais e α_T e σ_T os finais de $\alpha(t)$ e $\sigma(t)$, transcorridas **T** iterações de treinamento.

Esta rede SOM permite que ela aprenda o mapeamento entrada-saída estático. Para isso, a saída atual pode ser mapeada somente em função da entrada atual (Walter e Ritter, 1996):

$$\mathbf{y}(t) = G(\mathbf{u}(t)) \tag{6.11}$$

onde $\mathbf{u}(t) \in \Re^n$ e $\mathbf{y}(t) \in \Re^m$ denotam, respectivamente, o vetor de entrada e o vetor de saída do sistema. O mapeamento $G(\cdot)$ é chamado de mapeamento direto (entrada $\rightarrow$ saída), sendo em geral, desconhecido e não-linear.

Por outro lado, em muitos problemas de sistemas dinâmicos deseja-se saber o mapeamento inverso $G^{-1}(\cdot)$. Porém, este pode ser de difícil aproximação, já que normalmente é definido como "mal-colocado" no sentido de *Hadamard*.

A formulação proposta no mapeamento estático consiste em modificar o vetor de entrada da rede SOM, ***x(t)***, de modo que ele passe a ter duas partes:

- $\mathbf{x}^{in}(t)$: que corresponde à informação de entrada do mapeamento que se quer aproximar.
- $\mathbf{x}^{out}(t)$: que corresponde à informação da saída deste mesmo mapeamento.

Como conseqüência, os vetores dos neurônios *i* têm também suas dimensões aumentadas. Matematicamente, tem-se que:

$$\mathbf{x}(t) = \begin{pmatrix} \mathbf{x}^{in}(t) \\ \mathbf{x}^{out}(t) \end{pmatrix} \tag{6.12}$$

Assim, pode-se utilizar a rede SOM para aprender o mapeamento mostrado na Equação 6.11, definindo os vetores $\mathbf{x}^{in}(t)$ e $\mathbf{x}^{out}(t)$ da seguinte maneira:

$$\begin{pmatrix} \mathbf{x}^{in}(t) \\ \mathbf{x}^{out}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{u}(t) \\ \mathbf{y}(t) \end{pmatrix} \tag{6.13}$$

Entretanto, o maior interesse deste trabalho é utilizar uma formulação para mapeamentos dinâmicos, o que justifica a utilização da formulação proposta por (Walter e Ritter, 1996), onde mostraram que esta formulação pôde ser estendida para aproximar também mapeamentos dinâmicos. E este tipo de mapeamento passa a ser genericamente descrito como:

$$\mathbf{y}(t+1) = F[\mathbf{y}(t),...,\mathbf{y}(t-n_y+1);\mathbf{u}(t),...,\mathbf{u}(t-n_u+1)] \tag{6.14}$$

onde n_u e n_y definem as ordens da memória de entrada e de saída, respectivamente, e o mapeamento $F(\cdot)$ é não-linear e desconhecida. De acordo com a Equação 6.14, a saída do sistema no instante $t+1$, depende dos n_y valores passados da saída e dos n_u valores passados da entrada. A representação matemática de sistemas dinâmicos mostrada na Equação 6.14 é genericamente conhecida como modelo NARX (do inglês, *Nonlinear AutoRegressive model with eXogenous variables*) (Sjöberg et al., 1995) (Aguirre, 2000) (Norgaard et al., 2000).

Para a rede SOM ser capaz de aprender mapeamentos dinâmicos, ela deve possuir algum tipo de mecanismo de memória. Ou seja, ela deve ser capaz de armazenar informação passada sobre os vetores de entrada e saída do sistema. Estes mecanismos de memória permitem que a rede SOM seja capaz de processar as séries temporais das variáveis de entrada e saída, $[\mathbf{u}(t),\mathbf{y}(t)]$, $t = 1,..., N$, que serão usados para construir o modelo. Dentre os possíveis mecanismos de memória que vêm sendo utilizados para construir versões temporais da rede SOM, apenas linhas de atraso (*delay line*, em inglês) serão utilizados neste trabalho.

A questão que se coloca neste instante passa então a ser como utilizar a rede SOM para obter uma aproximação do mapeamento $F(\cdot)$ de posse apenas das séries temporais das variáveis de entrada e saída.

Uma possível solução para esta questão utiliza um raciocínio semelhante ao do caso estático, ou seja, escolher adequadamente as variáveis que compõem os vetores $\mathbf{x}^{in}(t)$ e $\mathbf{x}^{out}(t)$.

Desta maneira, na busca pela aproximação do mapeamento representado na Equação 6.14, define-se:

$$\mathbf{x}^{in}(t) = [\mathbf{y}(t),...,\mathbf{y}(t-n_y+1);\mathbf{u}(t),...,\mathbf{u}(t-n_u+1)] \quad \textbf{(6.15)}$$

$$\mathbf{x}^{out}(t) = \mathbf{y}(t+1) \quad \textbf{(6.16)}$$

Apesar de parecer uma simples extensão da formulação estática de (Walter e Ritter, 1996), é importante ressaltar um aspecto que torna esta abordagem mais complexa. Como conseqüência direta do uso de atrasadores, acrescentou-se ao problema de aproximação dos mapeamentos dois outros graus de liberdade relativos à determinação das ordens n_y e n_u do modelo. Estas variáveis influenciam diretamente na precisão da aproximação e na complexidade (número de pesos) do modelo.

Desta maneira, denotando por $\boldsymbol{P}(t)$ o padrão de entrada da rede SOM no instante t, tem-se as seguintes definições:

$$\mathbf{P}(t) = \begin{pmatrix} \mathbf{x}^{in}(t) \\ \mathbf{x}^{out}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{y}(t),...,\mathbf{y}(t-n_y+1);\mathbf{u}(t),...,\mathbf{u}(t-n_u+1) \\ \mathbf{y}(n_y+t) \end{pmatrix} \quad \textbf{(6.17)}$$

Os padrões ***P(t)*** assim gerados podem ser apresentados seqüencialmente ou aleatoriamente para treinar a rede SOM.

Durante o treinamento, o neurônio vencedor, *i*(t)*, é encontrado usando apenas a porção correspondente a $\mathbf{x}^{in}(t)$:

$$i^*(t) = \arg\min_{i \in A} \left\| \mathbf{x}^{in}(t) - \mathbf{w}_i^{in}(t) \right\| \quad \textbf{(6.18)}$$

Na atualização dos pesos, ambos $\mathbf{w}^{in}(t)$ e $\mathbf{w}^{out}(t)$ são utilizados:

$$\mathbf{w}_i^{in}(t+1) = \mathbf{w}_i^{in}(t) + \alpha(t) \cdot h(i^*,i;t) \cdot [\mathbf{x}^{in}(t) - \mathbf{w}_i^{in}(t)] \quad \textbf{(6.19)}$$

$$\mathbf{w}_i^{out}(t+1) = \mathbf{w}_i^{out}(t) + \alpha(t) \cdot h(i^*,i;t) \cdot [\mathbf{x}^{out}(t) - \mathbf{w}_i^{out}(t)] \quad \textbf{(6.20)}$$

Ainda durante a fase de treinamento, a rede SOM aprende a associar a saída do mapeamento $\mathbf{x}^{out}(t)$, com a entrada correspondente $\mathbf{x}^{in}(t)$, além de realizar também a quantização vetorial do espaço de entrada e do espaço de saída.

E é justamente a quantização vetorial que permite uma representação compacta dos dados. Por isso, esta técnica é chamada de Memória Associativa Temporal por Quantização Vetorial (MATQV).

O modelo neural utilizado neste trabalho, por ser baseado na rede SOM e implementar uma versão não-supervisionada do modelo NARX, é então chamado de rede SONARX.

Após o treinamento, a rede SONARX é utilizada para obter estimativas dos valores de saída dos mapeamentos, a partir do vetor de pesos $\mathbf{w}_{i^*}^{out}(t)$ do neurônio vencedor. Definindo a saída da rede no instante *t* por $\hat{\mathbf{y}}(t+1)$, onde o neurônio vencedor *i*(t)* é determinado segundo a Equação 6.18, tem-se:

$$\hat{\mathbf{y}}(t+1) \cong \mathbf{w}_{i^*}^{out}(t) \quad \textbf{(6.21)}$$

Assim, este processo de estimação se repete até que a rede convirja para o estado de melhor aproximação do mapeamento em questão, sendo importante ressaltar ao menos dois aspectos importantes do método MATQV.

Em primeiro lugar, o neurônio vencedor $i^*(t)$ é o elemento responsável por fazer a ligação entre a os dados de entrada $\mathbf{x}^{in}(t)$, com os dados de saída $\mathbf{x}^{out}(t)$, do mapeamento. Esta associação fica codificada nos respectivos vetores de pesos, $\mathbf{w}_{i^*}^{in}(t)$ e $\mathbf{w}_{i^*}^{out}(t)$.

E em segundo lugar, existe uma diferença fundamental da estratégia de treinamento associativa do método MATQV em relação àquela comumente usada no treinamento de redes supervisionadas (MLP ou RBF), que é baseada na redução do erro de aproximação. Nestas redes, o vetor $\mathbf{x}^{in}(t)$ é utilizado na entrada da rede, enquanto o vetor $\mathbf{x}^{out}(t)$ é utilizado na saída para calcular o erro de aproximação usado para guiar o ajuste dos pesos da rede. Por outro lado, quando se usa a técnica MATQV, o vetor $\mathbf{x}^{out}(t)$ é apresentado na entrada da rede juntamente com o vetor $\mathbf{x}^{in}(t)$, sem o cálculo explícito do erro de aproximação.

Porém, o método MATQV é também baseado na redução do erro de aproximação, mas de forma implícita, ou seja, o erro é usado indiretamente para o ajustar os pesos de saída, $\mathbf{w}_{i}^{out}(t)$.

6.1.3 Análise da Convergência do Modelo SONARX

Nesta seção deseja-se saber se a rede SONARX, descrita como sendo a aplicação da técnica MATQV sobre o modelo de rede SOM, realmente aproxima de forma assintótica um determinado mapeamento dinâmico. Esta verificação será feita através de uma análise teórica do processo de convergência da rede SONARX.

Esta análise será feita com base no método de aproximação estocástica de (Robbins e Monro, 1951) apud (Barreto, 2003). Ela visa mostrar que o erro de estimação

decresce assintoticamente à medida que o treinamento da rede evolui, convergindo para um estado de equilíbrio estável, considerando que a função a ser mapeada é a definida pela Equação 6.14 e que a única fonte de informação disponível consiste em pares entrada-saída observados [$\mathbf{u}(t)$,$y(t)$],$t = 1,...,N$.

De acordo com as definições da seção anterior, a rede SONARX pode ser usada para prover uma aproximação, $\hat{F}(\cdot)$, da função não-linear $F(\cdot)$ usando mecanismos de memória associativa. Neste caso, a rede deve fornecer uma estimativa da saída do sistema, $\hat{\mathbf{y}}(t+1)$, dado o vetor de entrada $\mathbf{x}^{in}(t)$. Sendo representado matematicamente da seguinte forma:

$$\hat{\mathbf{y}}(t+1) \cong \mathbf{w}_{i*}^{out}(t) = \hat{F}[\mathbf{x}^{in}(t); i*(t)] \tag{6.22}$$

Podendo ser definida uma medida para o valor absoluto do erro de previsão da saída $\mathbf{y}(t+1)$ futura do sistema:

$$\in(t) = \left\|\mathbf{y}(t+1) - \hat{\mathbf{y}}(t+1)\right\| = \left\|\mathbf{x}^{out}(t) - \mathbf{w}_{i*}^{out}(t)\right\| \tag{6.23}$$

Através desta definição pode-se agora provar que o método não-supervisionado proposto para aproximação de função é, na verdade, um procedimento "auto-supervisionado" (do inglês, *self-supervised*) de aprendizagem baseada no erro de aproximação.

A análise que se segue é válida apenas para a fase de convergência do processo de aprendizado, fase esta em que a função vizinhança é, em geral, invariante no tempo. Para levar em conta, a interação entre o neurônio vencedor e seus vizinhos durante o processo de convergência, define-se a seguinte função J:

$$J(t) = \sum_{i=1}^{L} h(i*,i,t) \in^2 (t) = \sum_{i=1}^{L} h(i*,i,t)\left\|\mathbf{x}^{out}(t) - \mathbf{w}_{i}^{out}(t)\right\|^2 \tag{6.24}$$

onde L é o número de neurônios da rede.

A convergência da rede pode ser avaliada através do valor esperado do erro médio quadrático de aproximação, dado por:

$$\widetilde{J} = E\left\{\sum_{i\in L} h(i^*,i) \in^2\right\} = \int_{\mathrm{X}} \sum_{i\in L} h(i^*,i)\left\|\mathbf{x}^{out} - \mathbf{w}_i^{out}\right\|^2 p(\mathrm{X})d\mathrm{X} \qquad \textbf{(6.25)}$$

onde se assume que o valor esperado é tomado sobre uma seqüência infinita, $\mathbf{X}=[\mathbf{x}^{\mathrm{out}}(\mathrm{t})\ ;\ t = 1, 2, \ldots, \infty]$ de amostras estocásticas sobre a qual o estimador é formado. O termo $p(\mathbf{X})$ refere-se à função densidade de probabilidade da seqüência **X**, enquanto o termo $d\mathbf{X}$ é um "volume infinitesimal" no espaço em que **X** está definida. O índice t é omitido para simplificar a notação, pois ele já está implícito na ordem temporal da seqüência **X**.

Visto que a função densidade de probabilidade $p(\mathbf{X})$ é desconhecida e que a seqüência **X** é finita para os casos reais, a técnica de aproximação estocástica de Robins-Monro será utilizada para a minimização de $\widetilde{J}$ visando a obtenção de um valor ótimo para o vetor de pesos $\mathbf{w}_i^{out}$. A idéia básica desta técnica é tentar diminuir o valor da função $J(t)$ a cada novo passo t, caminhando na direção oposta ao gradiente desta função em relação ao vetor $\mathbf{w}_i^{out}(t)$ atual. A formulação recursiva desta idéia é dada por:

$$\mathbf{w}_i^{out}(t+1) = \mathbf{w}_i^{out}(t) - \frac{1}{2}\alpha(t)\frac{\partial J(t)}{\partial \mathbf{w}_i^{out}(t)} \qquad \textbf{(6.26)}$$

A derivada parcial de *J(t)* na Eq. 4.24 é dada por:

$$\frac{\partial J(t)}{\partial \mathbf{w}_i^{out}(t)} = -2\cdot h(i^*,i,t)[\mathbf{x}^{out}(t) - \mathbf{w}_i^{out}(t)] \qquad \textbf{(6.27)}$$

Logo, a Equação 6.26 passa a ser escrita como:

$$\mathbf{w}_i^{out}(t+1) = \mathbf{w}_i^{out}(t) - \alpha(t)h(i^*,i,t)[\mathbf{x}^{out}(t) - \mathbf{w}_i^{out}(t)] \qquad \textbf{(6.28)}$$

Nota-se que esta é exatamente a regra de aprendizagem mostrada na Equação 6.20. Logo, conclui-se que a partir de condições iniciais arbitrárias, a seqüência de valores do vetor $\mathbf{w}_i^{out}(t)$ converge para a vizinhança do valor ótimo $\widetilde{\mathbf{w}}_i^{out}(t)$, correspondente ao mínimo de $\widetilde{J}$.

É importante ressaltar que esta técnica tende a convergir para mínimos locais de *J*, dependendo das condições iniciais de $\mathbf{w}_i^{out}(t=0)$. Entretanto, para a rede SOM unidimensional, foi mostrado que existe apenas um único estado de equilíbrio estável, representando um mínimo global (Cottrell, Fort e Pages, 1998), podendo este resultado ser aplicado diretamente à rede SONARX.

6.1.4 Representação Espaço de Estados da Rede SONARX

Para caracterizar a rede SONARX em termos da representação Espaço de Estado, é necessário definir um vetor de estados, que leve em conta o tipo de mecanismo de memória usado pelo modelo para processar padrões dinâmicos. Neste caso, para a rede SONARX, há duas fontes de memória que estão representadas pelo vetor de entrada, $\mathbf{x}^{in}(t)$, e pelo vetor de saída, $\mathbf{x}^{out}(t)$.

Para o caso em que se aproxima o modelo do sistema, pode-se definir o seguinte vetor de estados:

$$\mathbf{s}(t) = \mathbf{x}^{in}(t) = \{\mathbf{y}(t) \oplus \cdots \oplus \mathbf{y}(t-n_y+1)\} \oplus \{\mathbf{u}(t) \oplus \cdots \oplus \mathbf{u}(t-n_u+1)\} \quad \textbf{(6.29)}$$

Sendo também necessária a definição do conjunto de parâmetros adaptáveis do modelo, que incluem as matrizes de pesos $\mathbf{w}_i^{in}(t)$ e $\mathbf{w}_i^{out}(t)$:

$$\mathbf{W}(t) = \begin{pmatrix} \mathbf{w}_i^{in}(t) \\ \mathbf{w}_i^{out}(t) \end{pmatrix}, \; i = 1,\ldots,N \quad \textbf{(6.30)}$$

Conforme mostrado na Equação 6.21, a saída da rede SONARX é uma estimativa da próxima saída do sistema:

$$\mathbf{y}(t) = f_y(\mathbf{s}(t), \mathbf{W}) \cong \mathbf{w}_i^{out}(t) \quad \textbf{(6.31)}$$

Os valores iniciais dos parâmetros são dados por:

$$\mathbf{W}(0) = \begin{pmatrix} \mathbf{w}_i^{in}(0) \\ \mathbf{w}_i^{out}(0) \end{pmatrix} = \begin{pmatrix} \text{rand}(0,1) \\ \text{rand}(0,1) \end{pmatrix}, \forall i \tag{6.32}$$

em que rand(0, 1) é uma função que atribui aos componentes de **x** números aleatórios extraídos de uma distribuição uniforme entre 0 e 1. Por fim, a função de ajuste de **W** é dada por:

$$\Delta\mathbf{W}(t) = \begin{pmatrix} \Delta\mathbf{w}_i^{in}(t) \\ \Delta\mathbf{w}_i^{out}(t) \end{pmatrix} = \begin{pmatrix} \eta(t)h(i^*,i,t)[\mathbf{x}^{in}(t) - \mathbf{w}_i^{in}(t)] \\ \eta(t)h(i^*,i,t)[\mathbf{x}^{out}(t) - \mathbf{w}_i^{out}(t)] \end{pmatrix} \tag{6.33}$$

conforme calculados pelas Equações 6.19 e 6.20.

Na próxima seção se discutirá como a rede SONARX pode ser usada em uma ampla gama de aplicações, tais como identificação de sistemas não-lineares, aprendizagem de trajetórias complexas, controle preditivo não-linear e principalmente na previsão de séries temporais, que é o enfoque deste trabalho de pesquisa.

6.2 Rede SONARX-RBF

O modelo de uma rede neural supervisionada pode ser entendido de diferentes formas. O algoritmo "back-propagation" para o modelo de rede multicamada MLP (sob supervisão) pode ser visto como uma aplicação de um método de otimização, conhecido em estatística como aproximação estocástica. Neste trabalho foi considerada uma abordagem diferente ao observar o modelo de uma rede neural como um "problema de ajuste de curva" (aproximação) em um espaço de alta dimensão.

De acordo com este ponto de vista, aprender é equivalente a encontrar uma superfície em um espaço multidimensional que define o melhor ajuste para os dados de treinamento, cujo critério de "melhor ajuste" é medido por algum parâmetro estatístico. De forma correspondente, a generalização é equivalente ao uso desta superfície multidimensional para

interpolar os padrões do conjunto teste de dados. Este ponto de vista é de fato a motivação por traz do método de funções de base radial, o que conduz ao trabalho de pesquisa de interpolação tradicional em espaços multidimensionais.

No contexto de uma rede neural, as unidades escondidas fornecem um ajuste de funções que constituem uma base arbitrária para os padrões de entrada (vetores) quando eles expandidos no espaço de unidade-escondida; estas funções são chamadas de *funções de base radial*. As funções de base radial foram introduzidas pela primeira vez na interpolação de problemas reais multivariados. O primeiro trabalho neste contexto foi investigado por (Powell, 1985), sendo hoje um dos principais campos da pesquisa em análise numérica.

(Broomhead e Lowe, 1988) foram os primeiros a explorar o uso de funções de base radial na construção de redes neurais. Dentre outras grandes contribuições para a teoria, modelagem e aplicação de redes com funções de base radial estão trabalhos como (Moody e Darken, 1989), (Renals, 1989) e (Poggio e Girosi, 1990). Este último trabalho enfatiza o uso da teoria da regularização aplicada a estas classes de redes neurais como um método de generalização de dados novos.

A arquitetura básica de uma rede com Função de Base Radial (RBF) possui apenas três diferentes camadas. A camada de entrada é constituída por nós fontes (unidades sensoriais). A segunda camada é a camada escondida de dimensão suficientemente grande, a qual executa um papel diferente daquela da rede Perceptron de Multicamadas (MLP). A camada de saída é responsável pela resposta da rede para os padrões de ativação aplicados à camada de entrada. A transformação a partir do espaço de entrada para o espaço de uma das unidades escondidas é não-linear, enquanto que a transformação do espaço da camada escondida para o espaço de saída é linear. Uma justificativa matemática racional pode ser verificada no trabalho de (Cover, 1965). Em particular, este trabalho mostrou que o problema de classificação de padrões num espaço de grandes dimensões é melhor linearmente separável

do que em problemas de baixas dimensões; sendo esta a razão da construção de uma rede RBF com muitas unidades na camada intermediária. Entretanto, através de uma modelagem cuidadosa, é possível reduzir a dimensão do espaço da camada escondida, especialmente se os centros das unidades escondidas forem ajustados de forma adaptativa.

6.2.1 Teorema de Cover: Separabilidade dos Padrões

Quando uma rede com função de base radial (RBF) padrão é usada em tarefas de classificação de padrões complexos, o problema é basicamente resolvido pela transformação deste dentro de um espaço de grande dimensão de maneira não-linear. O teorema de Cover, sobre a separabilidade dos padrões, é a justificativa básica para isto, o qual estabelece que um problema complexo de classificação de padrões levado a um espaço de alta dimensão é melhor linearmente separável do que num espaço de baixa dimensão (Cover, 1965). Em outras palavras, havendo padrões linearmente separáveis, o problema de classificação será relativamente fácil de resolver.

Dessa maneira, (Haykin, 1998) desenvolveu uma enorme investigação na operação de uma rede RBF como um classificador através do estudo da separabilidade de padrões.

Considere uma família de superfícies, em que cada uma delas divide o espaço de entradas em duas regiões. Denotando X como um conjunto de N padrões (pontos) $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_N$, onde cada um deles é separado dentre as duas classes X^+ e X^-. Esta *dicotomia* (partição binária) dos pontos diz para separá-los com respeito à família de superfícies, desde que exista uma família de superfície que separe os pontos da classe X^+ dos da classe X^-. Para cada padrão $\mathbf{x}$ ε X, define-se um vetor composto por um conjunto de funções de valores reais $[\varphi_i(\mathbf{x})|i = 1, 2, \ldots, M]$, como mostrado por:

$$\varphi(\mathrm{x}) = [\varphi_1(\mathrm{x}), \varphi_2(\mathrm{x}), \ldots, \varphi_M(\mathrm{x})]^T \qquad \textbf{(6.34)}$$

seja que o padrão **x** é um vetor *p*-dimensional do espaço de entrada. O vetor φ (**x**) então mapeia os pontos do espaço de entrada *p*-dimensional aos pontos correspondentes do novo espaço de dimensão *M*. Refere-se então à φ_i(**x**) como uma *função escondida*, já que ela atua de forma similar a uma unidade escondida de uma rede neural feedforward.

A dicotomia $[X^+, X^-]$ de *X* é dita *φ-separável* se existir um vetor *m*-dimensional ***w*** que permita escrever:

$$\mathrm{w}^T \varphi(\mathrm{x}) \geq 0, \qquad \mathrm{x} \in X^+ \tag{6.35}$$

$$\mathrm{w}^T \varphi(\mathrm{x}) < 0, \qquad \mathrm{x} \in X^- \tag{6.36}$$

sendo que o hiperplano definido pela equação:

$$\mathrm{w}^T \varphi(\mathrm{x}) = 0 \tag{6.37}$$

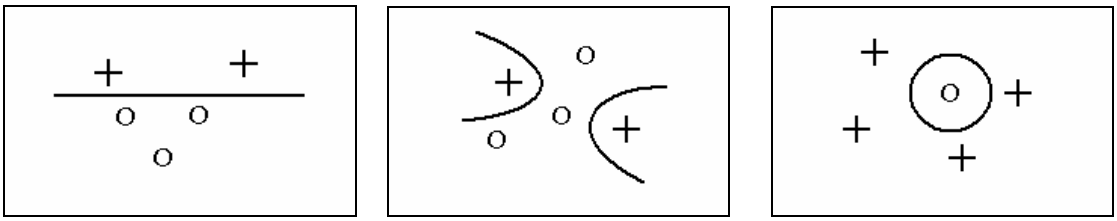
define a superfície de separação do espaço de entrada.

Considere uma classe natural de mapeamentos obtidos pelo uso de uma combinação linear dos elementos do vetor de padrões. As superfícies de separação correspondentes a cada mapeamento são referenciadas como as *variedades racionais de r-ésima ordem*. Uma variedade racional de ordem *r*, do espaço de dimensão *p*, é descrita por uma equação homogênea de grau *r* dos elementos do vetor de entrada **x**, como mostrado por:

$$\sum_{0 \leq i_1 \leq i_2 \leq \cdots \leq i_r \leq p} a_{i_1 i_2 \cdots i_r} x_{i_1} x_{i_2} \cdots x_{i_r} = 0 \tag{6.38}$$

onde x_i é o i-ésimo componente do vetor de entrada **x**, e x_0 é ajustado à unidade afim de expressar a equação na forma homogênea. Exemplos de superfícies de separação deste tipo são os *hiperplanos* (variedades racionais de primeira ordem), *parabolóides* (variedades de segunda ordem), e *hiperesferas* (variedades de segunda ordem com certas restrições lineares nos coeficientes). Estes exemplos são mostrados na Figura 6.3 para uma configuração de cinco pontos em um espaço de entrada bi-dimensional. Vale ressaltar que, em geral, os casos

de separabilidade linear implicam tanto em separabilidade esférica, como em separabilidade quadrática; entretanto, o inverso não é necessariamente verdadeiro.



(a) **(b)** **(c)**

Figura 6.3 - Três exemplos da dicotomia de separabilidade-φ para diferentes configurações de cinco pontos em duas dimensões: (a) separabilidade linear; (b) separabilidade quadrática; (c) separabilidade esférica.

Separabilidade polinomial, descrita anteriormente, pode ser vista como uma generalização natural da separabilidade linear. O ponto importante a ser notado aqui é que, dado um conjunto de padrões **x** do espaço de entrada de dimensão arbitrária p, pode-se encontrar um mapeamento não-linear $\varphi(\mathbf{x})$ de dimensão M, o qual possui separabilidade linear no espaço φ.

6.2.2 Problema de Interpolação

O ponto importante que surge do teorema de Cover sobre a separabilidade de padrões é que, resolvendo um problema de classificação de padrões não-linearmente separável, em geral, há benefício prático a ser ganho no mapeamento do espaço de entrada em um novo espaço de dimensão suficientemente grande. Basicamente, um mapeamento não-linear é usado para transformar um problema de classificação não-linearmente separável em um linearmente separável. Da mesma forma, pode-se usar um mapeamento não-linear para transformar um complexo problema de filtragem não-linear em um problema simples que envolve filtragem linear.

Considere uma rede neural de três camadas, com um neurônio na camada de saída. Ela foi modelada para fazer um mapeamento não-linear entre um espaço de entrada e o espaço da camada oculta (intermediária), seguida de um mapeamento linear a partir deste espaço da camada oculta para o espaço da camada de saída. Denotando p como a dimensão do espaço de entrada, esta rede passa a representar um mapeamento entre o espaço de entrada p-dimensional e o espaço de saída unidimensional, escrito como:

$$s : \Re^{p} \rightarrow \Re^{1} \qquad \textbf{(6.39)}$$

Este mapeamento s pode ser pensado como um hiperplano $\Gamma \subset \Re^{p+1}$, e visto como um gráfico multidimensional da saída como função da entrada. Na prática, o hiperplano Γ é desconhecido e os dados de treinamento são normalmente contaminados por ruídos. Assim, as fases de treinamento e de generalização do processo de aprendizagem podem ser vistas respectivamente como:

- A fase de treinamento consiste na otimização do procedimento de ajuste da superfície Γ, com base nos dados conhecidos e apresentados à rede na forma de dados entrada-saída (padrões).
- A fase de generalização é o mesmo que interpolar entre pontos dos dados treinados, ou seja, o mesmo que interpolar a partir do hiperplano gerado no processo de ajuste de maneira a aproximar ao máximo o tal hiperplano Γ desconhecido.

O problema de interpolação pode ser definido matematicamente como:

Dado um conjunto de N diferentes pontos $[\mathbf{x}_i \in \Re^{p} \mid i = 1, 2, ..., N]$ e um correspondente conjunto de N números reais $[d_i \in \Re^{1} \mid i = 1, 2, ..., N]$, encontrar uma função F: $\Re^{n} \rightarrow \Re^{1}$ que satisfaça a condição de interpolação:

$$F(\mathrm{x}_i) = d_i, \qquad i = 1, 2, \ldots, N \qquad \textbf{(6.40)}$$

Para que se obtenha uma boa interpolação, a função F (ou o hiperplano Γ) deve ser construída de forma a passar por todos os pontos de dados do treinamento.

A técnica de Funções de Base Radial (RBF) consiste na escolha de uma função F que tenha a seguinte forma (Powell, 1988):

$$F(\mathrm{x}) = \sum_{i=1}^{N} w_i \varphi\left(\left\|\mathrm{x} - \mathrm{x}_i\right\|\right) \tag{6.41}$$

onde $[\varphi(\|\mathbf{x} - \mathbf{x}_i\|) \mid i = 1, 2, ..., N]$ é um conjunto de funções quaisquer (geralmente não-lineares), conhecidas como *funções de base radial*, e $\|\cdot\|$ denota a *norma Euclidiana*, normalmente a mais utilizada. Os pontos de dados conhecidos $\mathbf{x}_i \in \Re^p \mid i = 1, 2, ..., N$ são tomados como os *centros* das funções de base radial.

Inserindo as condições de interpolação da Equação 6.40 na Equação 6.41, obtém-se o seguinte conjunto de equações lineares para os coeficientes (pesos) desconhecidos da expansão $[w_i]$:

$$\begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1N} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_{N1} & \varphi_{N2} & \cdots & \varphi_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \tag{6.42}$$

onde:

$$\varphi_{ji} = \varphi\left(\left\|\mathrm{x}_j - \mathrm{x}_i\right\|\right), \qquad j, i = 1, 2, \ldots, N \tag{6.43}$$

sendo que:

$$\mathrm{d} = [d_1, d_2, \ldots, d_N]^T \tag{6.44}$$

$$\mathrm{w} = [w_1, w_2, \ldots, w_N]^T \tag{6.45}$$

onde os vetores **d** e **w**, de dimensão N por 1, representam o *vetor de resposta desejada* e o *vetor de pesos lineares*, respectivamente.

Sendo ainda que $\mathbf{\Phi}$ denota a matriz N por N com elementos φ_{ji}:

$$\Phi = \{\varphi_{ji} \mid j, i = 1, 2, \ldots, N\} \tag{6.46}$$

chamada de *matriz de interpolação*.

Assim, pode-se reescrever a Equação 6.42 de forma compacta:

$$\Phi \mathrm{w} = \mathrm{d} \tag{6.47}$$

Existe uma classe de funções de base radial que seguem estas propriedades, segundo o teorema (Light, 1992):

"Sejam $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_N$ pontos distintos em $\Re^p$. Então a matriz de interpolação $\mathbf{\Phi}$, N por N, cujo j-ésimo elemento $\varphi_{ji} = \varphi(\|\mathbf{x}_j - \mathbf{x}_i\|)$, é definido positivo."

Este teorema se aplica às seguintes funções:

1. Multiquadrática inversa

$$\varphi(r) = \frac{1}{(r^2 + c^2)^{1/2}} \qquad \text{para algum } c > 0, \text{ e } r \geq 0 \tag{6.48}$$

2. Funções Gaussianas

$$\varphi(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right) \qquad \text{para } \sigma > 0, \text{ e } r \geq 0 \tag{6.49}$$

Investigações teóricas e resultados práticos, entretanto, mostram que o tipo de não-linearidade de $\varphi(\cdot)$ não é crucial para a performance da rede RBF (Powell, 1988).

Retornando ao teorema de Light, nota-se que a matriz de interpolação $\mathbf{\Phi}$ é positiva definida, e então pode-se resolver a Equação 6.47 para a obtenção dos pesos $\mathbf{w}$, da seguinte maneira:

$$\mathrm{w} = \Phi^{-1}\mathrm{d} \tag{6.50}$$

onde $\mathbf{\Phi}^{-1}$ é a inversa da matriz de interpolação $\mathbf{\Phi}$.

6.2.3 Modelagem da Rede RBF

Esta secção descreve um modelo de rede neural artificial inspirado na resposta localmente sintonizada (locally tuned response) observada em neurônios biológicos. Neurônios com esta característica podem ser encontrados em muitos sistemas nervosos. Estas células nervosas têm uma resposta característica, que é seletiva para um espectro finito do espaço do sinal de entrada.

A rede RBF possui uma estrutura *feedforward* constituída de uma única camada escondida de J unidades de sintonia local, as quais são totalmente conectadas à camada de saída com L unidades lineares, como na Figura 6.4. Todas as unidades escondidas recebem simultaneamente o vetor de entrada **x** n-dimensional.

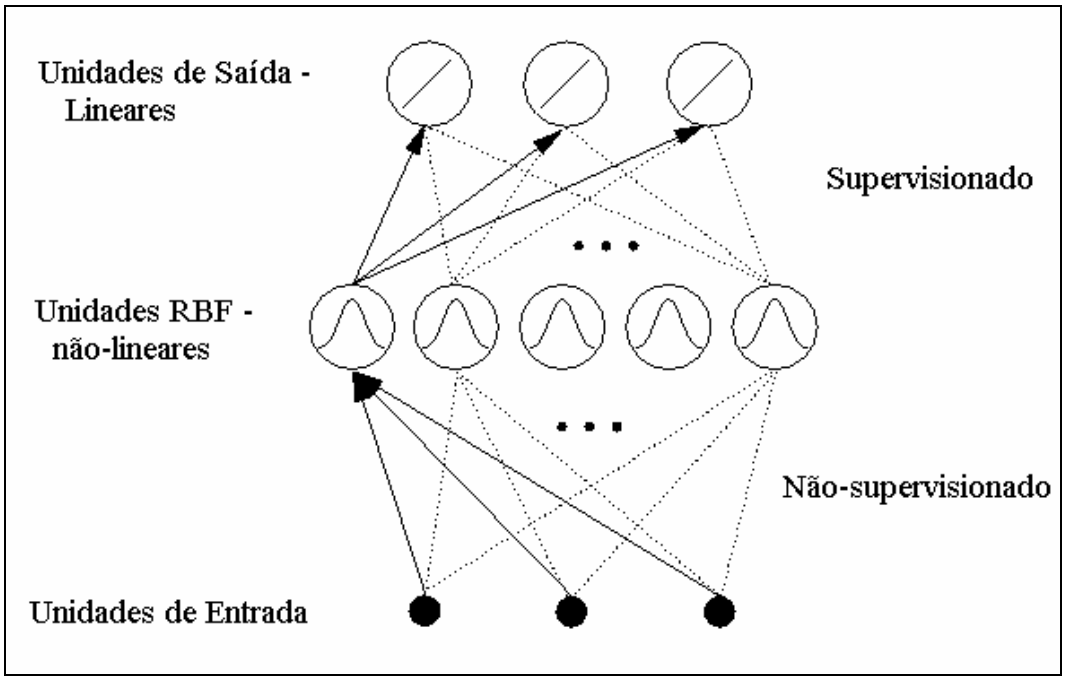


Figura 6.4 – Arquitetura de uma Rede Neural com Função de Base Radial (RBF)

Assim, as saídas φ_j de cada unidade escondida são obtidas em função da proximidade entre a entrada **x** e o parâmetro μ_j (definida como o centro de cada função radial). Esta proximidade é avaliada por uma função radial, como a Equação 6.49, função Gaussiana:

$$\varphi_j(\mathrm{x}\,;\mu_j\,;\sigma_j) = \exp\left(-\frac{\left\|\mathrm{x}-\mu_j\right\|^2}{2\sigma_j^2}\right) \tag{6.51}$$

onde o parâmetro σ_j é a largura do campo receptivo do espaço de entrada da unidade j. Isto implica que φ_j apresenta um valor considerável somente quando a distância $\|\mathbf{x} - \boldsymbol{\mu}_j\|$ for menor que a largura σ_j. Dado um vetor de entrada $\mathbf{x}$, a saída da rede RBF passa então a ser definida por:

$$\mathrm{y}(\mathrm{x}) = \sum_{j=1}^{J} w_{lj}\varphi_j(\mathrm{x}) \tag{6.52}$$

Redes RBF são muito boas para aproximar mapeamentos contínuos, ou contínuos por parte, do tipo $f: \Re^n \rightarrow \Re^L$, quando n é suficientemente pequeno. De acordo com as Equações 6.51 e 6.52, a rede RBF pode ser vista como uma aproximação de uma função desejada $f(\mathbf{x})$ pela superposição de funções de base radial não-ortogonais, podendo ser classificada como um aproximador universal. O grau de precisão pode ser controlado por três parâmetros: o número de funções de base, sua localização, e sua largura.

Durante a fase de treinamento procura-se reduzir o erro na saída da rede, que neste trabalho foi considerado o MSE, a valores aceitáveis através da adaptação dos parâmetros livres na rede RBF: os centros dos campos receptivos $\boldsymbol{\mu}_j$, as larguras dos campos receptivos σ_j e os pesos entre a camada escondida e a camada de saída w_{jl}. Este treinamento pode utilizar a aprendizagem supervisionada, não-supervisionada ou híbrida, sendo esta última a mais empregada e, inclusive, utilizada neste trabalho.

O treinamento híbrido combina aprendizagem não-supervisionada com supervisionada. A primeira treina a camada escondida, definindo os parâmetros desta camada, como localização dos centros e larguras dos campos receptivos. A segunda etapa define os valores dos pesos entre as camadas escondidas e de saída. Nesta segunda fase, os parâmetros definidos na primeira fase não se modificam.

Na etapa de treinamento não-supervisionado para a camada escondida são então determinados os centros das funções de base radial e as larguras dos campos receptivos. Dentre os métodos para determinação dos centros das funções de base radial, temos:

- Seleção Aleatória

  Os centros são vetores de entrada aleatoriamente selecionados. Esta técnica demanda que os padrões de treinamento representem precisamente todo o espaço de soluções. Este método é simples e direto, no entanto pode exigir grande número de unidades intermediárias, e escolher centros muito próximos uns dos outros pode acarretar mau funcionamento da rede.

- Fixação de grade regular

  Neste caso, os centros são fixados em uma grade regular, cobrindo todo o espaço dos vetores de entrada. Em geral, este método exige muitas unidades intermediárias para vetores de entrada com muitas dimensões, o que provoca um crescimento exponencial do número das unidades escondidas, fenômeno conhecido como maldição da dimensionalidade.

- Técnicas de agrupamento

  Os centros são definidos por técnicas de agrupamento. Entre as quais se destacam os algoritmos de k-médias e de mapas auto-organizáveis (SOM). Os mapas auto-organizáveis caracterizam-se por agrupar padrões espacialmente próximos que compartilhem micro-características. O centro que apresentar maior produto escalar com um dado vetor de entrada, adiciona este vetor de entrada ao seu grupo. Esta foi a técnica empregada neste trabalho de pesquisa.

A técnica dos mapas auto-organizáveis se mostra bastante eficiente e define os centros receptivos das funções de base radial de forma automatizada, durante esta etapa de treinamento não-supervisionado.

Neste trabalho os centros foram então definidos pela rede SONARX, caracterizada com uma rede de Mapa Auto-Organizável (SOM) de Kohonen com Modelos de Sistemas Dinâmicos, em outras palavras, uma versão não-supervisionada do modelo NARX.

Ainda na etapa de treinamento não-supervisionado da rede RBF, precisa-se definir uma heurística para a determinação da largura do campo receptivo σ_j, que pode ser único para todas as unidades ou pode ser diferente para cada unidade escolhida. Algumas das principais heurísticas são:

- Utilização da distância euclidiana média entre centros:

$$\sigma = \frac{1}{ng}\sum_{j=1}^{ng}\left\|\mu_j - \mu_{j(mprox)}\right\| \qquad \textbf{(6.53)}$$

onde ng é o número de grupos formados e $\mu_{j(mprox)}$ é o centro com menor distância euclidiana com relação ao centro μ_j. Este método produz um único valor de raio.

- Utilização da distância euclidiana entre centro e vetor de entrada:

$$\sigma_j^2 = \frac{1}{nv}\sum_{\mathrm{x}\in\psi_j}\left\|\mu_j - \mathrm{x}\right\|^2 \qquad \textbf{(6.54)}$$

onde ψ_j é o conjunto dos nv vetores de entrada com menor distância euclidiana para o centro μ_j.

- Utilização da distância euclidiana entre os centros determinados pelo método k-médias:

$$\sigma_j^2 = \frac{1}{nv}\sum_{\mathrm{x}\in S_j}\left\|\mu_j - \mathrm{x}\right\|^2 \qquad \textbf{(6.55)}$$

onde S_j é o agrupamento com *nv* vetores de entrada pertencentes a ele.

- Utilização da distância euclidiana entre centros:

$$\sigma_j = \alpha \left\| \mu_j - \mu_{j(mprox)} \right\| \tag{6.56}$$

onde $\mu_{j(mprox)}$ é o centro com menor distância euclidiana com relação ao centro μ_j e o parâmetro geralmente se situa $1{,}0 \leq \alpha \leq 1{,}5$.

Devido ao espalhamento não uniforme dos centros, definidos pela rede SONARX no espaço de entrada, a heurística utilizada neste trabalho foi a que utiliza a distância euclidiana entre centros (apresentada pela Equação 6.56), já que este método permite definir larguras distintas dos campos receptivos para cada centro, conseqüentemente para cada função de base radial, caracterizando de forma suave todo o espaço de entrada.

Já na etapa do treinamento supervisionado para a camada de saída passam a ser determinados os pesos entre a camada escondida e a de saída. A primeira etapa neste processo é o cálculo do erro. Dentre as maneiras de se calcular o erro, optou-se por:

- Erro quadrático médio (MSE):

$$MSE = \frac{1}{N} \sum_{t=1}^{N} [y(t) - \hat{y}(t)]^2 \tag{6.57}$$

onde $y(t)$ e $\hat{y}(t)$ são respectivamente os resultados desejado e obtidos pela rede RBF, e N é o número total de padrões.

Calculado o erro, este deve ser minimizado por algum procedimento que altere e defina valores "ótimos" para os pesos w_{lj}; dentre os quais destacam-se: o método dos mínimos quadrados, a matriz pseudo-inversa (ambos classificados como métodos lineares), e a regra delta. Neste trabalho, optou-se pela utilização da matriz pseudo-inversa dos sinais de ativação da camada escondida, uma vez que a tarefa de determinar os pesos da camada de saída

durante a etapa de treinamento supervisionado é caracterizada como um problema de regressão linear.

Assim, depois que as funções de base radial estão todas definidas, os pesos "ótimos" são calculados e definidos por (Haykin, 1998):

$$W = \left(\Phi^T \Phi\right)^{-1} \Phi^T d \qquad \textbf{(6.58)}$$

onde a matriz Φ é formada pelos sinais de ativação $\varphi(\cdot)$ das funções da base radial para todos os padrões de entrada, e o vetor d é o vetor de resposta desejada.

6.3 Fundamentos dos Sistemas "Neuro-Fuzzy"

Num passado recente, temos testemunhado um rápido crescimento do emprego da Lógica *Fuzzy* (LF) em diversas aplicações, tais como câmeras fotográficas, filmadoras, máquinas de lavar, inclusive em controle de processos industriais, instrumentação médica, sistemas de suporte a decisão, seleção de portifólio, etc.

A Lógica *Fuzzy* (LF) está fundamentada na teoria dos Conjuntos *Fuzzy* (CF), que, diferentemente da teoria clássica de conjuntos, se baseia em classes de conjuntos com fronteiras "abertas", onde cada elemento possui um grau de pertinência relativo a vários conjuntos *fuzzy*.

Além disto, na LF a variável possui valores semânticos e não numéricos, como na teoria clássica. Assim, a LF pode ser vista como uma metodologia que processa palavras (ou idéias) ao invés de números.

O que torna a LF interessante, apesar da aparente imprecisão inerente que este tipo de informação tem, em relação às variáveis numéricas, é que o uso dos CF se mostra bastante intuitivo e, ao processar palavras, tem-se uma grande tolerância à imprecisão, o que diminui

bastante o esforço computacional. Além disto, a LF é bastante eficiente na ponderação entre *importância* e *precisão*, algo que os seres humanos já tratam a milhares de anos.

Para isso, a LF utiliza regras *fuzzy* do tipo *if-then*, o que pode ser chamado simplesmente de Regras *Fuzzy* (RF). Estas RF podem ser definidas como uma maneira dos Sistemas de Inferência *Fuzzy* (*Fuzzy Inference Systems* – FIS, do inglês) tratar os antecedentes e conseqüentes *fuzzy*, o que será melhor discutido nas seções 6.3.1 e 6.3.2.

Um aspecto importante notado recentemente é o crescimento da aplicação da LF em combinação com Redes Neurais Artificiais (RNA) e Algoritmos Genéticos (AG). Os Sistemas Inteligentes fazem parte da área da ciência que trata justamente destas ferramentas, e de que maneira elas possam atuar juntas.

Os Sistemas Inteligentes tem como objetivo principal tratar as imprecisões presentes nas diversas aplicações no mundo real, explorando características como a *imprecisão*, *incerteza* e *verdade parcial*, na tentativa de atingir tratabilidade, robustez e baixo esforço computacional.

Neste sentido, dentre as várias combinações destas técnicas inteligentes, a que obteve maior destaque foi a que uniu LF e RNA, o que pode ser chamado de Sistemas Neuro-*Fuzzy*. E o exemplo mais conhecido deste tipo de Sistema Inteligente Híbrido é o Sistema de Inferência Neuro-*Fuzzy* Adaptativo (ANFIS), proposto por (Jang, 1993). Método que será apresentado na seção 6.3.3.

6.3.1 Fundamentos da Lógica "Fuzzy"

A Lógica *Fuzzy* (LF) surgiu com o trabalho "*Fuzzy Sets*" de (Zadeh, 1965), Professor da Universidade da Califórnia em Berkeley, onde coordena um importante grupo de pesquisa em Sistemas Inteligentes, mais precisamente em "*Soft Computing*".

A LF fornece uma maneira de se definir um algoritmo de processamento de informação que lida com dados vagos e ambíguos dentro de um domínio de incertezas. Isto possibilita a representação de sistemas de inferência e controle que "simulam" o conhecimento e raciocínio humano em problemas físicos, uma vez que o processo de inferência humano lida de forma inconsciente com as incertezas e ambigüidades do mundo real.

Esta teoria fornece uma boa aproximação matemática baseada na definição de Conjuntos *Fuzzy* (CF) para representar as incertezas dos sistemas. Com esta aproximação, pode-se assumir variáveis como elementos pertencentes parcialmente a um determinado conjunto e também assumir operadores *booleanos* aplicados para processar conjuntos de informações.

O resultado disto é uma transição suave e contínua entre as possíveis operações, durante o processamento dos CF. Tal característica é bastante apropriada para modelagem e controle de sistemas dinâmicos complexos (Yager e Filev, 1994a).

Os Conjuntos Clássicos ou *Crisp* usam um processo de modelagem que faz uso de operadores aritméticos, cujas transições entre um membro e um não membro de um dado conjunto são abruptas e bem definidas.

Em se tratando dos CF, a Função de Pertinência (FP) é uma função característica de mapeamento dos elementos de um universo de discurso X de um conjunto *crip* $A \subset X$ *(Ross, 1995)*, ou seja:

$$\mu_A(x) = \{0;1\} \tag{6.59}$$

onde cada elemento do universo de discurso apresenta exclusão completa $\mu_A(x) = 0$ ou pertinência completa $\mu_A(x) = 1$ ao conjunto *crisp A*.

O mapeamento é um importante conceito no relacionamento da teoria de conjuntos às representações por funções de informação, o qual, para conjuntos clássicos, pode ser representado como na Figura 6.5.

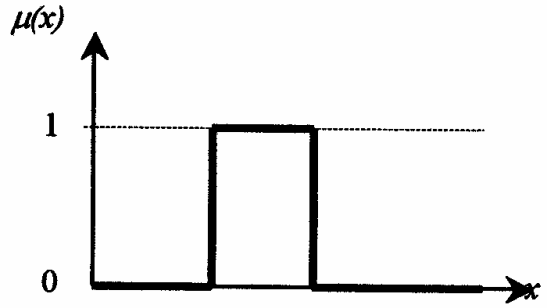


Figura 6.5 – Função de pertinência para um conjunto clássico

Já os Conjuntos *Fuzzy* utilizam conectivos do tipo lógico *booleano*, tais como *AND*, *OR* e *IF – THEN*, e apresentam uma transição gradual, no qual aos elementos do universo de discurso são atribuídos graus contínuos de pertinência entre 0 (exclusão completa) e 1 (pertinência completa) a cada conjunto definido. Estes valores de pertinência expressam os graus com os quais cada elemento é compatível com as propriedades dos conjuntos do universo e no qual pode ser representado pela função de pertinência da Equação 6.60:

$$\mu_A(x) = [0;1] \qquad \textbf{(6.60)}$$

A Figura 6.6 é uma representação típica de função de pertinência da Equação 6.60, com a forma triangular.

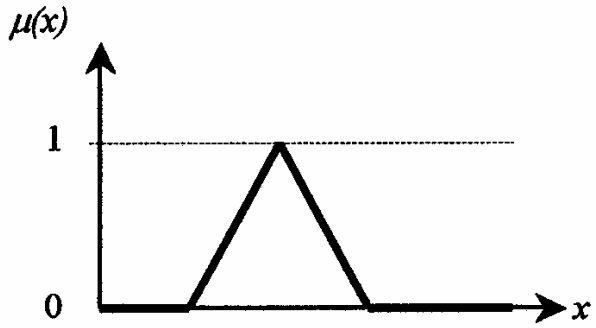


Figura 6.6 – Função de pertinência para um conjunto *fuzzy* (tipo triangular)

A Figura 6.7 ilustra a diferença básica entre os conjuntos clássicos e *fuzzy*. Nos conjuntos clássicos a transição entre um membro (elemento *b*) e um não membro (elemento *a*) é abrupta e bem definida, simplesmente pertencendo ou não ao conjunto *A*. Enquanto que nos conjuntos *fuzzy* a transição é gradual, onde por exemplo, o elemento *c* pertence parcialmente ao conjunto *A* (Yager e Filev, 1994a).

A função de pertinência provê uma relação do valor exato (*crisp*) do mundo real para uma representação num conjunto *fuzzy*. Este processo de conversão de valores exatos para valores *fuzzy* é chamado de *fuzzificação*.

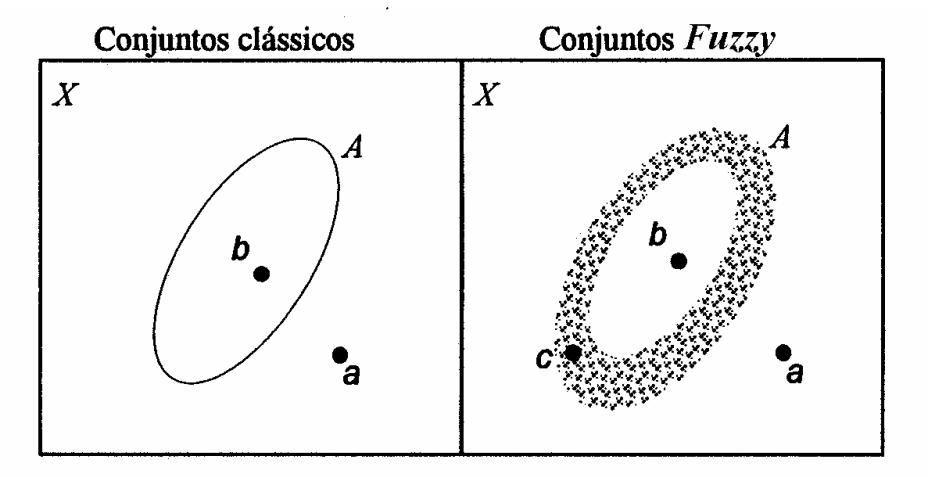


Figura 6.7 – Representação de conjuntos clássicos e *fuzzy*

Com relação aos operadores *fuzzy*, podem ser definidas: a intersecção *fuzzy* (*AND* ou min); união *fuzzy* (*OR* ou máx); e o complemento *fuzzy* (*NOT* ou complemento aditivo).

A intersecção de dois conjuntos *fuzzy A* e *B* é definida em geral por um mapeamento binário *T*, que agrega duas funções de pertinência como se segue:

$$\mu_{A \cap B}(x) = T(\mu_A(x), \mu_B(x)) \qquad \textbf{(6.61)}$$

Este operador binário T pode representar, por exemplo, a multiplicação entre $\mu_A(x)$ e $\mu_B(x)$. Estes operadores de intersecção *fuzzy*, que são comumente chamados de operadores *T*-norma (norma Triangular), devem satisfazer algumas condições.

Um operador *T*-norma é um mapeamento binário $T(\cdot,\cdot)$, se satisfizer:

- limitante: $T(0,0) = 0$, $T(a,1) = T(1,a) = a$
- monotonicidade: $T(a,b) \leq T(c,d)$ se $a \leq c$ e $b \leq d$
- comutatividade: $T(a,b) = T(b,a)$
- associatividade: $T(a,T(b,c)) = T(T(a,b),c)$

De maneira análoga, o operador união é definido por um mapeamento binário S, do tipo:

$$\mu_{A \cup B}(x) = S(\mu_A(x), \mu_B(x)) \qquad \textbf{(6.62)}$$

Já o operador binário *S*, por exemplo, pode representar a adição de $\mu_A(x)$ e $\mu_B(x)$. Os operadores de união *fuzzy*, que são chamados de operadores *S*-norma (ou *T*-conorma), também devem satisfazer algumas condições.

Um operador *S*-norma (ou *T*-conorma) é um mapeamento binário $T(\cdot,\cdot)$, se satisfizer:

- limitante: $S(1,1) = 1$, $S(a,0) = T(0,a) = a$
- monotonicidade: $S(a,b) \leq S(c,d)$ se $a \leq c$ e $b \leq d$
- comutatividade: $S(a,b) = S(b,a)$
- associatividade: $S(a,S(b,c)) = S(S(a,b),c)$

Evidentemente que a Lógica *Fuzzy* (LF) não é a solução para todos os problemas. Uma de suas características é a de promover um bom mapeamento entre o espaço de entradas e o espaço de saídas de um dado sistema real. Caso este mapeamento já seja conhecido, ou caso uma solução simples para ele já exista, fica claro que o uso da LF seria totalmente desnecessário (Tsoukalas e Uhrig, 1997).

Mesmo que vários controladores e sistemas operem de forma satisfatória sem usarem a Lógica *Fuzzy*, ela pode ser uma poderosa ferramenta para o tratamento rápido e eficiente das *imprecisões* e *não linearidades* dos problemas reais (Wang, 1994).

6.3.2 Sistemas de Inferência "Fuzzy"

Conforme dito anteriormente, Lógica *Fuzzy* (LF) utiliza regras *fuzzy* do tipo *if-then*, o que pode ser chamado simplesmente de Regras *Fuzzy* (RF). Estas RF podem ser definidas como uma maneira dos Sistemas de Inferência *Fuzzy* (FIS) tratar os antecedentes e conseqüentes *fuzzy*, tendo como base a teoria dos Conjuntos *Fuzzy* proposta por (Zadeh, 1965) e os Operadores F*uzzy* (*T*-norma e *S*-norma).

As Regras *Fuzzy* do tipo *If-Then* são expressões da seguinte forma:

IF <antecedente> THEN <conseqüente> **(6.63)**

onde o <*antecedente*> assume valores lingüísticos definidos por funções de pertinência, enquanto que o <*conseqüente*> pode tanto assumir valores lingüísticos quanto valores *não fuzzy*, ou seja, valores numéricos.

Este tipo de RF com *conseqüentes* do tipo *fuzzy* ou *não fuzzy* vêm sendo amplamente utilizados tanto em modelagem quanto em controle de sistemas reais. Através do uso de rótulos lingüísticos e funções de pertinência, as RF do tipo *If-Then* são capazes de capturar facilmente o espírito da "regra geral" usada inconscientemente por seres humanos.

Por outro lado, devido aos qualificadores dos *antecedentes*, cada RF do tipo *If-Then* pode ser vista como uma descrição local do sistema real. Logo, as Regras *Fuzzy* do tipo *If-Then* formam a parte central dos Sistemas de Inferência *Fuzzy* (FIS).

Os Sistemas de Inferência *Fuzzy* (FIS) são também chamados de Modelos Fuzzy, Memórias Associativas Fuzzy, ou Controladores Fuzzy quando usados como controladores. Basicamente um FIS é composto por cinco blocos funcionais, como podem ser visto na Figura 6.8.

Sendo que:

- a Base de Regras contém todas as Regras *Fuzzy* do tipo *If-Then*;
- a Base de Dados define as Funções de Pertinência dos Conjuntos *Fuzzy* usados nas Regra *Fuzzy*;
- a Base de Conhecimento se refere ao conjunto formado pela Base de Regras e Base de Dados;
- a Unidade de Tomada de Decisão promove as operações de inferência, com base nas Regras *Fuzzy*;
- a Interface de *Fuzzyficação* transforma as entradas *crisp* em graus de pertinências com relação aos Conjuntos *Fuzzy*, ou seja, transforma em valores *fuzzy*;
- a Interface de *Defuzzificação* transforma os resultados da inferência *fuzzy* em valores de saída *crisp*.

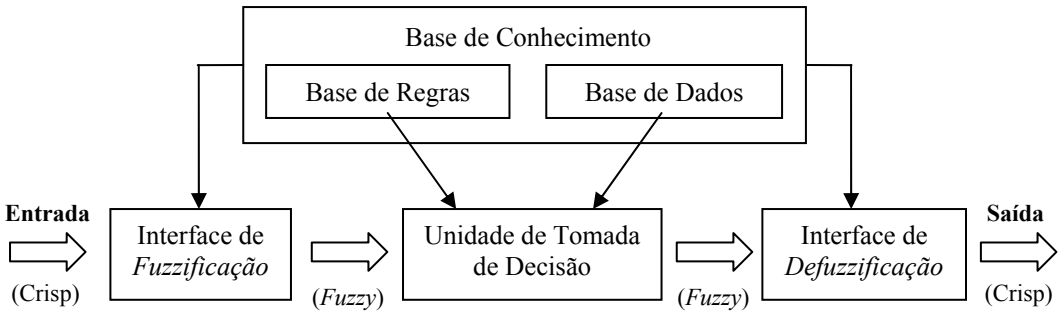


Figura 6.8 – Sistema de Inferência Fuzzy - FIS

O processamento da informação pelo Sistema de Inferência *Fuzzy* (FIS) pode ser dividido em 4 (quatro) passos:

1- Aplica as variáveis de entrada nas funções de pertinência dos antecedentes para calcular os respectivos graus de pertinência à cada um dos conjuntos *fuzzy*, obtendo-se valores *fuzzy* que servirão de entrada para a Base de Regras (*Fuzzificação*);

2- Calcula, através de um operador *T*-norma (normalmente *min* ou *multiplicação*), o valor de pertinência dos antecedentes que irá definir o *nível de disparo* de cada Regra *Fuzzy*;

3- Calcula o valor de cada *conseqüente* ativo, *crisp* ou *fuzzy*, de cada Regra *Fuzzy*, em função de cada *nível de disparo*;

4- Agrega os *conseqüentes* ativos para produzir uma saída *crisp* (*Defuzzyficação*).

Vários tipos de abordagens de FIS vêm sendo propostas na literatura (Jang, 1993). Dependendo do tipo de proposta *fuzzy* e do tipo de Regra *Fuzzy* utilizada, os FIS podem ser classificados em 3 (três) grupos:

Tipo 1: A saída total é a média ponderada das saídas *crisp*, z_i , de cada uma das Regras *Fuzzy*, definidas pelas funções de pertinência de saída e pelos *níveis de disparo* de cada RF (níveis definidos por um operador *T*-norma, normalmente *min* ou *multiplicação*), pelos próprios *níveis de disparo*, w_i , de cada RF. As funções de pertinência de saída usadas neste tipo de FIS devem ser funções monotônicas (Tsukamoto, 1979) apud (Jang, 1993).

Tipo 2: A saída total *fuzzy* é definida pela aplicação do operador *máx* nas saídas *fuzzy* das Regras *Fuzzy*, ativas pelos *níveis de disparo* de cada RF (níveis definidos por um operador *T*-norma, normalmente *min* ou *multiplicação*). Vários esquemas vêm sendo propostos para definir a saída final *crisp* em função da saída total *fuzzy*: *centro de área*; *bisseção de área*; *média dos máximos*; etc. (Mandani e Assilian, 1975) apud (Jang, Sun e Mizutani, 1997), (Lee, 1990a) (Lee, 1990b) apud (Jang, 1993).

Tipo 3: A saída de cada Regra *Fuzzy* é uma combinação linear da variáveis de entrada mais um termo constante c_i , e saída final é média ponderada de cada saída das RF. Sendo utilizadas portanto, as RF do tipo *If-Then* propostas por (Takagi e Sugeno, 1983) (Takagi e Sugeno, 1985) apud (Jang, 1993).

A Figura 6.9 mostra estes 3 (três) principais tipos de FIS, exemplificados num sistema de 2 (duas) variáveis de entrada, 2 (duas) RF e com apenas 1 (uma) variável de saída. O que

permite observar que as diferenças entre eles se concentram nas funções de pertinência do conseqüente e no cálculo da variável de saída.

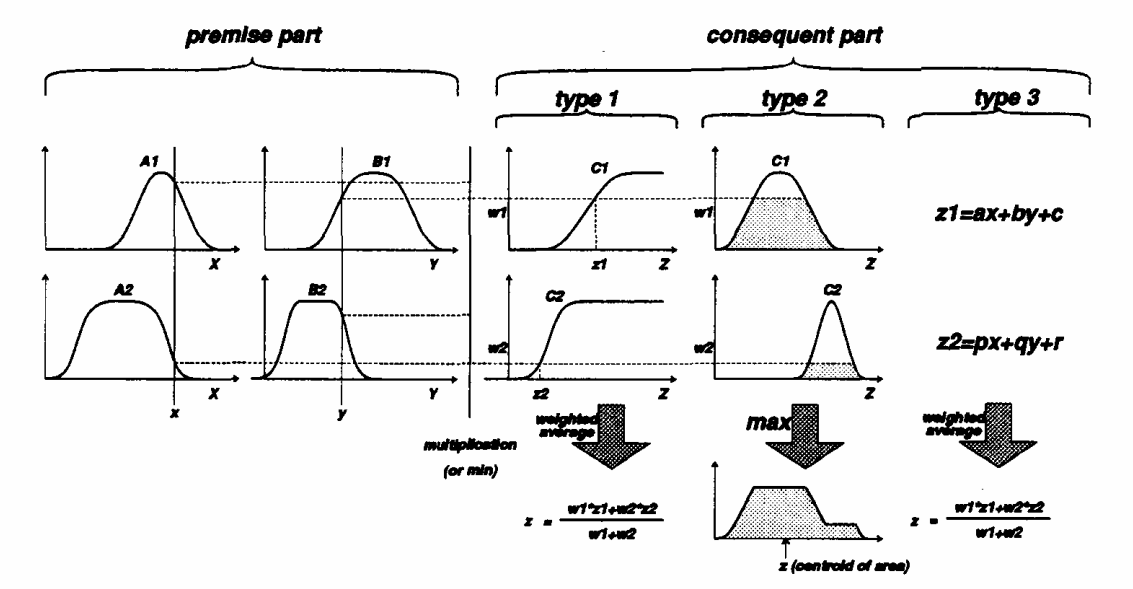


Figura 6.9 – Os 3 principais tipos de Sistemas de Inferência Fuzzy (FIS)

6.3.3 Sistema de Inferência "Neuro-Fuzzy" Adaptativo – ANFIS

Durante a década de 1990, várias redes neuro-*fuzzy* apareceram na literatura. As mais importantes foram: Sistema de Inferência Neuro-*Fuzzy* Adaptativo (ANFIS) desenvolvido por (Jang, 1993); Rede de Controle com Aprendizagem *Fuzzy* Adaptativa (FALCON) desenvolvida por (Lin e Lee, 1991); Controle Neuro-*Fuzzy* (NEFCON) proposto por (Nauck, Klawonn e Kruse, 1997); Controle Inteligente com Abordagem de Aproximação Generalizada (GARIC) desenvolvido por (Berenji e Khedkar, 1992). Estes desenvolvimentos deram origem a várias outras propostas.

A idéia deste tipo de abordagem é definir os formatos das funções de pertinência dos Sistemas de Inferência *Fuzzy* (FIS) de forma eficiente, através da propriedade adaptativa das Redes Neurais. Os primeiros a utilizarem esta idéia foram (Takagi e Sugeno, 1985) (Sugeno e

Kang, 1988). Posteriormente, (Jang, 1993) desenvolveu uma abordagem sistemática para a adaptação e ajuste dos parâmetros dos FIS, a qual se mostrou bastante eficiente na solução dos problemas por ele proposto.

Nesta seção é apresenta a arquitetura neuro-*fuzzy* utilizada neste trabalho de pesquisa: Sistema de Inferência Neuro-*Fuzzy* Adaptativo (ANFIS) desenvolvido por (Jang, 1993). Este modelo permite que os Sistemas de Inferência *Fuzzy* (FIS) ajuste seus parâmetros de forma adaptativa, ou seja, automaticamente durante o treinamento.

Através do uso de um algoritmo de aprendizagem híbrido, o ANFIS pode construir um mapeamento entrada-saída com base tanto no conhecimento humano quanto em dados numéricos.

Arquitetura ANFIS

Por simplicidade, mas sem perda de generalização, assume-se um FIS com 2 (duas) entradas, x e y , 1 (uma) saída, z . Para o modelo FIS de *Takagi-Sugeno*, a regra *fuzzy* normalmente utilizada é do tipo *If-Then*, como se segue:

Regra 1: *If x is A_1 and y is B_1. then* $f_1 = p_1 \cdot x + q_1 \cdot y + r_1$

Regra 2: *If x is A_2 and y is B_2. then* $f_2 = p_2 \cdot x + q_2 \cdot y + r_2$

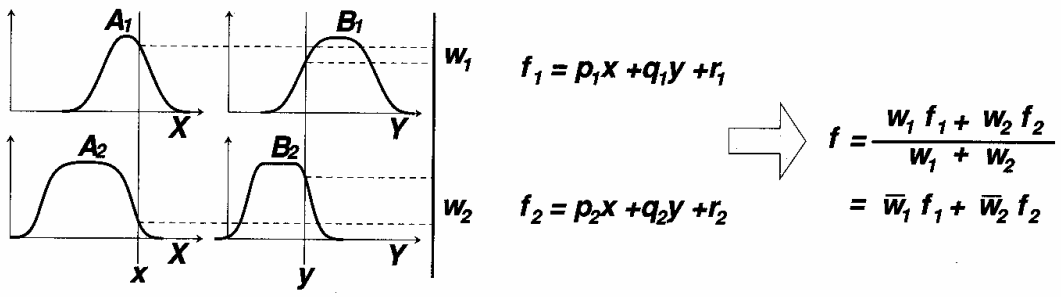


Figura 6.10 – modelo FIS de Takagi-Sugeno de 2 entradas e 1 saída

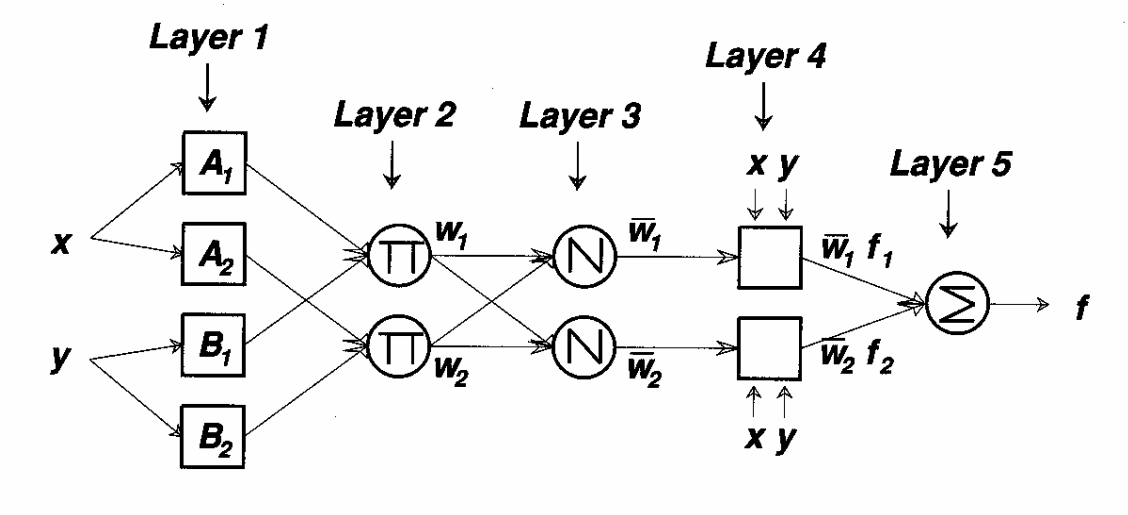


Figura 6.11 - arquitetura ANFIS de 2 entradas e 1 saída

A Figura 6.10 ilustra o mecanismo do modelo FIS de Takagi-Sugeno; a arquitetura ANFIS equivalente a este modelo é apresentada na Figura 6.11. E na seqüência, é descrito o funcionamento de cada uma das 5 (cinco) camadas da arquitetura ANFIS.

Camada 1: Todo nó *i* desta camada é um nó adaptativo cuja função é:

$$O_{1,i} = \mu_{Ai}(x) \qquad \text{para } x \in \boldsymbol{X}$$

$$O_{1,i} = \mu_{Bi}(y) \qquad \text{para } y \in \boldsymbol{Y} \qquad \textbf{(6.64)}$$

onde x e y são as entradas do i-ésimo nó, e A_i e B_i são rótulos lingüísticos associados à este nó. Em outras palavras, $O_{1,i}$ é o grau de pertinência do conjunto *fuzzy* $A = (A_1 , A_2)$ e $B = (B_1 , B_2)$, o que especifica o grau com que a entrada x e y satisfazem o quantificador A e B. Neste trabalho de pesquisa foi utilizada a função de pertinência do tipo Gaussiana:

$$\mu_A(x) = \exp\left(\frac{-(x-c)^2}{2 \cdot \sigma^2}\right) \qquad \textbf{(6.65)}$$

onde $[\sigma,c]$ são os parâmetros à serem ajustados. À medida que estes parâmetros se ajustam, o formato da função Gaussiana se modifica, definindo vários formatos

diferentes para o Conjunto *Fuzzy A* e *B*. Os parâmetros desta camada são chamados de *parâmetros dos antecedentes*.

Camada 2: Cada nó desta camada é rotulado como nó Π, cuja saída é o produto de todos os sinais de entrada:

$$O_{2,i} = w_i = \mu_{Ai}(x) \cdot \mu_{Bi}(y) \tag{6.66}$$

A saída de cada nó representa o *nível de disparo* de cada regra. Em geral, qualquer outro operador *T*-norma (por exemplo: *AND*) pode ser usado nesta camada.

Camada 3: Cada nó desta camada é rotulado como nó N. O *i*-ésimo nó define o termo, $\overline{w}_i$, usado na ponderação do conseqüente, f_i, de cada regra *fuzzy*, como se segue:

$$O_{3,1} = \overline{w}_i = \frac{w_i}{\sum_i w_i} \tag{6.67}$$

As saídas desta camada, pois, são chamados de *níveis de disparo normalizados*.

Camada 4: O *i*-ésimo nó desta camada é um nó adaptativo com a seguinte função:

$$O_{4,i} = \overline{w}_i \cdot f_i = \overline{w}_i \cdot (p_i \cdot x + q_i \cdot y + r_i) \tag{6.68}$$

onde $\overline{w}_i$ é o nível de disparo normalizado vindo da camada 3 e $[p_i, q_i, r_i]$ é o conjunto de parâmetros do *i*-ésimo nó. Estes parâmetros são definidos como os *parâmetros dos conseqüentes*.

Camada 5: O nó desta camada é rotulado como nó Σ, e calcula a saída total como o somatório de todos os sinais de cada regras *fuzzy*, ponderados por $\overline{w}_i$:

$$O_{5,1} = f = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i \cdot f_i}{\sum_i w_i} \tag{6.69}$$

Algoritmo de Aprendizagem Híbrida

Para a arquitetura ANFIS apresentada na Figura 6.11, observa-se que quando os valores dos parâmetros dos antecedentes [σ,c] são fixados, a saída total, f, pode ser expressa como uma combinação linear dos parâmetros dos conseqüentes, como segue:

$$\begin{aligned} f &= \sum_i \overline{w}_i \cdot f_i \\ &= \overline{w}_1 \cdot (p_1 \cdot x + q_1 \cdot y + r_1) + \overline{w}_2 \cdot (p_2 \cdot x + q_2 \cdot y + r_2) \\ &= (\overline{w}_1 \cdot x) \cdot p_1 + (\overline{w}_1 \cdot y) \cdot q_1 + (\overline{w}_1) \cdot r_1 + (\overline{w}_2 \cdot x) \cdot p_2 + (\overline{w}_2 \cdot y) \cdot q_2 + (\overline{w}_2) \cdot r_2 \end{aligned} \tag{6.70}$$

Com isto, o algoritmo de aprendizagem híbrida desta arquitetura ANFIS pode ser dividido em 2 (dois) passos: <u>*passo forward*</u> e <u>*passo backward*</u>. No passo *forward* (para frente), as saídas de cada nó fluem para frente até a camada 4 e os parâmetros dos conseqüentes são ajustados pelo método dos mínimos quadrados. Enquanto que no passo *backward* (para trás), os sinais de erro são propagados para trás e os parâmetros dos antecedentes são ajustados pelo método do gradiente descendente. A Tabela 6.1 ilustra bem as atividades de cada um destes passos:

Tabela 6.1 - Algoritmo de aprendizagem híbrida, em 2 passos

	Passo forward	Passo backward
Parâmetros dos antecedentes	Fixo	Gradiente Descendente
Parâmetros dos conseqüentes	Estimador de Mínimos Quadrados	Fixo
Sinais	Saídas dos nós	Sinais de erro

De acordo com (Jang, Sun e Mizutani, 1997), os parâmetros dos conseqüentes serão ajustados de forma ótima, desde que os parâmetros dos antecedentes sejam fixados. Além disto, a abordagem híbrida converge muito mais rápido, por reduzir a dimensão do espaço de

busca para os parâmetros dos antecedentes, em comparação ao algoritmo *backpropagation* original.

Particionamento do Espaço de Entrada

O espírito dos Sistemas de Inferência *Fuzzy* (FIS) nos faz lembrar o ditado: "dividir para conquistar" O antecedente da regra *fuzzy* define uma região local *fuzzy* através de suas Funções de Pertinência associadas. Enquanto que o conseqüente descreve o comportamento dentro desta região através de vários componentes, podendo ser de vários tipos, como ilustrado pela Figura 6.12.

Diferentes componentes dos conseqüentes resultam em diferentes FIS, mas seus antecedentes são sempre os mesmos. Logo, os métodos para o particionamento do espaço de entrada se mostram muito importantes e úteis para todos os modelos de FIS, inclusive o modelo ANFIS, objeto de estudo deste trabalho de pesquisa.

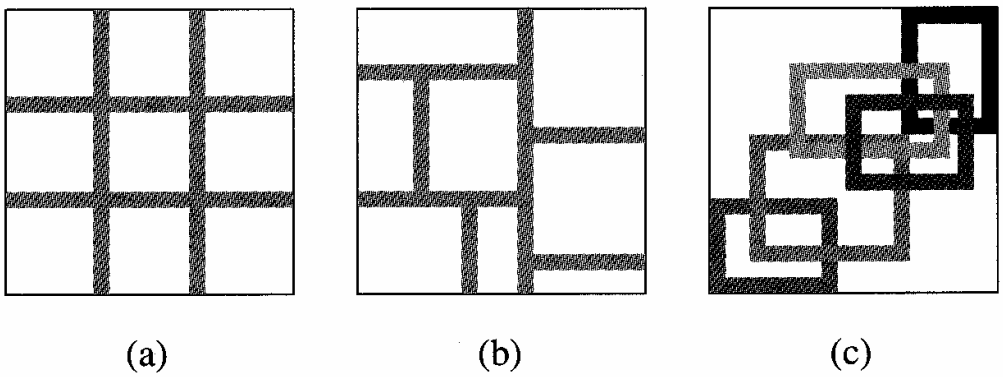


Figura 6.12 - Métodos de particionamento do espaço de entrada: (a) particionamento em *grid*, (b) particionamento em *tree*, (c) particionamento em *scatter*

Particionamento em *Grid*: A Figura 6.12(a) ilustra um típico particionamento em *grid* num espaço de entrada bi-dimensional. Este método de particionamento geralmente é usado no desenvolvimento de um controlador fuzzy, que trate somente

de variáveis de estado como variáveis de entrada do controlador. Este método de particionamento exige um número pequeno de Funções de Pertinência para cada entrada. Entretanto, este método encontra dificuldade para problemas com um número razoavelmente grade de variáveis de entrada. Por exemplo, um modelo *fuzzy* com 10 (dez) entradas e 2 (duas) Funções de Pertinência para cada uma delas irá resultar em 2^{10} = 1.024 regras *fuzzy* do tipo *If-Then*, o que seria proibitivo computacionalmente. Este problema é chamado de maldição da dimensionalidade, e pode ser tratado e contornado por outras estratégias de particionamento.

Particionamento em *Tree*: A Figura 6.12(b) mostra um típico particionamento em *tree*, no qual cada região é definida de forma única por uma árvore de decisão. O particionamento em *tree* contorna o problema do crescimento exponencial no número de regras. Entretanto, são necessárias mais Funções de Pertinência para cada entrada para definir cada região *fuzzy*, além de não ser possível a identificação das funções por rótulos lingüísticos.

Particionamento em *Scatter*: Como ilustrado na Figura 6.12(c), o particionamento em *scatter* é capaz de limitar o número de regras *fuzzy* à um número razoável, uma vez que este particionamento cobre parte de todo o espaço de entrada. Entretanto, o particionamento em *scatter* é normalmente determinado por pares de entrada-saída desejados e, conseqüentemente, normalmente não apresentam ortogonalidade. Isto faz com que seja bastante difícil se obter uma boa estimativa do mapeamento total com base nos conseqüentes de cada regra.

***Clusterização* Subtrativa no Particionamento do Espaço de Entrada**

A *Clusterização* divide um conjunto de dados em vários grupos de tal maneira que a similaridade dentro de um grupo é maior do que entre os grupos. Para isto é necessária a utilização de uma métrica de similaridade que retorne um valor que reflita tal similaridade em função dos dados de entrada. Como a maioria das métricas de similaridade é sensível aos universos de discurso de cada variável de entrada, é preciso que o conjunto de variáveis de entrada seja todo normalizado, no intervalo [0,1].

As técnicas de *Clusterização* são usadas nos Sistemas de Inferência *Fuzzy* (FIS), para determinarem a localização da Regras *Fuzzy* do tipo *If-Then*. Dentre as técnicas mais utilizadas para a *clusterização* off-line aplicadas em FIS, podemos destacar: C-médias (ou K-médias), *Fuzzy* C-médias, Método da Montanha, e *Clusterização* Subtrativa.

As técnicas por C-médias começam os seus processos com uma seleção aleatória dos centros iniciais, e não podem garantir a convergência para um ajuste ótimo destes centros, durante o processo iterativo. A performance do algoritmo de C-médias se mostra muito sensível e dependente da posição inicial dos centros dos clusters, o que é uma grande limitação do método.

O Método da Montanha, proposto por (Yager e Filev, 1994b), é uma técnica relativamente simples e eficiente para a definição aproximada dos centros de clusters com base em uma medida de densidade, chamada de *função da montanha*. Este método é baseado no que o ser humano faz visualmente, alocando centros onde há uma maior concentração de dados.

Entretanto, o custo computacional do Método da Montanha cresce exponencialmente com a dimensão do problema, porque o método precisa avaliar a *função montanha* em todos os pontos do *grid*, que forma o espaço de entrada. Por exemplo, um problema de

clusterização com 4 (quarto) variáveis de entrada e com uma resolução de 10 (dez) linha de grade para cada dimensão resulta em 10^4 pontos do *grid* para serem avaliados, à cada iteração.

Uma alternativa pode ser o método de *Clusterização* Subtrativa, proposta por (Chiu, 1994), na qual os próprios dados de entrada passam a ser candidatos a centro de *cluster*, ao invés dos pontos do *grid*. Com isto, o custo computacional passa a ser proporcional ao número de dados de entrada, e não da dimensão do problema.

A *Clusterização* Subtrativa pode ser definida pelo seguinte algoritmo:

Passo 1: Considere um conjunto de dados com n pontos $[\mathbf{x}_1 , ... , \mathbf{x}_n]$ num espaço M-dimensional, e que os dados estejam normalizados. Haja vista que cada dado de entrada é um candidato a centro de cluster, uma medida de densidade de cada ponto é definida por:

$$D_i = \sum_{j=1}^{n} \exp\left(-\frac{\left\| x_i - x_j \right\|^2}{(r_a / 2)^2} \right) \qquad \textbf{(6.71)}$$

onde r_a define o raio de vizinhança; um dado de entrada $\mathbf{x}_i$ terá um valor alto de densidade D_i se tiver um grande número de vizinhos, e um dado de entrada que esteja fora do raio r_a irá contribuir muito pouco no computo da densidade D_i.

Passo 2: Depois de calculada a medida de densidade de todos os pontos dos dados de entrada, o ponto que apresentar a maior densidade é definido como sendo o centro do primeiro cluster. Seja $\mathbf{x}_{c1}$ o ponto selecionado e D_{c1} sua densidade.

Passo 3: Em seguida, a medida de densidade de todos os pontos deve ser recalculada segundo a fórmula:

$$D_i = D_i - D_{c1} \cdot \exp\left(-\frac{\|x_i - x_{c1}\|^2}{(r_b / 2)^2} \right) \quad \textbf{(6.72)}$$

onde r_b define o raio de redução de densidade; os pontos perto do primeiro centro $\mathbf{x}_{c1}$ terão uma grande redução na medida de densidade, fazendo com que estes pontos vizinhos não sejam os próximos candidatos a centro de cluster. O raio de redução normalmente é: $r_b = 1{,}5 \cdot r_a$ (Chiu, 1994).

Depois que cada ponto teve sua medida de densidade recalculada, o próximo centro de cluster $\mathbf{x}_{c2}$ deve ser definido (Passo 2); e em seguida, todas as medidas de densidade deverão ser recalculadas segundo a Equação 6.72 (Passo 3).

Este processo deve ser repetido até que um número suficiente de centros de clusters seja definido.

6.4 Rede "Echo State Network" – ESN

Este item apresenta um algoritmo de aprendizagem extremamente útil para redes neurais recorrentes, no qual apenas os pesos das unidades de saída não modificados durante o processo de treinamento.

Sob certas condições, o estado de ativação $\mathbf{x}(n)$ de uma rede neural recorrente (RNN) é uma função do histórico dos dados de entrada [$\mathbf{u}(n)$; $\mathbf{u}(n\text{-}1)$; ...] apresentados à RNN, isto é, existe uma função $\mathbf{E}$ tal que $\mathbf{x}(n) = \mathbf{E}(\mathbf{u}(n); \mathbf{u}(n\text{-}1);\ldots)$. Metaforicamente, o estado $\mathbf{x}(n)$ pode ser entendido como um "echo" do histórico dos dados de entrada (Jaeger, 2001).

Em (Jaeger, 2001), Jaeger investigou qual poderia ser o benefício se os estados da RNN fossem interpretados como "echo states". Especificamente, sob quais condições

aparecem os “echo states” e, descreve como a RNN pode ser treinada explorando-se os “echo states”.

6.4.1 Redes Neurais Recorrentes

Redes Neurais Recorrentes (RNN) são redes neurais que apresentam uma ou mais retro-alimentações, distinguindo-se das redes neurais “feedforward”. A presença de retro-alimentações tem um profundo impacto na capacidade de aprendizagem e performance da rede (Haykin, 1998).

Tendo a rede Perceptron de Múltiplas Camadas (MLP) como arquitetura básica, existem várias formas de aplicar retro-alimentação. Podemos ter retro-alimentações partindo dos neurônios da camada de saída para os neurônios da camada de entrada. Outra possibilidade de retro-alimentação pode ser dos neurônios da camada escondida para a camada de entrada, ou ainda, para os neurônios da própria camada escondida.

Basicamente, existem duas aplicações para redes recorrentes:

- Memórias associativas
- Redes de Mapeamento entrada-saída

Para a aplicação como redes de mapeamento entrada-saída, a RNN reponde “temporariamente” a um dado sinal de entrada. Além disto, o uso de retro-alimentação possibilita as redes recorrentes a adquirir representação de “state”, o que as tornam ferramentas flexíveis para diversas aplicações como predição e modelagem não-linear, controlador não-linear, processamento de som, equalização adaptativa de canais de comunicação, etc.

Em tais aplicações, o uso de retro-alimentações pode ser capaz de reduzir significativamente a necessidade de memória (Haykin, 1998).

Apesar das RNN apresentarem vários tipos de arquitetura, todas elas seguem as seguintes características comuns:

- Incorporam uma camada (ou múltiplas camada) da MLP "estática"
- Exploram a capacidade de mapeamento não-linear da MLP

A Figura 6.13 mostra o bloco de diagrama de uma Rede Recorrente Simples (SRN). Também conhecida por Rede Elman, a SRN contém conexões recorrentes a partir dos neurônios da camada escondida à camada de entrada, após sofrer operações de atraso z^{-1}. Assim, a camada escondida possui um tipo de registro de suas ativações prévias, o que permite que a rede desempenhe tarefas de aprendizagem ao longo do tempo. Os neurônios da camada escondida também alimentam os neurônios da camada de saída que fornecem a resposta da rede aos estímulos recebidos.

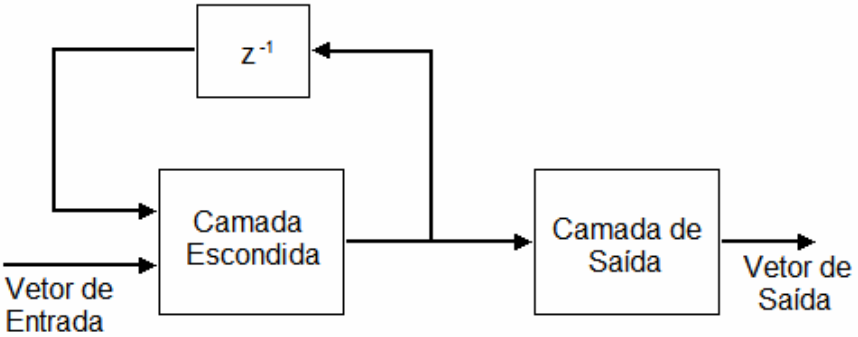


Figura 6.13 – Rede Recorrente Simples (SRN)

Devido a estas características, os neurônios da camada escondida reciclam continuamente a informação que passa pela rede por vários instantes de tempo, e assim descobrem representações abstratas do tempo.

Outra arquitetura importante de Rede Neural Recorrente (RNN) é a Perceptron Multicamada Recorrente (RMLP), que pode possuir uma ou mais camadas escondidas. Cada camada da RMLP possui retro-alimentações em torno de si, como mostra a Figura 6.14.

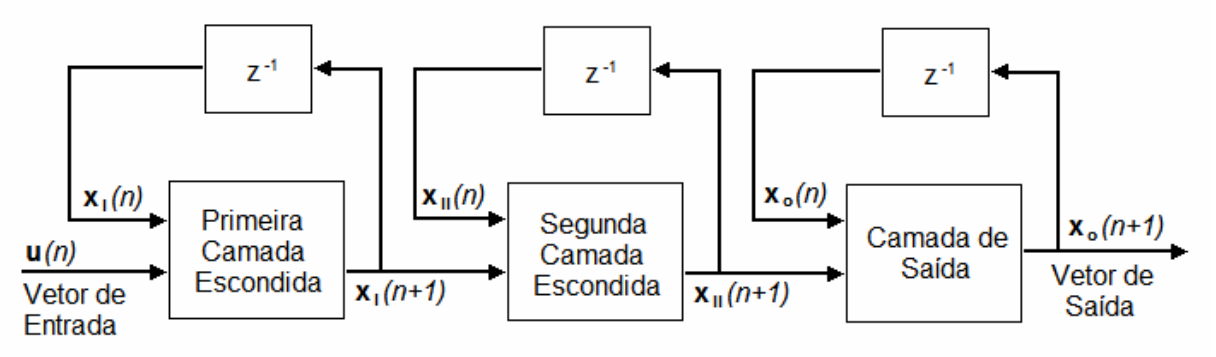


Figura 6.14 - Perceptron Multicamada Recorrente (RMLP)

O vetor $\mathbf{x}_I(n)$ denota a saída da primeira camada escondida, $\mathbf{x}_{II}(n)$ denota a saída da segunda camada escondida, e assim por diante. Já o vetor $\mathbf{x}_o(n)$ denota a saída da camada de saída. Assim, o comportamento dinâmico da RMLP, em geral, em resposta a um vetor de entrada $\mathbf{u}(n)$ é descrito pelo sistema de equações acopladas que se segue:

$$\begin{aligned} x_I(n+1) &= \varphi_I[x_I(n), u(n)] \\ x_{II}(n+1) &= \varphi_{II}[x_{II}(n), x_I(n+1)] \\ &\vdots \\ x_o(n+1) &= \varphi_o[x_o(n), x_K(n+1)] \end{aligned} \tag{6.73}$$

onde $\boldsymbol{\varphi}_I(\cdot,\cdot)$, $\boldsymbol{\varphi}_{II}(\cdot,\cdot)$, ... , $\boldsymbol{\varphi}_o(\cdot,\cdot)$ são as funções de ativação da primeira camada escondida, da segunda camada escondida, ... , e da camada de saída da RMLP, respectivamente; e $\boldsymbol{K}$ é o número de camadas escondidas da rede.

O algoritmo mais utilizado para o treinamento das Redes Neurais Recorrentes (RNN) é o "*Back-Propagation Through Time*" (BPTT), que trabalha como uma extensão do algoritmo back-propagation padrão. O BPTT pode ser entendido como o desdobramento das conexões de retro-alimentação da RNN em uma rede "*feedforward*" em camadas sucessivas, cuja topologia cresce de camada em camada a cada instante de tempo do processo de treinamento.

Sendo n_0 o instante inicial de uma época de treinamento e n_1 o seu instante final. Dado esta época, pode-se definir a função de custo:

$$\varepsilon_{total}(n_0, n_1) = \frac{1}{2}\sum_{n_0}^{n_1}\sum_{j\in A} e_j^2(n) \tag{6.74}$$

onde A é o conjunto dos neurônios j da rede que tem suas saídas desejadas especificadas, e $e_j(n)$ é o sinal de erro na saída destes neurônios.

O algoritmo BPTT trabalha da seguinte forma (Haykin, 1998):

1. Primeiro, a rede processa de modo "*forward*" o conjunto de dados de entrada desde o instante inicial até o atual, cujo intervalo é (n_0 , n_1).
2. Em seguida, num processamento "*backward*" são computados todos os gradientes locais dos pesos sinápticos da função custo da Equação 6.74:

$$\delta_j(n) = -\frac{\partial \varepsilon_{total}(n_0, n_1)}{\partial v_j(n)} \tag{6.75}$$

para todo $j \in A$ e $n_0 < n \leq n_1$. O que pode ser calculado pela fórmula:

$$\delta_j(n) = \begin{cases} \varphi'(v_j(n)) \cdot e_j(n) & \text{para } n = n_1 \\ \varphi'(v_j(n)) \cdot \left[e_j(n) + \sum_{k\in A} w_{jk}\delta_k(n+1) \right] & \text{para } n_0 < n < n_1 \end{cases} \tag{6.76}$$

onde $\varphi'(\cdot)$ é a derivada da função de ativação com relação ao seu argumento, e $v_j(n)$ é o campo de atração local do neurônio j. A Equação 6.76 é repetida, partindo do instante de tempo n_1 e caminhando para trás, instante a instante – camada por camada, até o instante inicial n_0 .

3. Quando o processo de "*back-propagation*" calcula o gradiente local $\delta_j(n+1)$ da camada n_0+1, o peso sináptico w_{ji} do neurônio j pode ser ajustado da seguinte forma:

$$\Delta w_{ji} = -\eta \frac{\partial \varepsilon_{total}(n_0, n_1)}{\partial w_{ji}} = \eta \sum_{n=n_0+1}^{n_1} \delta_j(n) \cdot x_i(n-1) \tag{6.77}$$

onde η é a taxa de aprendizagem e $x_i(n-1)$ é o i-ésimo sinal de entrada do neurônio j no instante de tempo $n-1$.

6.4.2 Modelagem da Rede ESN

Redes Neurais Recorrentes (RNN) são exemplos de modelos computacionais neuro-inspirados que lidam com o tempo sem a necessidade do pré-processamento com linhas de atraso. RNN possuem conexões recorrentes entre os seus neurônios, criando internamente a memória necessária para armazenar o histórico dos padrões de entrada. RNN têm sido amplamente utilizadas em várias aplicações tais como identificação de sistemas e controle de sistemas dinâmicos (Kechriotis, Zervas e Manolakos, 1994; Puskorius e Feldkamp, 1994; Delgado, Kambhampati e Warwick, 1995).

O principal problema das RNN é a dificuldade de adaptação dos pesos sinápticos. O algoritmo de treinamento das RNN normalmente apresenta uma série de problemas: treinamento lento como resultado da complexidade computacional, superfície de avaliação complexa, a possibilidade de instabilidade, e o decaimento dos gradientes através da topologia e do tempo (Haykin, 1998).

Recentemente, uma nova utilização das topologias de redes recorrentes que pretende lidar com as dificuldades do treinamento da RNN foi introduzido por Jaeger sob o nome de "*Echo State Network*" (ESN) (Jaeger, 2001; Jaeger e Hass, 2004).

A idéia da ESN é dividir a arquitetura da RNN em duas partes distintas: uma topologia recorrente de neurônios não-lineares, chamada de reservatório dinâmico; e uma camada linear sem-memória, chamada de "*readout*". O estado do reservatório dinâmico é chamado "*echo states*" e eles contêm informação sobre o histórico dos padrões de entrada. As saídas dos neurônios internos são fornecidas à camada de saída "*readout*" que lê o reservatório e produz

a saída da rede. A propriedade interessante da ESN é que apenas a camada linear sem memória de saída "*readout*" é treinada enquanto que a topologia recorrente permanece com suas conexões fixas.

Considere a rede neural discreta recorrente apresentada na Figura 6.15 com M unidades de entrada, N neurônios internos e L neurônios de saída. O valor da unidade de entrada no instante de tempo n é $u(n)$, das unidades internas são $x(n)$, e das unidades de saída são $y(n)$.

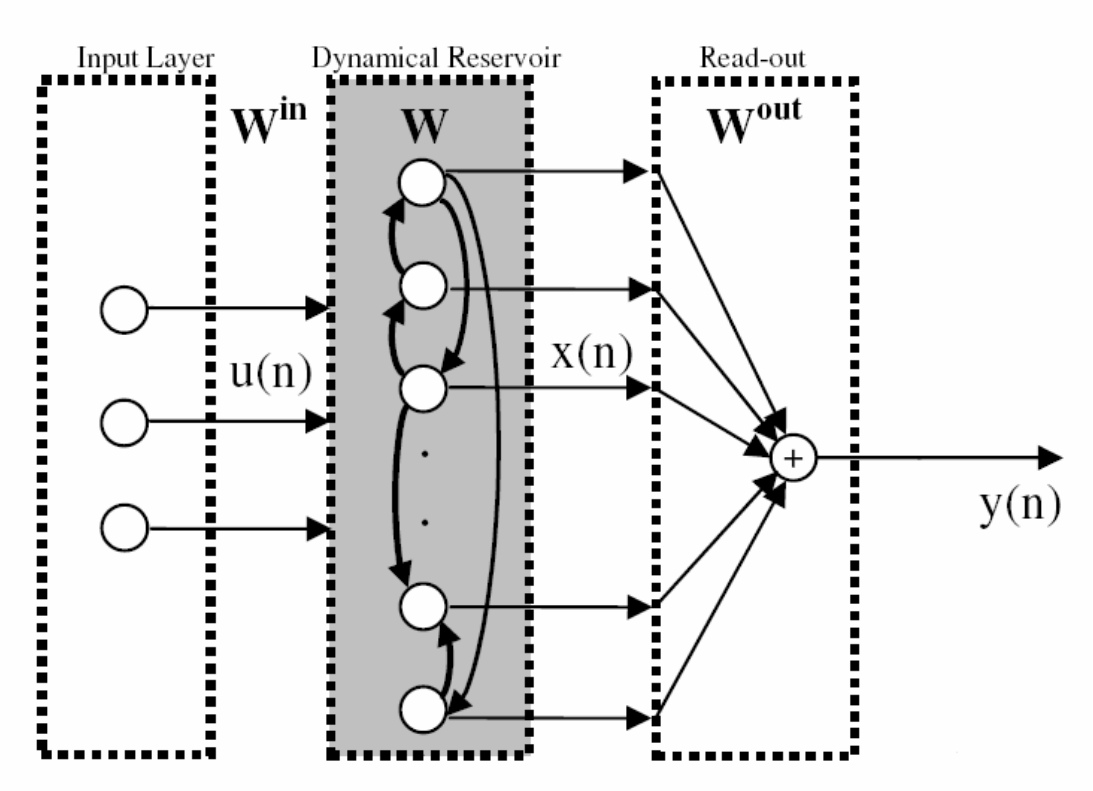


Figura 6.15 - Diagrama de uma "Echo State Nerwork" (ESN)

A ativação dos neurônios internos ("echo states") é atualizada de acordo com:

$$\mathbf{x}(n+1) = \mathbf{f}(\mathbf{W}^{\mathbf{in}} \cdot \mathbf{u}(n+1) + \mathbf{W} \cdot \mathbf{x}(n)) \qquad \textbf{(6.78)}$$

onde $\mathbf{x}$(n) é o vetor de "echo state", $\mathbf{W}$ é a matriz de pesos das conexões recorrentes, $\mathbf{u}$(n) é o sinal de entrada, $\mathbf{W}^{\mathbf{in}}$ são os pesos entre a camada de entrada e os neurônios internos da

camada escondida. Neste caso, **f** representa a função de ativação das unidades internas, normalmente definida como a função tangente hiperbólica.

A condição de "echo state" é definida em termos do raio espectral da matriz de pesos do reservatório dinâmico ($|\mathbf{W}|<1$), que está relacionada com as condições sob as quais os estados das redes recorrentes estão fortemente associados com o histórico dos padrões de entrada. Mesmo que esta condição seja bastante útil para definir a região dos parâmetros resultantes em "echo states", ela não especifica precisamente nenhum princípio para a construção de ESNs na tarefa de aproximação funcional.

A saída da ESN é computada de acordo com:

$$\mathbf{y}(n) = \mathbf{W}^{\mathbf{out}} \cdot \mathbf{x}(n) \tag{6.79}$$

onde $E[.]$, P, **x** e **d** denotam o operador valor esperado, número de dados de entrada, os dados de entrada e o sinal desejado, respectivamente.

Neste trabalho de pesquisa, a matriz de pesos da camada de entrada $\mathbf{W}^{\mathbf{in}}$ foi definida totalmente conectada e com seus valores de pesos ±1, de mesma probabilidade. A matriz de conexões recorrentes **W** foi gerada a partir de uma distribuição Gaussiana aleatória de média zero. Os valores de **W** foram normalizados para que apresentassem um raio espectral de 0,8. Foram desconsideradas as conexões de retro-alimentação de saída e as de conexão direta da entrada para a saída.

Capítulo 7

Resultados da Previsão de Vazões Afluentes

Nesta etapa do trabalho de pesquisa foi escolhida a Usina Hidrelétrica de Furnas, pertencente à Bacia do Rio Paraná – Brasil, para o estudo de caso devido a sua posição relativa na cascata, um reservatório bem à montante, conseqüentemente, sem influência operativa, com grande capacidade de armazenamento e regularização das vazões afluentes.

Foram testados e avaliados vários modelos de sistemas inteligentes na tarefa de prever um-passo-a-frente as futuras vazões afluentes médias mensais ao reservatório da UHE Furnas. Dentre eles temos: o modelo Auto-Regressivo Não-linear Auto- Organizável (SONARX) (Barreto e Araújo, 2001b), a rede com Função de Base Radial (SONARX-RBF) (Sacchi, Carneiro e Araújo, 2004), o modelo Sistema de Inferência Neuro-Fuzzy Adaptativo (ANFIS) (Jang, 1993) e a rede "Echo State Network" (ESN) (Jaeger, 2001; Jaeger e Hass, 2004).

Estes modelos de Sistemas Inteligentes foram todos implementados e testados através do *Software MatLab*® , usando um tipo de programação mista entre a própria codificação das modelagens por parte do Pesquisador quanto pela utilização das "*Toolbox*" do *MatLab*® . Com isto, o trabalho de desenvolvimento das modelagens por prototipagem ficou bastante facilitado e eficiente.

Sendo ainda necessária uma avaliação global deste tipo de modelagem, no que se refere aos erros de predição, define-se então o Erro Percentual Médio (EPM), o Erro Quadrático Médio (*Mean Squared Error* – MSE, em inglês) e o Erro Relativo Médio (*Mean Relative Error* – MRE, em inglês), dados pelas seguintes expressões:

$$EPM = \frac{100}{N}\sum_{t=1}^{N}\frac{|y(t)-\hat{y}(t)|}{y(t)} \quad \textbf{(7.1)}$$

$$MSE = \frac{1}{N}\sum_{t=1}^{N}[y(t)-\hat{y}(t)]^2 \quad \textbf{(7.2)}$$

$$MRE = \frac{1}{N}\sum_{t+1}^{N}\left[\frac{|y(t)-\hat{y}(t)|}{y(t)}\right]^2 \quad \textbf{(7.3)}$$

onde N é o tamanho da seqüência a ser prevista pelo modelo de previsão.

7.1 Avaliações Iniciais da Previsão de VA utilizando RNA

Num primeiro momento, houve a necessidade de um pré-processamento de todos os dados envolvidos neste processo de treinamento e teste de arquitetura da rede proposta. Como as redes neurais trabalham com dados de natureza contínua, optou-se por uma normalização destes dados, a fim de garantir a estabilidade numérica das RNA através da utilização de dados limitados entre 0 e 1. Esta normalização foi feita para toda a série histórica de vazões afluentes.

Tomou-se como limite superior o recorde histórico já registrado para a UHE de Furnas. De posse desta vazão afluente histórica máxima, dividiu-se todo o histórico de vazões por este pico máximo, mês a mês, e usina por usina. Gerando um "histórico" de afluências normalizado pelo pico histórico de cada usina hidrelétrica.

Os dados são referentes à UH de Furnas, mais precisamente os dados de afluência deslocados no tempo de t, $t-1$, $t-2$, constituindo os 3 (três) dados de entrada. Essa janela do passado foi assim definida com a intenção de fornecer ao modelo condições de distinguir e capturar a tendência hidrológica dentro do próprio ano.

Já o dado de saída é fornecido por 1 (uma) unidade de saída, sendo que ela está relacionada aos dados de previsão 1 (um) passo a frente, ou seja $t+1$; mas vale lembrar que as RNA são capazes de prever k-passos a frente, $t+k$.

Para o treinamento dos Sistemas Inteligentes (SI), todo o histórico de vazões afluentes normalizado, foi dividido em 2 (dois) períodos, sendo o primeiro correspondente ao período de janeiro de 1972 até dezembro de 1976, considerado um período de vazões afluentes medianas, ou seja, caracterizando o comportamento hidrológico "médio" para este caso. O segundo referente ao período de janeiro de 1933 até dezembro de 1994, a menos é claro, do período mencionado anteriormente.

Desta forma, tem-se um período de 5 (cinco) anos, destinado para o teste do desempenho de cada modelo treinado. E para o treinamento, ou seja, os ajustes paramétricos dos modelos, tem-se um conjunto de dados de vazões afluentes que corresponde a 57 anos de um histórico de 64 anos. Haja visto que o período de teste não faz parte do período de treinamento, além de não serem utilizados os dois primeiros anos do histórico.

Os resultados da previsão de vazões afluentes da UH de Furnas são apresentados na Tabela 7.1, a seguir:

Tabela 7.1 - Resultados da previsão de vazões afluentes da UH de Furnas: período de 1972 a 1976, histórico normalizado e 3 dados de entrada (t , t-1 , t-2)

		SONARX	SONARX-RBF	ANFIS	
				Grid Partition	Sub Cluster
		36 neurônios	24 neurônios	2 fç de pert.	3 clusters
				8 regras	3 regras
MSE (10^{-3})	treino	13,6	9,8	9,6	10,1
	teste	16,4	6,9	7,1	6,3
EPM (%)	treino	26,98	26,04	26,18	26,71
	teste	37,94	29,35	29,90	29,04

Com base nos resultados obtidos, pode-se notar que a rede SONARX teve o pior desempenho, pois apresenta erros elevados nesta tarefa de previsão de Vazões Afluentes. Isto já era de se esperar, pois o modelo SONARX possui um erro inerente a este tipo de modelagem que trabalha com prototipagem, levando à um mapeamento discreto no espaço de saída, o que já foi discutido anteriormente.

Pôde-se observar uma melhora sensível no desempenho das previsões, com a utilização da rede SONARX-RBF, uma vez que ela produz uma interpolação dos dados de saída. Observa-se uma melhora de **57.9 %** para o Erro Quadrático Médio (MSE), e uma melhora de **22.6 %** para o Erro Percentual Médio (EPM).

Já com relação ao desempenho do modelo ANFIS, nota-se que este tipo de abordagem apresenta um desempenho similar, mas um pouco melhor do que a rede SONARX-RBF. Com maior destaque para o modelo ANFIS que utiliza o particionamento em cluster, por apresentar um desempenho um pouco melhor, para uma arquitetura bem menor (mais "enxuta"), o que leva evidentemente a um treinamento bem menos custoso computacionalmente.

Em outras palavras, para este caso, o modelo ANFIS com particionamento em *grid* possui 2 (duas) funções de pertinência para cada variável de entrada, levando um total de 8 (oito) regras *fuzzy*. Enquanto que o modelo ANFIS com particionamento em *cluster* possui apenas 3 (três) *clusters* e conseqüentemente somente 3 (três) regras *fuzzy*.

Este modelo ANFIS com particionamento em *cluster* apresenta uma melhora de **61.6 %** para o Erro Quadrático Médio (MSE), e uma melhora de **23.5 %** para o Erro Percentual Médio (EPM), em relação ao desempenho da rede SONARX.

7.2 Previsões utilizando Análise dos Componentes Principais

Posteriormente a esta bateria de testes, foi feita uma nova seqüência de teste também para a previsão de futuras vazões afluentes à UH de Furnas. Mas neste caso, foi aplicada uma técnica para o pré-processamento dos dados do histórico de vazões afluentes.

A Análise dos Componentes Principais (PCA), foi aplicada neste caso, com o intuito de se definir de forma automática a ordem de memória a ser adotada, como parâmetro para se definir o número de variáveis de entrada aos SI. Ou seja, com o uso desta metodologia, não

seria mais necessário que o usuário, com a ajuda de um especialista, defina quais deveriam ser os dados de vazões do passado a serem levados em consideração para o pretendido mapeamento das correlações, para a previsão de futuras vazões afluentes. Esta tarefa ficaria a cargo da técnica de PCA.

Para a implementação da PCA, foram utilizados padrões com dados dos últimos 2 (dois) anos, ou seja, padrões com uma ordem de memória de 24 meses passados. Como a PCA faz uma transformação linear do espaço original, e cria um novo espaço ortogonal de Componentes Principais de mesma dimensão e escalonado em ordem decrescente com relação a variância de cada um deles, é possível definir um certo grau de preservação da informação contida em poucos Componentes Principais, desprezando todos os demais.

Desta maneira, foi possível reduzir bastante a dimensão do problema, de um espaço 24-dimensional para um espaço 4-dimensional. Isto porque havia um elevado grau de redundância de informação presente nos padrões.

Mas para que isto pudesse ser feito, foi preciso padronizar a série toda de vazões afluentes. Foi preciso retirar a não-estacionaridade da série, tornando-a uma série estacionária, ou seja, sem o comportamento sazonal anual, característico desta série.

Em seguida, o novo conjunto de padrões foi finalmente apresentado aos SI utilizados neste trabalho de pesquisa. E os resultados da previsão de vazões afluentes da UH de Furnas são apresentados na Tabela 7.2, a seguir:

Tabela 7.2 - Resultados da previsão de vazões afluentes da UH de Furnas: período de 1972 a 1976, histórico padronizado e 4 dados de entrada (PCA)

		SONARX	SONARX-RBF	ANFIS	
				Grid Partition	Sub Cluster
		24 neurônios	60 neurônios	2 fç de pert.	3 clusters
				16 regras	3 regras
MSE (10^{-3})	treino	12,2	7,8	7,6	9,4
	teste	4,6	3,9	6,3	4,4
EPM (%)	treino	24,91	21,81	21,43	23,38
	teste	24,77	22,71	25,02	21,40

Com base nos resultados das Tabelas 7.1 e 7.2, pode-se notar uma sensível melhora de desempenho por parte de todos os SI, o que nos leva a concluir que a técnica de PCA se mostrou bastante eficiente em reduzir a dimensão do problema, preservando boa parte da informação.

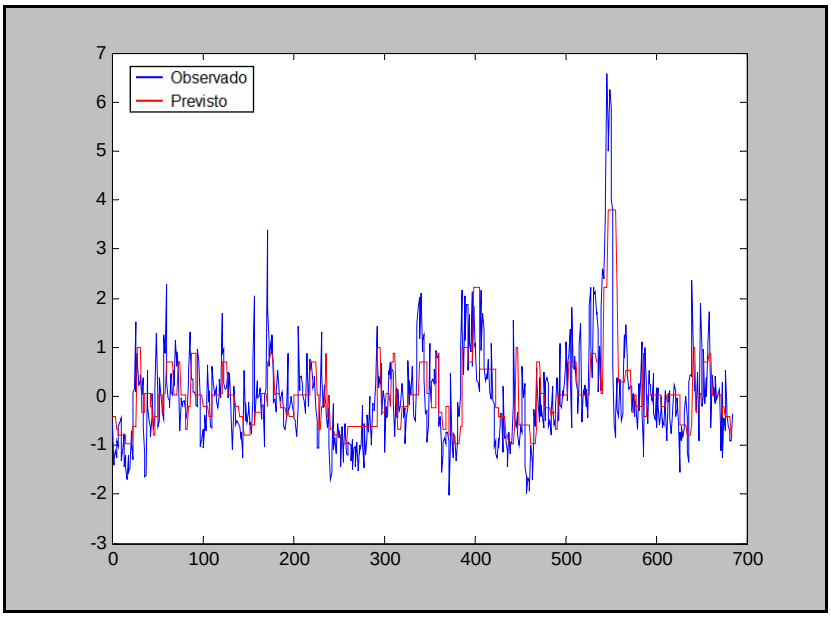


Figura 7.1 – Recuperação das vazões afluentes de treinamento: UH Furnas, série padronizada, rede SONARX

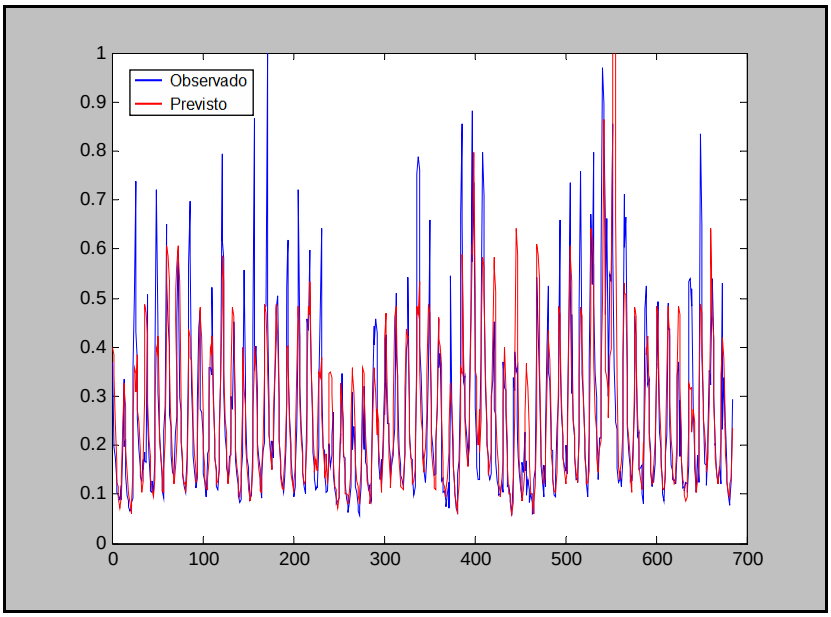


Figura 7.2 – Recuperação das vazões afluentes de treinamento: UH Furnas, série normalizada, rede SONARX

Novamente, a rede SONARX teve o pior desempenho, pois apresenta erros elevados nesta tarefa de previsão de Vazões Afluentes. Entretanto, observa-se uma melhora de **71.9 %** para o Erro Quadrático Médio (MSE), e uma melhora de **34.7 %** para o Erro Percentual Médio (EPM), se comparado ao caso anterior, sem a técnica de PCA.

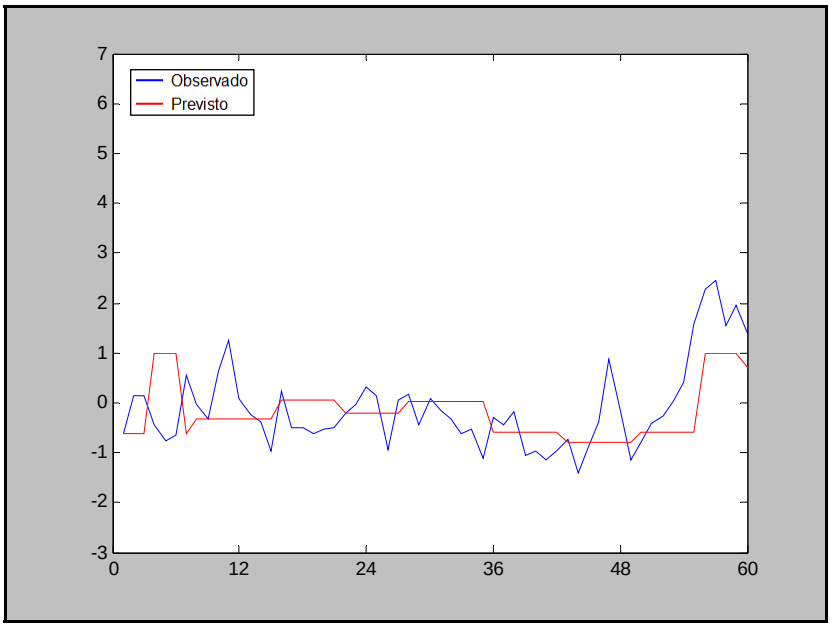


Figura 7.3 – Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série padronizada, rede SONARX

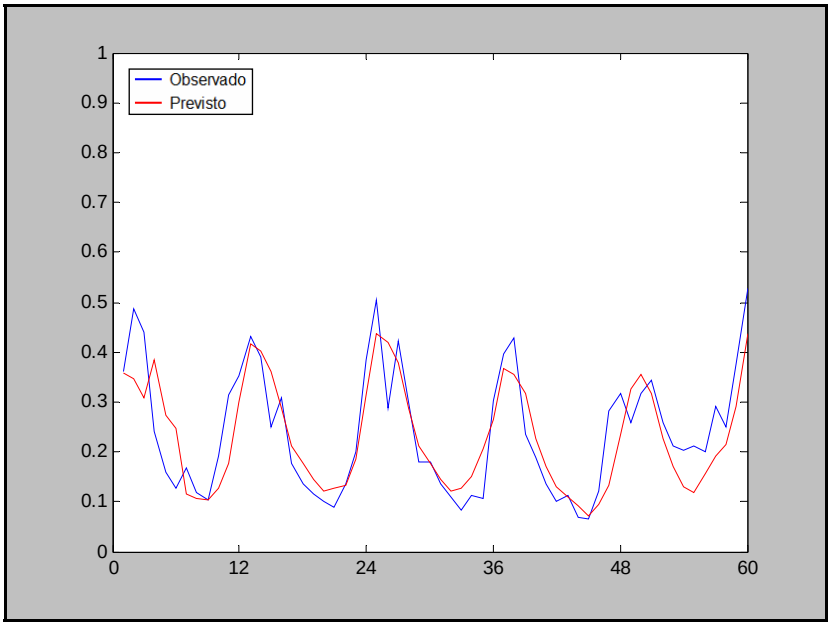


Figura 7.4 – Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série normalizada, rede SONARX

O mesmo ocorre para os demais modelos, podendo ser destacado para o modelo ANFIS com particionamento em cluster, que apresenta uma melhora de **30.2 %** para o Erro Quadrático Médio (MSE), e uma melhora de **26.3 %** para o Erro Percentual Médio (EPM), em relação ao caso sem a técnica de PCA.

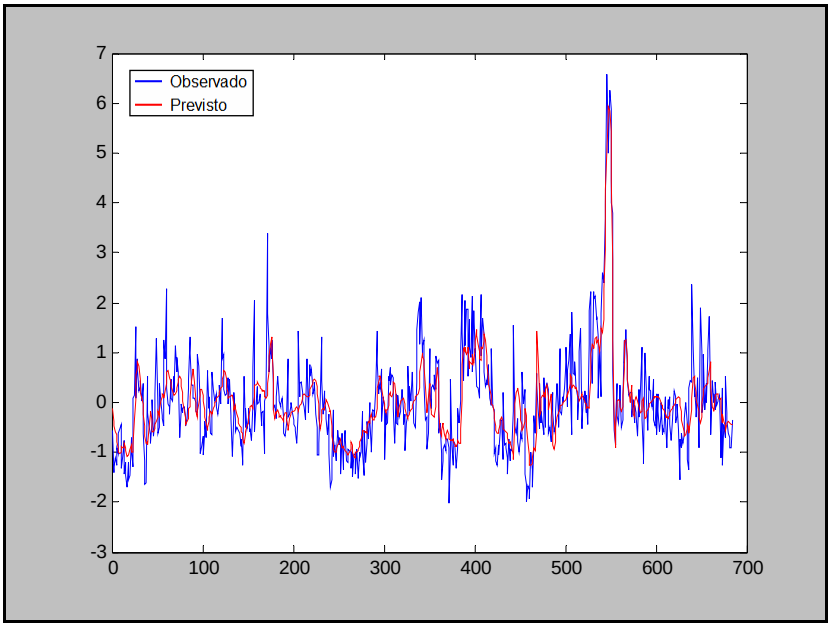


Figura 7.5 – Recuperação das vazões afluentes de treinamento: UH Furnas, série padronizada, rede SONARX-RBF

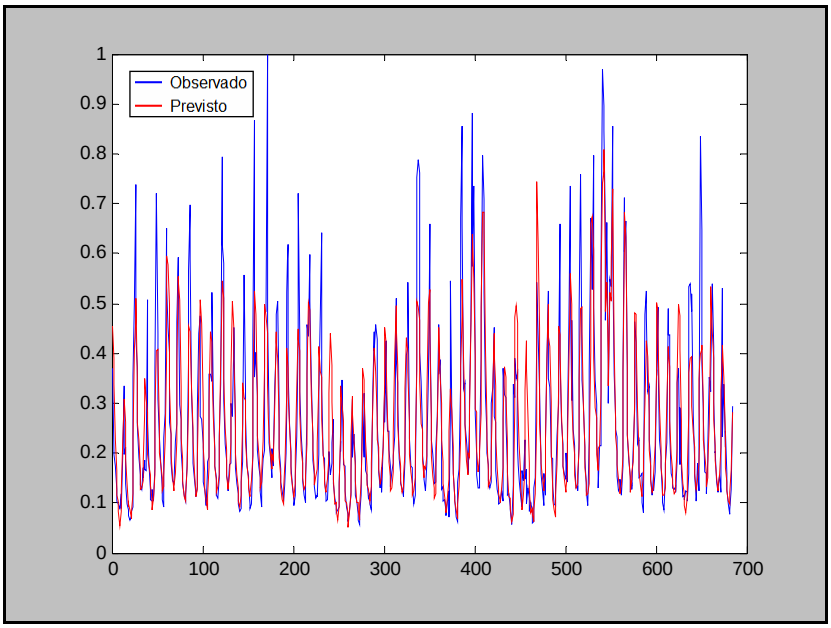


Figura 7.6 – Recuperação das vazões afluentes de treinamento: UH Furnas, série normalizada, rede SONARX-RBF

Nos gráficos da Figura 7.1 até a Figura 7.4, são apresentados os resultados de recuperação, para os anos de treinamento, e previsão, para os anos de teste, das vazões afluentes a UH de Furnas, usando a rede SONARX.

Observa-se que em ambos os casos o desempenho foi razoável, mostrando uma certa capacidade de predição da rede SONARX.

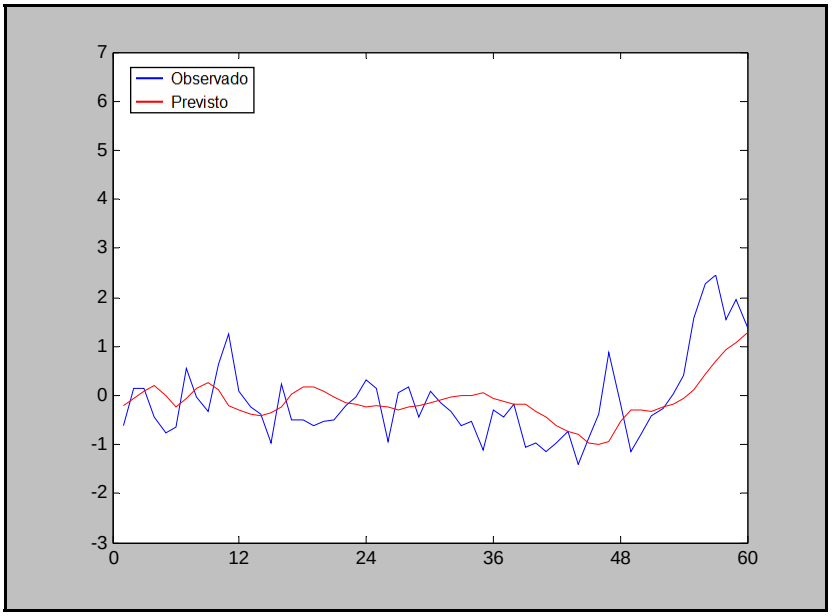


Figura 7.7 – Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série padronizada, rede SONARX-RBF

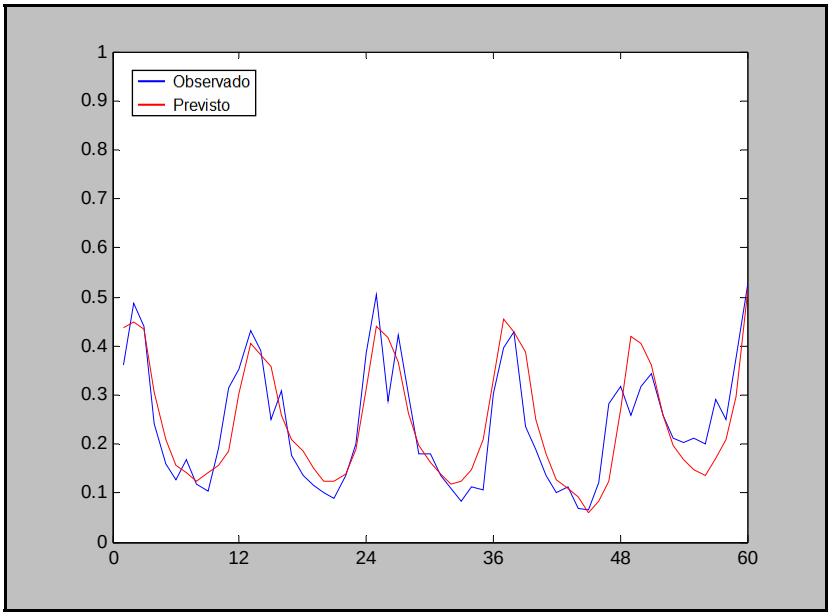


Figura 7.8 – Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série normalizada, rede SONARX-RBF

Nos gráficos da Figura 7.5 até a Figura 7.8, são apresentados os resultados de recuperação, para os anos de treinamento, e previsão, para os anos de teste, das vazões afluentes a UH de Furnas, usando a rede SONARX-RBF.

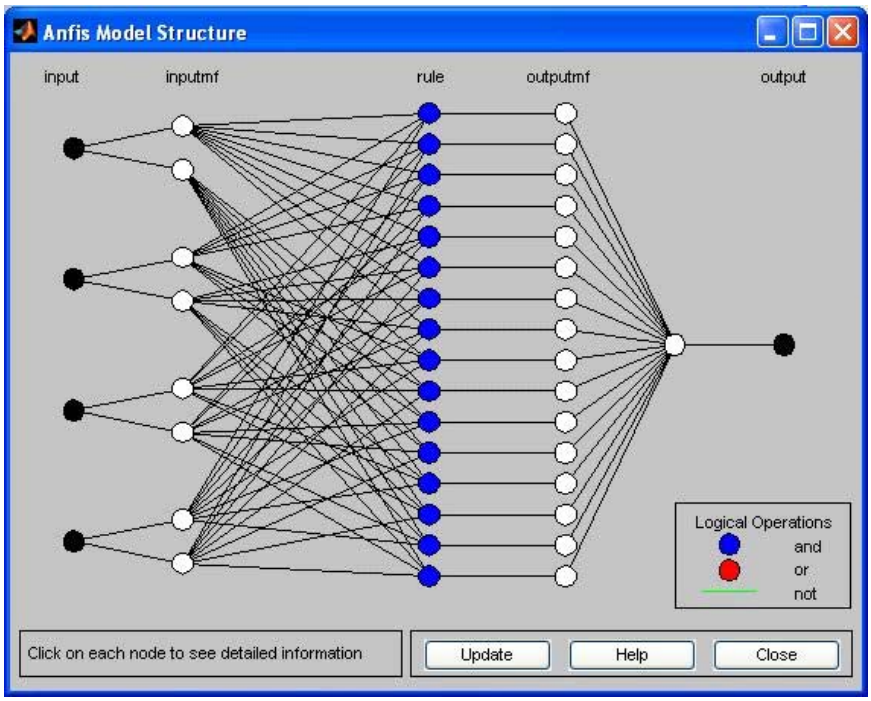


Figura 7.9 – Arquitetura ANFIS com particionamento em *grid*: 16 regras *fuzzy*

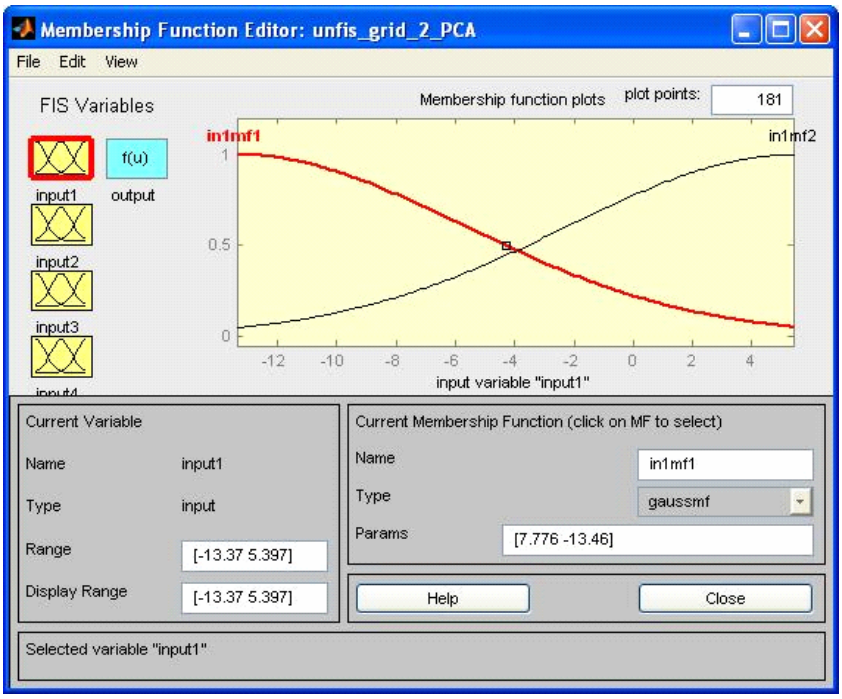


Figura 7.10 – Funções de pertinência do modelo ANFIS, particionamento em *grid*: 2 FP

Visualmente, pode-se notar uma melhora de desempenho da rede SONARX-RBF, suavizando a trajetória da série, decorrente da interpolação.

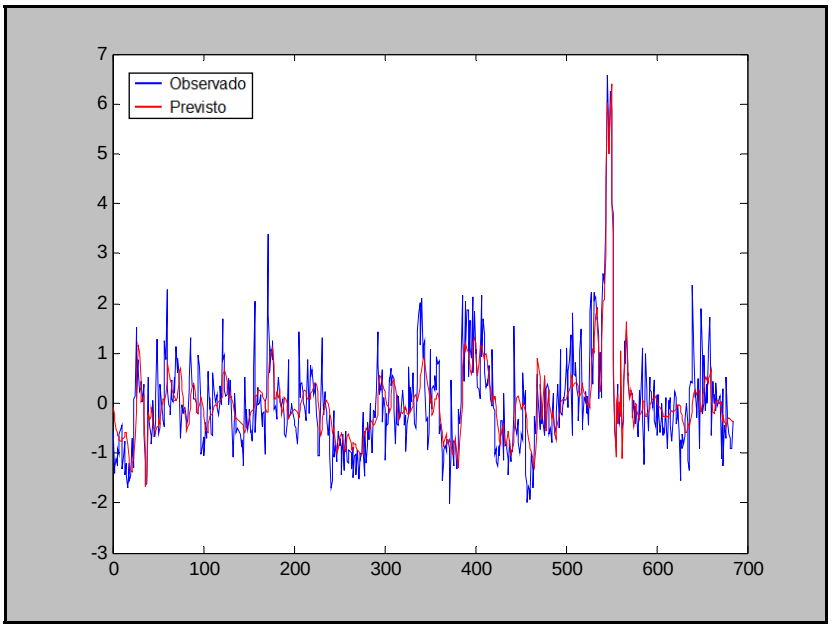


Figura 7.11 – Recuperação das vazões afluentes de treinamento: UH Furnas, série padronizada, modelo ANFIS – particionamento em *grid*

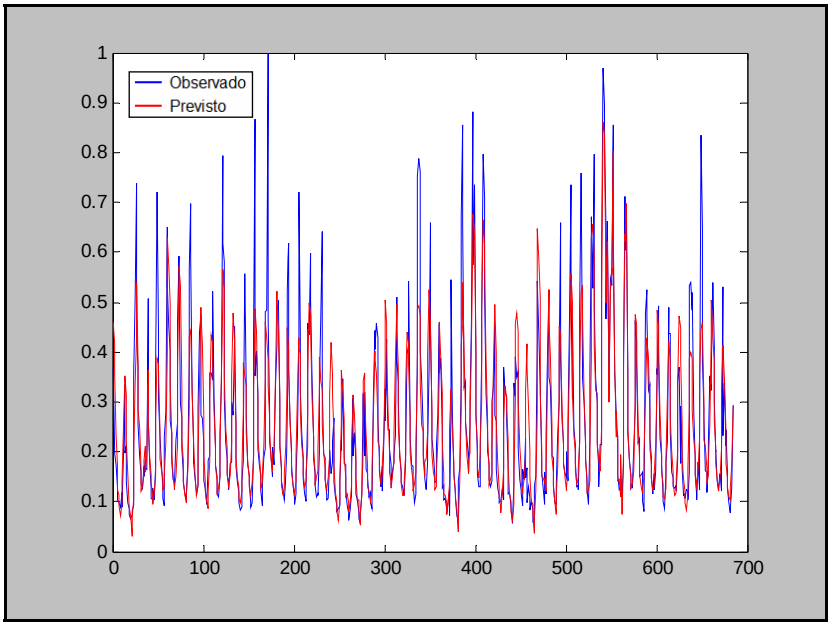


Figura 7.12 – Recuperação das vazões afluentes de treinamento: UH Furnas, série normalizada, modelo ANFIS – particionamento em *grid*

Já a Figura 7.9 ilustra a arquitetura ANFIS com particionamento em *grid*. Neste caso, já começa a aparecer o problema do crescimento exponencial do número de regras *fuzzy*: 2^4 =

16 regras. A Figura 7.10 ilustra as duas funções de pertinência associadas a primeira variável de entrada, z_t .

Nos gráficos da Figura 7.11 até a Figura 7.14, são apresentados os resultados de recuperação, para os anos de treinamento, e previsão, para os anos de teste, das vazões afluentes a UH de Furnas, usando o modelo ANFIS com particionamento em *grid*.

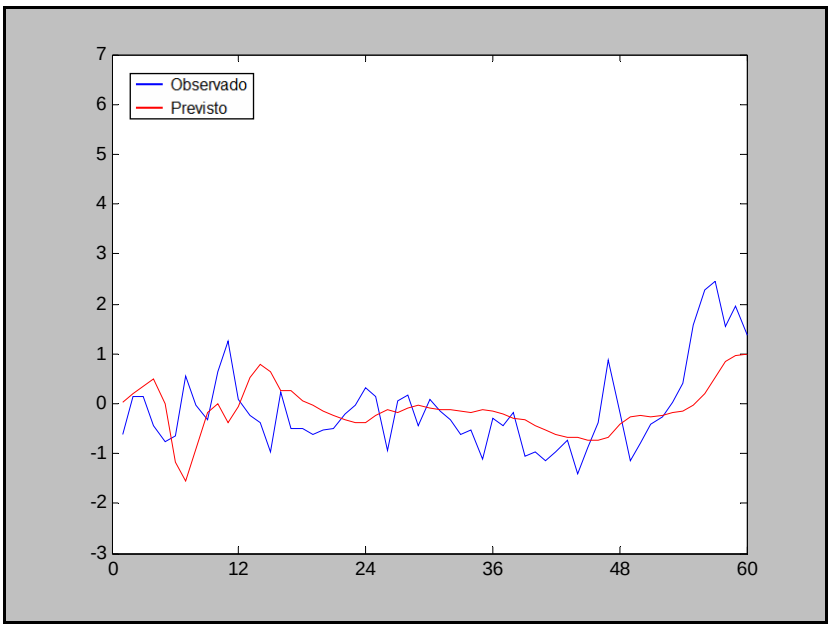


Figura 7.13 – Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série padronizada, modelo ANFIS – particionamento em *grid*

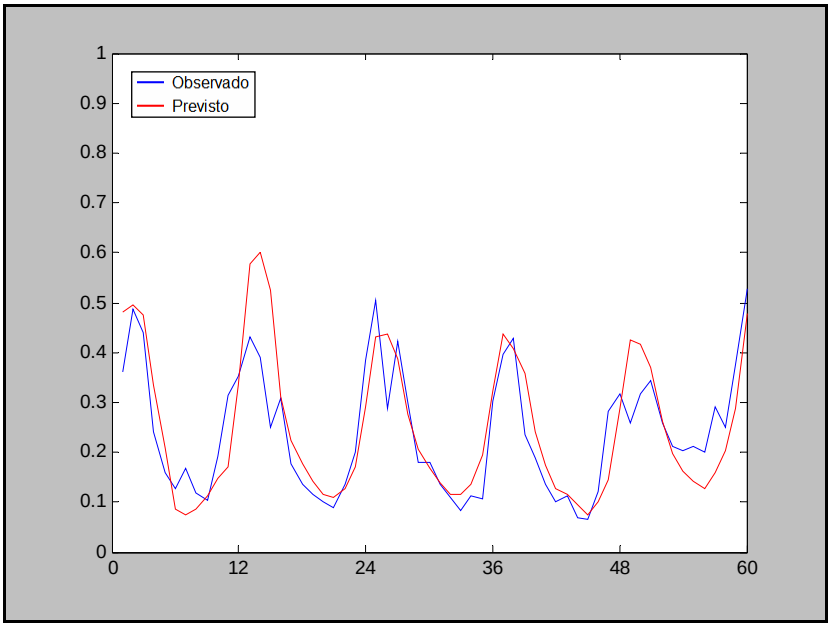


Figura 7.14 - Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série normalizada, modelo ANFIS – particionamento em *grid*

Visualmente, pode-se notar um bom desempenho do modelo ANFIS, comparável ao da rede SONARX-RBF.

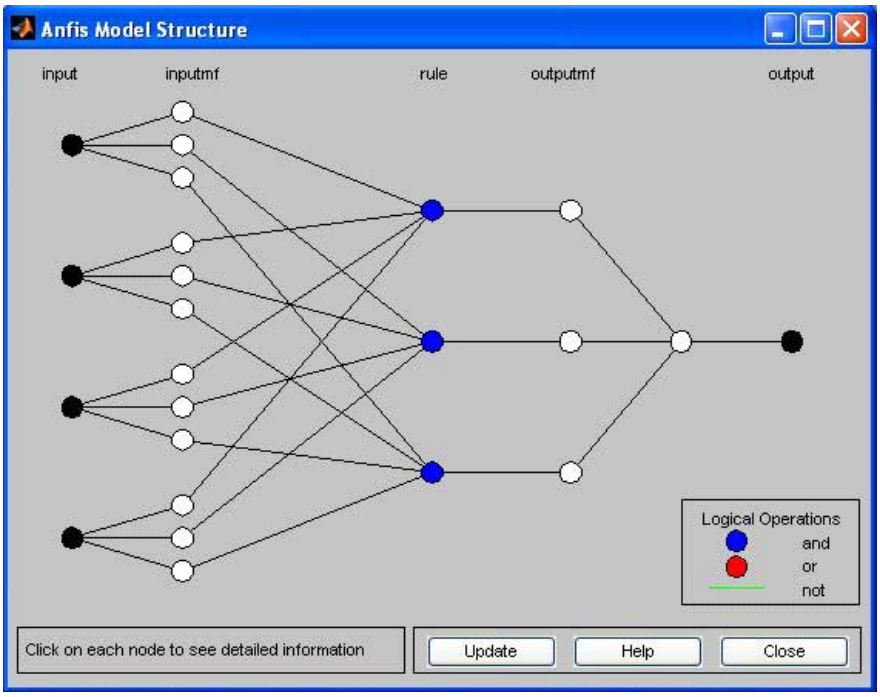


Figura 7.15 - Arquitetura ANFIS com particionamento em cluster: 3 clusters e 3 regras *fuzzy*

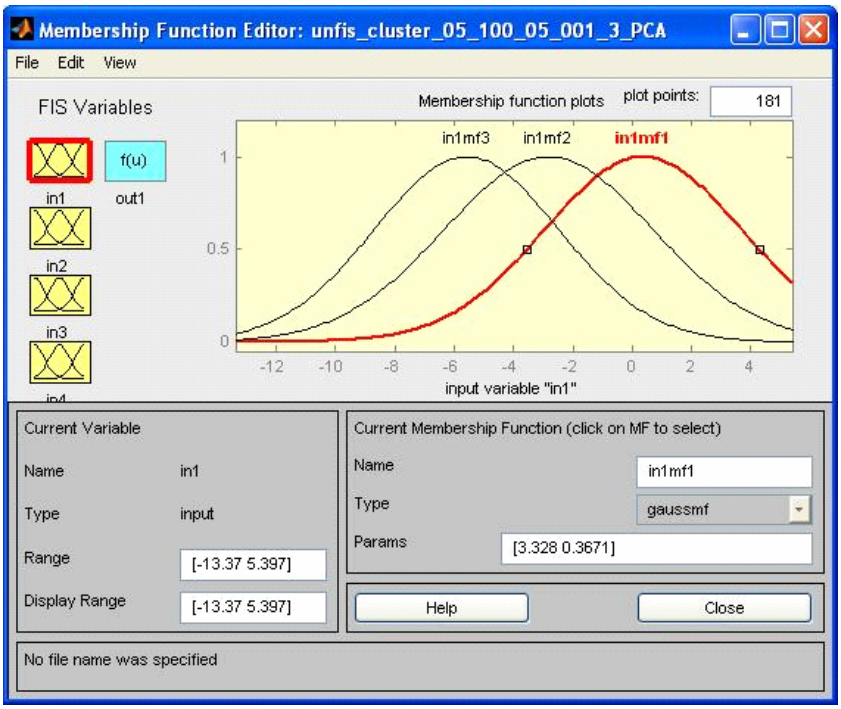


Figura 7.16 – Funções de pertinência do modelo ANFIS, particionamento em *cluster*: 3 FP

As Figura 7.15 ilustra a arquitetura ANFIS com particionamento em *cluster*. Esta arquitetura se mostra bem mais compacta que a do caso anterior, exibindo apenas 3 *clusters* e conseqüentemente apenas 3 regras *fuzzy*. Na Figura 7.16 aparecem as 3 funções de pertinência associadas a primeira variável de entrada, z_t .

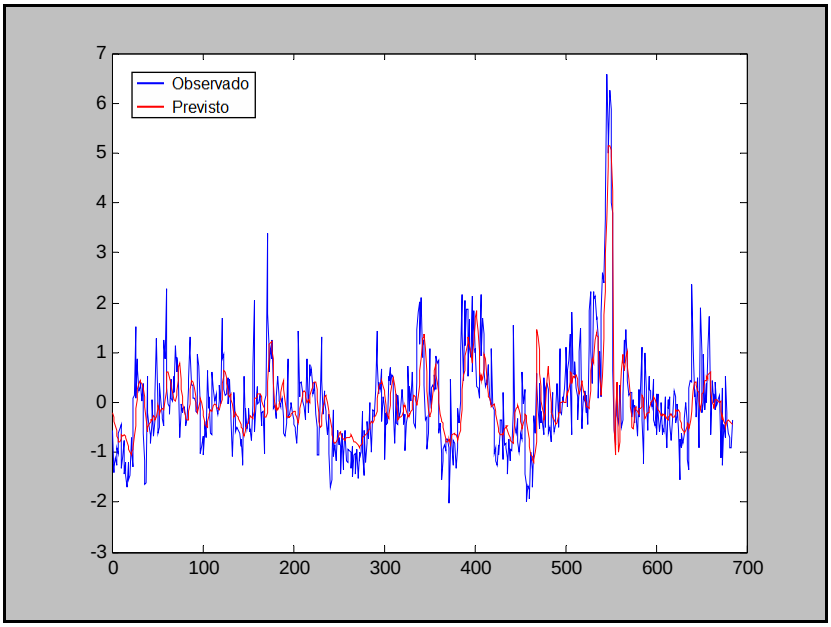


Figura 7.17 – Recuperação das vazões afluentes de treinamento: UH Furnas, série padronizada, modelo ANFIS – particionamento em *cluster*

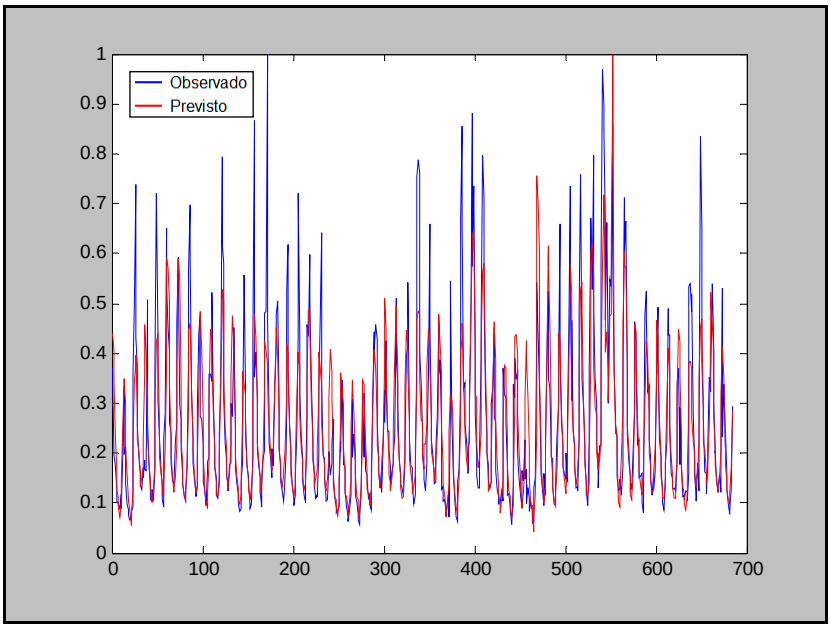


Figura 7.18 - Recuperação das vazões afluentes de treinamento: UH Furnas, série normalizada, modelo ANFIS – particionamento em *cluster*

Nos gráficos da Figura 7.17 até a Figura 7.20, são apresentados os resultados de recuperação, para os anos de treinamento, e previsão, para os anos de teste, das vazões afluentes a UH de Furnas, usando o modelo ANFIS com particionamento em *grid*.

Visualmente, pode-se notar um bom desempenho do modelo ANFIS, perfeitamente aceitável para utilização prática.

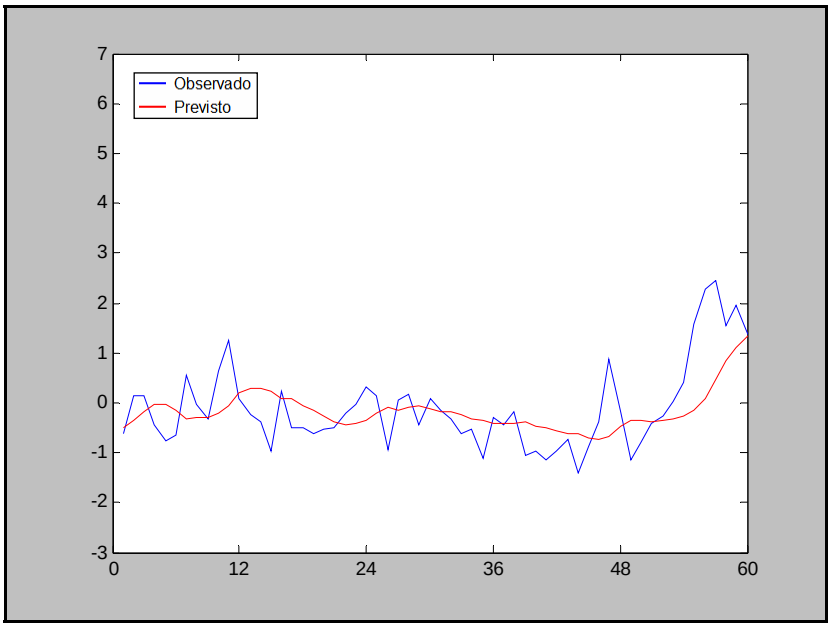


Figura 7.19 - Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série padronizada, modelo ANFIS – particionamento em *cluster*

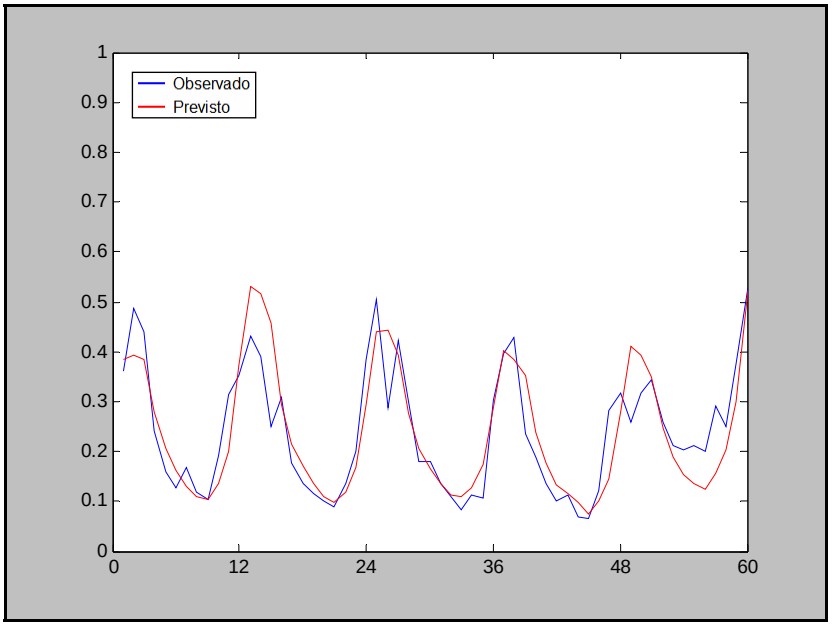


Figura 7.20 - Previsão das vazões afluentes de teste (1972 a 1976): UH Furnas, série normalizada, modelo ANFIS – particionamento em *cluster*

7.3 Previsões utilizando a rede "Echo State Network"

Usualmente, a maioria dos modelos de previsão, incluindo os modelos com os quais a ESN esta sendo comparado, necessita da escolha prévia do espaço de entrada "ótimo" para o previsor, selecionando as mais importantes dentre as medidas anteriores.

Foram escolhidos para todos os modelos aqui utilizados, exceto para a ESN, três opções de espaço de entrada. A primeira opção foi escolhida como uma memória de atraso de curta duração, definida como:

$$x_1(t) = va(t)$$
$$x_2(t) = va(t-1)$$
$$x_3(t) = va(t-2)$$

onde, $t = 1,...,T$ denota o mês atual, va a vazão afluente, e x_i a variável de entrada para o modelo de previsão.

A segunda opção para o espaço de entrada foi definida como uma memória de atraso de média duração, contendo provavelmente uma quantidade maior de informação, definida como:

$$x_1(t) = va(t)$$
$$x_2(t) = va(t-1)$$
$$x_3(t) = va(t-2)$$
$$x_4(t) = va(t-3)$$
$$x_5(t) = va(t-4)$$
$$x_6(t) = va(t-5)$$

E, a terceira opção foi definida como a combinação de dois períodos curtos, um relativo a um passado recente e outro relativo ao mesmo período do ano anterior. Desta maneira, pode-se levar em consideração não somente as informações mensais mas também as informações das variações anuais. Sendo então definida como:

$$x_1(t) = va(t)$$

$$x_2(t) = va(t-1)$$

$$x_3(t) = va(t-2)$$

$$x_4(t) = va(t-10)$$

$$x_5(t) = va(t-11)$$

$$x_6(t) = va(t-12)$$

No entanto, como mencionado anteriormente, a ESN não necessita do uso de nenhuma destas opções de espaço de entrada, uma vez que a ESN possui uma memória embutida como resultado das conexões recorrentes.

Além disso, como o objetivo, neste trabalho, foi o de obter a previsão um-passo-a-frente, a saída desejada para todos os modelos de previsão foi definida como:

$$y_d(t) = va(t+1)$$

Através dos experimentos realizados, decidiu-se utilizar uma arquitetura de ESN com N = 36 elementos de processamentos. A matriz de pesos da camada de entrada $\mathbf{W}^{\mathbf{in}}$ foi definida totalmente conectada e com seus valores de pesos ±1, de mesma probabilidade. A matriz de conexões recorrentes **W** foi gerada a partir de uma distribuição Gaussiana aleatória de média zero. Os valores de **W** foram normalizados para que apresentassem um raio espectral de 0,8. Foram desconsideradas as conexões de retro-alimentação de saída e as de conexão direta da entrada para a saída.

A figura 7.21 mostra os "echo states" resultantes do treinamento da ESN com os dados de treinamento. Pode-se notar que os "ESN states" apresentam um amplo comportamento dinâmico para este problema, para os instantes em que foi necessária uma maior cobertura do espaço de características, durante todo o período de treinamento.

Esta variedade de "ESN states" representa um conjunto de bases funcionais construído dinamicamente pelos dados de entrada, enquanto que a camada de saída simplesmente projeta a resposta desejada neste espaço de representação.

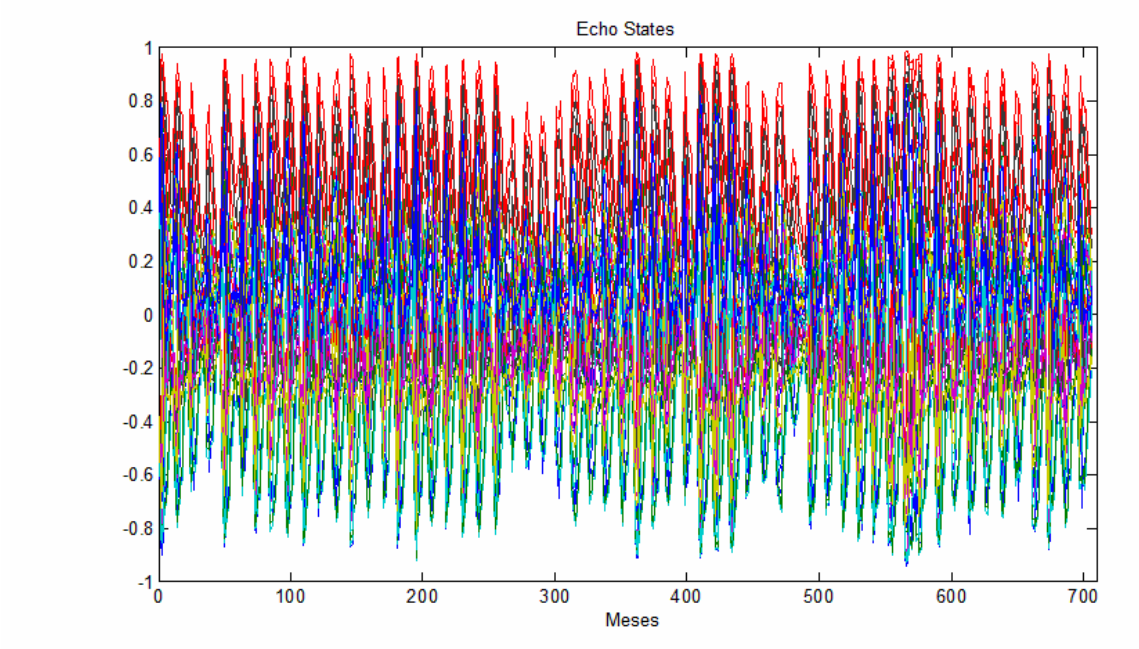


Figura 7.21 - "Echo States" do modelo ESN treinado com histórico de vazões afluentes mensais

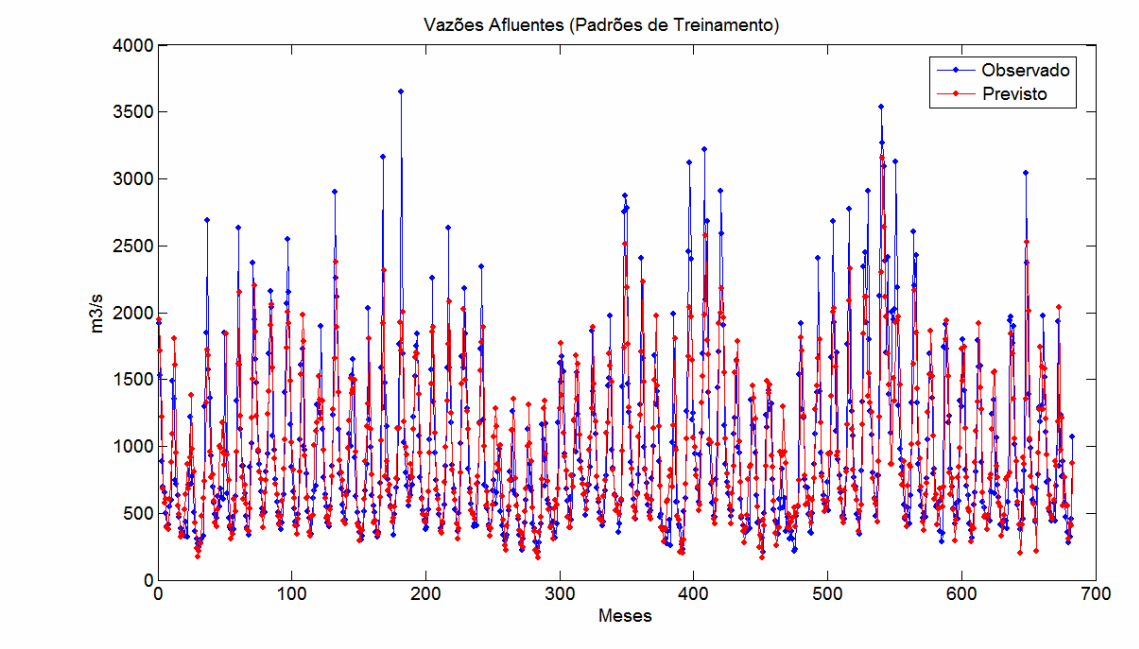


Figura 7.22 – Padrões de Treinamento: Vazões Afluentes a Usina Hidrelétrica de Furnas

Pode ser visto, na Figura 7.22, o resultado da recuperação das vazões afluentes para os dados de treinamento. A linha azul corresponde às medições observadas desta série temporal, e a linha vermelha representa os valores previstos das vazões afluentes.

Com base na Figura 7.22, pode-se dizer que o modelo de previsão foi capaz de aprender e capturar a maior parte das variações presentes nesta série temporal; o que parece, mas não é uma tarefa fácil para nenhum previsor.

A Figura 7.23 mostra a previsão de vazões afluentes para o período de 1972-1976. O previsor ESN apresenta um ótimo resultado, onde, pode-se observar uma grande semelhança entre as duas séries temporais. A linha vermelha, representando a previsão das vazões afluentes, apresenta um comportamento bem próximo da linha azul, os valores reais observados.

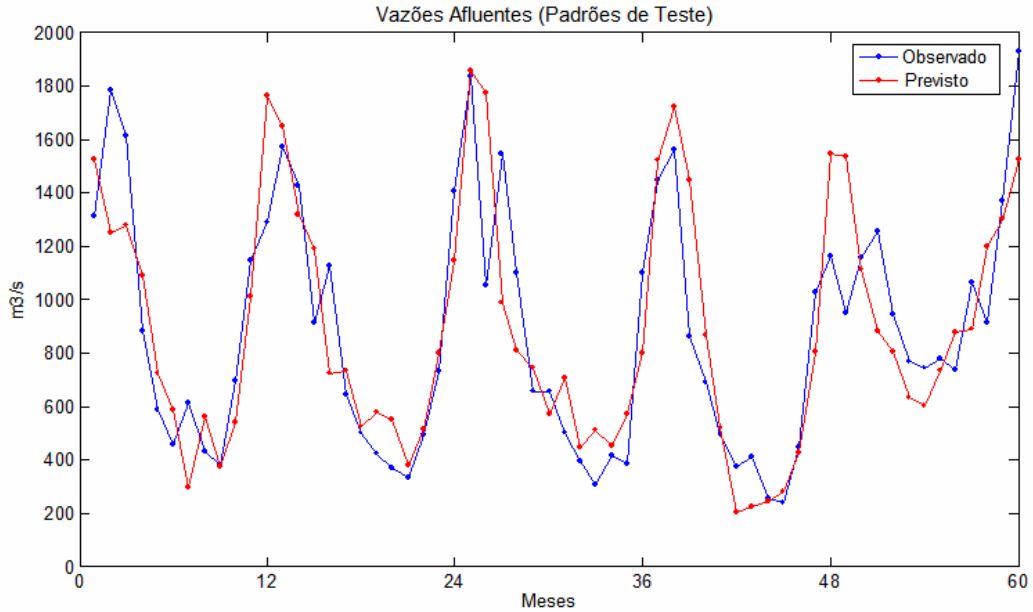


Figura 7.23 - Padrões de Teste (1972 - 1976): Vazões Afluentes a Usina Hidrelétrica de Furnas

Mesmo no último ano, 1976, é possível notar que o previsor ESN foi capaz de prever um maior (mais úmido) período de seca, no meio daquele ano.

Tabela 7.3 - Erros de Previsão das VA da Usina Hidrelétrica de Furnas

Modelos	*Espaço de Entrada*	*MSE (x10^4)*	*RMSE*	*MAD*	*MPE (%)*
Echo State Network	embutido	5.97	244.45	177.41	21.04
SONARX	t,t-1,t-2	21.78	466.73	320.13	37.94
	t,t-1,t-2,t-3,t-4,t-5	12.20	349.30	253.18	28.78
	t,t-1,t-2,t-10,t-11,t-12	13.54	367.90	261.04	28.32
SONARX-RBF	t,t-1,t-2	9.24	303.94	233.24	29.35
	t,t-1,t-2,t-3,t-4,t-5	6.35	252.02	187.53	22.41
	t,t-1,t-2,t-10,t-11,t-12	7.13	266.94	202.56	24.35
ANFIS	t,t-1,t-2	9.45	307.47	235.67	29.89
	t,t-1,t-2,t-3,t-4,t-5	7.07	265.97	198.52	24.27
	t,t-1,t-2,t-10,t-11,t-12	6.72	259.25	205.99	25.02

A Tabela 7.3 mostra os erros de previsão obtidos pelos modelos de previsão usados e comparados neste trabalho de pesquisa: "Echo State Network", SONARX, SONARX-RBF e o modelo ANFIS. Nota-se que a ESN, SONARX-RBF e ANFIS apresentam melhor performance em comparação ao modelo SONARX, uma vez que este modelo apresenta um erro intrínseco por basear-se na Memória Associativa Temporal via Quantização-Vetorial (VQTAM). Além disso, os modelos SONARX-RBF e ANFIS apresentam praticamente o mesmo desempenho.

Finalmente, o modelo ESN apresenta uma performance melhor do que todos os outros modelos, apesar dele possuir um algoritmo de treinamento bem mais simples e rápido. Estes resultados demonstram que a ESN é uma poderosa e promissora ferramenta para este tipo de aplicação.

7.4 Previsões utilizando "Analytic Signal" e "Dynamic Modelling"

A abordagem proposta nos Capítulos 4 e 5 combina a metodologia de pré-processamento, a decomposição "*Analytic Signal*", com o uso de Redes Neurais Artificiais (RNN) para o "Dynamic Modelling" Não-linear da série temporal de VA.

Nesta seção, o modelo SONARX acoplado a rede com Função de Base Radial (SONARX-RBF) (Sacchi, Carneiro e Araújo, 2004) foi aplicada como um modelo de previsão de VA mensais de vários-passos-a-frente, para a UHE de Furnas.

Um novo histórico de Vazões Afluentes mensais, para o período de 1931-2006, foi utilizado neste estudo. O período de 2000-2006 foi selecionado como dados de teste, enquanto que o período de 1931-1999 foi utilizado como dados de treinamento.

Novamente foram escolhidas três opções de espaço de entrada. A primeira opção foi escolhida como uma memória de atraso de curta duração, definida como:

$$x_1(t) = va(t)$$
$$x_2(t) = va(t-1)$$
$$x_3(t) = va(t-2)$$

onde, $t = 1,\ldots,T$ denota o mês atual, va a vazão afluente, e x_i a variável de entrada para o modelo de previsão.

A segunda opção para o espaço de entrada foi definida como uma memória de atraso de longa duração, contendo provavelmente uma quantidade maior de informação, definida como:

$$x_1(t) = va(t)$$
$$x_2(t) = va(t-1)$$
$$x_3(t) = va(t-2)$$
$$x_4(t) = va(t-3)$$
$$x_5(t) = va(t-4)$$
$$x_6(t) = va(t-5)$$

E, a terceira opção foi definida como a combinação de dois períodos curtos, um relativo a um passado recente e outro relativo ao mesmo período do ano anterior. Desta maneira, pode-se levar em consideração não somente as informações mensais mas também as informações das variações anuais. Sendo então definida como:

$$x_1(t) = va(t)$$

$$x_2(t) = va(t-1)$$

$$x_3(t) = va(t-2)$$

$$x_4(t) = va(t-10)$$

$$x_5(t) = va(t-11)$$

$$x_6(t) = va(t-12)$$

O resultado destas três opções de espaço de entrada para o modelo SONARX-RBF está na Tabela 7.4. É possível observar que a terceira opção, apresentou o melhor resultado em comparação com as outras duas.

Depois disto, foi definido e avaliado o "embedding" ótimo de acordo com a proposta de "*Dynamic Modelling*" deste trabalho de pesquisa. Com base nos resultados dos Capítulos 4 e 5, a série de Vazões Afluentes foi decomposta nas séries de Amplitude (Envelope) e Ângulo de Fase, pela representação da "*Analytic Signal*". Logo, o modelo de previsão SONARX-RBF foi treinado duas vezes com diferentes "*embedding*": os vetores de "*time delay*" para a série de Amplitude e para a série de Ângulo de Fase foram, respectivamente:

$$\vec{y}^{A}(n) = [A(t)\,,\, A(t-4)\,,\, A(t-8)\,,\, A(t-12)]$$

e

$$\vec{y}^{\phi}(n) = [\phi(t)\,,\, \phi(t-8)\,,\, \phi(t-16)\,,\, \phi(t-24)]$$

Após prever a Amplitude e o Ângulo de Fase, os dados da série de Vazões Afluentes poderam ser recuperados pela Equação 4.10 do Capítulo 4, reescrita como:

$$VA[t+T] = A[t+T] \cdot e^{j\phi[t+T]}$$

As Figuras 7.24, 7.25 e 7.26 apresentam os resultados das previsões do Envelope, Ângulo de Fase e Vazões Afluentes para os dados de teste da UHE de Furnas: Jan/2000-Dez/2006, obtidas pelo modelo SONARX-RBF.

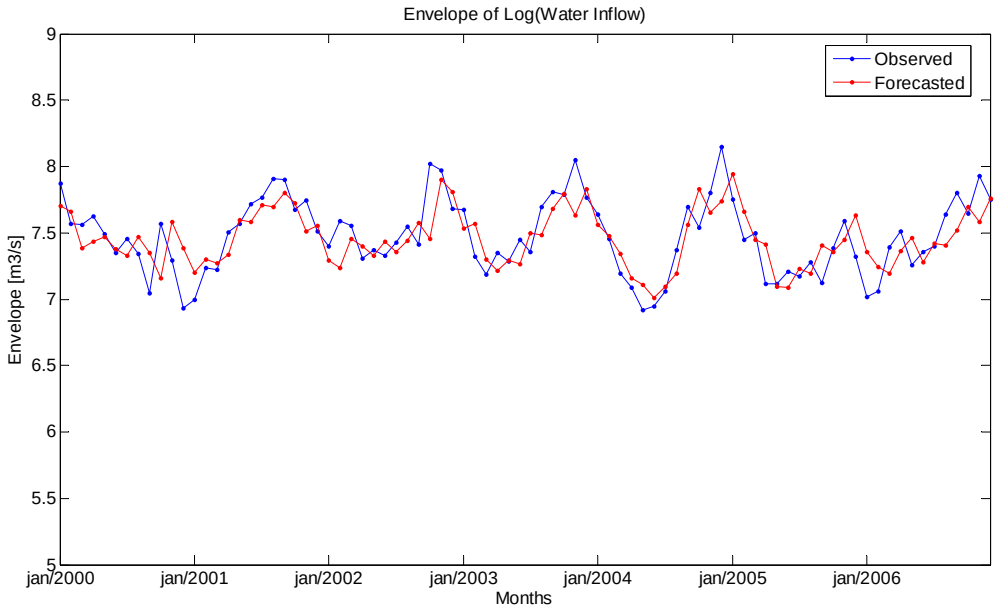


Figura 7.24 – Previsão do Envelope do Log(Vazão Afluente) da UHE de Furnas

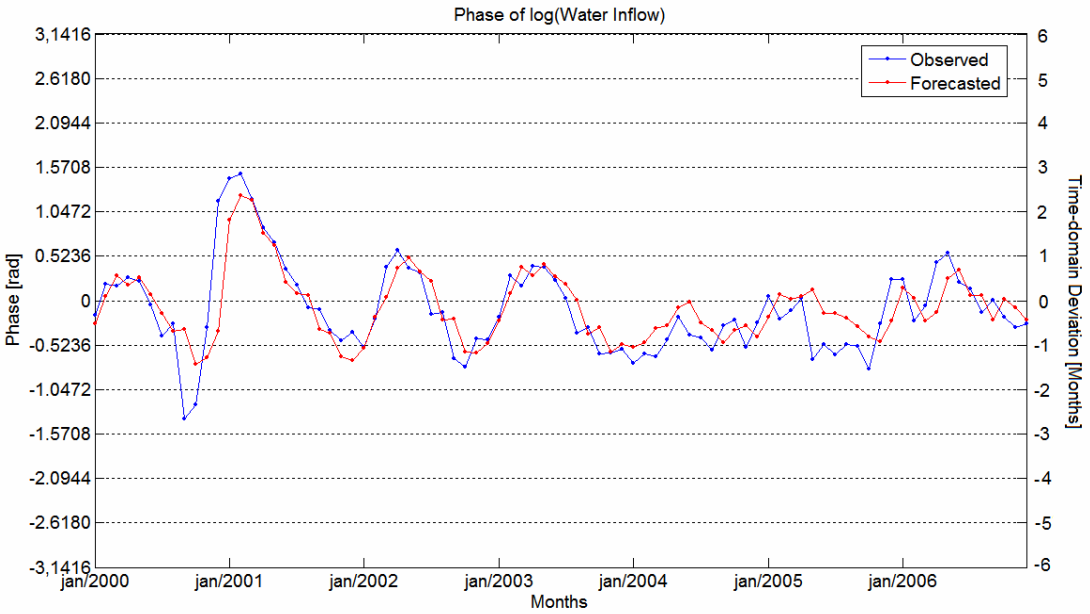


Figura 7.25 – Previsão do Ângulo de Fase do Log(Vazão Afluente) da UHE de Furnas

A Figura 7.24 mostra os resultados de previsão do Envelope para o período de teste. A linha azul corresponde às amostras observadas, e a linha vermelha representa os valores previstos do Envelope da série de Vazões Afluentes.

Os resultados de previsão do Ângulo de Fase estão na Figura 7.25, e a linha azul também corresponde às amostras observadas, e a linha vermelha representa os valores previstos do Ângulo de Fase. Como observado, o modelo SONARX-RBF foi capaz de seguir o comportamento observado de ambas as séries.

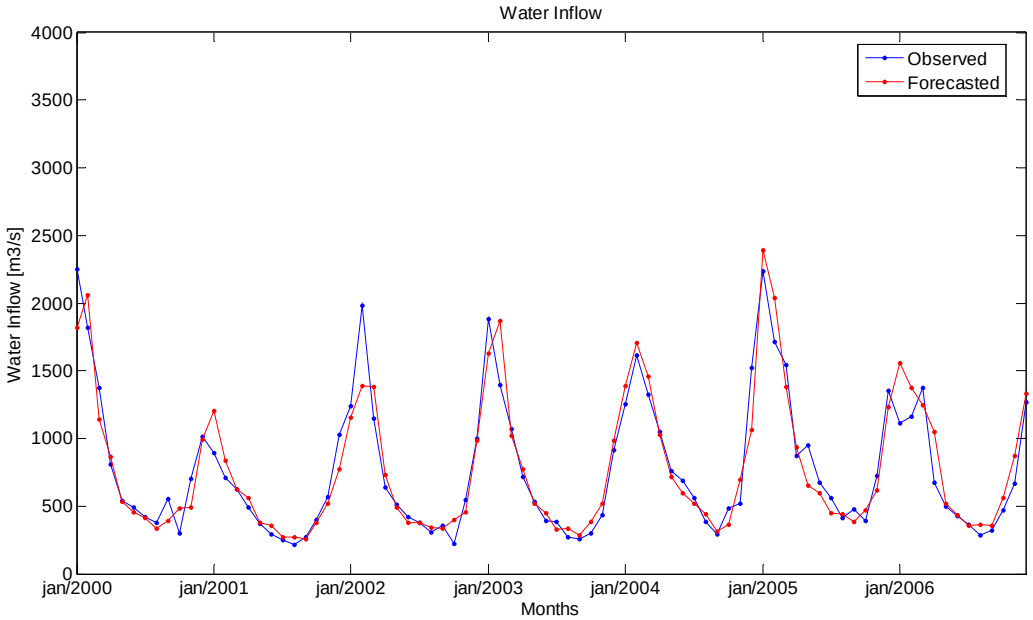


Figura 7.26 – Previsão de Vazão Afluente da UHE de Furnas

E a Figura 7.26 mostra os resultados de previsão das Vazões Afluentes para o período de 2000-2006. O modelo SONARX-RBF apresentou um desempenho bastante aderente, onde é possível ver uma grande similaridade entre os dados observados, em azul, e os valores previstos, em vermelho.

Como mencionado anteriormente, a Tabela 7.4 apresenta os erros de previsão obtidos pelo modelo SONARX-RBF usando três "*embedding*" arbitrários. Além disto, a Tabela 7.4 apresenta a performance da abordagem proposta neste trabalho.

Tabela 7.4 – Erros de previsão de Vazões Afluentes da UHE de Furnas

Approach	*Embedding*	*MSE (x10^4)*	*RMSE*	*MAD*	*MPE (%)*
SONARX-RBF	t,t-1,t-2	7,98	282,50	204,85	30,37
	t,t-1,t-2,t-3,t-4,t-5	6,54	255,69	180,34	26,22
	t,t-1,t-2,t-10,t-11,t-12	5,93	243,47	162,23	21,59
"AS + AMI + FNN" SONARX-RBF	Envelope: t,t-4,t-8,t-12	2,94	171,53	118,78	15,62
	Phase: t,t-8,t-16,t-24				

É possível observar que a abordagem proposta significativamente melhora o desempenho do modelo de previsão, não apenas por conta do pré-processamento via "*Analytic Signal*", mas também devido a definição automática do "*embedding*" usando "*Average Mutual Information*" e "*False Nearest Neighbors*".

Outra informação importante para o Planejamento da Operação é a previsão das VA vários-passos-a-frente. A abordagem proposta também é capaz de fornecer tais previsões.

A Figura 7.27 apresenta os valores das Vazões Afluentes previstos para os próximos doze meses para o mesmo período de teste: 2000-2008, começando em Janeiro e terminado em Dezembro de cada ano. É possível ver que a previsão de vários-passos-a-frente apresentou uma performance muito boa, e não apresentou degradação na previsão.

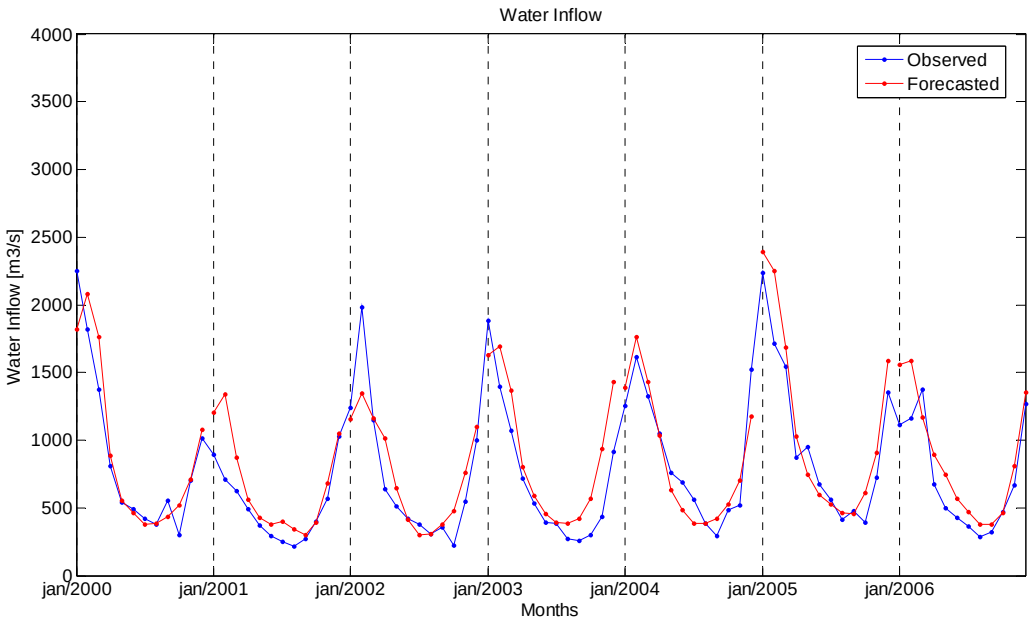


Figura 7.27 – Previsão de VA da UHE de Furnas: vários-passos-a-frente, 12 meses

Como no Brasil há um planejamento da operação pluri-anual, esta informação de previsão anual é muito importante. A Figura 7.28 apresenta o resultado das previsões médias anuais em comparação aos dados anuais médios observados.

Como é possível ver, o resultado na Figura 7.28 mostra que a abordagem proposta é muito útil para fornecer uma boa idéia do que se espera para o ano todo, relativo ao montante total disponível do recurso hídrico para o sistema hidrelétrico.

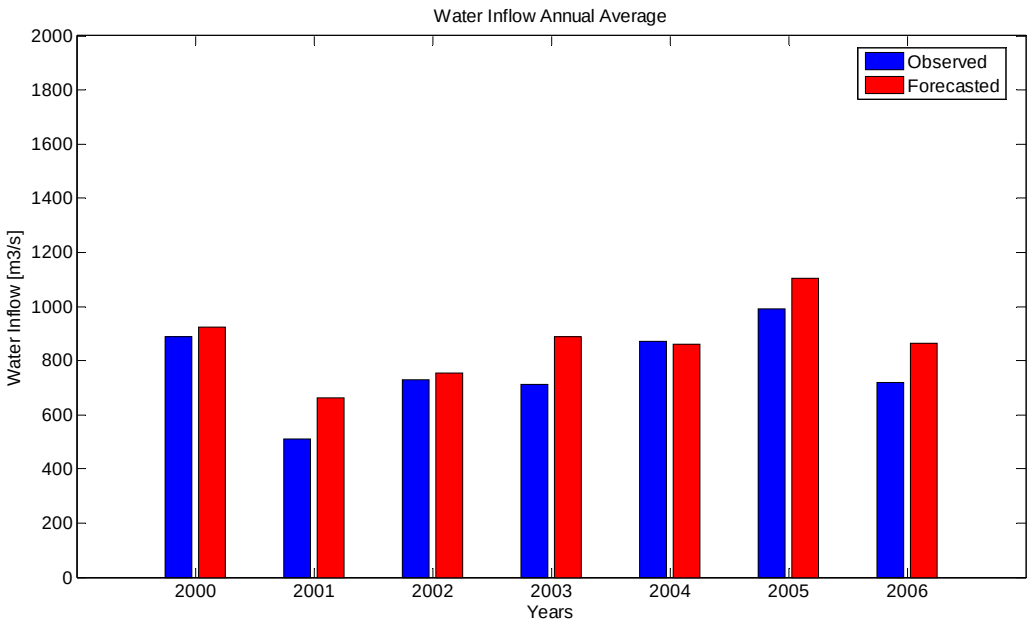


Figura 7.28 – Previsões anuais de VA à UHE de Furnas

A Tabela 7.5 apresenta os erros de previsão anuais obtidos pelo previsor de vários-passos-a-frente. Estes resultados ilustram a grande habilidade da "*Dynamic Modelling*" para tratar desta série temporal não-linear extremamente complexa.

Tabela 7.5 – Erros de Previsão Anuais da UHE de Furnas

Approach	*Embedding*	*MSE (x10^4)*	*RMSE*	*MAD*	*MPE (%)*
"AS + AMI + FNN" SONARX-RBF	Envelope: t,t-4,t-8,t-12 Phase: t,t-8,t-16,t-24	1,28	113,28	94,01	13,53

Capítulo 8

Política de Operação POLFC com Termo Inercial

Os resultados da comparação de desempenho entre PDE e a POLFC feita por (Martinez e Soares Filho, 2002) mostrou maior geração hidrelétrica média para a POLFC. Apesar deste fato, o custo total de operação ficou maior, por causa da maior flutuação da geração térmica complementar, i.e. um maior desvio padrão.

A grande vantagem da Política POLFC é que ela pode ser aplicada numa representação individualizada do sistema hidrelétrico, preservando a informação de cada característica construtiva. Isto é possível porque esta abordagem utiliza um modelo de otimização determinístico, tratando o aspecto estocástico do problema separadamente através do modelo de previsão.

Com base nisto, o presente trabalho de pesquisa tratou de estabilizar a geração térmica complementar, e conseqüentemente, reduziu o custo total de operação para o mais próximo possível da operação ótima absoluta. A solução mais estável foi obtida usando dois modelos de otimização não-lineares determinísticos, com representação individual de cada UHE, e aplicando o Termo Inercial (TI), responsável por estabilizar a geração termelétrica. Através de experimentos realizados para vários valores, decidiu-se utilizar o termo inercial it = 0,8.

Para efeito de comparação do desempenho da política operativa proposta por este trabalho de pesquisa (POLFC Estabilizada via Termo Inercial), foram também implementadas e testadas: a Política de Operação Ingênua (*Naive Feedback Control* – NFC, em inglês); a

Política de Operação POLFC original; e a Operação Ótima Determinística – considerada a solução ideal.

Para a realização de todos estes estudos, foi necessária a implementação de duas Modelagens Matemáticas de Otimização Não-Linear, através do Software LINGO®, além da implementação de um Simulador de Operação, através do Software MatLab®. Com isto, o trabalho de desenvolvimento das modelagens por prototipagem ficou bastante facilitado e eficiente.

8.1 O Sistema Hidrelétrico de Teste

O sistema teste utilizado é composto por 7 (sete) importantes usinas do Sistema Sudeste Brasileiro: as usinas de Emborcação, Itumbiara e São Simão, localizadas ao longo do rio Paranaíba; enquanto as usinas de Furnas, Marimbondo e Água Vermelha, localizadas ao longo do rio Grande; e finalmente a usina de Ilha Solteira, localizada no rio Paraná. Este sistema está interligado em cascata formando um Y como mostra a Figura 8.1.

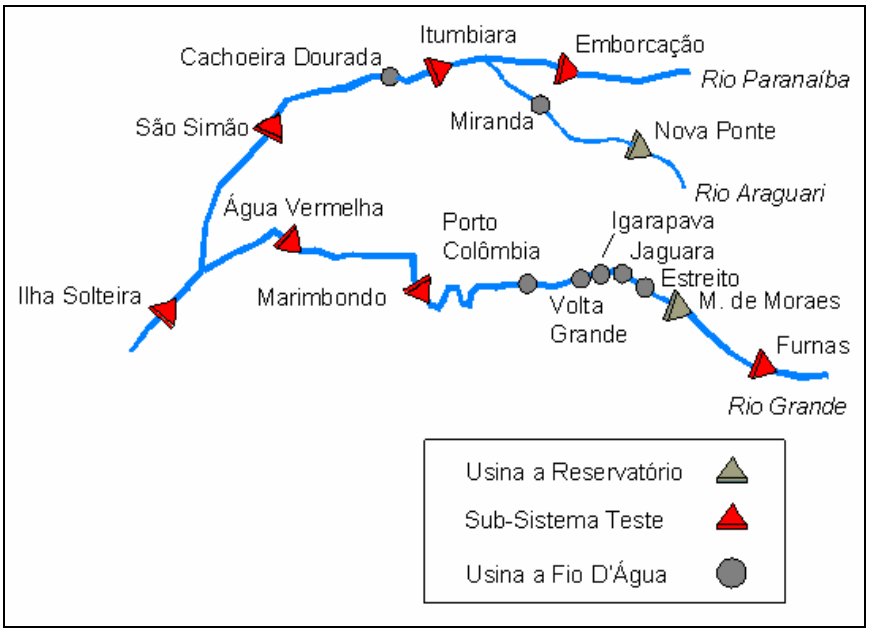


Figura 8.1 - Sistema teste, parte da Bacia do Rio Paraná

Este sistema, com uma potência instalada de 12.572 MW, apresenta uma configuração muito interessante para o estudo em questão, tendo em vista a interligação hidráulica, com conseqüente complexidade operativa das usinas quando analisadas em conjunto.

8.1.1 Solução da Otimização Determinística para o Histórico de VA

Durante a otimização determinística utilizou-se uma demanda de 12.570 MW, que corresponde a 100% da potência instalada deste parque hidrelétrico. Este montante de demanda foi assim definido para que o parque termelétrico complementar pudesse atuar gerando uma certa quantidade de energia, a qual deveria ser minimizada neste processo de otimização, como pode ser visto na Figura 8.3, enquanto na Figura 8.2 pode ser observada a queda do custo de operação durante este processo.

A otimização deste sistema define de forma determinística a operação ótima do sistema, já que são conhecidas as vazões afluentes desde janeiro de 1931 até dezembro de 1994.

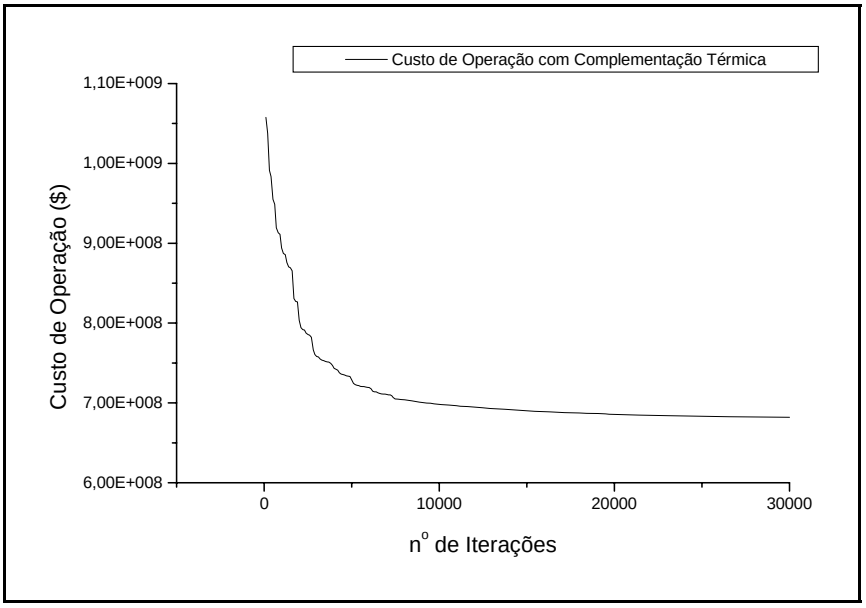


Figura 8.2 - Evolução do custo de operação durante o processo de otimização

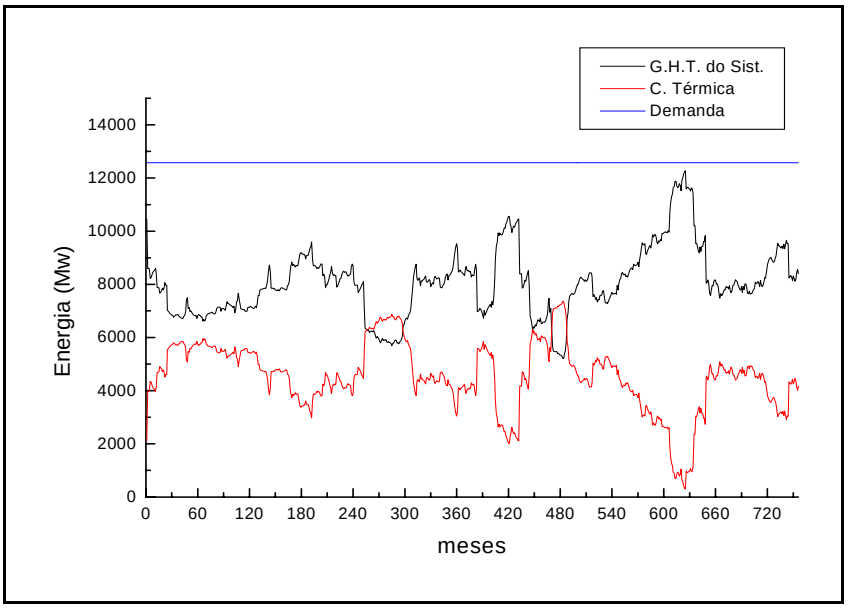


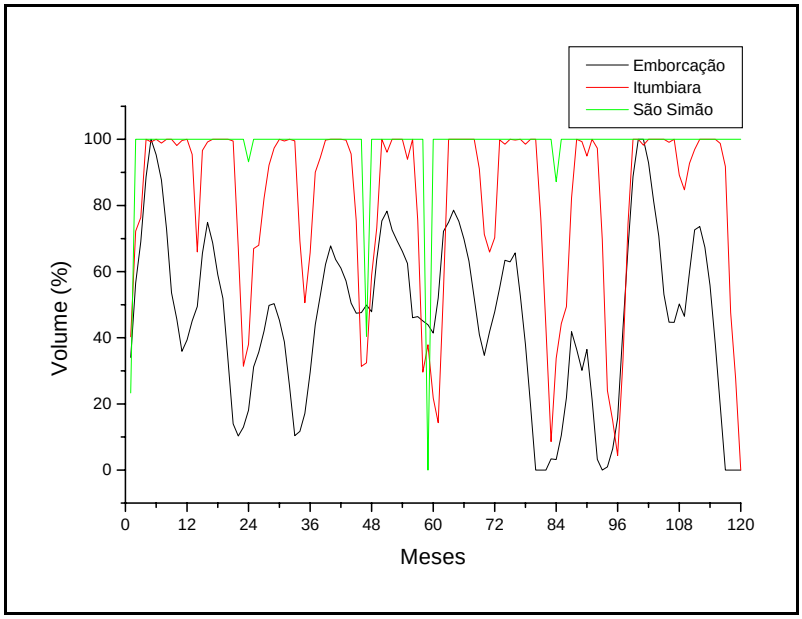
Figura 8.3 - Trajetória de mercado ao longo do período de otimização

Visto que o ano hidrológico, definido como o período compreendido entre o início da estação seca até o final da estação chuvosa, na bacia hidrográfica do rio Paraná começa em maio e termina em abril do ano seguinte, tivemos que fazer a otimização, levando em consideração somente o período compreendido entre maio de 1931 até abril de 1994, totalizando 63 anos de otimização, o que corresponde a 756 meses. Posteriormente, este resultado deverá ser pré-processado e servir como conjunto de padrões para o "treinamento" dos Sistemas Inteligentes (SI).

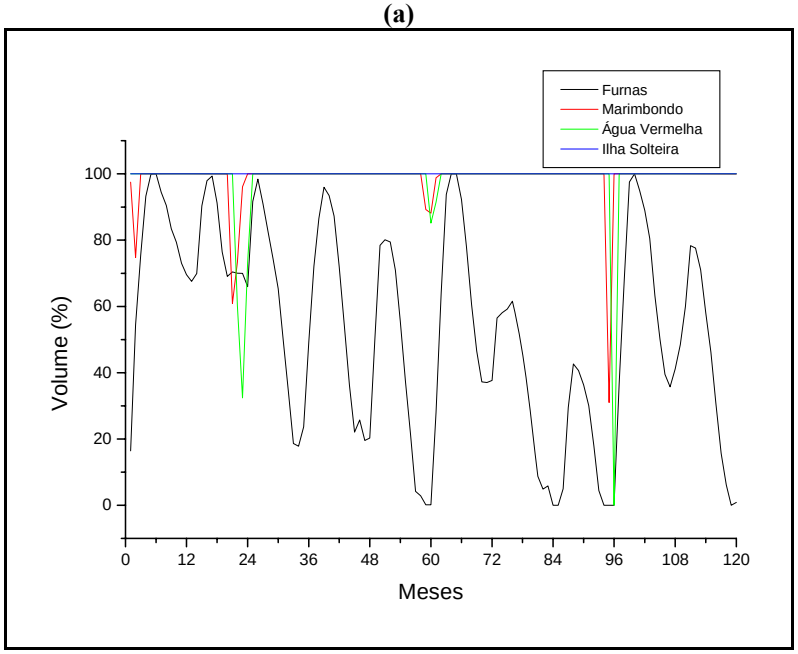
Como os SI dependem de dados do passado para fazer suas previsões futuras, houve a necessidade da exclusão de alguns anos iniciais e finais deste conjunto de dados. Desta maneira, tivemos um conjunto de padrões para treinamento e teste, correspondente ao período histórico de janeiro de 1935 até dezembro de 1992, totalizando 58 anos de otimização, o que corresponde a 696 meses. Como pode ser observado da Figura 8.4 até a Figura 8.9.

As usinas em cascata têm a tendência de valorizar a água, de tal forma que o processo de deplecionamento é feito de montante à jusante, e o processo de enchimento de jusante à montante. Com isso, as usinas de jusante tendem a permanecer como usinas a fio d'água, sem alterações significativas nos seus volumes, cabendo aos reservatórios de cabeceira a regulação

das vazões afluentes ao sistema, o que faz com que, a produtividade das usinas permaneça a maior possível.



(a)

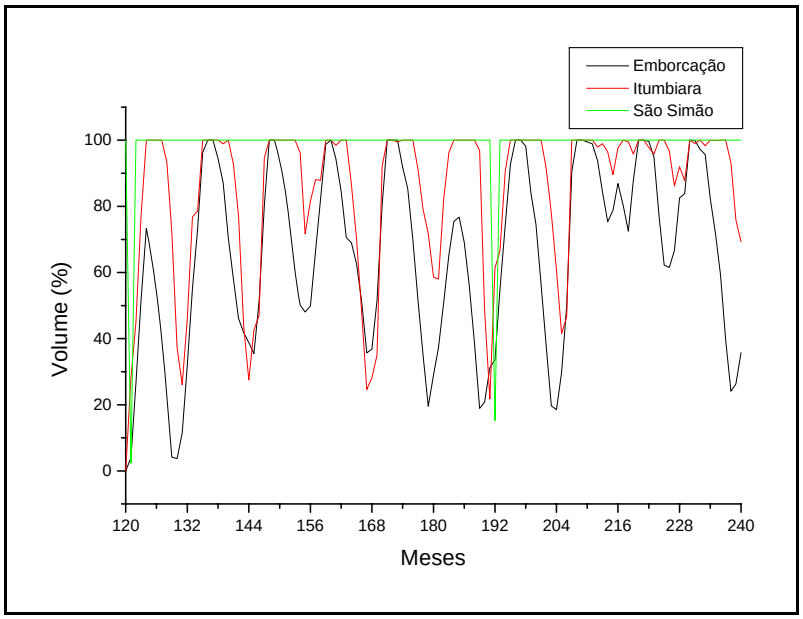


(b)

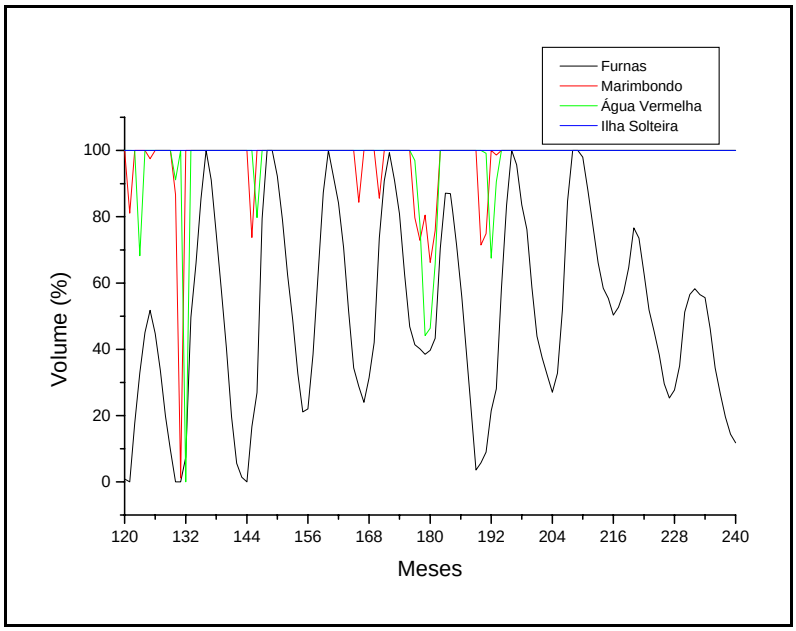
Figura 8.4 - Trajetória dos volumes ótimos para o período de Jan/1935 até Dez/1944

Como pode ser visto nestas figuras, que descrevem as trajetórias operativas ótimas, no caso da usina de Emborcação, definida como a trajetória de cor preta em (a), seu reservatório é praticamente todo utilizado ano após anos, na tarefa de regularizar as vazões afluentes à usina de jusante Itumbiara. Mas como o seu reservatório não é grande o suficiente para isto, a

usina de Itumbiara é "requisitada" a auxiliar neste papel regulador. Observando a trajetória de cor vermelha em (a), nota-se que o reservatório de Itumbiara depleciona seu volume em praticamente todos os anos, tão logo o reservatório de Emborcação esteja no mínimo operativo ou quase no mínimo.

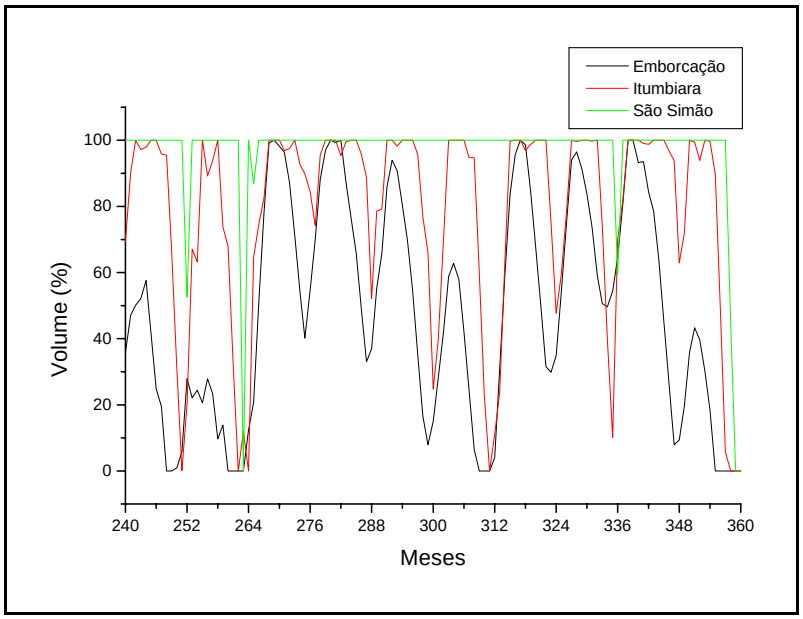


(a)

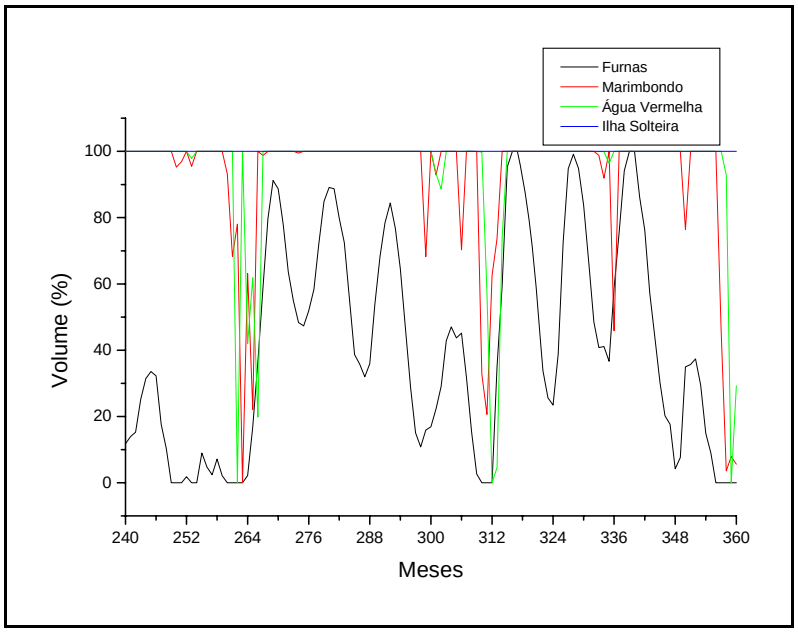


(b)

Figura 8.5 - Trajetória dos volumes ótimos para o período de Jan/1945 até Dez/1954



(a)

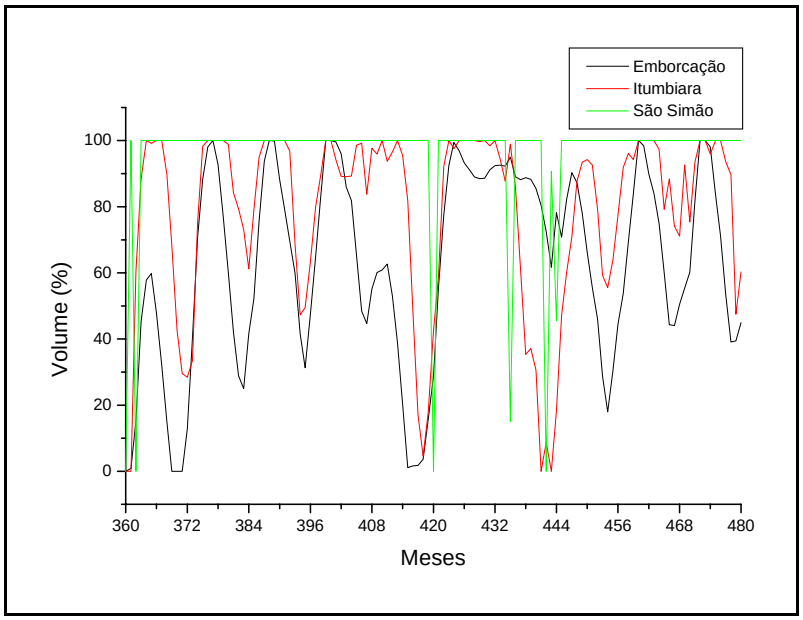


(b)

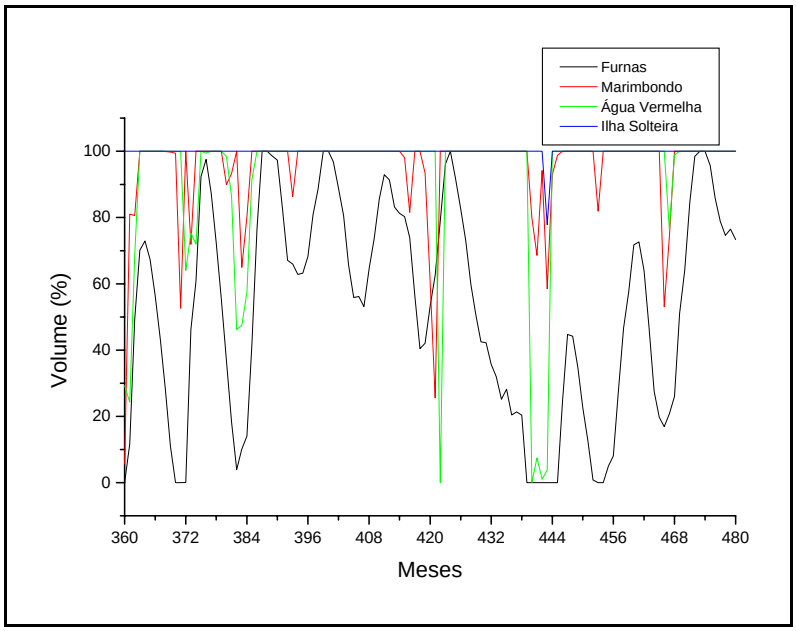
Figura 8.6 - Trajetória dos volumes ótimos para o período de Jan/1955 até Dez/1964

Ainda com relação às 3 (três) usinas do rio Paranaíba, logo abaixo tem-se a usina de São Simão, que permaneceu quase que o tempo todo operando a fio d'água, como pode ser observado na trajetória de cor verde em (a). Esta decisão operativa, como já foi discutido, busca preservar altura de queda desta usina, mantendo uma maior produtividade, o que corresponde a uma eficiência sistêmica, já que toda a água defluida por Emborcação e

Itumbiara invariavelmente passa por São Simão. Entretanto, nas ocasiões que São Simão deplecionou seu reservatório, ela o fez antevendo um período chuvoso à sua frente, evitando assim futuros vertimentos, o que corresponde a um desperdício energético muito grande.

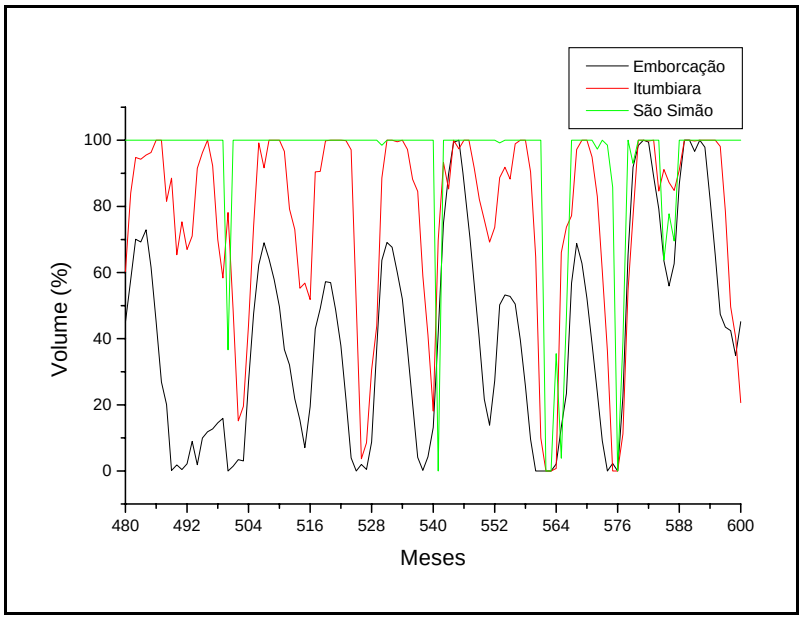


(a)

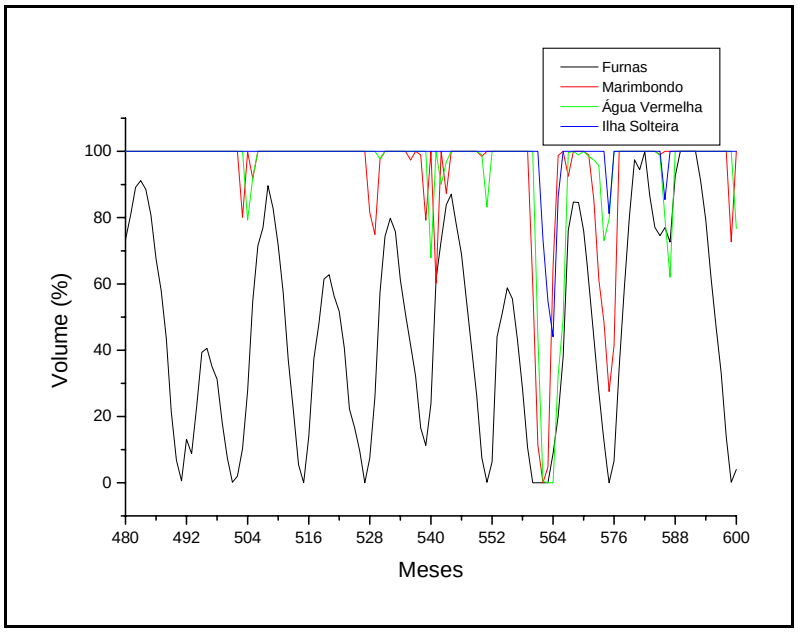


(b)

Figura 8.7 - Trajetória dos volumes ótimos para o período de Jan/1965 até Dez/1974



(a)

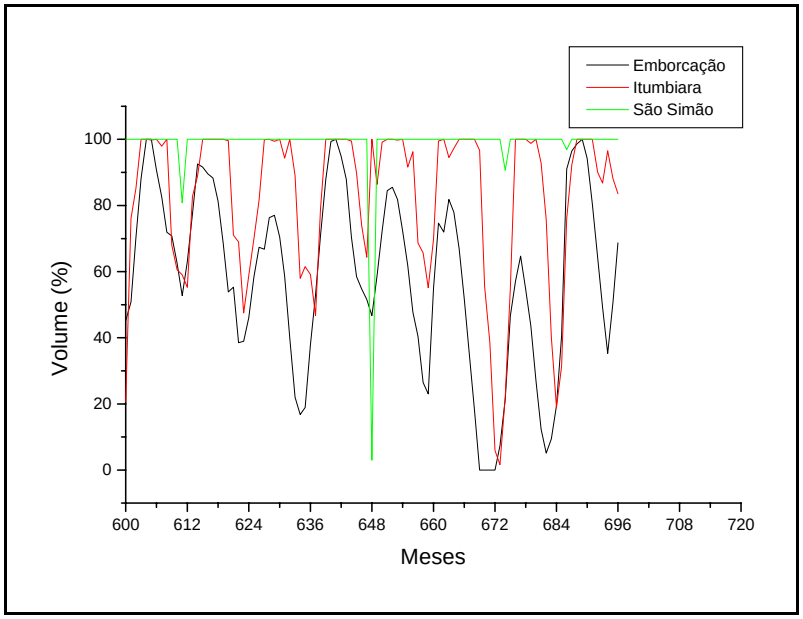


(b)

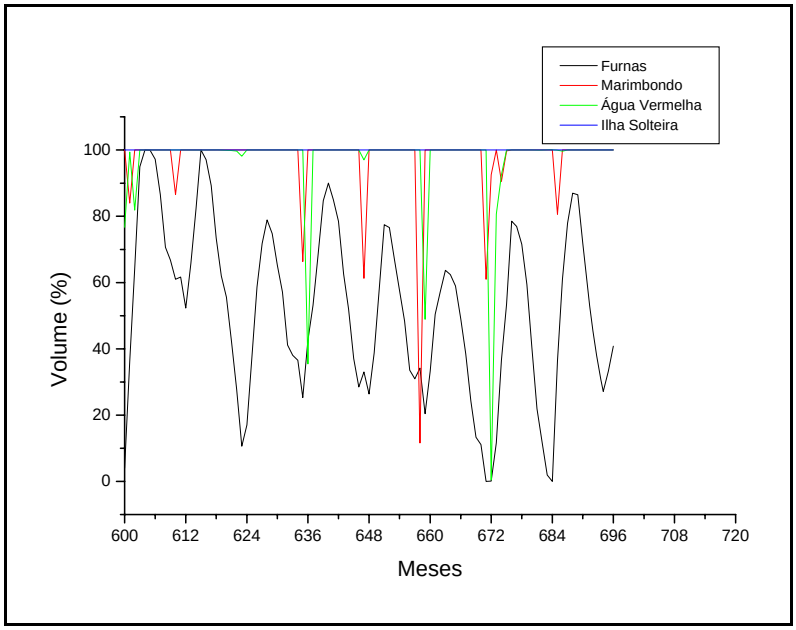
Figura 8.8 - Trajetória dos volumes ótimos para o período de Jan/1975 até Dez/1984

Com relação às 3 (três) usinas pertencentes ao rio Grande, a usina de cabeceira, Furnas, também exerce o papel regulador desta cascata. Como o seu reservatório é bem grande, ela consegue regularizar boa parte das vazões afluentes, ficando para Marimbondo deplecionar e regularizar as vazões em poucos meses do período de otimização, como pode ser visto nas trajetórias de cor preta e vermelha em (b), respectivamente. Já a usina de Água

Vermelha, permanece quase que o tempo todo operando a fio d'água, assim como São Simão, segundo a trajetória de cor verde em (b).



(a)



(b)

Figura 8.9 - Trajetória dos volumes ótimos para o período de Jan/1985 até Dez/1992

Finalmente temos a usina de Ilha Solteira, pertencente ao rio Paraná, tida com a usina mais a jusante do sistema teste, por isso e pelas razões já comentadas de eficiência energética,

Ilha Solteira teve uma trajetória de operação ótima, quase que totalmente "flat", com pouquíssimos meses em que isto não ocorreu.

Uma outra característica observada foi a maneira como as usinas se comportam para manterem a máxima produtividade possível. As usinas utilizam somente a parcela de água dos reservatórios possível de ser recuperada nos próximos períodos chuvosos, e ainda evitando a ocorrência de vertimentos. Assim, nos períodos de menor vazão afluente, menor é o deplecionamento dos reservatórios das usinas, isto porque, com o conhecimento da pouca afluência futura, o sistema prefere a utilização das térmicas, poupando água para o enchimento futuro dos reservatórios para a obtenção de produtividade.

Assim, a partir destes resultados, observou-se certos comportamentos para os reservatórios das usinas de acordo com a sua posição na cascata, comprovando o que já foi dito, de que os reservatórios das usinas possuem comportamentos diferenciados e estes devem ser levados em consideração durante a operação.

Os resultados confirmam o papel que cada tipo de usina possui no sistema, de acordo com a sua localização e sua importância. As usinas de montante com incumbência de absorver a sazonalidade das afluências e da demanda, são as que mais oscilam, variando o seu volume. As usinas de jusante, com o papel de manter a máxima produtividade, oscilam o menos possível o seu volume. E as usinas intermediárias possuem um comportamento misto entre os dois tipos citados anteriormente.

8.2 POLFC-IT aplicado a uma Usina Hidrelétrica: UHE de Furnas

Nesta etapa do trabalho de pesquisa optou-se inicialmente pela operação de apenas uma Usina Hidrelétrica (UHE) isolada, somada a um parque termelétrico complementar para o atendimento ao consumo de energia elétrica. Opção esta para facilitar a implementação e

melhor sensibilidade da metodologia proposta por parte do Pesquisador. Para tanto, foi escolhida a UHE de Furnas, pertencente à Bacia do Rio Paraná – Brasil, para o estudo de caso, um reservatório de montante e de grande capacidade de regularização hidrológica.

Com isto, a Geração Termelétrica total e a Geração Hidrelétrica total para o próximo mês *GT(t)* e *GH(t)*, respectivamente, foram adotadas como as variáveis de decisão do primeiro modelo de otimização não-linear determinístico, com representação individualizada de cada Usina Hidrelétrica (UHE). Este modelo foi proposto com base nos modelos apresentados por (Read, 1982; Lyra, Tavares e Soares Filho, 1984; Carvalho e Soares Filho, 1987).

Com relação a seqüência de Vazões Afluentes Previstas, optou-se inicialmente em trabalhar com a seqüência do histórico de VA ruidoso, através da adição de um ruído gaussiano que representaria o Erro de Previsão. Com base no trabalho (Ballini, Soares Filho e Andrade Filho, 2001), onde foram realizadas previsões de vários-passos-a-frente de VA, observou-se que as RNA seriam capazes de prever com 1 passo-a-frente com erro em torno de 20%, enquanto que com 12 passos-a-frente com erro em torno de 30%. Conseqüentemente, estes foram os ruídos então adicionados ao histórico de VA, sendo utilizado como VA Prevista no Simulador da Operação.

É justamente a diferença entre as Vazões Previstas e Verificadas que causa a sub-otimalidade deste tipo de Política, como a POLFC. Esta sub-otimalidade se expressa em termos do comportamento operativo através da instabilidade resultante da Geração Termelétrica, conforme comentada em (Martinez e Soares Filho, 2002).

Voltando ao primeiro estudo de caso, da UHE de Furnas operando isoladamente que conta apenas com a complementação da Geração Termelétrica para atender toda a Demanda, os resultados operativos das três Políticas de Operação aqui avaliadas, além da solução ótima

determinística que é tida como a solução ideal, são apresentados de forma gráfica pelas Figuras 8.10 a 8.13.

A Figura 8.10 apresenta o comportamento operativo do Volume Armazenado durante a operação simulada para o ciclo hidrológico de maio/1953 a abril/1954. Seu volume inicial foi estabelecido como cheio. Nela é possível observar que para a solução ótima a UHE de Furnas deve ter seu armazenamento deplecionado até o mês 7 (Novembro/1953) – final do período seco; a partir do mês 8 (Dezembro/1953) até o mês 12 (Abril/1954) – período chuvoso, a UHE deve ter seu armazenamento reenchido.

Vale observar que ao final do reenchimento do reservatório, no final do ciclo hidrológico, o volume armazenado é exatamente o de volume máximo armazenável. Isto só foi possível porque o otimizador determinístico, sabendo da Vazão Afluente total de todo o período chuvoso, permitiu o deplecionamento durante o período seco suficiente para uma completa recuperação posterior de seu armazenamento.

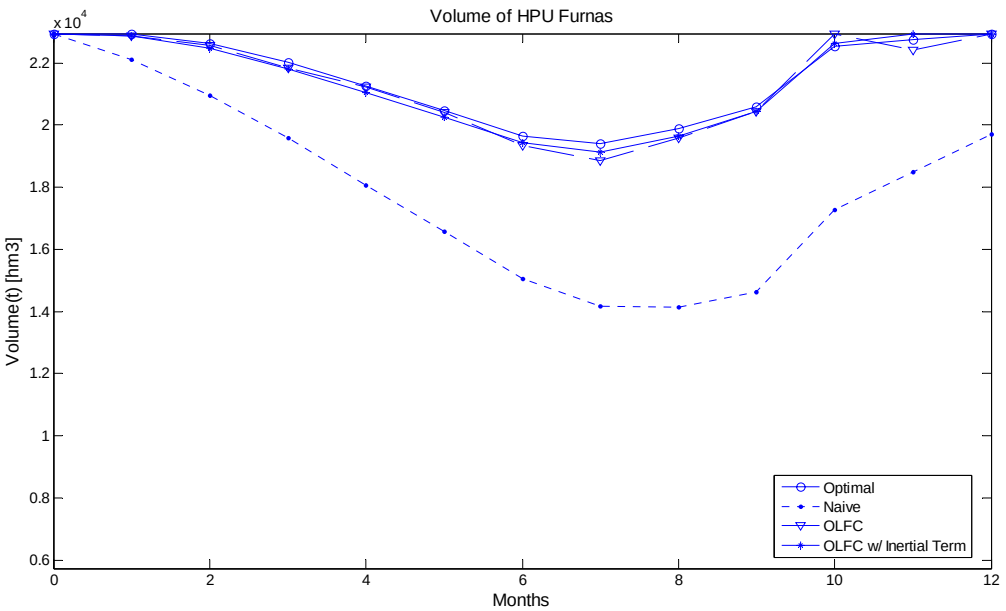


Figura 8.10 - Trajetória de Volumes Armazenados da UHE de Furnas: Maio/1953 a Abril/1954

1 UHE Isolada

A Política de Operação Ingênua (Naive Feedback Control - NFC) apresentou um comportamento diferente. Nela, a UHE de Furnas teve um deplecionamento muito maior, o que não permitiu a total recuperação do armazenamento ao final do perído chuvoso, no mês 12 (Abril/1954). Isto se deve ao fato desta política estar sempre na espectativa de futuras VA médias, ou seja iguais a Média de Longo Termo (MLT), e nestes anos de 1953-54, as VA foram bem abaixo da média, levando à decisões equivocadas e não-ótimas.

Já as Políticas de Controle de Malha Parcialmente Aberta (POLFC) e POLFC estabilizada via Termo Inercial (POLFC-IT) apresentaram um comportamento muito próximo do ideal, como pode ser visto ainda na Figura 8.10. Isto porque ambas as políticas trabalham com um Modelo de Previsão de VA, na tentativa de prever como será o período chuvoso enquanto definem os níveis de deplecionamento necessários e que permitam a total recuperação ao final do ciclo hidrológico.

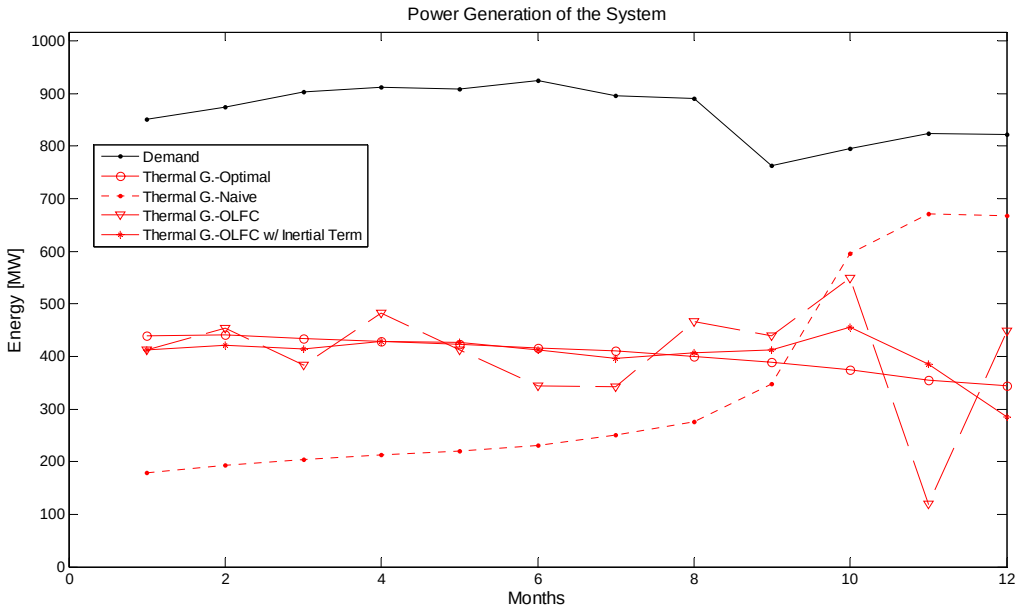


Figura 8.11 - – Trajetórias das Gerações Termelétricas e da Demanda: Maio/1953 a Abril/1954
1 UHE Isolada

Todo este comportamento de variação do volume armazenado durante o ciclo hidrológico é necessário para proporcionar uma Geração Termelétrica o mais constante

possível ao longo do período de planejamento (Read, 1982), como pode ser observado na Geração Termelétrica da solução ótima determinística da Figura 8.11.

A Política de Operação Ingênua (Naive) apresentou um comportamento bem diferente. Nos primeiros meses, ela apresenta uma Geração Termelétrica muito abaixo da ideal para este ciclo hidrológico de 1953-54, abaixo da média, fato que acabou comprometendo a disponibilidade de hidreletricidade para o período chuvoso, o que obrigou um grande aumento da Geração Termelétrica para o período chuvoso, na tentativa de preservar o parque Hidrelétrico.

Já a Política POLFC, na Figura 8.11, apresenta um nível de Geração Termelétrica mais parecido com a solução ótima, devido às previsões de VA. Entretanto, ela apresenta uma grande flutuação, ou seja, uma grande instabilidade da Geração Termelétrica em torno da solução ótima praticamente constante ao longo do horizonte de planejamento. Esta instabilidade é causada pelos erros de previsão das VA.

Foi justamente, com o objetivo de mitigar esta instabilidade e suas conseqüências operativas e econômicas que este trabalho introduziu a utilização do Termo Inercial na Política POLFC. Como pode ser visto na Figura 8.11, a Política POLFC-IT apresentou um comportamento bem próximo da solução ideal, muito mais estável que as outras duas políticas de operação, o que demonstra o sucesso desta abordagem aqui proposta.

Já a Figura 8.12 apresenta o comportamento das Gerações Hidrelétricas da UHE de Furnas para o ciclo hidrológico de 1953-54. Nela é possível notar que para a solução ótima determinística, a Geração Hidrelétrica deve acompanhar a variabilidade sazonal de Demanda energética, uma vez que pela Figura 8.13 a Geração Termelétrica deve ser quase que constante.

Comportamento este que não foi apresentado pela Política Ingênua (Naive), na Figura 8.13. Como conseqüência dá expectativa por futuras VA iguais a MLT, o despacho de

hidreletricidade não foi poupado nos primeiros meses como o deveria ter sido. Logo, no período chuvoso foi necessário diminuir bastante a Geração da UHE de Furnas na tentativa de recuperar o armazenamento de seu reservatório.

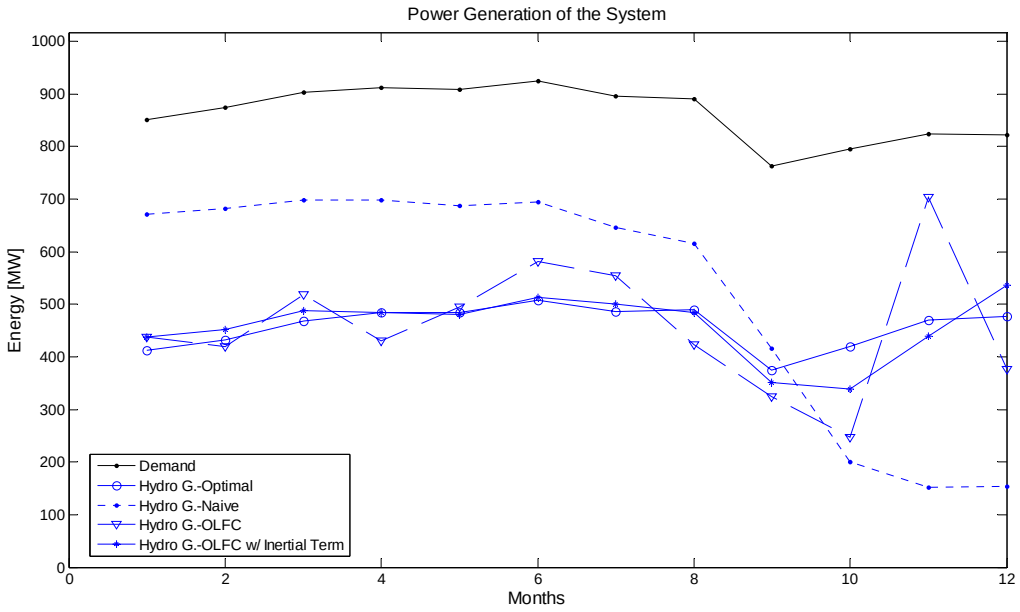


Figura 8.12 - Trajetórias das Gerações Hidrelétricas e da Demanda: Maio/1953 a Abril/1954
1 UHE Isolada

A Política POLFC apresentou, na Figura 8.12, também uma grande instabilidade da Geração Hidrelétrica em torno da solução ótima como conseqüência do comportamento da complementação Termelétrica da Figura 8.11.

Já a Política POLFC-IT, na Figura 8.12, apresentou conseqüentemente um comportamento bem próximo da solução ideal, muito mais estável que as outras duas políticas de operação e acompanhando o comportamento sazonal da Demanda.

Como conseqüência das trajetórias da Figura 8.11, a Figura 8.13 apresenta o comportamento do Custo Marginal de Operação (Preço Spot ou Preço de Curto Prazo), uma vez que o CMO é definido pelo despacho do Parque Termelétrico.

Na Figura 8.13 é possível observar um comportamento estável, quase que constante e ligeiramente descendente do CMO para a solução ótima determinística, devido ao Efeito

Queda. Comportamento este que propicia um ambiente de Mercado extremamente favorável e tranqüilo para que os Agentes de Mercado possam se posicionar de forma planejada e confiável.

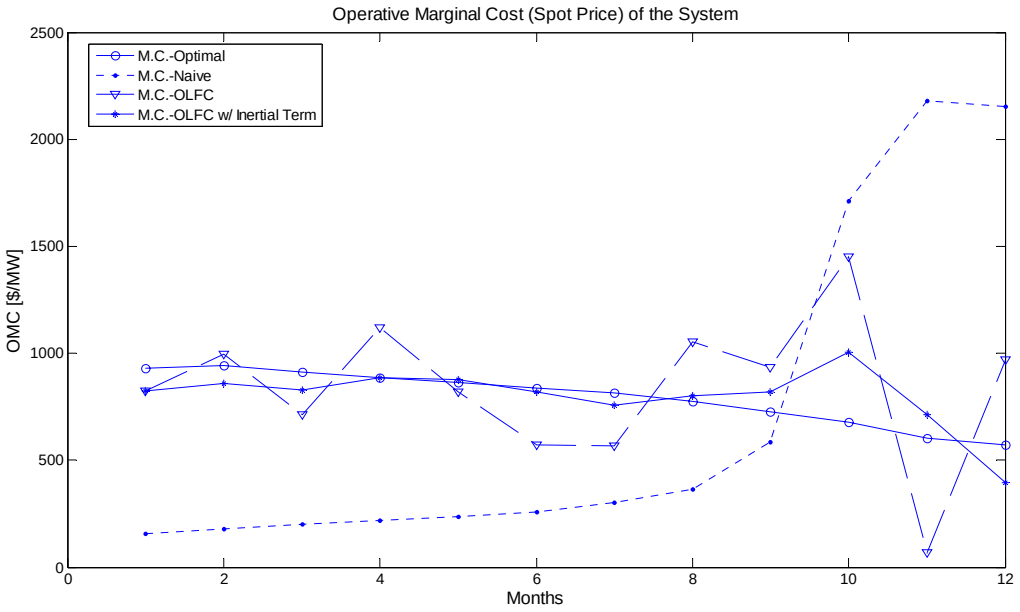


Figura 8.13 - Trajetórias dos Custos Marginais de Operação (CMO): Maio/1953 a Abril/1954
1 UHE Isolada

Estas características de estabilidade e constância não são observadas nem pela Política Ingênua (Naive) nem pela POLFC. Na Figura 8.13 pode-se notar que a Política Naive definiu valores de CMO iniciais bem menores do que o ideal, causando uma elevação abrupta do Preço da Energia de curto prazo no período chuvoso, uma vez que as VA médias esperadas não se confirmaram em 1953-54.

A Política POLFC apresentou um comportamento de extrema instabilidade para os valores de CMO, uma vez que os valores oscilaram muito em torno dos valores ideais. Estas oscilações são extremamente inconvenientes para um Mercado de Energia de curto prazo, como é o caso brasileiro. Estas grandes flutuações do CMO trazem desconfiança e insegurança para os Agentes de Mercado, prejudicando a consolidação deste tipo de modelo setorial.

Já a Política de Operação POLFC-IT foi capaz de estabilizar o comportamento do CMO, fazendo com que se aproximasse dos valores ideais. Com uma menor volatilidade nos Preços, os Agentes tendem a se sentirem mais confortáveis e seguros em suas decisões estratégicas de mercado, atraindo assim mais "*players*" e investidores para o Setor Elétrico Brasileiro.

Depois das apresentações gráficas dos resultados operativos das três Políticas de Operação, são apresentados na Tabela 8.1 alguns resultados numéricos. Nela é possível observar o desempenho operativo destas políticas em comparação a solução ideal ótima determinística.

Tabela 8.1 - Tabela comparativa entre as Políticas de Operação: Maio/1953 a Abril/1954
1 UHE Isolada

Política de Operação	G. Termelétrica [MW]		CMO [$/MWh]		C. de Operação	
	Média	Desvio Padrão	Média	Desvio Padrão	Total [$ milhões]	Desvio [%]
NAIVE	337,69	183,41	712,17	767,33	1.111	16,96
POLFC	404,74	102,17	840,35	329,69	1.089	14,65
POLFC c/ Termo Inercial	405,03	39,73	798,76	139,90	961	1,14
Ótimo Determinístico	404,70	31,45	794,61	119,79	950	----

Na Tabela 8.1, nota-se que a Política de Operação Ingênua apresentou a menor média para Geração Termelétrica e CMO, conseqüência do não re-enchimento total do reservatório da UHE de Furnas ao final do ciclo hidrológico, enquanto que as outras políticas o conseguiram.

Tabela 8.2 - Tabela Custo de Operação com relação ao Ótimo Determinístico: Maio/1953 a Abril/1954
1 UHE Isolada

Política de Operação	Custo de Operação [$ milhões]	Custo Pós-Operativo Ótimo [$ milhões]	Desvio [%]
NAIVE	1.111	513	116,48

Conseqüentemente, para uma melhor avaliação comparativa do seu desempenho operativo foi necessário avaliar o Custo Pós-Operativo Ótimo, considerando o estado final de armazenamento desta política. Para tanto, foi necessário executar uma nova otimização

determinística com a redefinição do volume final Vol_{FINAL} , cujo resultado pode ser observado na Tabela 8.2.

De volta a Tabela 8.1, tanto as Políticas de Operação Naive quanto a POLFC apresentaram grandes valores de Desvio Padrão para a Geração Termelétrica e para o CMO, o que não é nada conveniente para o Mercado de Energia como já foi comentado. O que acabou encarecendo o Custo de Operação destas políticas.

Já a Política POLFC-IT conseguiu apresentar tanto valores Médios como de Desvio Padrão bastante próximos do ideal ótimo determinístico. Logo, apresentou um Custo Operativo bastante baixo e próximo do ótimo.

Todos estes resultados mostram a eficácia e aplicabilidade da Política de Operação POLFC com Termo Inercial na solução do Problema de Planejamento da Operação Hidro-Térmico de Potência.

8.3 POLFC-IT aplicado a Subsistemas de Usinas Hidrelétricas

Num segundo momento, foi testada a operação conjunta das UHE de Furnas, Marimbondo e Água Vermelha, três das principais UHE do Rio Grande – Bacia do Rio Paraná e dispostas em cascata, somadas ao parque termelétrico complementar. Desta forma, foi possível analisar o comportamento operativo de um conjunto de UHE dispostas em série numa cascata.

8.3.1 Três UHE em Cascata: Furnas, Marimbondo e Água Vermelha

Os resultados serão apresentados de forma gráfica pelas Figuras 8.14 a 8.19, sendo que as Figuras 8.14, 8.15 e 8.16 mostram, respectivamente, o comportamento operativo do

Volume Armazenado das UHE de Furnas, Marimbondo e Água Vermelha durante a operação simulada para o ciclo hidrológico de maio/1953 a abril/1954. Novamente, seus volumes iniciais foram estabelecidos como cheios.

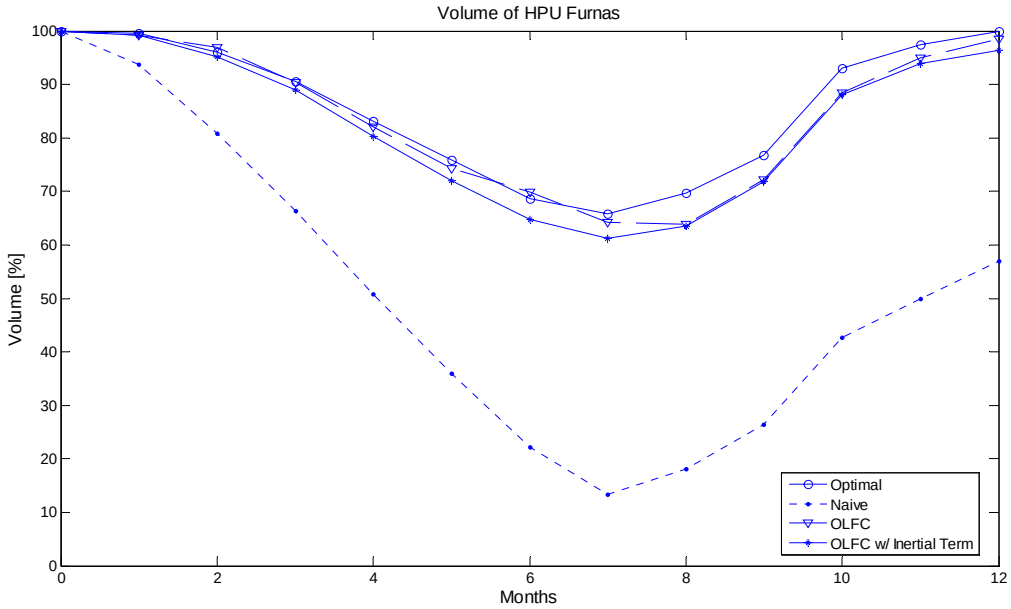


Figura 8.14 - Trajetória de Volumes Armazenados da UHE de Furnas: Maio/1953 a Abril/1954
3 UHE: Furnas, Marimbondo e Água Vermelha

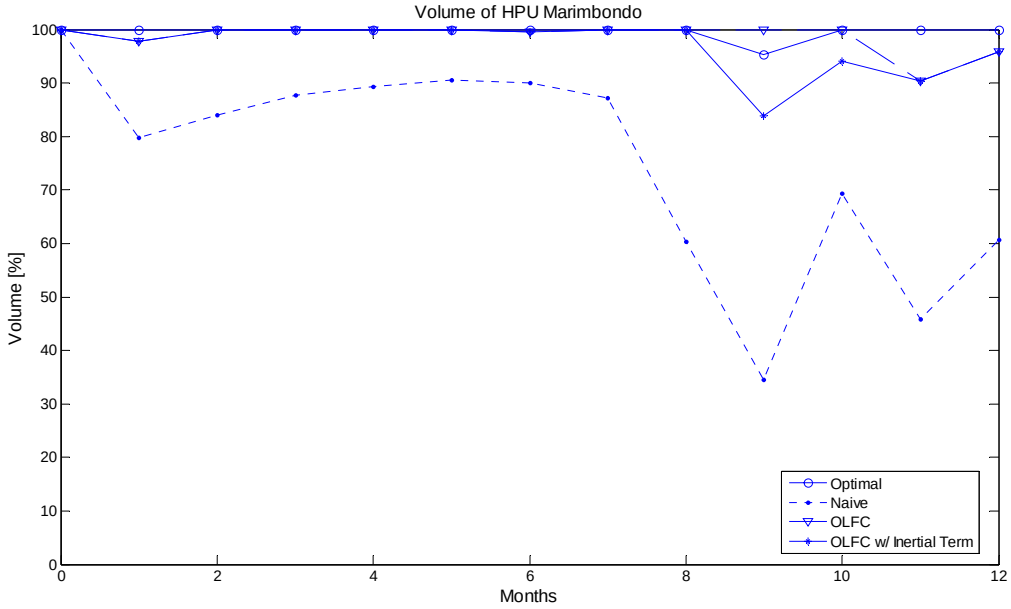


Figura 8.15 - Trajetória de Volumes Armazenados da UHE de Marimbondo: Maio/1953 a Abril/1954
3 UHE: Furnas, Marimbondo e Água Vermelha

Na Figura 8.14 é possível observar que para a solução ótima a UHE de Furnas deve ter seu armazenamento deplecionado em maior proporção, isto se deve ao fato dela ser a UHE mais a montante desta cascata.

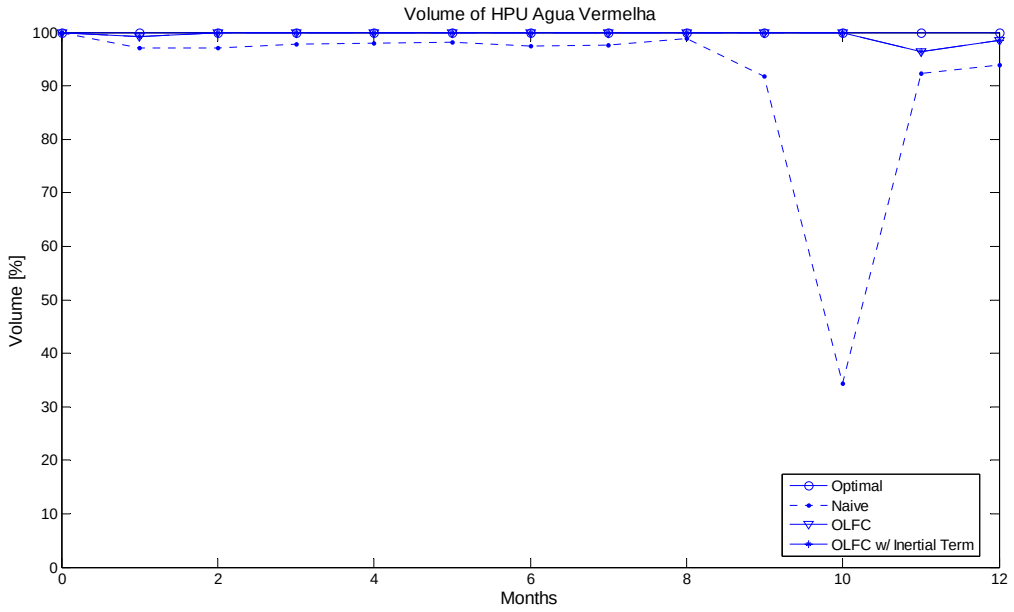


Figura 8.16 - Trajetória de Volumes Armazenados da UHE de Água Vermelha: Maio/1953 a Abril/1954 3 UHE: Furnas, Marimbondo e Água Vermelha

A UHE de Marimbondo já deve ter seu reservatório um pouco poupado de deplecionar, pois ocupa uma posição intermediária na cascata, conforme a Figura 8.15. Já a Figura 8.16, mostra que a UHE de Água Vermelha de manter seu nível de armazenamento o maior possível, garantindo assim uma maior produtividade nesta que é a UHE mais a jusante desta cascata.

A Política de Operação Ingênua (Naive) se mostra ciente deste compromisso, mas incapaz de executá-lo, pois, como já mencionado, usa a MLT como projeção das futuras VA, o que não ocorreu para este ciclo hidrológico de 1953-54. Enquanto que as Políticas de Operação POLFC e POLFC-IT apresentam comportamentos muito próximos do ótimo determinístico, demonstrando que o segundo modelo de otimização executou de forma ótima, como deveria, o despacho hidrelétrico individual das UHE da cascata.

De modo similar ao estudo anterior, este comportamento de variação do volume armazenado durante o ciclo hidrológico é necessário para proporcionar uma Geração Termelétrica o mais constante possível ao longo do período de planejamento, que também pode ser observado na Geração Termelétrica da solução ótima determinística da Figura 8.17.

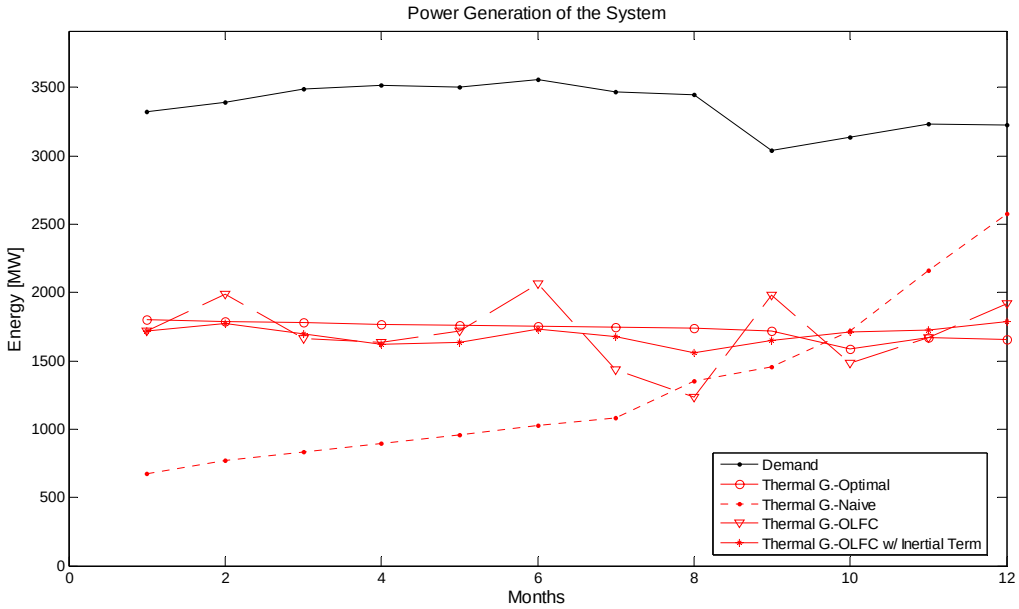


Figura 8.17 - Trajetórias das Gerações Termelétricas e da Demanda: Maio/1953 a Abril/1954
3 UHE: Furnas, Marimbondo e Água Vermelha

Novamente, nos primeiros meses a Política de Operação Ingênua (Naive) apresentou uma Geração Termelétrica muito abaixo da ideal para este ciclo hidrológico de 1953-54, abaixo da média, fato que acabou comprometendo a disponibilidade de hidreletricidade para o período chuvoso, o que obrigou um grande aumento da Geração Termelétrica para o período chuvoso, na tentativa de preservar o parque Hidrelétrico.

Já a Política POLFC, novamente apresenta uma grande instabilidade em torno da solução ótima praticamente constante ao longo do horizonte de planejamento. Esta instabilidade é causada pelos erros de previsão das VA.

Enquanto que a Política POLFC-IT apresentou um comportamento bem próximo da solução ideal, muito mais estável que as outras duas políticas de operação.

A Figura 8.18 apresenta o comportamento da Geração Hidrelétrica total das UHE de Furnas, Marimbondo e Água Vermelha. Nela é possível notar que como no estudo anterior, para a solução ótima determinística, a Geração Hidrelétrica deve acompanhar a variabilidade sazonal de Demanda energética.

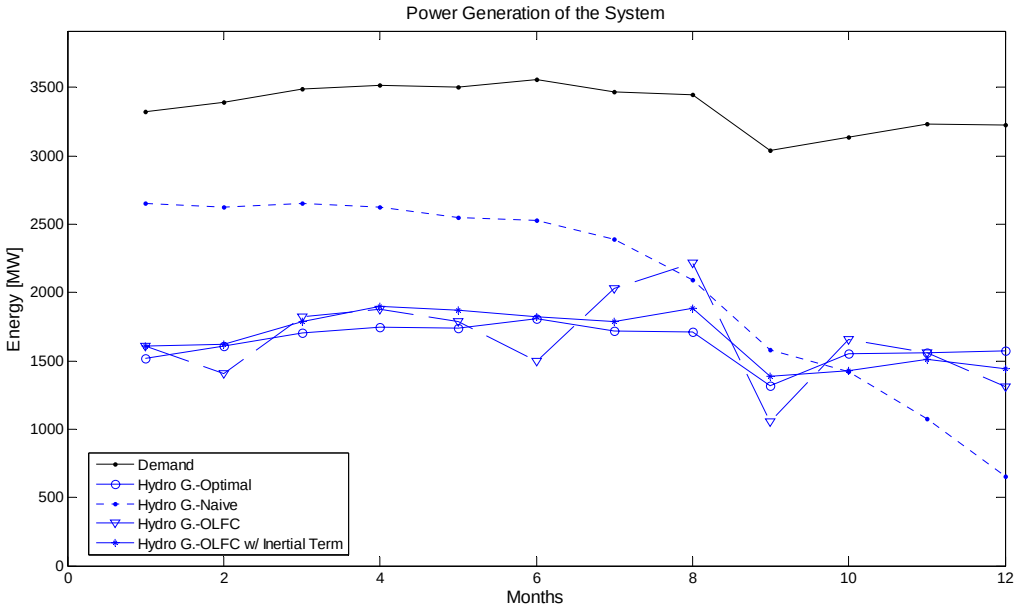


Figura 8.18 - Trajetórias das Gerações Hidrelétricas e da Demanda: Maio/1953 a Abril/1954
3 UHE: Furnas, Marimbondo e Água Vermelha

A Política Ingênua (Naive) apresentou comportamento muito similar ao do estudo anterior, onde o despacho de hidreletricidade não foi poupado nos primeiros meses como o deveria ter sido. Logo, no período chuvoso foi necessário diminuir bastante a Geração Hidrelétrica na tentativa de recuperar o armazenamento dos reservatórios.

A Política POLFC apresentou, novamente uma grande instabilidade da Geração Hidrelétrica em torno da solução ótima como conseqüência do comportamento da complementação Termelétrica da Figura 8.17.

Já a Política POLFC-IT, apresentou conseqüentemente um comportamento bem próximo da solução ideal, muito mais estável que as outras duas políticas de operação e acompanhado o comportamento sazonal da Demanda, a exemplo do estudo anterior.

A Figura 8.19 apresenta o comportamento do Custo Marginal de Operação (Preço Spot ou Preço de Curto Prazo) como conseqüência das trajetórias da Figura 8.17, que é definido pela Geração Termelétrica.

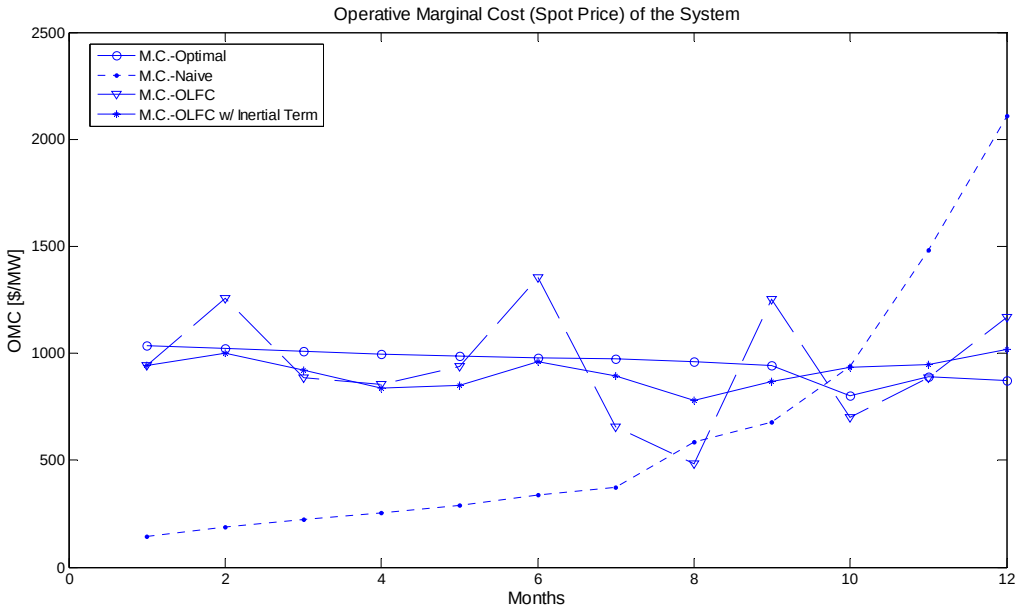


Figura 8.19 - Trajetórias dos Custos Marginais de Operação (CMO): Maio/1953 a Abril/1954
3 UHE: Furnas, Marimbondo e Água Vermelha

É possível observar na Figura 8.19 o comportamento estável, quase que constante e ligeiramente descendente do CMO para a solução ótima determinística, comportamento que é favorável ao Mercado de Energia, conforme foi comentado nas análises do estudo anterior.

A Política "Naive" definiu valores de CMO iniciais bem menores do que o ideal, causando uma elevação abrupta do Preço da Energia de curto prazo no período chuvoso, a exemplo do estudo anterior.

Novamente, a Política POLFC apresentou um comportamento de extrema instabilidade para os valores de CMO, uma vez que os valores oscilaram muito em torno dos valores ideais.

Já a Política de Operação POLFC-IT apresentou um comportamento muito mais estável de CMO, e bem próximo dos valores ótimos.

A exemplo do estudo anterior, depois de apresentados os resultados de forma gráfica, a Tabela 8.3 mostra alguns resultados numéricos. Nela é possível observar o desempenho operativo destas políticas em comparação a solução ideal ótima determinística.

Tabela 8.3 - Tabela comparativa entre as Políticas de Operação: Maio/1953 a Abril/1954
3 UHE: Furnas, Marimbondo e Água Vermelha

Política de Operação	G. Termelétrica [MW]		CMO [$/MWh]		C. de Operação	
	Média	Desvio Padrão	Média	Desvio Padrão	Total [$ milhões]	Desvio [%]
NAIVE	1.290,30	566,97	633,43	579,39	3.333	-31,16
POLFC	1.708,40	237,75	948,73	254,75	4.909	1,40
POLFC c/ Termo Inercial	1.690,10	62,08	912,13	66,44	4.514	-6,77
Ótimo Determinístico	1.730,30	61,40	955,88	66,46	4.841	----

Na Tabela 8.3, nota-se que a Política de Operação Ingênua apresentou a menor média para Geração Termelétrica e CMO, conseqüência do não re-enchimento total dos reservatórios das UHE ao final do ciclo hidrológico. Entretanto, as outras duas políticas quase que o fizeram, mas também não conseguiram reencher completamente os três reservatórios.

Logo, para uma melhor avaliação do desempenho operativo foi necessário avaliar o Custo Pós-Operativo Ótimo para cada umas das condições terminais de cada política, como pode ser observado na Tabela 8.4. Desta maneira, ficou comprovado que a Política POLFC-IT é extremamente eficiente e capaz de fornecer decisões operativas muito boas.

Tabela 8.4 - Tabela Custo de Operação com relação ao Ótimo Determinístico: Maio/1953 a Abril/1954
3 UHE: Furnas, Marimbondo e Água Vermelha

Política de Operação	Custo de Operação [$ milhões]	Custo Pós-Operativo Ótimo [$ milhões]	Desvio [%]
NAIVE	3.333	1.829	82,17
POLFC	4.909	4.650	5,58
POLFC c/ Termo Inercial	4.514	4.483	0,67

De volta a Tabela 8.3, tanto as Políticas de Operação Naive quanto a POLFC apresentaram grandes valores de Desvio Padrão para a Geração Termelétrica e para o CMO, o que não é nada conveniente para o Mercado de Energia como já foi comentado, o que acabou encarecendo o Custo de Operação destas políticas na Tabela 8.4.

Já a Política POLFC-IT conseguiu apresentar tanto valores Médios como de Desvio Padrão bastante próximos do ideal ótimo determinístico. Logo, apresentou um Custo Operativo bastante baixo e próximo do ótimo na Tabela 8.4.

Todos estes resultados ratificam a eficácia e aplicabilidade da Política de Operação POLFC com Termo Inercial na solução do Problema de Planejamento da Operação Hidro-Térmico de Potência.

Capítulo 9

Conclusões

O objetivo do Planejamento da Operação é definir uma política de operação que atenda a demanda de energia elétrica do sistema, com confiabilidade e pelo menor custo operativo. Entretanto esta não é uma tarefa fácil, em se tratando do planejamento do sistema hidrotérmico brasileiro.

Devido à complexidade do planejamento e aos diferentes aspectos que devem ser abordados, o mesmo é dividido em horizonte de médio, curto e curtíssimo prazos. O planejamento de médio prazo foi executado utilizando-se a Programação Dinâmica Estocástica (PDE) durante mais de duas décadas, sendo substituída por dois modelos baseados na Programação Dual Estocástica (PDDE), os modelos NEWAVE / DECOMP, e em uso pelo setor elétrico desde o final da década de 1990.

Apesar destas abordagens representarem a estocasticidade do problema, aleatoriedade das Vazões Afluentes, de forma explícita, ambas sofrem da "maldição da dimensionalidade", exigindo uma representação dos sistema hidrelétrico de forma agregada. Além disso, os modelos de otimização são linearizados e, o modelo NEWAVE, para o planejamento de médio prazo, utiliza uma modelagem hidrológica linear. E o modelo DECOMP consulta Funções de Custo Futuro linearizada, definida pelo NEWAVE.

Todos estes aspectos comprometem a qualidade das decisões operativas obtidas com esta abordagem, além de sinalizações inadequadas de Custo Marginal de Operação (CMO). Notadamente tem-se observado na última década uma grande volatilidade do preço de curto

prazo da energia elétrica, o que tem causado bastante insatisfação por parte dos agentes setoriais.

A ponto de em maio de 2008, a Câmara de Comercialização de Energia Elétrica (CCEE) organizar um Workshop Internacional para tratar do tema, e encontrar maneiras de diminuir tamanha volatilidade do Preço de Liquidação das Diferenças (PLD). Além desta iniciativa, vários agentes setoriais têm formado Grupos de Estudos dedicados ao tema. E mais recentemente, no final de 2008, a Agencia Nacional de Energia Elétrica (ANEEL) lançou chamada pública de P&D Estragégicos – projetos cooperados entre empresas do setor elétrico, onde um dos temas de Pesquisa foi justamente o desenvolvimento de novos modelos de otimização do despacho hidrotérmico.

Alternativamente a estas abordagens, a Política de Controle de Malha Parcialmente Aberta (POLFC) também tem sido aplicada, que combina um modelo de otimização determinístico do problema de programação hidrotérmica com as previsões de vazões afluentes geradas por um modelo de previsão. A grande vantagem da Política POLFC é que ela pode ser aplicada numa representação individualizada do sistema hidrelétrico, preservando a informação de cada característica construtiva.

Os resultados da comparação de desempenho entre PDE e a POLFC mostraram maior geração hidrelétrica média para a POLFC. Apesar deste fato, o custo total de operação ficou maior, por causa da maior flutuação da geração térmica complementar, i.e. um maior desvio padrão.

Com base nisto, o presente trabalho de pesquisa tratou de estabilizar a geração térmica complementar, e conseqüentemente, reduziu o custo total de operação para o mais próximo possível da operação ótima absoluta. Uma solução mais estável foi obtida pela Política de Operação Preditiva estabilizada via Termo Inercial, usando dois modelos de otimização não-

lineares determinísticos, com representação individual de cada UHE, e aplicando o Termo Inercial (TI), responsável por estabilizar a geração termelétrica.

9.1 Contribuições do Trabalho

Este trabalho de pesquisa teve duas linhas de investigação. Uma tratou do problema de previsão de vazões afluentes mensais, na busca por abordagens e técnicas que definissem bons modelos de previsão.

A outra linha de pesquisa tratou de encontrar uma nova Política de Operação, dentro do problema de Planejamento da Operação, que fosse capaz de definir uma seqüência de decisões operativas mais estáveis, confiáveis e de menor custo operativo.

Na primeira linha de pesquisa, investigou-se três aspectos importantes na definição um modelo de previsão: técnicas de pré-processamento dos dados, definição automática do espaço de entrada e, avaliação do desempenho de alguns modelos de Redes Neurais e Sistemas "*Fuzzy*" como modelos de previsão.

Nestes três aspectos foram investigadas a utilização da Análise dos Componentes Principais (PCA), e o tratamento da série temporal de Vazões Afluentes como um sinal discreto, utilizando-se a representação "*Analytic Signal*". Para a definição do espaço de entrada utilizou-se a abordagem da "*Dynamic Modelling*", empregando-se a "*Average Mutual Information*" e "*False Nearest Neighbors*" para uma definição do espaço de entrada automatizado. Para implementação dos modelos de previsão foram estudados e avaliados quatro modelos inteligentes: rede SONARX, rede SONARX-RBF, modelo ANFIS e a rede ESN.

Após todo este trabalho de investigação, foi possível encontrar muito bons resultados para a previsão de Vazões Afluentes mensais de um passo-a-frente e de vários-passos-a-frente.

Já a outra linha de pesquisa, de Política de Operação, foi proposta uma abordagem capaz de estabilizar os despachos termelétricos e conseqüentemente o CMO. A Política de Operação Preditiva Estabilizada via Termo Inercial (POLFC-IT) produziu excelentes resultados operativos, melhorando ainda mais a performance da Política Preditiva (POLFC).

9.2 Trabalhos Futuros

Dentre os possíveis encaminhamentos de trabalhos futuros deste trabalho de pesquisa podemos destacar:

Para a linha de pesquisa sobre modelos de previsão, poder-se-ia aplicar as abordagens propostas para outras UHE de outros subsistemas além do subsistema Sudeste, onde se encontra a UHE de Furnas, objeto deste trabalho.

Com relação a Política de Operação POLFC-IT, poder-se-ia avaliar seu comportamento frente as mais variadas condições hidrológicas. Desta maneira, poderiam ser feitas análises comparativas entre esta Política de Operação e diversas outras, como por exemplo a Programação Dinâmica Dual Estocástica (base do modelo “Newave”), a qual vem sendo aplicada no Setor Elétrico Brasileiro.

BIBLIOGRAFIA

Abarbanel, H. D. I. (1996). Analysis of Observed Chaotic Data. New York - NY, USA, Springer.

Aguirre, L. A. (2000). Introdução à Identificação de Sistemas. Belo Horizonte - MG, Editora UFMG.

Arvanitidis, N. V. e J. Rosing (1970a). "Composite Representation of a Multireservoir Hydroelectric Power System." IEEE Transactions on Power Apparatus and System **PAS-89**(2): 319-326.

Arvanitidis, N. V. e J. Rosing (1970b). "Optimal Operation of Multireservoir System Using a Composite Representation." IEEE Transactions on Power Apparatus and System **PAS-89**(2): 327-335.

Ashby, R. (1947). "Principles of the Self-Organizing Dynamic System." Journal of General Psychology(37): 125-128.

Ashby, R. (1952). Design for a Brain: The Origin of Adaptive Behavior. New York, Wiley.

Atiya, A. F., et al. (1999). "A comparison between neural-network forecasting techniques – Case study: River flow forecasting." IEEE Transactions on Neural Networks **10**(2): 402–409.

Ballini, R. (2000). Análise e Previsão de Vazões utilizando Modelos de Séries Temporais, Redes Neurais e Redes Neurais Nebulosas. Campinas, UNICAMP. **Doutorado:** 169 p.

Ballini, R., S. Soares Filho e M. G. Andrade Filho (2001). Multi-Step-Ahead Monthly Streamflow Forecasting by a Neurofuzzy Network Model. Joint 9th IFSA World Congress and 20th NAFIPS International Conference, Vancouver, BC, Canada.

Barreto, G. d. A. (2003). Redes Neurais Não-Supervisionadas Temporais para Identificação e Controle de Sistemas Dinâmicos. Engenharia Elétrica. São Carlos-SP, Universidade de São Paulo**:** 237.

Barreto, G. d. A. e A. F. R. Araújo (2001a). "Time in self-organizing maps: An overview of models." International Journal of Computer Research, Special Issue on Neural Networks: Past, Present and Future **10**(2): 139-179.

Barreto, G. d. A. e A. F. R. Araújo (2001b). A self-organizing NARX network and its application to prediction of chaotic time series. International Joint Conference on Neural Networks - IJCNN '01, Washington D.C., USA.

Barros, M. T. L., et al. (2003). "Optimization of Large-Scale Hydropower System Operations." Journal of Water Resources Planning and Management **129**(3): 178-188.

Becker, L. e W. W.-G. Yeh (1974). "Optimization of Real Time Operation of a Multiple-Reservoir System." Water Resources Research **10**(6): 1107-1112.

Bellman, R. (1962). Dynamic Programming. Princeton-NJ, USA, Princeton Univ. Press.

Berenji, H. R. e P. Khedkar (1992). "Learning and Tuning Fuzzy Logic Controllers through Reinforcements." IEEE Transactions on Neural Networks **3**(5): pp.724-740.

Bertsekas, D. P. (1976). Dynamic Programming and Stochastic Control. New York-NY, USA, Academic Press.

Bertsekas, D. P. (1995). Dynamic Programming and Optimal Control. Belmont-MA, USA, Athena Scientific.

Box, G. E. P., G. M. Jenkins e G. C. Reinsel (1994). Times Series Analysis: Forecasting and Control. San Francisco, Prentice Hall.

Braga Jr., B. P. F., et al. (1991). "Stochastic Optimization of Multiple-Reservoir System Operation." Journal of Water Resources Planning and Management **117**(4): 471-481.

Broomhead, D. S. e D. Lowe (1988). Multivariable functional interpolation and adaptive networks. Complex Systems. **2:** 321-355.

Carneiro, A. A. F. M. e M. Kadowaki (1996). Regras de Operação para Grandes Sistemas Hidroelétricos em Cascata. XI Congresso Brasileiro de Automática, São Paulo-SP, Brazil.

Carneiro, A. A. F. M. e S. Soares Filho (1989). An Adaptive Approach for Hydrothermal Scheduling. IFAC Symposium on Power Systems and Power Plant Control, Seoul, Korea.

Carneiro, A. A. F. M., S. Soares Filho e P. S. Bond (1990). "A Large Scale Application of an Optimal Deterministic Hydrothermal Scheduling Algorithm." IEEE Transactions on Power System **5**(1).

Carvalho, M. F. H. e S. Soares Filho (1987). "An Efficient Hydrothermal Scheduling Algorithm." IEEE Transactions on Power System **PWRS-2**(3): 537-542.

Casdagli, M. (1989). "Nonlinear Prediction of Chaotic Time Series." Physica D **35**(3): 335-356.

CEPEL (1977). Modelo de Programação Dinâmica Estocástica para a Operação de Sistemas Hidrotérmicos, Relatório Técnico no. 144/77.

CEPEL (2000). Modelo NEWAVE - Manual de Referência. Rio de Janeiro - RJ, Brasil, CEPEL**:** 103.

CEPEL e ELETROBRÁS (1980). Modelo a Sistema Equivalente - Descrição Geral, Relatório Técnico.

Chiu, S. L. (1994). "Fuzzy Model Identification based on Cluster Estimation." Journal of Intelligent & Fuzzy Systems **2**(3).

Cottrell, M., J. C. Fort e G. Pages (1998). "Theoretical Aspects of the SOM Algorithm." Neurocomputing **21**: 119-138.

Coulibaly, P. e F. Anctil (1999). Real-Time Short-Term Natural Water Inflows Forecasting using Recurrent Neural Networks. IEEE International Joint Conference on Neural Networks - IJCNN'99, Washington-DC, USA.

Cover, T. M. (1965). "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition." IEEE Transactions on Electronic Computers **14**: 326-334.

Cruz Jr., G. d. e S. Soares Filho (1999). General Composite Representation of Hydroelectric Systems. 21st IEEE International Conference of Power Industry Computer Applications.

da Silva, E. L. e E. C. Finardi (2001). Planning of Hydrothermal Systems Using a Power Plant Individualistic Representation. IEEE Power Tech Conference 2001, Porto, Portugal.

Delgado, A., C. Kambhampati e K. Warwick (1995). "Dynamic recurrent neural network for system identification and control." IEEE Proceedings of Control Theory and Applications **142**(4): 307–314.

Durán, H., et al. (1985). "Optimal Operation of Multireservoir Systems Using an Aggregation-Decomposition Approach." IEEE Transactions on Power Apparatus and System **PAS-104**(8): 2086-2090.

Figueiredo, M., et al. (2004). "Learning Algorithms for a Class of Neurofuzzy Network and Application." IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews **34**(3): 293-301.

Fortunato, L. A. M., et al. (1985). Operating Planning Studies of the Brazilian Generation System. IFAC Symposium on Planning and Operation of Electric Energy System - Preprints, Rio de Janeiro.

Fraser, A. M. e H. L. Swinney (1986). "Independent Coordinates for Strange Attractors from Mutual Information." Physical Review A **33**(2): 1134-1140.

Gallager, R. G. (1968). Information Theory and Reliable Communication. New York, John Wiley and Sons.

Garbor, D. (1946). "Theory of Communications." Journal of the Institution of Electrical Engineers **93**(3): 429-457.

Grossberg, S. (1986). Adaptive self-organization of serial order in behavior: speech, language, and motor control. Pattern Recognition by Humans and Machines: Speech Perception. E. Schwab and H. Nusbaum, Academic Press. **1:** 187-294.

Haykin, S. (1998). Neural Networks: A Comprehensive Foundation. NY, Prentice Hall.

Haykin, S. e J. C. Principe (1998). "Dynamic modeling of chaotic time series with neural networks " IEEE Signal Processing Magazine(May): 66-81.

Hebb, D. O. (1949). The Organization of Behavior. Wiley.

Hilborn, R. C. (2000). Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers. New York - NY, USA, Oxford University Press.

Hotelling, H. (1933). "Analysis of a complex of statistical variables into principal components." Journal of Educational Psychology **24**(6,7): 417-441,498-520.

Jaeger, H. (2001). The echo state approach to analyzing and training recurrent neural networks. Technical Report GMD Report 148, German National Research Center for Information Technology.

Jaeger, H. e H. Hass (2004). "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication." Science **304**(5667): 78-80.

Jang, J.-S. R. (1993). "ANFIS: Adaptive-Network-based Fuzzy Inference Systems." IEEE Transactions on Systems, Man and Cybernetics **23**(03): pp. 665-685.

Jang, J.-S. R., C.-T. Sun e E. Mizutani (1997). Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. New Jersey, Prentice Hall.

Kaplan, D. e L. Glass (1995). Undestanding Nonlinear Dynamics. New York-NY, USA, Springer-Verlag.

Kaski, S., J. Kangas e T. Kohonen (1998). "Bibliografy of the Self-Organizing Map (SOM) Papers: 1981-1997." Neural Computing Surveys **1**: 102-350.

Kechriotis, G., E. Zervas e E. S. Manolakos (1994). "Using recurrent neural networks for adaptive communication channel equalization." IEEE Transactions on Neural Networks **5**(2): 267–278.

Kendall, M. (1975). Multivariate Analysis. London, Charles Griffin & Company.

Kennel, M. B., R. Brown e H. D. I. Abarbanel (1992). "Determining Minimum Embedding Dimension using a Geometrical Construction." Physical Review A **45**(6): 3403-3411.

Kim, Y.-O., D. Jeong e I. H. Ko (2006). "Combining Rainfall-Runoff Model Outputs for Improving Ensemble Streamflow Prediction." Journal of Hydrologic Engineering **11**(6): 578-588.

Kohonen, T. (1989). Self-Organization and Associative Memory. Berlin, Springer-Verlag.

Lee, C.-C. (1990a). "Fuzzy Logic in Control Systems: Fuzzy Logic Controller - Part I." IEEE Transactions on Systems, Man and Cybernetics **20**(2): pp. 404-418.

Lee, C.-C. (1990b). "Fuzzy Logic in Control Systems: Fuzzy Logic Controller - Part II." IEEE Transactions on Systems, Man and Cybernetics **20**(2): pp.419-435.

Light, W. A. (1992). Some aspects of radial basis function approximation. Approximation Theory, Spline Functions and Applications, Boston, MA.

Lin, C. T. e C. S. G. Lee (1991). "Neural-Network-based Fuzzy Logic Control and Decision System." IEEE Transactions on Computers **40**(12): pp. 1320-1336.

Lyra, C., H. Tavares e S. Soares Filho (1984). "Modelling and optimization of hydrothermal generation scheduling." IEEE Transactions on Power Apparatus and System **103**(8): 2126-2133.

Mandani, E. H. e S. Assilian (1975). "An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller." International Journal of Man-Machine Studies **7**(1): pp.1-13.

Manly, B. F. J. (1986). Multivariate Statistical Methods: A PRIMER. London, Chapman and Hall.

Marple Jr., S. L. (1987). Digital Spectral Analysis. Englewood Cliffs, NJ, Prentice-Hall.

Marple Jr., S. L. (1999). "Computing the Discrete-Time "Analytic" Signal via FFT." IEEE Transactions on Signal Processing **47**(9): 2600-2603.

Martinez, L. e S. Soares Filho (2002). "Comparison Between Closed-Loop and Partial Open-Loop Feedback Control Policies in Long Term Hydrothermal Scheduling." IEEE Transactions on Power System **17**(2): 330-336.

MecLeod, A. I. (1994). "Diagnostic checking of periodic autoregression." Journal of Time Series Analysis **15**(2): 221-223.

Moody, J. e C. J. Darken (1989). "Fast learning in networks of locally-tuned processing units." Neural Computation **1**: pp. 281-294.

Nauck, D., F. Klawonn e R. Kruse (1997). Foudations of Neuro-Fuzzy Systems. New York, Wiley.

Norgaard, M., et al. (2000). Neural Networks for Modelling and Control of Dynamic Systems, Springer-Verlag.

Oppenheim, A. V. e R. W. Schafer (1999). Discrete-Time Signal Processing. New Jersey - NJ, USA, Prentice Hall.

Pearson, K. (1901). "On lines and planes of closest fit to systems of points in space." Philosophical Magazine **2**: 559-572.

Pereira, M. V. F. (1985). Optimal Scheduling of Hydrothermal System - An Overview. IFAC - Electric Energy Systems, Rio de Janeiro, Brazil.

Pereira, M. V. F. (1989). "Optimal Stochastic Operations Scheduling of Large Hydroelectric System." Electrical Power & Energy Systems **11**(3): 161-169.

Pereira, M. V. F. e L. M. V. G. Pinto (1985). "Stochastic Optimization of a Multireservoir Hydroelectric System : A Decomposition Approach." Water Resources Research **21**(6): 779-792.

Poggio, T. e F. Girosi (1990). "Networks for approximation and learning." PROCEEDINGS OF THE IEEE **78**: pp. 1471-1497.

Powell, M. J. D. (1985). Radial basis functions for multivariable interpolation: A review. IMA Conference on Algorithms for the Approximation of Functions and Data, Shrivenham, UK.

Powell, M. J. D. (1988). Radial basis function approximations to polynomials. Numerical Analysis, Dundee, UK.

Principe, J. C. (1997). "Neural Networks for Dynamic Modeling." IEEE Digital Signal Processing Magazine **14**(6): 33-35.

Principe, J. C., N. R. Euliano e W. C. Lefebvre (1999). Neural and Adaptive Systems: Fundamentals through Simulations, Wiley.

Principe, J. C., L.-X. Wang e J.-M. Kuo (1998). Non-Linear Dynamic Modelling with Neural Networks. Signal Analysis and Prediction. A. Prochazka, J. Uhlir, P. J. W. Rayner and N. G. Kingsbury**:** 275-290.

Principe, J. C. e L. Wang (1995). Non-Linear Time Series Modeling with Self-Organization Feature Maps. IEEE Intl. Work. on Neural Networks for Signal Processing, Cambridge, USA.

Puskorius, G. V. e L. A. Feldkamp (1994). "Neurocontrol of nonlinear dynamical systems with kalman filter trained recurrent networks." IEEE Transactions on Neural Networks **5**(2): 279–297.

RE-SEB, P. (1998). Projeto de Reestruturação do Setor Elétrico Brasileiro, ELETROBRÁS-BIRD.

Read, E. G. (1982). Economic Principles of Reservoir Operation 1: Perfect Foresight, University of Tennessee**:** 30.

Reilly, A., G. Frazer e B. Boashash (1994). "Analytic signal generation-Tips and traps." IEEE Transactions on Signal Processing **42**: 3241-3245.

Renals, S. (1989). "Radial basis function network for speech pattern classification." Electronics Letters **25**: pp. 437-439.

Rezende, S. O. (2005). Sistemas Inteligentes: Fundamentos e Aplicações. Barueri - SP, Manole.

Robbins, H. e S. Monro (1951). "A Stochastic Approximation Method." Annals of Mathematical Statistics **22**: 400-407.

Ross, T. J. (1995). Fuzzy Logic with Engineering Applications. New York, McGraw-Hill.

Rotting, T. A. e A. Gjelsvik (1992). "Stochastic Dual Dynamic Programming for Seasonal Scheduling in the Norwegian power System." IEEE Transactions on Power System **7**(1): 273-279.

Sacchi, R., A. A. F. M. Carneiro e A. F. R. Araújo (2004). A RBF Network Trained by the SONARX Model and Applied to Obtain the Operation Policies of the Hydropower Systems. VIII Brazilian Symposium on Neural Networks - SBRN, São Luis, MA.

Sacchi, R., R. F. Marques e R. Castro (2008). Possíveis Aprimoramentos na Formação do PLD – Preço de Liquidação das Diferenças no Mercado Brasileiro. Workshop Internacional sobre Formação de Preço de Energia Elétrica no Mercado de Curto Prazo, São Paulo-SP, Câmara de Comercialização de Energia Elétrica - CCEE.

Sacchi, R., et al. (2007). Water Inflow Forecasting using the Echo State Network: a Brazilian Case Study. International Joint Conference on Neural Networks - IJCNN'07 / INNS - IEEE, Orlando, FL - USA.

Salas, J. D., D. C. Boes e R. A. Smith (1982). "Estimation for arma models with seasonal parameters." Water Resources Research **18**(4): 1006-1010.

Sandberg, I. W. e L. Xu (1997). "Uniform Approximation and Gama Networks." Neural Networks **10**: 781-784.

Sauer, T., J. A. Yorke e M. Casdagli (1991). "Embedology." Journal of Statistical Physics **65**(3-4): 579-616.

Silva Filho, D., et al. (2000). Influência das Regras de Operação Sobre Parâmetros de Dimensionamento de Usinas Hidroelétricas. VII SEPOPE - Symposium of Specialists in Electric Operational and Expansion Planning, Curitiba, Brazil.

Sjöberg, J., et al. (1995). "Nonlinear Black-Box Modeling in System Identification: A Unified Overview." Automatica **31**(12): 1691-1724.

Soares Filho, S. (1987). "Planejamento da Operação de Sistemas Hidrotérmicos." SBA - Controle e Automação **1**(2): 122-131.

Soares Filho, S. e A. A. F. M. Carneiro (1991). "Optimal Operation of Reservoirs for Electric Generation." IEEE Transactions on Power Delivery **6**(3): 1101-1107.

Stokelj, T., D. Paravan e R. Golob (2002). "Enhanced Artificial Neural Network Inflow Forecasting Algorithm for Run-of-River Hydropower Plants." Journal of Water Resources Planning & Management - ASCE **128**(6): 415–423.

Sugeno, M. e G. T. Kang (1988). "Structure Identification on Fussy Model." Fuzzy Sets and Systems **28**: pp.15-33.

Takagi, T. e M. Sugeno (1983). Derivation on Fuzzy Control Rules from Human Operator's Control Actions. Proceedings of the IFAC Symposium on Fuzzy Information, Knowledge Representation and Decision Analysis.

Takagi, T. e M. Sugeno (1985). "Fuzzy Identification of Systems and Its Applications to Modeling and Control." IEEE Transactions on Systems, Man and Cybernetics **15**: pp.116-132.

Takens, F. (1981). Detecting strange attractors in turbulence. Dynamical Systems and Turbulence, Lecture Notes in Mathematics. D. A. Rand and L.-S. Young, Springer-Verlag. **898:** 366–381.

Takens, F. (1985). On the Numerical Determination of the Dimension of an Attractor. Dynamical Systems and Bifurcations. D. A. Rand and L.-S. Young, Springer-Verlag. **1125:** 99-106.

Toscano, A. E. (2009). Comparação entre os modelos NEWAVE e ODIN no planejamento energético do Sistema Interligado Nacional. Faculdade de Engenharia Elétrica e de Computação. Campinas, Universidade de Campinas. **Mestrado:** 112.

Tsoukalas, L. H. e R. E. Uhrig (1997). Fuzzy and Neural Approaches in Engineering. New York, John Wiley & Sons.

Tsukamoto, Y. (1979). An Approach to Fuzzy Reasoning Method. Advances in Fuzzy Set Theory and Applications. M. M. Gupta, R. K. Ragade, R. R. Yager and Eds. Amsterdam, North-Holland**:** pp. 137-149.

Turgeon, A. e R. Charbonneau (1998). "An Aggregation-Disaggregation Approach to Long-Term Reservoir Management." Water Resources Research **34**(12): 3585-3594.

Turing, A. (1952). The Chemical Basis of Morphogenesis. Phil. Trans. of the Royal Soc.

Valenca, M., T. B. Ludermir e A. Valenca (2005). River Flow Forecasting for Reservoir management through Neural Networks. International Conference on Hybrid Intelligent Systems - HIS'05.

Ville, J. A. (1948). "Theorie et application de la notion du signal analytique." Cables Transmission **2**: 61-74.

Walter, J. e H. Ritter (1996). "Rapid Learning with Parametrized Self-Organizing Maps." Neurocomputing **12**: 131-153.

Wang, L.-X. (1994). Adaptive Fuzzy Systems and Control. New Jersey, PTR Prentice Hall.

Weigend, A. S., D. E. Rumelhart e B. A. Huberman (1991). Generalization by weightelimination applied to currency exchange rate prediction. IEEE International Joint Conference on Neural Networks, Seattle, USA.

Wood, A. J. e B. F. Wollenberg (1984). Power Generation, Operation and Control. New York, John Wiley and Sons.

Yager, R. e D. Filev (1994a). Essentials of Fuzzy Modeling and Control. New York, John Wiley & Sons.

Yager, R. e D. Filev (1994b). "Generation of Fuzzy Rules by Mountain Clustering." Journal of Intelligent & Fuzzy Systems **2**(3): pp. 209-219.

Yeh, W. W.-G. (1985). "Reservoir Management and Operations Models: A State-of-the-Art Review." Water Resources Research **21**(12): 1797-1818.

Yeh, W. W.-G., et al. (1992). "Optimization of Real-Time Hydrothermal System Operation." Journal of Water Resources Planning and Management **118**(6): 636-653.

Zadeh, L. A. (1965). "Fuzzy Sets." Information and Control **8**: pp. 338-353.

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Administração
Baixar livros de Agronomia
Baixar livros de Arquitetura
Baixar livros de Artes
Baixar livros de Astronomia
Baixar livros de Biologia Geral
Baixar livros de Ciência da Computação
Baixar livros de Ciência da Informação
Baixar livros de Ciência Política
Baixar livros de Ciências da Saúde
Baixar livros de Comunicação
Baixar livros do Conselho Nacional de Educação - CNE
Baixar livros de Defesa civil
Baixar livros de Direito
Baixar livros de Direitos humanos
Baixar livros de Economia
Baixar livros de Economia Doméstica
Baixar livros de Educação
Baixar livros de Educação - Trânsito
Baixar livros de Educação Física
Baixar livros de Engenharia Aeroespacial
Baixar livros de Farmácia
Baixar livros de Filosofia
Baixar livros de Física
Baixar livros de Geociências
Baixar livros de Geografia
Baixar livros de História
Baixar livros de Línguas

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo