

**UNIVERSIDADE FEDERAL DE ITAJUBÁ**

**TESE DE DOUTORADO**

---

**PREVISOR NEURAL DE CARGA ELÉTRICA BASEADO EM  
ANÁLISE DE MULTIRESOLUÇÃO VIA WAVELETS E  
TÉCNICAS DE RECONSTRUÇÃO DO ESPAÇO-FASE**

---

Candidato: Agnaldo José da Rocha Reis.

Orientador: Alexandre Pinto Alves da Silva, Ph.D.

Tese submetida ao programa de pós-graduação em Engenharia Elétrica da Universidade Federal de Itajubá, como requisito parcial para a obtenção do grau de Doutor em Ciências em Engenharia Elétrica

Itajubá, novembro de 2003.

“Avançamos, não quando somos melhores do que os outros. E sim, quando a cada dia progredimos no que somos”. *Dámaso Cerruti*

Dedico este trabalho a todas aquelas pessoas que acreditam e/ou investem em ciência e tecnologia. Em especial, aos meus pais Adriene e Joaquim, a minha avó Dorvalina (in memoriam), a Luci Mary e ao nosso pequeno Alexander.

# Agradecimentos

A Universidade Federal de Itajubá e a família GESis<sup>1</sup> pela acolhida.

A todos os professores, membros da banca examinadora, amigas e amigos que, direta ou indiretamente, contribuíram para o meu crescimento pessoal e profissional.

A agência de fomento CNPq pelo suporte financeiro.

Aos meus pais e a Luci Mary pelo incessante apoio e compreensão.

A minha avó Dorvalina pelo seu carinho e sua visão de futuro.

A Deus, por nos ajudar a transformar nossos sonhos em realidade.

---

<sup>1</sup> GESis – Grupo de Engenharia de Sistemas

# Agenda

<b>Resumo</b>	viii
<i>Abstract</i>	ix
<b>Lista de Figuras</b>	x
<b>Lista de Tabelas</b>	xii
<b>Lista de Abreviaturas</b>	xiii
<b>Capítulo 1 - Introdução.....</b>	<b>1</b>
1.1 Previsão de Carga a Curto Prazo (PCCP).....	1
1.2 Objetivos.....	2
1.2.1 Motivação.....	2
1.2.2 Proposta.....	2
1.2.3 Revisão bibliográfica sobre wavelets aplicada a PCCP .....	4
1.2.4 Contribuições.....	6
1.3 Organização da tese.....	7
<b>Capítulo 2 - Previsão de Carga Elétrica a Curto Prazo.....</b>	<b>8</b>
2.1 Visão geral do problema.....	8
2.2 A Importância da PCCP .....	10
2.3 Características gerais das cargas elétricas .....	11
2.4 Classificação quanto ao tipo de modelo de carga .....	13
2.4.1 Modelos de pico de carga .....	13
2.4.2 Modelos de curva de carga .....	13
2.5 Técnicas de previsão .....	15
2.5.1 Métodos convencionais .....	15
2.5.1.1 Modelos baseados em Regressão .....	16
2.5.1.2 Modelos de Alisamento Exponencial.....	16
2.5.1.3 Séries Temporais Estocásticas (Modelos Box & Jenkins) .....	18
2.5.2 Modelos não convencionais.....	22
2.6 Redes Neurais Artificiais como solução para o problema de previsão de carga .....	23
2.7 Revisão bibliográfica sobre a PCCP via RNA.....	23
<b>Capítulo 3 - A Transformada Wavelet Discreta .....</b>	<b>28</b>
3.1 Motivação para se utilizar a transformada wavelet.....	28
3.2 Transformada Wavelet Contínua e Discreta.....	28
3.3 Análise de Multiresolução e o Algoritmo Piramidal de Mallat .....	30

3.4 Minimizando o problema de distorção nas bordas.....	32
3.5 Exemplos práticos .....	34
3.5.1 Descrição estatística de uma série de carga via AMR.....	34
3.5.2 Identificando-se <i>Outliers</i> e Dados Faltantes .....	36
3.5.3 Espectro de Potência.....	38
3.6 Aplicações de Wavelets em SEP.....	38
<b>Capítulo 4 – Técnicas de Reconstrução do Espaço-Fase .....</b>	<b>40</b>
4.1 Introdução .....	40
4.2 Reconstrução do Espaço-Fase.....	43
4.2.1 Falsos vizinhos no espaço-fase .....	46
4.2.2 Informação Mútua Média .....	46
4.3 Reconstrução do Espaço-Fase aplicada à PCCP.....	47
4.3.1 Implementação .....	48
4.3.1.1 Cálculo do valor em atraso através da IMM .....	50
4.3.1.2 Cálculo da dimensão de imersão através do algoritmo de falsos vizinhos .....	51
4.3.1.3 Reconstrução do espaço-fase via método das coordenadas em atraso – Um exemplo prático.....	52
<b>Capítulo 5 - Testes e Resultados .....</b>	<b>56</b>
5.1 Definição das séries históricas de carga e temperatura .....	56
5.2 Agrupamento de dados de previsão .....	57
5.3 Identificação das variáveis explicativas do processo gerador da série histórica .....	57
5.4 Definição do número de pontos utilizados no processo de treinamento/validação .....	58
5.5 Etapa de pré-processamento dos dados disponíveis.....	59
5.6 Partição ótima dos dados .....	59
5.7 Definição da estrutura do PMC.....	62
5.8 Algoritmo de aprendizagem.....	62
5.9 Modelos de PCCP via PMC.....	64
5.9.1 Modelo 1 (M1) .....	64
5.9.2 Modelo 2 (M2) .....	65
5.9.3 Modelo 3 (M3) .....	67
5.9.4 Modelo 4 (M4) .....	69
5.9.5 Modelo 5 (M5) .....	71
5.9.6 Modelo 6 (M6) .....	72

5.10 Etapas de treinamento/validação.....	73
5.11 Tipos de previsão .....	75
5.12 Índices de desempenho.....	75
5.13 Testes.....	77
5.13.1 Caso A.....	77
5.13.2 Caso B.....	83
5.13.3 Caso C.....	84
<b>Capítulo 6 – Conclusões e trabalhos futuros.....</b>	<b>86</b>
6.1 Conclusões.....	86
6.2 Trabalhos futuros.....	89
6.3 Contribuições .....	89
<b>Referências .....</b>	<b>90</b>
<b>Anexo I – Teoria sobre Redes Neurais Artificiais.....</b>	<b>98</b>
I.1 Introdução .....	98
I.2 O Neurônio Biológico.....	98
I.3 Representação Matemática de um Neurônio.....	100
I.4 Tipos de Redes Neurais Artificiais.....	102
I.5 Perceptrons de Múltiplas Camadas.....	103
I.6 Discussão .....	116
I.7 Aplicações de RNA em Sistemas Elétricos de Potência.....	117
I.8 Principais Referências de Aplicações de RNA em SEP.....	120
I.9 Resumo .....	122
<b>Anexo II – Domínio da Frequência .....</b>	<b>123</b>
II Introdução ao Domínio da Frequência.....	123
II.1 Análise de Fourier para sinais discretos .....	123
II.2 O Espectro de Potência.....	125
II.3 Transformada Inversa de Fourier .....	126
<b>Anexo III – Fundamentos da Wavelet de Haar .....</b>	<b>127</b>
III.1 Wavelet de Haar.....	127
III.2 Representação em Wavelets de funções geradas por um conjunto de dados .....	129
III.3 Um exemplo numérico .....	129

III.4 Subrotina implementada .....	131
<b>Anexo IV - Publicações</b> .....	133

## Resumo

A importância da previsão de carga a curto prazo tem crescido ultimamente. Com a desregulamentação e a competição advinda desse processo, a previsão do preço de energia se transformou em uma atividade bastante lucrativa. A previsão das cargas das barras é essencial para alimentar métodos analíticos utilizados para determinar os preços de energia. A variabilidade e a não estacionariedade das cargas estão ficando cada vez piores devido à dinâmica dos preços de energia. Além disso, o número de cargas nodais a serem previstas não permite interações frequentes com os especialistas em previsão de carga. Portanto, previsores de carga mais autônomos são necessários nesse novo cenário competitivo.

Esta tese apresenta duas linhas de pesquisa diferentes. Na primeira delas, duas estratégias para a utilização da transformada wavelet na previsão de carga via redes neurais são apresentadas. A primeira estratégia é nova. Ela consiste na criação de um modelo de previsão de carga cujas entradas são baseadas na informação da série de carga original e na informação fornecida pelas subséries no domínio wavelet. Já na segunda estratégia, o comportamento futuro da carga é conseguido através da combinação de previsões independentes de cada subsérie no domínio wavelet.

A segunda linha de pesquisa investiga a aplicabilidade de uma metodologia não linear baseada no método de coordenadas em atraso para a seleção das variáveis de entrada mais significativas para previsores neurais. Esse critério é comparado com um outro critério linear baseado na função de autocorrelação.

Com a utilização das metodologias supraditas, objetiva-se o desenvolvimento de previsores de carga mais robustos. Para testá-las, dados horários reais de carga e temperatura de uma concessionária de energia elétrica norte-americana são utilizados.

*Palavras-chave*—Previsão de carga; Redes neurais; Reconstrução do espaço-fase; Transformada wavelet.



## ***Abstract***

The importance of short-term load forecasting has been increasing lately. With deregulation and competition, energy price forecasting has become a big business. Bus-load forecasting is essential to feed analytical methods utilized for determining energy prices. The variability and non-stationarity of loads are becoming worse due to the dynamics of energy prices. Besides, the number of nodal loads to be predicted does not allow frequent interactions with load forecasting experts. More autonomous load predictors are needed in the new competitive scenario.

This thesis deals with two main research lines. In the first one, two different strategies for successfully embedding the Discrete Wavelet Transform into Artificial Neural Networks-based short-term load forecasting is presented. The first strategy is new. It consists of creating a model for load forecasting whose inputs are based on information from the original load sequence and from wavelet domain subseries, as well. The second alternative predicts the load's future behavior by independently forecasting each subseries in the wavelet domain.

The other research line evaluates the feasibility of a nonlinear criterion based on the method of delay coordinates for determining the best set of input variables for a neural forecaster. This criterion is fully compared to another linear criterion based on the autocorrelation function.

The main goal of this work is to develop more robust load forecasting algorithms. Hourly load and temperature data for a North-American electric utility are used to test the proposed methodologies.

Index Terms-- Load forecasting; Neural networks; Phase-space reconstruction; Wavelet transform.

## Lista de Figuras

Figura 1.1 Decomposição do problema de previsão de carga.....	4
Figura 1.2 Diagrama esquemático da wavenet utilizada.....	6
Figura 2.1 Fluxograma do modelo ARIMA de Box & Jenkins.....	21
Figura 3.1 Exemplos de wavelet mãe .....	30
Figura 3.2 Análise de resolução simples via algoritmo de Mallat.....	31
Figura 3.3 Esquema de decomposição de múltiplos níveis.....	32
Figura 3.4 Desvios de 4 esquemas de enchimento em relação ao enchimento ideal.....	33
Figura 3.5 Wavelets da família Daubechies .....	35
Figura 3.6 Segmento de curva de carga decomposto em 3 níveis de resolução.....	37
Figura 3.7 Seg. de curva de carga corrompido com 2 outliers e 12 dados faltantes.....	37
Figura 3.8 Espectro de potência para os segmentos de curva de carga considerados .....	38
Figura 3.9 Principais áreas de aplicação da TW em SEP.....	39
Figura 4.1 Exemplo de atratores .....	41
Figura 4.2 Exemplo de um atrator tridimensional .....	41
Figura 4.3 Determinação da dimensão Euclideana de alguns objetos .....	42
Figura 4.4 Determinação da dimensão de Hausdorff de um objeto fractal.....	43
Figura 4.5 Diagrama esquemático da técnica de REF .....	48
Figura 4.6 Segmento de série de carga considerado .....	49
Figura 4.7 Série de carga suavizada (componente A3).....	49
Figura 4.8 Um comparativo entre as séries pré-processadas.....	50
Figura 4.9 Inf. Mútua Média em função do valor em atraso para a série normalizada ....	51
Figura 4.10 Fração de falsos vizinhos mais próximos em função da dimensão.....	51
Figura 4.11 Repres. do EF para valores em atraso de 1 e 8h (série normalizada).....	53
Figura 4.12 Repres. do EF para valores em atraso de 1 a 8h (dados suavizados) .....	54
Figura 5.1 estrutura típica de um PMC .....	54
Figura 5.2 Algoritmo para a implementação de $\eta$ adaptativa .....	61
Figura 5.3 Autocorrelograma típico de um segmento de série de carga.....	63
Figura 5.4 Diagrama para M1.....	65
Figura 5.5 Diagrama esquemático do modelo 2 .....	65
Figura 5.6 Diagrama esquemático do modelo 3 .....	66
Figura 5.7 Diagrama esquemático do modelo 4 .....	68
Figura 5.8 Desempenho dos modelos M4 e M6 para previsões recursivas (Dom.) .....	71
Figura 5.9 Desempenho dos modelos M4 e M6 para previsões recursivas (feriado).....	81

Figura 5.10 Desempenho dos modelos M1 e M5 para previsões recursivas (Dom.) .....	81
Figura 5.11 Desempenho dos modelos M1 e M5 para previsões recursivas (feriado).....	82
Figura 5.12 Carga real e as respectivas previsões (+) obtidas com M4.....	83
Fig. I.1. Partes constituintes de um neurônio biológico.....	99
Fig. I.2. Representação Matemática de uma Unidade [Russel & Norvig, 1995].....	100
Fig. I.3. Três diferentes funções de ativação.....	101
Fig. I.4 Estrutura de Perceptron simples.....	102
Fig. I.5. Rede de Hopfield e Máquinas de Boltzman. ....	102
Fig. I.6. Perceptron de Múltiplas Camadas [Russel & Norvig, 1995]. ....	103
Fig. I.7. Superfície de erro característica de problemas linearmente separáveis.....	110
Fig. I.8. Superfície de erro multimodal.....	111
Fig. I.9 Generalização em aprendizagem de RNA. ....	113
Fig. I.10. Problema proposto. ....	114
Fig. I.11. “Underfitting”. ....	115
Fig. I.12. “Overfitting”. ....	115
Fig. I.13. Generalização adequada. ....	115
Fig. I.14. Principais áreas de aplicação em abril de 1993. ....	118
Fig. I.15 Modelos de RNA usados em problemas de SEP (1993). ....	118
Fig. I.16. Principais áreas de aplicação em 1995. ....	119
Fig. I.17. Modelos de RNA usados em problemas de SEP (1995). ....	119
Fig. III.1. Representação gráfica dos dados.....	130

## Lista de Tabelas

Tabela 2.1 Comparação entre alguns trabalhos de PCCP via RNA .....	28
Tabela 5.1 Informações sobre os dados de carga e temperatura utilizados.....	57
Tabela 5.2 Variáveis de entrada para M2.....	67
Tabela 5.3 Variáveis de entrada para M3 .....	68
Tabela 5.4 Variáveis de entrada para M4 .....	70
Tabela 5.5 Variáveis de entrada selecionadas para M5 .....	72
Tabela 5.6 Variáveis de entrada para M6.....	73
Tabela 5.7 MAPE horário para os 6 modelos (1-24 passos à frente).....	80
Tabela 5.8 Avaliação global média (1-24 passos à frente) .....	80
Tabela 5.9 MAPE horário para os 6 modelos (1 passo à frente).....	80
Tabela 5.10 Avaliação global média (1 passo à frente).....	81
Tabela 5.11 Avaliação global média (recursiva) .....	85
Tabela 5.12 Avaliação global média (1 passo à frente) .....	85
Tabela I.1 RNA no ISAP 99.....	120
Tabela II.1 Aplicação da DFT.....	124

## Lista de Abreviaturas

<b>ACF</b>	Função de Autocorrelação
<b>AMR</b>	Análise de Multiresolução
<b>ANNSTLF</b>	<i>Artificial Neural Network Short-Term Load Forecaster</i>
<b>AR</b>	Autoregressivo
<b>ARMA</b>	Autoregressivo, Média móvel
<b>CAG</b>	Controle Automático da Geração
<b>CPP</b>	Critério de Parada Prematura
<b>DE</b>	Despacho Econômico
<b>DFT</b>	Transformada de Fourier Discreta
<b>EBP</b>	Algoritmo de Retropropagação de Erros
<b>EDO</b>	Equação Diferencial Ordinária
<b>EE</b>	Energia Elétrica
<b>EESP</b>	Estimador de Estado do Sistema de Potência
<b>e.g.</b>	Por exemplo
<b>EPRI</b>	<i>Electric Power Research Institute</i>
<b>FFT</b>	Transformada Rápida de Fourier
<b>IMM</b>	Informação Mútua Média
<b>i.e.</b>	Isto é
<b>LF</b>	Lógica <i>Fuzzy</i>
<b>MA</b>	Média móvel
<b>MAPE</b>	Erro Percentual Médio Absoluto
<b>MÁX.</b>	Erro Máximo Médio
<b>ME</b>	Erro Médio
<b>MSE</b>	Erro Médio Quadrático
<b>PACF</b>	Função de Correlação Parcial
<b>PCCP</b>	Previsão de Carga a Curto Prazo

<b>PMC</b>	Perceptrons de Múltiplas Camadas
<b>QMF</b>	<i>Quadrature Mirror Filter</i>
<b><i>r</i></b>	Função de Autocorrelação
<b>REF</b>	Reconstrução do Espaço-Fase
<b>RNA</b>	Redes Neurais Artificiais
<b>RWE</b>	Rede Wavelet Evolucionária
<b>SAR</b>	Autoregressivo Sazonal
<b>SE</b>	Sistemas Especialistas
<b>SEP</b>	Sistemas Elétricos de Potência
<b>SMA</b>	Média Móvel Sazonal
<b>TWC</b>	Transformada Wavelet Contínua
<b>TWD</b>	Transformada Wavelet Discreta
<b>TWI</b>	Transformada Wavelet Inversa
<b>VC</b>	Validação Cruzada

# Capítulo 1 - Introdução

Neste capítulo, a descrição do tema previsão de carga, sua função e sua importância na operação de um sistema elétrico de potência serão brevemente discutidos. Os principais objetivos desta tese, a metodologia pela qual o problema será abordado e as principais contribuições deste trabalho também serão apresentadas.

## 1.1 Previsão de Carga a Curto Prazo

A importância da Previsão de Carga a Curto Prazo (PCCP) tem aumentado ultimamente. Do ponto de vista técnico, algumas das decisões e funções de controle relativas à operação de Sistemas Elétricos de Potência (SEP) dependem do conhecimento prévio do comportamento da carga. A título de exemplo, em estudos de análise de segurança, o conhecimento das cargas futuras para o período de 1 a 24 horas pode ser utilizado na detecção de períodos de vulnerabilidade do sistema e no planejamento da saída de linhas de transmissão, transformadores e geradores para manutenção. Já do ponto de vista econômico, com a desregulamentação do setor elétrico e a competição advinda desse processo, a previsão do preço da energia tem se transformado em uma atividade bastante atraente. A previsão de carga das diversas barras é essencial para alimentar métodos analíticos utilizados na determinação dos preços da energia. A variabilidade e a não estacionariedade das cargas estão ficando cada vez mais acentuadas devido à dinâmica das tarifas de energia.

Redes Neurais Artificiais (RNA) têm sido aplicadas à PCCP com muito êxito ([Park et al., 1991], [Lee et al., 1992], [Bakirtzis et al., 1996], [Papalexopoulos et al., 1994]). Muitas concessionárias de energia elétrica que anteriormente empregavam ferramentas para PCCP baseadas em técnicas estatísticas clássicas, agora estão usando previsores baseados em RNA [Khotanzad et al., 1998]. Uma revisão bastante abrangente da aplicação de RNA na PCCP pode ser encontrada em [Hippert et al., 2001]. Este é um daqueles casos raros onde a ciência se transformou em tecnologia após um curto período de desenvolvimento. No entanto, essa tecnologia está longe de alcançar os seus limites.

## 1.2 Objetivos

### 1.2.1 Motivação

Uma certa regularidade dos dados é uma importante pré-condição para se lograr êxito com a aplicação de redes neurais na previsão de carga [Bunn, 2000]. Quando técnicas estatísticas clássicas são utilizadas, considera-se que o processo a ser modelado é estacionário. Para séries de carga elétrica, tal consideração, via de regra, deve ser descartada. Além disso, deve-se considerar a existência de vários tipos de não estacionariedade.

Para tratar do problema de não estacionariedade, *wavelets* são utilizadas. Isso se deve ao fato de que uma decomposição via wavelets pode produzir uma boa representação local do sinal tanto no domínio do tempo quanto no domínio da frequência [Mallat, 1989]. Além disso, a decomposição via wavelets utilizada como uma ferramenta de análise em múltiplas escalas, pode ser empregada para identificar certas características da série de carga que não podem ser facilmente detectadas com o uso de ferramentas de análise tradicionais (e.g., Transformada de Fourier), e que, portanto, podem influenciar positivamente o desempenho de previsores neurais.

### 1.2.2 Proposta

Com o intuito de investigar a aplicabilidade da combinação entre redes neurais e wavelets, dados horários reais de carga e temperatura para cinco anos de uma concessionária norte americana são utilizados. O período de análise compreende os meses de janeiro de 1988 a outubro de 1992. As previsões realizadas foram dos tipos um passo à frente e recursiva de 1 a 24 passos à frente. Para tanto, dividiu-se o problema de previsão de carga em duas partes (Fig. 1.1), a saber:

- problema de representação;
- problema de aproximação funcional.

O problema de representação diz respeito à maneira pela qual a informação contida nos dados é codificada para uso futuro. Este problema é abordado em dois estágios. No primeiro deles, particiona-se uma série de carga via transformada wavelet em  $n+1$  componentes, sendo uma componente de baixa frequência (ou tendência) mais  $n$



componentes de alta frequência (ou detalhes). A partir da nova representação do sinal original, duas alternativas são investigadas. A primeira delas é nova e consiste em criar um modelo para PCCP cujas entradas são compostas de sinais de carga e sinais no domínio de wavelets. A segunda abordagem consiste em modelar a série de carga (no domínio do tempo) através da modelagem individual de cada componente no domínio de wavelets. Isto é, cada uma das  $n+1$  componentes é modelada separadamente e a previsão final é computada através da combinação das previsões fornecidas por cada um dos modelos. Essa idéia foi apresentada em [Ning & Yunping, 1998], [Zhang & Dong, 2001] e [Kim et al., 2002]. A diferença básica entre as duas alternativas é o modo pelo qual as características das subséries no domínio de wavelets são tratadas. Nos trabalhos anteriores, após prever-se cada subsérie de carga, a previsão final é obtida através de reconstrução, i.e., do somatório ponto a ponto de todas as subséries previstas. Porém, na alternativa aqui proposta, este tipo de reconstrução não se faz necessário e as características das diferentes subséries são combinadas em um mesmo modelo.

Já no segundo estágio do problema de representação, averigua-se a aplicabilidade de técnicas de reconstrução do espaço-fase de um sistema dinâmico. No caso de PCCP, está-se interessado em estabelecer um mapeamento entre um dado vetor de carga na hora corrente  $t$  e a carga na hora seguinte ( $t+1$ ). Em outras palavras, objetiva-se determinar quais seriam as melhores variáveis de entrada para um previsor neural.

Por outro lado, com relação ao problema de aproximação funcional, em linhas gerais, o mesmo pode ser definido como um mapeamento entre pares de entrada e saída realizado por algum dispositivo. A título de exemplo, um mapeamento não linear pode ser conseguido utilizando-se um Percéptron de Múltiplas Camadas (PMC) treinado com o algoritmo de retropropagação de erros padrão. Hornik et al. [Hornik et al., 1989] mostraram que tal dispositivo, usando uma função de transferência arbitrária, pode aproximar qualquer função mensurável de maneira precisa e satisfatória, se a camada oculta for composta por um número suficientemente grande de neurônios.

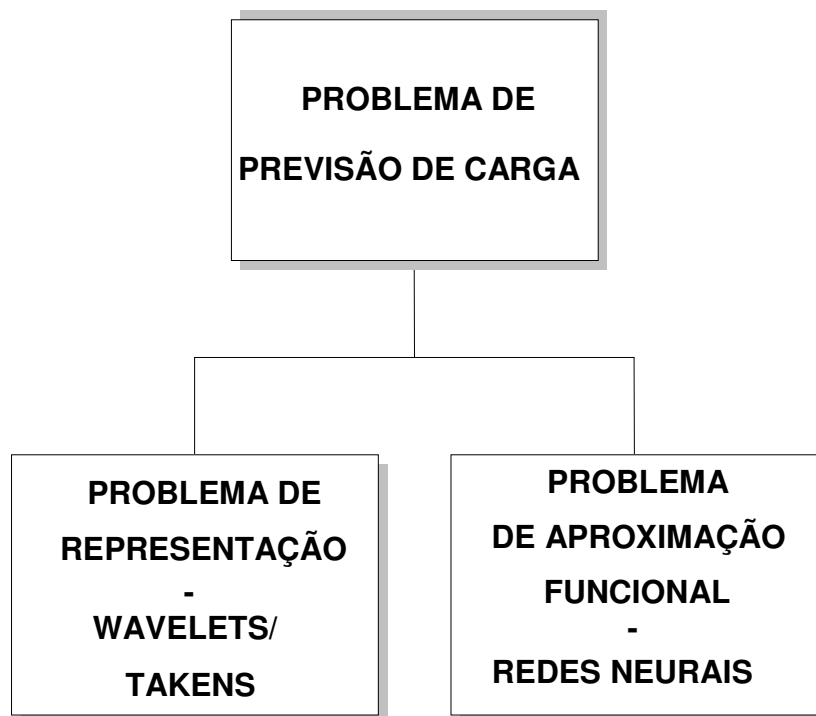


Fig. 1.1. Decomposição do problema de previsão de carga.

### 1.2.3 Revisão bibliográfica sobre wavelets aplicada à PCCP

Há alguns trabalhos publicados na literatura especializada que tratam da combinação entre wavelets e previsores neurais. De um lado, [Ning & Yunping, 1998], [Zhang & Dong, 2001] e [Kim et al., 2002], anteriormente citados, prevêm o comportamento futuro da carga através da previsão independente de cada subsérie no domínio de wavelets. Por outro lado, [Oonsivilai e El-Hawary, 1999] e [Huang & Yang, 2001] avaliam a aplicabilidade das chamadas *wavenets*<sup>1</sup> para fins de previsão. Uma breve revisão bibliográfica desses artigos é apresentada na seqüência.

Ning e Yunping apresentaram um método para previsão de carga a curto prazo baseado em RNA e na transformada wavelet [Ning e Yunping, 1998]. A série de carga é mapeada em algumas subséries com a transformada wavelet e então essas subséries são previstas por RNA. Fatores climáticos são levados em conta na previsão. Depois que todas as subséries são previstas, a série de carga prevista total pode ser obtida via reconstrução. Além disso, um novo algoritmo de retropropagação de erros é proposto

<sup>1</sup> Dispositivo híbrido entre redes neurais e wavelets

para acelerar o processo de treinamento e melhorar o processo de convergência das redes. Os resultados obtidos mostram a viabilidade das metodologias supracitadas.

B.-L. Zhang & Z.-Y. Dong [Zhang & Dong, 2001] propuseram um modelo de previsão de carga a curto prazo para mercados de energia elétrica competitivos. O modelo é uma combinação entre wavelets e redes neurais do tipo PMC. A metodologia proposta é testada usando dados de carga de uma concessionária de energia elétrica da área de Queensland (Austrália). Variáveis climáticas não foram incluídas no modelo.

C.-I. Kim et al. [Kim et al., 2002] apresentaram um modelo de PCCP combinando-se wavelets e redes neurais do tipo Kohonen. Primeiramente, os dados de carga são separados em quatro grupos usando redes neurais de Kohonen. Em seguida, wavelets da família Daubechies dos tipos daub2<sup>2</sup>, daub4 e daub10 são adotadas para prever a carga horária. Os coeficientes wavelets são ajustados usando o método de regressão múltipla convencional e as componentes de baixa e alta frequências são reconstruídas para prever as cargas finais através da transformada wavelet inversa. Cinco níveis de decomposição são considerados.

Em [Oonsivilai e El-Hawary, 1999], os autores propuseram uma abordagem para a previsão de cargas elétricas comerciais usando uma wavenet. As wavelets do tipo Morlet e “Chapéu de Mexicano” foram usadas para gerar as funções de transferência dos neurônios da camada oculta. Uma wavenet foi treinada para cada tipo de carga. Os resultados obtidos mostraram que sob certas circunstâncias as wavenets podem apresentar um desempenho melhor do que as arquiteturas tradicionais utilizadas na previsão de cargas elétricas.

C.-M. Huang & H.-T. Yang [Huang & Yang, 2001] desenvolveram um previsor de carga a curto prazo baseado em Redes Wavelets Evolucionárias [RWE]. As RWE têm uma estrutura de três camadas que contém nós-wavelet na camada de entrada, nós-ponderadores na camada intermediária e nós-somadores na camada de saída, respectivamente (ver Fig. 1.2). As redes evoluem ajustando-se os parâmetros de translação e dilatação nos nós-wavelets (camada de entrada), e os ponderadores na camada intermediária. Valendo-se da habilidade de busca global da computação evolucionária, assim como da propriedade de multiresolução das wavelets, as RWE (assim construídas) podem identificar as características não lineares presentes nas séries de carga. A abordagem proposta é verificada através de dados de carga e de algumas variáveis climáticas fornecidas pela *Taiwan Power Company*.

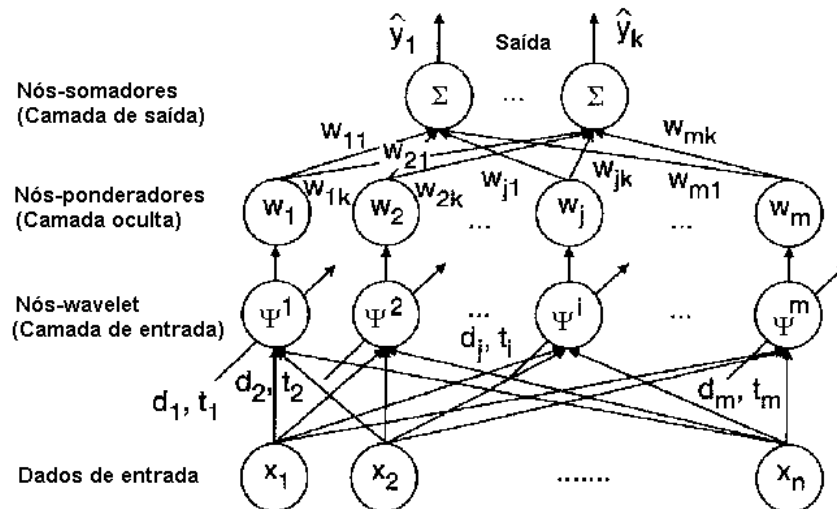


Fig. 1.2. Diagrama esquemático da wavenet utilizada em [Huang & Yang, 2001].

### 1.2.4 Contribuições

As principais contribuições deste trabalho são:

- 1) Melhoria da capacidade de generalização de uma rede neural através da extração das características mais relevantes de uma série de carga via análise de multiresolução – Espera-se que com uma melhor representação do conhecimento contido na base de dados de carga, o trabalho do previsor neural seja facilitado.
- 2) Inclusão da variável temperatura na PCCP via RNA e wavelets – A grande maioria dos artigos publicados referentes a esse tópico considera apenas a informação fornecida pela série de carga. É verdade que em [Zhang & Dong, 2001], os autores consideram a informação acerca da carga e da temperatura. No entanto, a abordagem adotada por eles é bastante distinta daquela tratada aqui.
- 3) Apresentação de um esquema prático de enchimento<sup>3</sup> para lidar com o problema conhecido como distorção de bordas – Este assunto tem sido sistematicamente ignorado na literatura.

<sup>2</sup> *daub2* significa “Wavelet Daubechies de ordem 2”.

<sup>3</sup> *Padding*, do inglês.

### **1.3 Organização da tese**

Esta tese é organizada como se segue. O capítulo 2 apresenta o tópico PCCP e suas aplicações. Além disso, uma revisão bibliográfica é relatada. A teoria sobre wavelets é assunto do capítulo 3. Discute-se a reconstrução de espaços de estados no capítulo 4. No capítulo 5 são apresentados os testes realizados, seus respectivos resultados e uma discussão dos mesmos. O capítulo 6 é dedicado a conclusões, a trabalhos futuros e contribuições. Finalmente, as redes neurais artificiais, a transformada de Fourier discreta e a wavelet de Haar são discutidas nos Anexos I, II e III, respectivamente.

## Capítulo 2 - Previsão de Carga Elétrica a Curto Prazo

Neste capítulo, apresentar-se-á uma visão geral do problema de previsão de carga. Uma breve revisão bibliográfica sobre a PCCP via RNA também será apresentada.

### 2.1 Visão geral do problema

É bem conhecido que grandes blocos de Energia Elétrica (EE) não podem ser armazenados eficientemente. Devido ao fato de que a carga elétrica só pode ser controlada pelas concessionárias de EE em uma pequena faixa (i.e., grandes variações de carga em curtos intervalos de tempo, via de regra, são difíceis ou até mesmo impossíveis de serem fornecidas), atividades de previsão de carga se tornam um tópico de grande interesse e importância para a geração e transmissão de EE.

A operação diária e as atividades de planejamento de uma concessionária de EE requerem a previsão da demanda de carga de seus consumidores. Em geral, a previsão de carga pode ser dividida em quatro tipos, a saber:

- previsão a curtíssimo prazo;
- previsão a curto prazo;
- previsão a médio prazo;
- previsão a longo prazo.

Na previsão a curtíssimo prazo, o horizonte de interesse é de uns poucos minutos a uma hora à frente. Já a previsão de carga a curto prazo refere-se à previsão horária da carga, i.e., das próximas 24 horas a uma semana à frente. A médio prazo, pode ser, ou uma previsão horária ou de pico de carga, para um horizonte de um a vários meses. Por conseguinte, a previsão a longo prazo se refere a previsão feita para períodos superiores a um ano. É importante salientar que esses horizontes de previsão têm um caráter orientativo, já que os mesmos variam de concessionária para concessionária, segundo sua conveniência.

Em geral, o comportamento da carga é influenciado por fatores típicos tais como:

- influências climáticas (e.g., temperatura, velocidade do vento, umidade relativa do ar);
- hora do dia, em um dia específico da semana;

- tarifação;
- estação do ano;
- eventos especiais, tais como, decisões de campeonatos, especiais de TV, etc.;
- fatores aleatórios desconhecidos.

Os objetivos para cada horizonte de previsão são diferentes. Na escala dos segundos, quando são observadas variações de carga aleatórias e de pequena magnitude, o Controle Automático de Geração (CAG) garante que a geração *on-line* atenda a carga. Além disso, para permitir uma avaliação segura das condições de operação em tempo real, o Estimador de Estado do Sistema de Potência (EESP) é usado para validar grandezas telemedidas e estimar a magnitude de tensão e o ângulo em cada barra. Estes valores podem ser utilizados para computar estimativas para a carga instantânea. Para períodos da ordem de minutos, onde variações de carga mais apreciáveis são sentidas, a função Despacho Econômico (DE) é usada para garantir o atendimento da carga de forma econômica considerando a geração disponível. Já na escala de tempo de horas e dias, variações de carga ainda maiores estão presentes, envolvendo a entrada/saída de unidades geradoras e o intercâmbio de potência com sistemas vizinhos (intercâmbio entre áreas), o que é realizado através de funções de controle da geração tais como programação da geração hidráulica, coordenação hidrotérmica, pré-despacho e programação de intercâmbios. Finalmente, em intervalos de semanas, onde se verificam oscilações de carga mais amplas, funções ligadas à programação energética e manutenção de equipamentos são executadas, buscando o atendimento da carga da forma mais econômica e segura possíveis.

Em linhas gerais, com a previsão a curtíssimo e a curto prazo, o que se objetiva é atuar no controle e operação *on-line* de sistemas elétricos de potência. Já na previsão a médio e a longo prazo, deseja-se estimar quanto será necessário investir no sistema elétrico<sup>1</sup>.

Outro aspecto importante se refere à qualidade da PCCP. Ela tem um impacto importante na operação econômica da concessionária, já que muitas das decisões tomadas, baseadas naquelas previsões, apresentam conseqüências econômicas significativas. Essas decisões incluem, e.g., programação da geração, programação das compras de combustível, avaliação da segurança do sistema, planejamento para transações de energia, entre outras.

## 2.2 A Importância da PCCP

A PCCP desempenha um papel chave na formulação de estratégias de operação econômica, confiável e segura para o sistema de potência. O principal objetivo da função PCCP é fornecer previsões de carga para:

- programação da geração básica;
- avaliação da segurança do sistema de potência em qualquer instante;
- o comprometimento de unidades;
- fornecer informação adequada para o despachante .

A aplicação primária da função PCCP é fornecer dados para as funções de programação que determinam o comprometimento mais econômico das fontes de geração de forma consistente com os requisitos de confiabilidade, restrições operacionais e limitações físicas, ambientais e de equipamentos. Para sistemas puramente hidráulicos, a previsão de carga é necessária para que a função de programação hidráulica determine o volume vertido pelos reservatórios e os níveis de geração nas usinas. Para sistemas puramente térmicos, a função pré-despacho necessita da previsão de carga para determinar as estratégias horárias de custo mínimo para a entrada e saída de unidades de modo a atender a carga prevista. Para sistemas mistos (hidráulicos e térmicos) as cargas previstas são utilizadas pela função de coordenação hidrotérmica na programação da operação horária das várias unidades geradoras de forma a minimizar os custos operacionais. Estas funções necessitam da carga prevista para o próximo dia ou para a próxima semana, de modo a determinar o planejamento da operação de custo mínimo sujeito às várias restrições impostas na operação do sistema. Uma tarefa intimamente associada à previsão de carga é a programação de intercâmbios. Para essa função, as previsões de carga a curto prazo são também usadas para determinar as políticas de intercâmbio com outras empresas.

A segunda aplicação da PCCP é para a avaliação preditiva da segurança do sistema de potência. A previsão de carga do sistema é um requisito essencial para a análise off-line, de forma a detectar as condições futuras sob as quais o sistema de potência pode estar vulnerável. Esta informação permite aos despachantes tomar as ações corretivas necessárias (e.g., entrada de unidades de pico, corte de carga, compra de energia, operações de chaveamento) com o objetivo de operar o sistema de potência da forma mais segura.

---

<sup>1</sup> Planejamento da expansão do sistema.



É importante salientar que existem funções de previsão auxiliares, tais como, previsão da carga por área e previsão por barra. Para empresas que possuem uma extensa classe de zonas geográficas ou áreas com diversidade climática, a função de previsão de carga por área fornece a previsão da carga total. Estas previsões de curto prazo por área são necessárias para a regulação de fluxos nas linhas de interligação entre áreas, para programação da geração por área e para as funções de previsão de carga por barramento, caso fatores de participação sejam utilizados. As previsões de carga por barra são necessárias para a análise da segurança tanto on-line quanto off-line.

## 2.3 Características gerais das cargas elétricas

As cargas de um sistema de potência são compostas por milhares de aparelhos e dispositivos eletro-eletrônicos que, quando usados individualmente, consomem energia de maneira bastante aleatória. Entretanto, quando considerados em conjunto, apresentam certos padrões de consumo que podem ser estatisticamente previstos. Tais padrões refletem a influência de uma série de fatores cujo entendimento é essencial para se conseguir modelos que representem as cargas do sistema com o propósito de se realizar previsões.

Pode-se dizer que a carga total do sistema é influenciada, a curto prazo, por fatores classificados nas seguintes categorias:

- temporais;
- meteorológicos;
- aleatórios;
- determinísticos.

As categorias relacionadas anteriormente serão descritas a seguir.

### *Fatores Temporais*

Basicamente, existem 3 fatores de origem temporal que influenciam diretamente o comportamento da carga: periodicidade diária-semanal, variações sazonais e ocorrência de feriados. A periodicidade diária-semanal é uma manifestação do ciclo trabalho-descanso da sociedade durante semanas típicas de determinada estação do ano. Já as variações sazonais refletem a influência das mudanças de estação ao longo do ano. Certas alterações no padrão

de comportamento da carga são graduais, como as que são reflexo do encurtamento/alongamento do período de luz natural ou aumento/declínio da temperatura. Outras são abruptas, como aquelas relativas à adoção do horário de verão, alteração significativa do nível de atividade em períodos de férias ou início do ano escolar, etc. Finalmente, a ocorrência de feriados reflete o comportamento não rotineiro dos consumidores e normalmente produz o efeito de reduzir os valores da carga a níveis bem inferiores aos verificados em dias comuns. Além disso, observa-se que tal mudança de comportamento se antecipa e se prolonga em relação à ocorrência do feriado, produzindo um “alongamento” do mesmo.

### ***Fatores Meteorológicos***

As condições meteorológicas são responsáveis por mudanças significativas no comportamento da carga. Com destaque para a temperatura e a umidade, os seguintes fatores meteorológicos exercem também influência sobre a carga: velocidade e direção do vento, nível de luminosidade, nebulosidade e precipitação.

### ***Fatores Aleatórios***

Nesta categoria estão os fatores de natureza aleatória que produzem alterações no comportamento da carga sem ter relação com a base temporal ou com as condições meteorológicas. Entre estes fatores estão: entrada/saída de grande consumidores, greve de ampla repercussão, transmissão de programa de televisão de interesse geral, etc. Embora a ocorrência de tais eventos possa ser prevista, seus efeitos sobre a carga são incertos.

### ***Fatores Determinísticos***

Além dos fatores citados anteriormente, é reconhecido que certas ações tomadas pelo operador do sistema tem efeitos totalmente previsíveis sobre a carga. Essas ações de controle caracterizam a chamada *gerência de demanda* que recentemente vem sendo considerada como fator de controle efetivo. Entre as ações de controle que qualificam o gerenciamento de demanda, pode-se mencionar: redução deliberada da tensão, implementação de tarifas horosazonais, celebração de contratos de suprimento com previsão de condições de interrupção, hierarquização geográfica da confiabilidade, apelo ao público, etc.

## 2.4 Classificação quanto ao tipo de modelo de carga

A classificação considera basicamente dois modelos: modelos de pico de carga e modelos de curva de carga.

### 2.4.1 Modelos de pico de carga

Usualmente é modelado o pico de carga horário ou semanal, geralmente, como uma função das condições meteorológicas [Goh et al., 1986]. O fator tempo não desempenha papel algum em tais modelos que são tipicamente da forma:

*pico de carga (P) = carga base (B) + componente dependente dos efeitos meteorológicos (F(CM));*

em que a carga base  $B$  é uma componente média independente das condições meteorológicas a qual será adicionada à componente dependente  $F(CM)$ . As variáveis meteorológicas podem incluir a temperatura na hora do pico de carga ou uma combinação de temperaturas previstas ou históricas. Umidade, intensidade luminosa, velocidade do vento e precipitação têm sido também consideradas em tais modelos.

As vantagens desse modelo são a sua estrutura simples e a baixa exigência de quantidade de dados para inicialização e atualização. Os parâmetros do modelo são estimados através de regressão linear e/ou não linear. As desvantagens de tais modelos são que eles não definem tanto a hora de ocorrência do pico e nem fornecem qualquer informação acerca do formato da curva de carga.

### 2.4.2 Modelos de curva de carga

Tais modelos descrevem a carga como uma série temporal discreta ao longo do intervalo de previsão. O intervalo de tempo considerado na amostragem da carga é tipicamente uma hora, enquanto que a quantidade medida é geralmente a energia consumida neste intervalo em MWh.

Basicamente, existem dois tipos de modelos de curva de carga: modelos hora-do-dia e dinâmicos. Combinações destes dois tipos também são possíveis.

### **Modelo Hora-do-dia**

O modelo hora-do-dia define a carga  $Y_t$  em cada instante  $t$  do período de previsão de duração  $T$  para uma série temporal. Em sua forma mais simples, este modelo armazena os  $T$  valores de carga baseados no comportamento prévio da carga observada. Algumas empresas ainda utilizam o padrão de carga verdadeiro da semana anterior como um modelo para prever a carga da semana corrente. Alternativamente, um conjunto de curvas é armazenada para semanas típicas do ano, e para condições meteorológicas típicas, tais como, dias úmidos, secos e nublados, que podem ser combinados com o padrão de carga semanal mais recente para realizar a previsão.

Um modelo hora-do-dia típico pode assumir a seguinte forma:

$$Y_t = \sum_{i=1}^N \beta_i f_i(t) + \varepsilon_t, \quad t \in \tau$$

em que a carga no instante  $t$ ,  $Y_t$ , é considerada como sendo uma soma de um número finito de funções periódicas  $f_i(t)$ , senos e cossenos de período 24 ou 168 h, dependendo do horizonte de previsão. Os coeficientes  $\beta$  são considerados como parâmetros do modelo e  $\varepsilon_t$  representa o erro da modelagem, admitido ser um ruído branco. Os parâmetros  $\beta_j$  são estimados através da aplicação de regressão linear simples ou alisamento exponencial a um histórico de carga  $\{ Y_t, t \in \tau_{\text{past}} \}$  onde  $\tau_{\text{past}}$  é um intervalo de tempo que cobre o passado recente da carga [Schweppe, 1973].

A vantagem dos mesmos é que eles são estruturalmente simples. Por outro lado, a adaptabilidade destes modelos é baseada em parâmetros de ajuste arbitrários e que geralmente não se modificam após a calibração inicial, o que faz com que previsões insatisfatórias sejam obtidas para condições não usuais.

### **Modelos Dinâmicos**

Os modelos dinâmicos reconhecem o fato de que a carga não é apenas uma função do parâmetro tempo, mas, também, função de entradas de informação meteorológica e de entradas aleatórias. Os modelos dinâmicos são basicamente de dois tipos: modelos de Box & Jenkins e modelos de Espaço de Estados.

## 2.5 Técnicas de previsão

Basicamente, as técnicas de previsão de carga podem ser divididas em convencionais e não convencionais. Por técnicas convencionais de previsão, entende-se aquelas técnicas baseadas em modelos estatísticos clássicos. Existem dois tipos principais, são eles:

- modelos de séries temporais;
- modelos de regressão.

No primeiro tipo, a previsão é baseada em valores passados de uma variável e/ou erros de previsão passados. O objetivo desses modelos é reconhecer ou identificar um padrão contido nos dados disponíveis e extrapolá-lo em direção ao futuro (e.g., alisamento e métodos de Box & Jenkins). Já os modelos de regressão consideram que as variáveis a serem previstas têm uma relação causa-efeito com uma ou mais variáveis independentes. O objetivo desse segundo modelo é descobrir a forma desta provável relação e usá-la para prever as variáveis dependentes através do uso dos valores futuros das independentes (e.g., regressão simples ou múltipla).

Por outro lado, as técnicas não convencionais de previsão se apresentam como uma alternativa àquelas baseadas em modelos estatísticos clássicos. Uma característica marcante dessas técnicas é sua adaptabilidade frente as alterações do ambiente onde se realizam as previsões. Os chamados Sistemas Especialistas (SE) e as redes neurais artificiais são dois exemplos de técnicas não convencionais de previsão que têm sido largamente empregados nessa última década (separadamente, ou em conjunto, quando recebem o nome de *Sistemas Híbridos*). Devido aos bons resultados que diversos pesquisadores têm conseguido, inúmeros artigos já foram publicados nas principais revistas e periódicos internacionais.

### 2.5.1 Métodos Convencionais

Nesta subseção serão descritos, de forma sucinta, os métodos estatísticos mais usados para a solução do problema da PCCP. Em [IEEE Committee report, 1981] podem ser encontradas revisões bibliográficas extensas, focalizando desde aspectos filosóficos e metodológicos gerais, até aspectos econômicos relativos ao tema.

### 2.5.1.1 Modelos Baseados em Regressão

A análise de regressão diz respeito a modelagem das relações entre variáveis. Esta análise quantifica como uma variável de resposta (dependente) está relacionada com um conjunto de variáveis explanatórias (independentes, preditoras). Por exemplo, a carga do sistema pode ser encontrada em termos das variáveis explanatórias, tais como variáveis climáticas e não climáticas que influenciam a carga elétrica.

Em sua forma mais geral, o modelo de regressão pode ser escrito como:

$$Y_t = f(x_t; \beta) + \varepsilon_t \quad \text{Eq.2.1}$$

onde  $f(x_t; \beta)$  é uma função matemática das  $p$  variáveis independentes  $x_t = (x_{t1}, x_{t2}, \dots, x_{tp})^t$  e dos parâmetros desconhecidos  $\beta = (\beta_1, \beta_2, \dots, \beta_m)^t$ , e  $Y_t$  é a série temporal. O modelo em (2.1) é probabilístico, uma vez que o erro  $\varepsilon_t$  é uma variável aleatória. É admitido que os erros  $\varepsilon_t$  são não correlacionados. Devido à natureza aleatória do erro, a variável independente  $Y_t$  também é uma variável aleatória. Um dos trabalhos pioneiros nessa área pode ser encontrado em [Heinemann et al., 1966].

Uma objeção para a utilização do método de regressão é que ele requer um número muito grande de dados para a estimação dos parâmetros do modelo. Essa metodologia também requer uma extensa análise off-line e a exatidão dos resultados depende fortemente do modelo inicialmente assumido. A aplicação de regressão à modelagem de séries temporais geralmente ignora as correlações seriais, produzindo modelos não parcimoniosos com coeficientes espúrios [Abraham & Ledolter, 1983].

### 2.5.1.2 Modelos de Alisamento Exponencial

Uma grande classe de métodos de previsão é a do tipo alisamento exponencial<sup>2</sup>. Técnicas específicas desse tipo assumem que os valores extremos da série representam a aleatoriedade e, assim, através do alisamento desses extremos pode se identificar o padrão básico.

Em geral, tais modelos podem ser escritos da seguinte forma:

$$Y_t = f(t; \beta) + \varepsilon_t \quad \text{Eq.2.2}$$

onde  $f(t; \beta)$  é uma função do tempo  $t$ , os coeficientes  $\beta$  desconhecidos e  $\varepsilon_t$  são erros não correlacionados.

Através da escolha apropriada das funções de ajuste  $f(\cdot)$ , pode-se representar uma variedade de séries sazonais e não sazonais. Essa função também pode representar a tendência de comportamento da série. No caso da carga elétrica,  $f(\cdot)$  é usualmente uma função polinomial do tempo.

Dentre os vários modelos de alisamento exponencial podemos citar:

- **Modelos para séries localmente constantes:** Médias Móveis Simples (MMS), Alisamento Exponencial Simples (AES) e Exponencial Adaptativo de Trigg & Leach;
- **Modelos para séries que apresentam tendência:** Alisamento Exponencial Linear de Brown (AELB), Método de Holt-2 parâmetros (MH2P), Alisamento Exponencial Quadrático de Brown (AEQ);
- **Modelos para séries sazonais:** Alisamento Exponencial Sazonal de Holt-Winters (HW), Alisamento Exponencial Geral (Método de Brown) (AEG).

Maiores detalhes sobre tais métodos podem ser vistos em [Abraham & Ledolter, 1983; Brockwell & Davis, 1996]. As seguintes desvantagens devem ser mencionadas: os erros do modelo ajustado são serialmente correlacionados - é difícil saber quantos harmônicos devem ser ajustados para uma dada série temporal sazonal; problemas quanto a utilização de uma única constante de alisamento; além disso, a exatidão das previsões depende fortemente da constante de alisamento. Essa metodologia não permite considerar mais de uma série temporal.

---

<sup>2</sup> *Exponential Smoothing*, do inglês.

### 2.5.1.3 Séries Temporais Estocásticas (Modelos Box & Jenkins)

A metodologia é baseada na análise dos dados, i.e., é permitido que os dados gerem o modelo a ser usado, embora, ainda requerendo conhecimento e experiência do analista.

Basicamente, a metodologia Box & Jenkins assume que qualquer série temporal estacionária  $Y_t$  pode ser modelada parcimoniosamente<sup>3</sup> por uma classe de modelos lineares. A aplicação do método a uma série qualquer é feita em quatro etapas distintas, a saber:

- a) identificação do(s) modelo(s) para a série;
- b) estimação dos parâmetros do modelo identificado;
- c) testes estatísticos para a verificação da adequabilidade do modelo;
- d) previsão dos valores futuros da série.

Cada uma destas etapas requer intervenção e análise do usuário, onerando o custo de utilização do mesmo, muito embora, já existam versões automáticas. Dada a natureza linear dos demais métodos, estes podem ser considerados casos particulares do método de Box & Jenkins. Por esta razão, o usuário deve, sempre que possível, utilizar esta metodologia em detrimento das outras, apesar do seu maior custo de utilização. A seguir serão analisados os modelos ARMA (*Autoregressive Moving Average*) e os modelos de Função de Transferência de Box & Jenkins.

#### **Modelos ARMA e ARIMA**

No processo ARMA, o valor corrente da série temporal  $Y_t$  é função linear de seus valores nos períodos anteriores ( $Y_{t-1}, Y_{t-2}, \dots$ ) e dos valores corrente e anteriores de um ruído branco ( $a_{t-p}, a_{t-2}, \dots$ ). Um processo ARMA de ordem p e q (i.e., ARMA(p,q)) pode ser escrito da seguinte forma:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad \text{Eq.2.3}$$

Introduzindo-se o operador de atraso  $B$  que define  $Y_{t-1} = B.Y_t$ , e conseqüentemente,  $Y_{t-m} = B^m Y_t$ , a equação anterior pode ser escrita da seguinte forma:



$$\phi(B)Y_t = \theta(B)a_t, \quad \text{Eq.2.4}$$

em que:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \quad \text{Eq.2.5}$$

é o polinômio **AR** simples de grau “p”;

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q, \quad \text{Eq.2.6}$$

é o polinômio **MA** simples de grau “q”,

e  $a_t$  é ruído branco.

No caso dos modelos puramente sazonais tem-se que:

$$(1 - \Phi_1 B^S - \dots - \Phi_{PS} B^{PS}) y_t = (1 - \Theta_1 B^S - \dots - \Theta_{QS} B^{QS}) \varepsilon_t, \quad \text{Eq.2.7}$$

onde  $\Phi$  e  $\Theta$  são os parâmetros do modelo sazonal, e S é o período sazonal (S = 24, para sazonalidade diária, e S = 168, para sazonalidade semanal). Como as séries temporais sazonais têm, em geral, componentes não sazonais, o modelo deve ser estimado como misto sazonal-não sazonal. A representação do modelo é feita como se segue:

$$(1 - \phi_1 B - \dots - \phi_p B^p) (1 - \Phi_1 B^S - \dots - \Phi_{PS} B^{PS}) y_t = (1 - \theta_1 B - \dots - \theta_q B^q) (1 - \Theta_1 B^S - \dots - \Theta_{QS} B^{QS}) \varepsilon_t \quad \text{Eq.2.8}$$

Assim, temos um modelo misto ARMA(p,q)(P,Q)<sub>s</sub>, onde p e q referem-se, respectivamente, à ordem autoregressiva e de média móvel, enquanto que P e Q referem-se, respectivamente, à ordem AutoRegressiva Sazonal (SAR) e de Média Móvel Sazonal (SMA).

Um dos requisitos para a utilização dos modelos supraditos é que a estacionaridade deve ser obtida antes de sua aplicação. A estacionaridade pode ser investigada tomando-se diferenças da série e analisando-se as funções de Autocorrelação (ACF) e Autocorrelação

<sup>3</sup> Representação adequada com poucos parâmetros

Parcial (PACF). Em muitas situações, diversas outras transformações devem ser consideradas (e.g., log, potência, Johnson, etc.). Portanto, se a série em estudo for não estacionária e apresentar períodos sazonais e não sazonais, deve-se considerar a utilização de um modelo do tipo ARIMA (p,d,q)(P,D,Q)<sub>s</sub>, onde *d* representa a ordem de diferenciação não sazonal e *D* a ordem de diferenciação sazonal. As diferenciações sazonal e não sazonal podem ser realizadas da seguinte forma:

$$\Delta^d y = y_t - y_{t-1} \text{ (diferença não sazonal),}$$

$$\Delta^D y = y_t - y_{t-s} \text{ (diferença sazonal } s\text{).}$$

Em situações normais, segundo [Moretin e Toloi, 1987], será suficiente tomar uma ou duas diferenças para que a série se torne estacionária. O número *d* de diferenças necessárias para tornar a série estacionária é denominado ordem de integração. Portanto, introduzindo-se o operador  $\nabla$ , uma série temporal diferenciada de ordem 1 pode ser escrita como  $\nabla Y_t = Y_t - Y_{t-1} = (I - B) Y_t$ , usando-se a definição do operador de atraso **B**. Consequentemente, uma série temporal diferenciada de ordem “*d*” é escrita como  $\nabla^d Y_t = (I - B)^d Y_t$ . Assim, o modelo de Box & Jenkins toma a forma ARIMA (p, d, q) e pode ser escrito da seguinte maneira:

$$\phi(B)\nabla^d Y_t = \theta(B)a_t \tag{Eq.2.9}$$

sendo “*d*” a ordem da série diferenciada.

Uma questão crucial a ser respondida em uma análise de série temporal univariada é: Qual o processo (modelo) que melhor explica a dinâmica da série temporal em análise? Além disso, qual a ordem de (p,d,q) e/ou (P,D,Q)<sub>s</sub>? Essas questões podem ser respondidas com base na metodologia de Box-Jenkins [Box & Jenkins, 1976] descrita abaixo:

1. **Identificação** - Consiste em descobrir qual dentre as várias versões dos modelos de Box-Jenkins, sejam eles sazonais ou não, descreve o comportamento da série. A identificação do modelo a ser estimado ocorre pelo comportamento das ACFs e PACFs e seus respectivos correlogramas. Outros detalhes referentes à obtenção dessas funções e a quais comportamentos representam os modelos anteriormente abordados podem ser encontrados em [Makridakis et al., 1998];

2. **Estimação** - Consiste em estimar os parâmetros  $\phi$  e  $\Phi$  do componente auto-regressivo, os parâmetros  $q$  e  $Q$  do componente de médias móveis e a variância de  $\varepsilon_t$ .

3. **Verificação** - Consiste em avaliar se o modelo estimado é adequado para descrever o comportamento dos dados.

Caso o modelo não seja adequado, o ciclo é repetido, voltando-se à fase de identificação (Fig.2.1). Um procedimento muito utilizado é identificar não só um único modelo, mas alguns modelos que serão então estimados e verificados. Quando se obtém um modelo satisfatório, passa-se para a última etapa da metodologia de Box-Jenkins, que trata da realização de previsões.

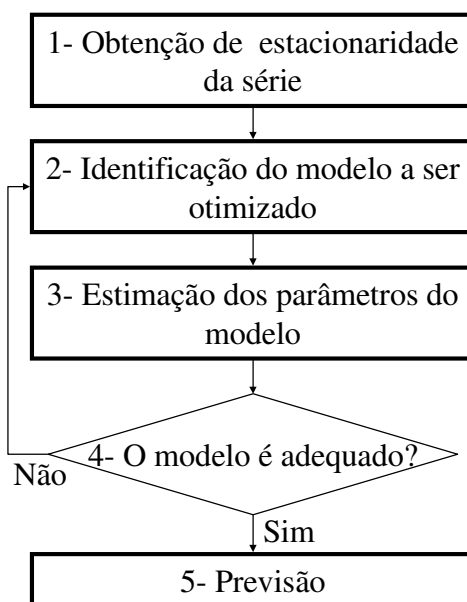


Fig.2.1. Fluxograma do modelo ARIMA de Box & Jenkins.

## Considerações Finais

A identificação dos parâmetros de um modelo ARMA é uma tarefa computacionalmente intensiva; entretanto, esse esforço extra é necessário com a finalidade de se obter um modelo mais robusto que incorpore os efeitos dinâmicos, climáticos e aleatórios [Schweppe, 1973].

Um problema com os modelos de Box & Jenkins descritos anteriormente é que as informações meteorológicas não são incluídas explicitamente, i.e., como entradas do modelo.

Modelos que não incluem as informações meteorológicas, automaticamente atualizam alguns parâmetros para levar em consideração o efeito das variações meteorológicas na carga. Esta abordagem, entretanto, não é satisfatória durante mudanças rápidas das condições climáticas sob as quais a hipótese de que o processo de carga é estacionário não é mais válido. A título de exemplo, em [Barakat et al., 1990], os autores encontraram para a previsão do pico de carga, erros variando de um mínimo de 2,31% a um máximo de 69,48%. Os modelos de Box & Jenkins que consideram explicitamente a informação meteorológica (conhecidos como Modelos de Função de Transferência) não serão descritos aqui.

### 2.5.2 Métodos Não Convencionais

Durante as últimas décadas, vários artigos sobre previsão de carga têm sido publicados. Alguns deles contam com sofisticados modelos que, de certa maneira, dependem da forma da curva de carga e de variáveis como temperatura, umidade, velocidade do vento, nebulosidade, entre outras. Apesar de que todas essas informações possam ser extraídas e atualizadas automaticamente através de estatísticas aplicadas ao histórico disponível, os modelos tradicionais de carga-base usam abordagens que são inflexíveis às várias mudanças no comportamento do consumidor.

Surge então, a necessidade de se conceber técnicas que não estejam somente baseadas em análises matemáticas complexas, mas que também possuam a capacidade de adaptação frente a alterações do ambiente onde se realizam as previsões.

Essas considerações, e o fato de que num passado recente *previsores humanos* terem gerado resultados aceitáveis, permitiram o desenvolvimento de ferramentas baseadas na emulação do conhecimento, experiência e lógica do especialista humano através do desenvolvimento dos Sistemas Especialistas.

Outra metodologia que tem-se destacado bastante ultimamente é aquela baseada em RNA. Essa abordagem tem alcançado sucesso na modelagem de séries temporais, devido a sua capacidade de adaptação a conjuntos de dados com diferentes características, muito além dos modelos lineares tradicionais. Além disso, tem-se mostrado extremamente robusta no tratamento das correlações existentes entre as variáveis e as interdependências empíricas entre as diferentes seqüências temporais.

## 2.6 Redes Neurais Artificiais como solução para o problema de previsão de carga

As redes neurais artificiais têm demonstrado serem capazes de encontrar representações internas de interdependências dentro de um histórico de dados disponível que não sejam explicitamente dadas ou mesmo conhecidas por um especialista humano [Niebur,1993]. Elas apresentam duas características bastante atrativas e que incentivaram vários pesquisadores a aplicarem-nas à tarefa de previsão de carga. São elas:

- as RNA são uma representação versátil. Ao invés de termos que escolher entre representações específicas AR, ou MA, ou ARMA, ou ARIMA, etc., basta lhe apresentar uma quantidade de dados representativos do problema em questão e deixar que ela extraia o conhecimento necessário para se fazer previsões diretamente desses dados;
- as RNA são muito mais tolerantes a dados com ruídos (e.g., *outliers*) do que a maioria dos métodos clássicos de previsão.

Muitos trabalhos apresentam a aplicação de RNA na PCCP. Várias concessionárias de EE ao redor do mundo utilizam previsores baseados em redes neurais em caráter de testes ou em operação [Mori, 1996; Papalexopoulos, 1996; Khotanzad et al., 1998].

## 2.7 Revisão Bibliográfica sobre a PCCP via RNA

Em [Park et al., 1991], uma nova abordagem para o problema de PCCP empregando-se uma RNA como componente principal do previsor foi apresentada. Os autores utilizaram uma rede neural do tipo PMC treinada com o algoritmo de retropropagação de erros padrão. Uma função do tipo logística foi utilizada como função de ativação nas camadas oculta e de saída. Os previsores via RNA foram usados na previsão da carga de pico diária, energia diária total, e carga diária horária. Para cada tipo de previsão, a mesma configuração de RNA, consistindo de uma camada oculta, foi utilizada. Todos os três previsores foram treinados com dados reais da concessionária de EE *Puget Sound Power and Light Company*, compreendendo um período de três meses. Dias atípicos e feriados foram excluídos dos conjuntos de treinamento das redes. Os erros médios observados foram 2,04%, 1,68% e 1,40% para carga de pico, carga total e carga horária, respectivamente.

Em [Papalexopoulos et al., 1994], os autores fazem uma avaliação comparativa entre um sofisticado modelo de previsão baseado em regressão e um novo modelo baseado em RNA. O modelo de regressão estava em operação nas concessionárias dos EUA por vários anos e representava o *estado da arte* em previsão de carga via modelos estatísticos tradicionais. No entanto, quando os modelos foram confrontados, o novo modelo neural superou o modelo de regressão. Além de produzir melhores resultados, tanto o tempo quanto o custo de desenvolvimento do novo modelo foram significativamente menores. Como consequência, o modelo neural substituiu o modelo tradicional vigente. Esse artigo é importante porque avalia o desempenho de previsores via RNA em condições reais de operação. No entanto, há muito pouca informação sobre a identificação e a construção dos modelos neurais. Também, a seleção de variáveis de entrada para os previsores, assim como a escolha dos parâmetros de treinamento, são feitas baseadas em anos de experiência no sistema elétrico em questão.

Em [Mohammed et al., 1995], os autores relatam o desenvolvimento de um sofisticado previsor neural para a concessionária de EE *Florida Power and Light*. Cada estação do ano é dividida em três tipos de dia, e cada dia é ainda dividido em cinco períodos. Para cada um dos períodos, uma rede neural é destinada. O número de entradas varia de 5 a 16, dependendo da estação, período e tipo de dia sendo previsto. A seleção das entradas para cada RNA é feita por tentativa e erro, e conta com a experiência dos operadores do sistema. Adaptações diárias, semanais e mensais são realizadas para atualizar o conjunto inicial de pesos de acordo com os resultados mais recentes e padrões temporais. Esse sistema bastante sofisticado foi necessário devido à grande variação climática na região de operação da concessionária. No entanto, a avaliação geral dos modelos não foi mencionada no artigo.

Em [Liu et al., 1996], três técnicas práticas – Lógica Fuzzy (LF), RNA e modelos autoregressivos - aplicadas ao problema de previsão de carga a curtíssimo prazo, são apresentadas e discutidas. A série de carga é discretizada minuto a minuto e o horizonte de previsão varia de 1 a 30 minutos. Para o previsor via RNA, dentre outras, foram utilizadas como entradas:

- os 30 valores de carga imediatamente anteriores ao instante de tempo  $k+1$ , i.e.,  $L_{k-29}$ ,  $L_{k-28}$ ,  $L_{k-27}$ , ...,  $L_{k-2}$ ,  $L_{k-1}$ ,  $L_k$  (ressalta-se que  $L_k$  é o valor da carga no instante  $k$ ).

As 30 saídas da rede são as previsões para os próximos 30 minutos. O desempenho das três técnicas é avaliado através de simulações. O estudo realizado mostra que é viável projetar um

previsor dinâmico para prever tendências de carga em modo on-line, devido ao fato dos previsores baseados em LF e RNA apresentarem um desempenho bem superior àquele obtido pelo modelo autoregressivo.

Em [Bakirtzis et al., 1996], os autores utilizam uma rede neural para prever um ano completo com base diária. A estrutura da rede consiste de 63 entradas, 24 neurônios na camada oculta e 24 neurônios na camada de saída, representando cada hora do dia. A rede é desenvolvida para PCCP do sistema de potência grego, que em 1993 tinha um pico de 5500 MW. Os autores usam o ano anterior inteiro no processo de treinamento da RNA (feriados e dias atípicos são excluídos). Embora a rede apresentasse, contabilizando-se os *bias*,  $(63+1) \times 24 + (24+1) \times 24 = 2136$  parâmetros livres e ter sido treinada com aproximadamente 365 padrões de entrada e saída, os resultados foram bastantes satisfatórios. O Erro Percentual Absoluto Médio (MAPE) para um dia à frente foi de 2,34%. Infelizmente, os autores não fornecem informações detalhadas sobre a estratégia de treinamento desenvolvida, que seria desejável tanto do ponto de vista teórico quanto prático. No modo de previsão, a rede é retreinada diariamente usando-se os 365 padrões de entrada e saída mais recentes. O artigo também descreve o procedimento utilizado na previsão de feriados e dias atípicos, assim como para o período de 2-7 dias à frente (MAPE variando entre 2,89% e 4,00%).

Em [Khotanzad et al., 1998], os autores descrevem a terceira geração de um sistema de previsão de carga a curto prazo com base horária conhecido como ANNSTLF. Este previsor tem recebido larga aceitação pelas concessionárias de EE dos EUA e Canadá e está sendo utilizado por 35 delas. A arquitetura do previsor da geração três é bastante diferente daquela utilizada na geração anterior. Ela apresenta apenas dois previsores via RNA, sendo que um deles prevê a carga-base e o outro as variações da carga. Ambos previsores tem a mesma topologia: 79 entradas e 24 saídas. As entradas são:

- os 24 valores de carga do dia anterior e suas respectivas medições de temperatura;
- os próximos 24 valores de temperatura;
- 7 indicadores do tipo do dia para cada dia da semana.

Como saída, o primeiro previsor é treinado para fornecer as cargas para as próximas 24 horas. Já o segundo, prevê a variação horária da carga entre o dia anterior e o dia corrente. A previsão final é computada através de uma combinação adaptativa dos dois previsores. O efeito da umidade e velocidade do vento são considerados através de uma transformação linear da temperatura. Um novo esquema de interpolação ponderada é desenvolvido para a previsão de feriados. Primeiramente, o pico de carga para o feriado é estimado e então a

previsão fornecida pelo ANNSTLF é remodelada com o novo pico previsto. O desempenho das duas gerações do ANNSTLF é comparado através de simulações para 10 concessionárias dos EUA e Canadá. Para previsões de 1 a 24 passos à frente, o MAPE médio observado (considerando-se todas as concessionárias) para a geração três foi de 2,05% contra 2,26% da geração anterior. Isso representa um ganho de 9,29% (i.e.,  $\frac{2,26\% - 2,05\%}{2,26\%} \times 100\%$ ).

Em [Hippert et al., 2001], os autores examinam uma coleção de artigos publicados entre 1991 e 1999 que tratam da aplicação de RNA no problema de PCCP. O objetivo principal dos autores é avaliar, de maneira crítica, os modos pelos quais as RNA propostas nesses artigos foram projetadas e testadas. Os autores concluem o artigo sugerindo um maior volume de pesquisa sobre o comportamento dos previsores baseados em redes neurais e a adoção de critérios de avaliação mais rigorosos.

Em [Marín et al., 2002], um modelo global de PCCP via RNA é proposto. O modelo prediz os valores horários de carga para as 24 horas do dia seguinte. O desenvolvimento do modelo se divide em três estágios. No primeiro deles, partindo-se dos dados históricos, cada dia é classificado de acordo com o seu perfil de carga através de mapas auto-organizáveis. O segundo estágio consiste na construção e no treinamento de uma rede neural recorrente para cada classe. Finalmente, as previsões são realizadas utilizando-se a rede neural previamente treinada. A base de dados usada na etapa de testes representa o consumo de EE na área central da Espanha no período de 1989 e 1999. Exaustivos testes mostraram o bom desempenho da metodologia proposta.

Em [Taylor e Buizza, 2002], investiga-se a incorporação de múltiplos cenários para uma dada variável climática na PCCP via RNA. As previsões são realizadas para um horizonte de 1 a 10 dias à frente. Os resultados mostraram que a média dos cenários de carga é uma estimativa de previsão melhor do que aquela conseguida com o uso de informação acerca das variáveis climáticas previstas de forma tradicional.

Finalizando, a Tab. 2.1 apresenta, de forma sintética, as principais características de alguns dos trabalhos apresentados acima. Como pode ser visto a partir dela, todas as aplicações que utilizam RNA como o componente principal do previsor, compartilham as seguintes propriedades:

- as RNA são do tipo PMC;
- utilizam o algoritmo de retropropagação de erro para fins de treinamento.



A grande diferença entre as diversas abordagens está no processo de seleção de variáveis de entrada e na determinação da estrutura das redes. Tais tarefas foram realizadas de maneira empírica nesses artigos.

**Tab.2.1.** Comparação entre alguns trabalhos de PCCP via RNA.

Autores	Tipo de Rede/ Aprendizagem	Topologia e   o1   o2   s	Dados de entrada	Previsão / Erro	Nota
Park et. al. (1991)	1. BP / RDG com momento	3   5   0   1	1. Temp. média.(d) 2. Temp. min. (d) 3. Temp. max. (d)	Pico de carga do dia d Média 2,04% carga total (d) média 1,68%	Feriados e finais de semana foram excluídos. Erros relativos absolutos foram utilizados. Nova etapa de treinamento para cada semana com dados de uma semana.
	2. BP / RDG com momento	6   10   0   1	1. Hora h 2. Carga (h-2) 3. Carga (h-1) 4. Temp. (h-2) 5. Temp (h-1) 6. Temp. prevista (h)	carga horária (h) 1 hora à frente média 1,40%	
Liu et. al. (1996)	1. BP / RDG com momento	38  30  25  16	1-4. $\tau_k$ (i = 1,...,4) 5. $NIDEV_k$ 6. $FREDEV_k$ 7. $TACE_k$ 8. $FACE_k$ 9-38. $L_{ki}$ (i=0,...,29)	Carga de 1 a 30 minutos à frente  RMS típico variando entre 1 e 1,5%	Os autores sugerem que se o RMS médio sobre um período de 24 h for maior que, e.g., 1%, uma nova etapa de treinamento do preditor será necessária.
	2. Lógica Fuzzy	-	-	-	
	3. Modelo AR	-	-	-	
Bakirtzis et. al. (1996)	BP / RDG	63  24   0   24	1-48. Carga (h-i) i = 1,...,48 49. Temp. A max. (d-1) 50. Temp. A min. (d-1) 51. Temp. B max. (d-1) 52. Temp. B min. (d-1) 53. Temp. A max. (d) 54. Temp. A min. (d) 55. Temp. B max. (d) 56. Temp. B min. (d) 57.-63. Codificação para o dia da semana	1-24. Carga (h+i) i = 1,..., 24.  1 dia à frente 2.34% 2-7 dias à frente 2.89% - 4.00%  (média para 1 ano)	Treinamento é feito com uma janela móvel de um ano. A rede é retreinada a cada dia. Há previsões para feriados.
Khotanzad et. al. (1998)	1. BP / RDG	79  ?   0   24	1-24. Carga (h-i) i = 1,...,24 25-48. Temp. (h-i) i = 1,...,24 49-72. Temp. Prev. (h+i) i = 1,...,24 73-79. Codificação para o dia da semana	1-24. Carga-base (h+i) i = 1,..., 24.	A previsão final é computada através de uma combinação adaptativa dos dois preditores. Para previsões de 1 a 24 passos à frente, o MAPE médio observado foi igual a 2,05%. Também, um novo esquema de interpolação ponderada é desenvolvido para a previsão de feriados.
	2. BP / RDG	79  ?   0   24	1-24. Carga (h-i) i = 1,...,24 25-48. Temp. (h-i) i = 1,...,24 49-72. Temp. Prev (h+i) i = 1,...,24 73-79. Codificação para o dia da semana	1-24. variação horária da carga entre o dia anterior e o dia corrente	

\* BP – Retropropagação de Erro; RDG – Regra Delta Generalizada; d – dia previsto; h – hora prevista; neurônios: e - entrada, o1,o2 - ocultos, s

- saída

## **Capítulo 3 - A Transformada Wavelet Discreta**

Wavelets podem ser descritas como um sinal de curta duração com energia finita e média zero. Elas são localizadas no tempo (ou espaço), diferentemente das funções trigonométricas. Essa característica torna-as adequadas para a análise de vários sinais não estacionários [Strang & Nguyen, 1996].

### **3.1 Motivação para se utilizar a transformada wavelet**

Quando se utiliza a transformada de Fourier na análise de um dado sinal, é praticamente impossível determinar quando um evento em particular ocorreu porque as funções-base usadas na análise de Fourier são precisamente localizadas em frequência, mas são aplicadas em todos os instantes. Esse inconveniente sugere que a análise de Fourier é indicada para tratar de frequências que não evoluam com o tempo.

Por outro lado, séries de carga, via de regra, apresentam várias características não estacionárias, tais como: tendências, mudanças de nível e/ou inclinação, sazonalidades, entre outras. Geralmente, essas características são as componentes mais importantes do sinal de carga e devem ser levadas em consideração quando da sua modelagem. Portanto, nota-se que tais características intrínsecas da carga dificultam a utilização da transformada de Fourier na análise de tais sinais.

A análise via wavelets supera as limitações dos métodos baseados em Fourier utilizando funções que são locais tanto no tempo quanto na frequência. Essas funções permitem o uso de intervalos de tempo maiores quando a informação de baixa frequência é necessária e intervalos mais curtos quando a informação de alta frequência é requerida. Pode-se dizer que as wavelets têm uma janela que se adapta para fornecer a resolução desejada [Misiti et al., 1996].

### **3.2 Transformada Wavelet Contínua e Discreta**

A análise via wavelets emprega uma função protótipo chamada wavelet mãe. Essa função tem média zero e decai bruscamente de maneira oscilatória. Os dados são representados via superposição de versões dilatadas e transladadas da wavelet mãe pré-

especificada. Matematicamente, a Transformada Wavelet Contínua (TWC) de um dado sinal  $x(t)$ , com relação a wavelet mãe  $g(t)$ , é definida como na Eq. 3.1, a saber:

$$TWC(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) g\left(\frac{t-b}{a}\right) dt, \quad \text{Eq. 3.1}$$

em que  $a$  é o fator escala e  $b$  é o fator de translação.

Um coeficiente  $TWC(a, b)$ , em uma dada escala e fator de translação, representa quão bem o sinal original  $x(t)$  e a wavelet mãe dilatada/transladada se assemelham. Assim, o conjunto de todos os coeficientes  $TWC(a, b)$ , associados com um dado sinal  $x(t)$ , é a representação wavelet do sinal com respeito a wavelet mãe  $g(t)$ .

A título de exemplo, considere que o fator de escala de uma wavelet seja igual a  $2^j$  (e.g., uma “Coiflet”). À medida que  $j$  cresce, o fator de escala decresce, há um “encolhimento” no tempo, as oscilações crescem e a wavelet exhibe “alta frequência”. Por outro lado, se  $j$  decresce (a escala cresce e há uma expansão no tempo), as oscilações tornam-se mais lentas, i.e., apresentam um comportamento de “baixa frequência”. Esta é uma característica marcante das wavelets - componentes de alta frequência são analisados em intervalos de tempo curtos enquanto que componentes de baixa frequência são analisados em intervalos de tempo longos. Os coeficientes de wavelets caracterizam o comportamento local de um sinal. A Fig. 3.1 ilustra algumas famílias de wavelets.

Desde que a TWC é conseguida dilatando-se e transladando-se a wavelet mãe continuamente, uma grande quantidade de informação redundante é gerada. Portanto, ao invés de proceder dessa maneira, a wavelet mãe pode ser dilatada e transladada usando escalas e posições especiais baseadas em potência de 2 (i.e, escalas e posições diádicas). Esse esquema é muito mais eficiente e tão eficaz quanto a TWC [Strang & Nguyen, 1996]. Ele é conhecido como a Transformada Wavelet Discreta (TWD) e definida como:

$$TWD(m, k) = \frac{1}{\sqrt{a_0^m}} \sum_n x(n) g\left(\frac{k - nb_0 a_0^m}{a_0^m}\right) \quad \text{Eq. 3.2}$$

Os parâmetros  $a$  e  $b$  (como mostrado na Eq. 3.1) são funções do parâmetro inteiro  $m$ , i.e.,  $a = a_0^m$  e  $b = nb_0a_0^m$ . Ainda na Eq. 3.2,  $k$  é uma variável inteira que se refere a uma dada amostra no sinal de entrada.

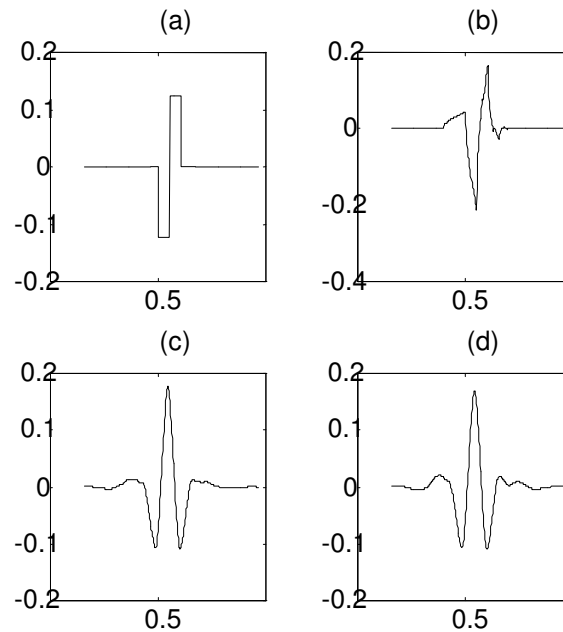


Fig. 3.1. Exemplos de wavelet mãe. (a) Haar; (b) Daubechies; (c) Coiflet; (d) Symmlet.

### 3.3. Análise de Multiresolução e Algoritmo Piramidal de Mallat

Um modo rápido e confiável de se implementar a TWD foi desenvolvido por Mallat em 1989 [Mallat, 1989]. Análise de Multiresolução (AMR) via algoritmo piramidal de Mallat refere-se ao procedimento de se obter “aproximações” e “detalhes” de um dado sinal. Uma aproximação é uma representação de baixa frequência do sinal original enquanto que um detalhe é a diferença entre duas aproximações sucessivas do sinal original. Uma aproximação mantém a tendência geral do sinal original enquanto que um detalhe mostra os componentes de alta frequência do mesmo.

O algoritmo piramidal apresenta dois estágios: um para o cálculo da TWD e outro para o cálculo da Transformada Wavelet Inversa (TWI). A multiresolução pode ser obtida utilizando-se um banco de filtros composto por  $L$ ,  $H$ ,  $L'$ , e  $H'$ , como mostrado na Fig. 3.2. Os filtros passa-baixa e passa-alta de decomposição ( $L$  e  $H$ ), juntamente com seus filtros de reconstrução associados ( $L'$  e  $H'$ ), formam um sistema chamado *Quadrature Mirror Filters*

(QMF). Uma discussão técnica sobre o projeto e a construção desses filtros pode ser encontrada em [Strang & Nguyen, 1996].

A partir do sinal  $S$  (i.e.,  $x(n)$ ), dois conjuntos de coeficientes podem ser produzidos a partir da TWD: coeficientes de aproximação  $cA_1$  e coeficientes de detalhe  $cD_1$ . Esses conjuntos são obtidos através da convolução de  $S$  com o filtro passa-baixa  $L$  para aproximação, e com o filtro passa-alta  $H$  para detalhe, seguido por decimação diádica, i.e., a cada duas saídas do filtro, descarta-se uma delas. Esse estágio, conhecido como decomposição (ou análise), determina o conjunto de coeficientes  $cA_1$  e  $cD_1$ .

Por outro lado, partindo-se de  $cA_1$  e  $cD_1$ , TWI reconstrói  $S$  através da inserção de zeros entre os coeficientes de wavelets (*up-sampling*), convoluindo-se o sinal resultante com os filtros de reconstrução  $L'$  e  $H'$ . Esse é o estágio de reconstrução (ou síntese).

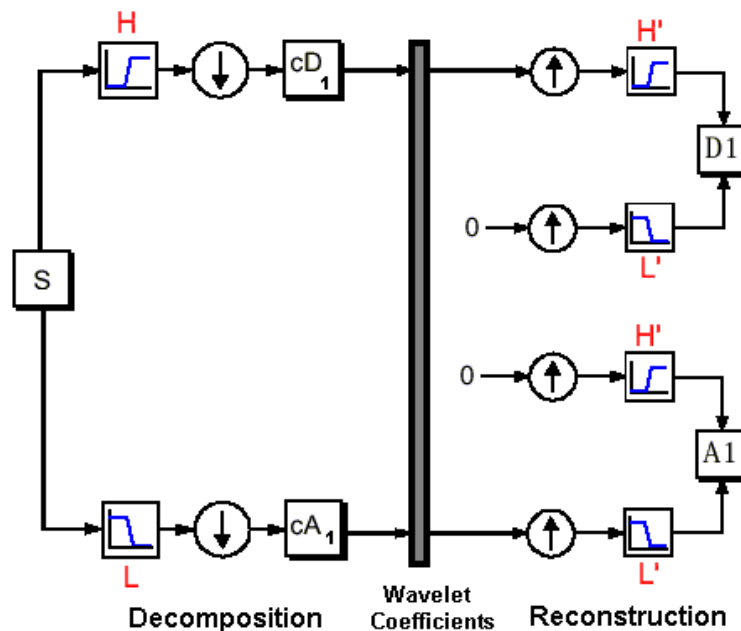


Fig. 3.2. Análise de resolução simples via algoritmo de Mallat ( $S=A_1+D_1$ ) (adaptado de [Misiti et al., 1996]).

Finalmente, um processo de decomposição de múltiplos níveis pode ser conseguido de acordo com a Fig. 3.3. Note que aproximações sucessivas são decompostas de modo que  $S$  é dividido em vários componentes de resolução mais baixa.

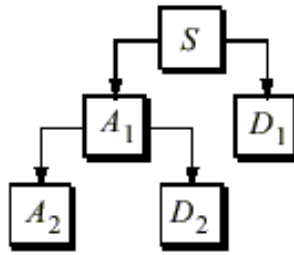


Fig. 3.3. Esquema de decomposição de múltiplos níveis ( $S=A_2+D_2+D_1$ ).

### 3.4 Minimizando o problema de distorção nas bordas

Geralmente, quando efetua-se uma operação de convolução em sinais de comprimento finito, há o surgimento de distorções nas extremidades [Misiti et al., 1996]. Para lidar com esse problema, algum tipo de extensão de sinal deve ser utilizado (conhecido como enchimento). Neste caso em particular, o objetivo é minimizar a quantidade de distorção em ambas extremidades de uma série de carga analisada via multiresolução.

Quatro esquemas de enchimento foram comparados neste trabalho. O primeiro deles, conhecido como enchimento com zeros (*zero padding*, em inglês), insere zeros no começo e no fim de uma série de carga. A segunda abordagem considera a replicação de borda simétrica (*symmetric boundary replication*). A principal desvantagem dos esquemas supraditos é que descontinuidades são artificialmente criadas nas bordas. Enchimento suavizado (*smooth padding*) é o terceiro esquema avaliado aqui e utiliza um ajuste de extensão linear para os dois primeiros e os dois últimos valores da série. Esse método proporciona uma transição mais suave do sinal para o enchimento. Esses esquemas de enchimento são discutidos em detalhes em [Misiti et al., 1996]. Finalmente, o esquema de enchimento proposto anexa valores de carga imediatamente anteriores no começo da série sob análise e valores previstos no seu fim. Utilizando-se essa abordagem, pode-se esperar nenhuma distorção no começo dos sinais decompostos e uma quantidade reduzida de distorção na outra extremidade. Deve-se ressaltar que quanto melhor for a qualidade das previsões, melhor será a qualidade do enchimento, e, conseqüentemente, menor será o nível de distorção nas bordas. Outra importante questão está relacionada com o comprimento da informação anexada (i.e., número de pontos que deverá ser anexado no começo e no final da série). Empiricamente, descobriu-se que o uso de no mínimo 72 pontos em cada extremidade da série de carga é suficiente para reduzir as distorções das bordas significativamente.

A Fig. 3.4 mostra os desvios desses quatro esquemas de enchimento em relação ao enchimento ideal, i.e., valores medidos anexados no começo e no final da série. Por conveniência, apenas a subsérie  $A_3$  é mostrada.

Os três primeiros esquemas de enchimento (os tradicionais) apresentam desvios bastantes pronunciados em ambos lados de  $A_3$ . O esquema proposto (prev) produz uma subsérie  $A_3$  bastante similar àquela gerada com o enchimento ideal (verifica-se nenhuma distorção no começo de  $A_3$  e uma reduzida quantidade de distorção na outra extremidade).

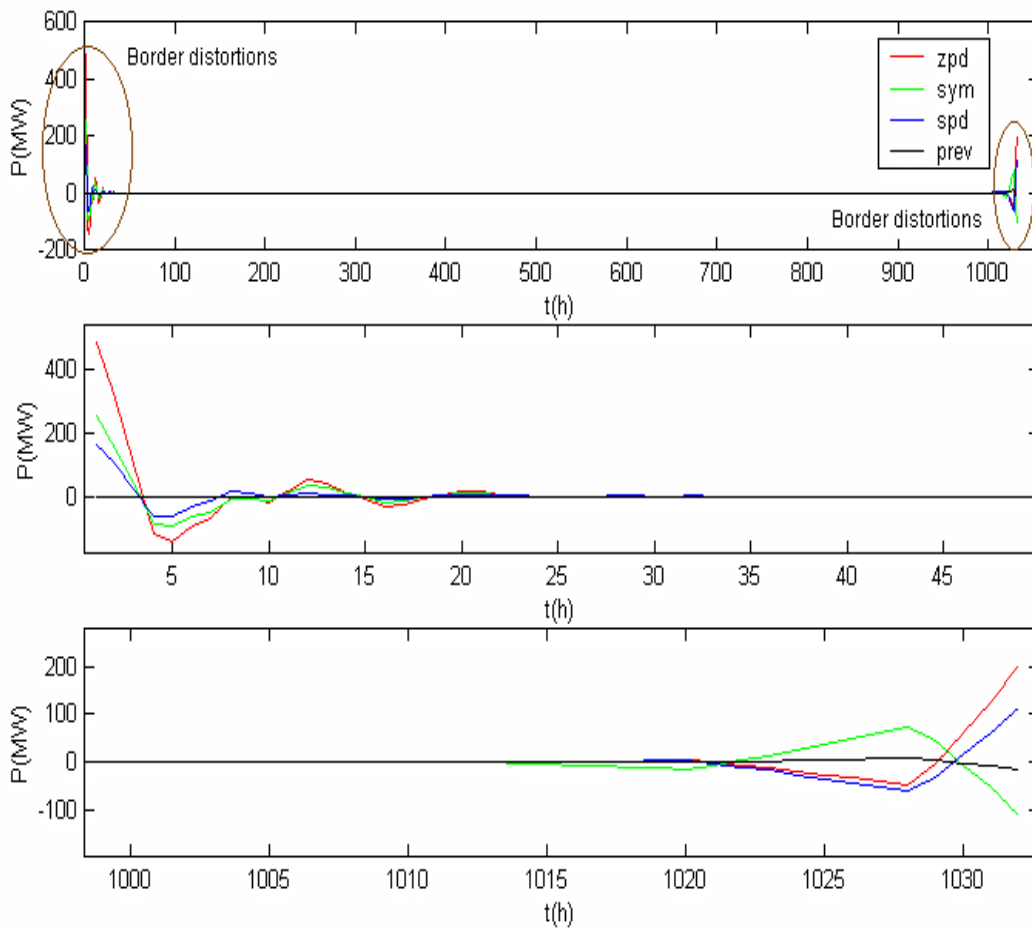


Fig. 3.4. Representação gráfica de  $A_3$ . Desvios dos quatro esquemas de enchimento em relação ao enchimento ideal. As técnicas avaliadas “zpd” (enchimento com zeros), “sym” (replicação de borda simétrica), “spd” (enchimento suavizado), e “prev” (esquema de enchimento proposto) aparecem na legenda no painel superior. Ampliações das extremidades são mostradas nos painéis intermediário e inferior.

## 3.5 Exemplos Práticos

### 3.5.1 Descrição estatística de uma série de carga via AMR

Encontra-se aqui um exemplo da análise de uma série de carga via AMR. O principal objetivo deste exemplo é mostrar as formas de onda típicas provenientes da AMR de uma série de carga normalizada entre  $[-0,5; 0,5]$ . Contudo, antes de efetuar a análise via wavelet, duas tarefas devem ser realizadas:

- seleção da wavelet mãe;
- definição do número de níveis de decomposição.

#### *Seleção da Wavelet Mãe*

Há muitos tipos de wavelet mãe que podem ser usados na prática. Para se escolher o tipo mais adequado para uma dada tarefa, os atributos da wavelet mãe e as características do sinal devem ser levadas em consideração (e.g., tendências, mudanças de nível e/ou inclinação, sazonalidades).

Em [Coifman & Wickerhauser, 1992], os autores criaram uma biblioteca de formas de onda modulada formadas por funções trigonométricas localizadas e pacotes de wavelet ortogonais, que apresentam propriedades de localização em tempo-frequência razoavelmente bem controladas. A idéia é retirar dessa biblioteca uma base ortonormal relativa a um dado sinal de interesse que apresente o menor custo de informação. Várias funções de custo podem ser utilizadas. A entropia de Shannon é uma das mais atrativas por propiciar interpretação geométrica neste contexto.

Ingrid Daubechies, uma das principais pesquisadoras sobre wavelets, inventou as chamadas wavelets ortonormais de suporte compacto [Daubechies, 1992]. Sua invenção fez com que a análise de sinais discretos via wavelets fosse praticável. Os nomes dessa família são escritos com a notação  $dbN$ , onde  $N$  é a ordem e  $db$  é o “sobrenome” da wavelet. A wavelet  $db1$  é também conhecida como wavelet de Haar e sua formulação matemática será apresentada no Anexo III desta tese. A Fig. 3.5 mostra os quatro primeiros membros desta família.

Em [Misiti et al., 1994], os autores decompõe segmentos de séries de carga (e.g., 12h-1h30 PM, 3h30-6h AM) via TWD e realizam análises locais. Várias famílias de wavelets



foram consideradas. O estudo apontou a wavelet da família Daubechies como a mais apropriada para tratar a série de carga considerada no estudo.

[Kim et al., 2002] apresentaram um modelo de PCCP combinando-se wavelets da família Daubechies e redes neurais do tipo Kohonen.

Portanto, valendo-se dos apelos teórico e prático da wavelet Daubechies, decidiu-se investigar sua aplicabilidade no contexto deste trabalho. Wavelets Daubechies de ordens 2, 3 e 4 foram avaliadas.

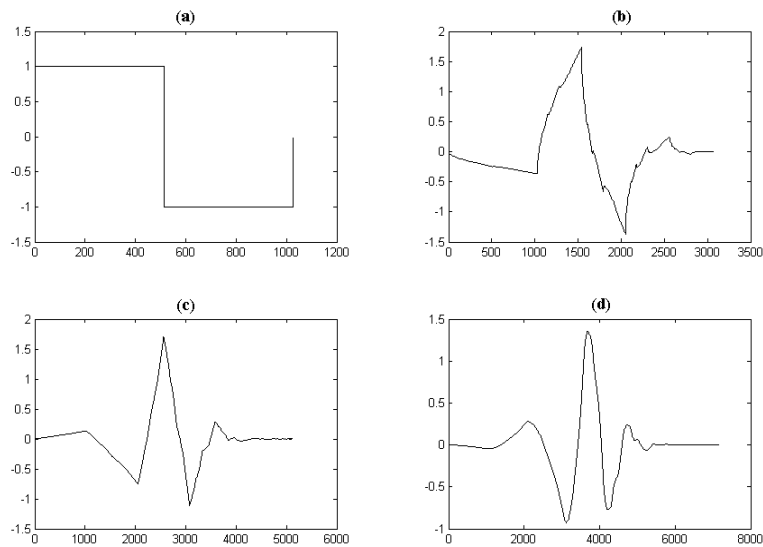


Fig. 3.5. Wavelets da família Daubechies. (a) Haar; (b) Daubechies de ordem 2; (c) Daubechies de ordem 3; (d) Daubechies de ordem 4.

### ***Definição do Número de Níveis de Decomposição***

Na prática, é aconselhável definir um número adequado de níveis de decomposição baseado na natureza do sinal ou em um determinado critério, e.g., entropia [Coifman & Wickerhauser, 1992].

Nesta tese, com base na dinâmica da carga apenas, três estratégias de análise via wavelets com 3, 4 e 5 níveis de decomposição foram consideradas. Em linhas gerais, o estudo consistiu dos seguintes passos:

1. Considerar o número de níveis de resolução  $K$  igual a 3.
2. Decompor a série de interesse via AMR.
3. Fazer uma inspeção visual da aproximação e dos detalhes.

4. Observar se a aproximação preserva as características mais importantes do sinal original (no caso, a sazonalidade diária era uma delas). Caso isso se verifique, este número de níveis de resolução pode ser uma boa escolha.
5. Faça  $K=K+1$  e retorne ao passo 1.

Concluído esse estudo, uma representação com 3 níveis de decomposição se mostrou como a escolha mais promissora porque ela descreve a série de carga de maneira mais significativa que as demais. Essa conclusão tem a ver principalmente com o conteúdo de baixa frequência (aproximação) que é a componente mais importante do sinal. Nota-se que o esquema de 3 níveis de decomposição enfatiza o comportamento regular da série de carga e revela padrões ocultos que não podem ser vistos com os demais esquemas de decomposição. Além disso, esse esquema mostra explicitamente a sazonalidade diária da série de carga. Essa característica não pôde ser verificada com os demais esquemas.

A Fig. 3.6 mostra um segmento de uma curva de carga real, sua aproximação  $A_3$  e os níveis de detalhe  $D_3$ ,  $D_2$  e  $D_1$ . Nota-se que  $A_3$  é uma boa aproximação do sinal. Ela representa a sazonalidade diária da carga e é uma versão suavizada da curva de carga. Os níveis de detalhe  $D_1$  e  $D_2$  contêm ruído e informação de alta frequência bastante útil. Essas componentes exibem alguns padrões e têm formas e valores médios bastante similares. Algumas irregularidades naqueles níveis de detalhe são provavelmente devidas a erros de medição e a eventuais cortes de carga. O nível de detalhe  $D_3$  mostra alguns picos que permitem a localização temporal de certos eventos, tais como, períodos de carga pesada.

### 3.5.2 Identificando-se *Outliers* e Dados Faltantes

O painel superior da Fig. 3.7 mostra o mesmo segmento de curva de carga apresentado na Fig. 3.6 corrompido por dois outliers (localizados em  $t=27h$  e  $t=80h$ ) e doze dados faltantes ( $t=153h$  a  $t=164h$ ). O efeito produzido nos detalhes fica evidenciado quando enfoca-se os níveis inferiores  $D_1$  e  $D_2$ . Pode-se notar que os outliers estão sincronizados com a carga naqueles níveis.  $D_3$  mostra o efeito dos outliers com um certo atraso. Com relação aos dados faltantes, a parte plana de  $D_1$ , indo de  $t=153h$  até  $t=164h$ , pode ser utilizada para detectá-los e corrigi-los posteriormente.

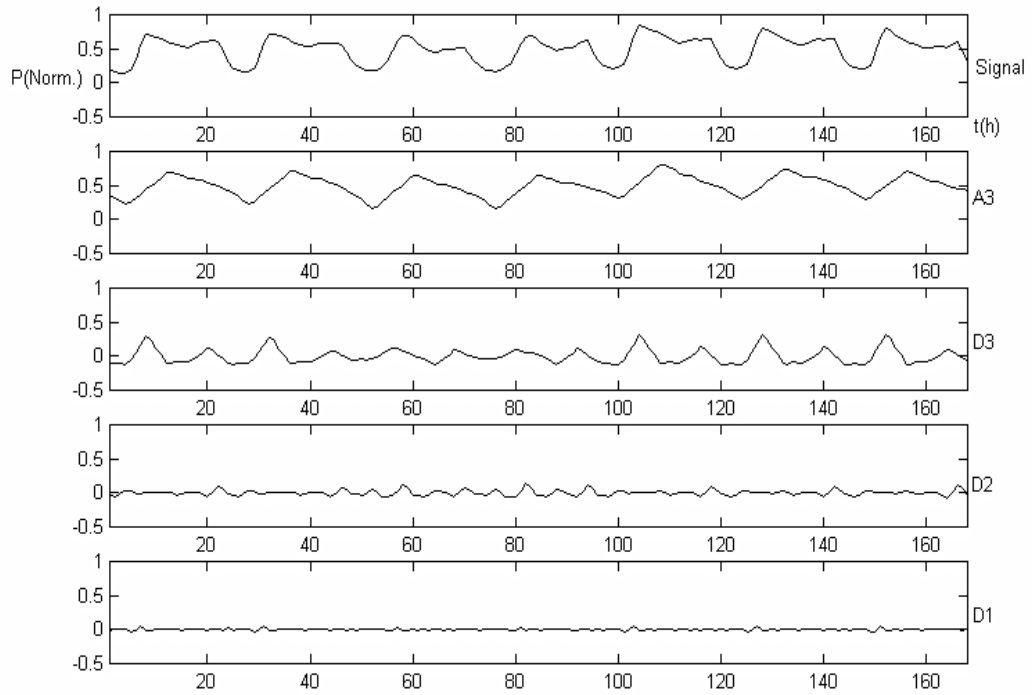


Fig. 3.6. Segmento de curva de carga decomposto em três níveis de resolução.

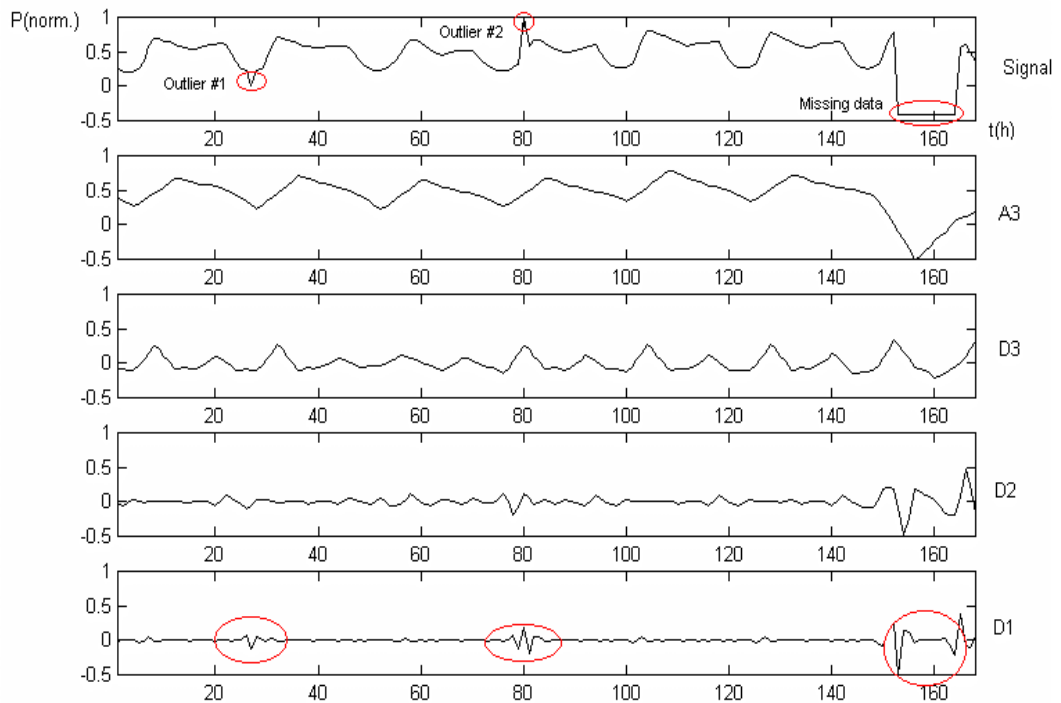


Fig. 3.7. Segmento de curva de carga corrompido com dois outliers e doze dados faltantes.

### 3.5.3 Espectro de Potência

Para muitos sinais, a análise de Fourier é extremamente útil porque o conteúdo de frequência do sinal é bastante conclusivo. Porém, a análise de Fourier tem um sério inconveniente. Ao se transformar um sinal no domínio do tempo para o domínio da frequência, a informação com relação a localização temporal de eventos é perdida. Ou seja, quando se analisa a transformada de Fourier de um sinal, é impossível de se dizer quando um evento de interesse ocorreu.

A título de exemplo, a Fig. 3.8 mostra os espectros de potências dos segmentos de curva de carga apresentados nas Figs. 3.6 e 3.7. Nota-se que eles são diferentes entre si devido as presenças dos outliers e dos dados faltantes. No entanto, os instantes de tempo onde esses eventos ocorreram não podem ser inferidos.

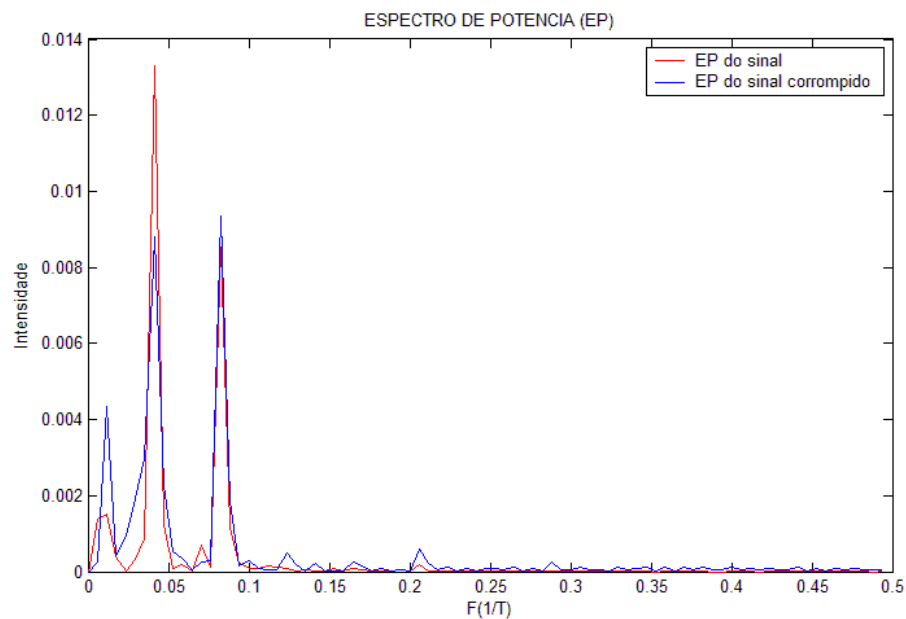


Fig. 3.8. Espectro de potência para os segmentos de curva de carga considerados nos itens 3.5.1 e 3.5.2.

## 3.6 Aplicações de Wavelets em SEP

Na década passada houve um grande número de publicações sobre a aplicação da transformada wavelet em sistemas elétricos de potência com resultados bastante promissores. De acordo com [Wildberger, 1994], pesquisadores do *Electric Power Research Institute* (EPRI) vêm pesquisando famílias de wavelets e aplicando técnicas de representação em wavelets em

problemas de interesse das concessionárias de energia elétrica. Por exemplo, simulação de combustão e poluição dinâmica, análise da estabilidade transitória em sistemas de potência e exploração de wavelets em tarefas de compressão e transmissão de dados em circuitos de comunicação instalados sobre as linhas de transmissão. Em [Kim & Aggarwal, 2001], os autores apresentam as principais áreas de aplicação da transformada de wavelet em SEP (ver Fig. 3.9).

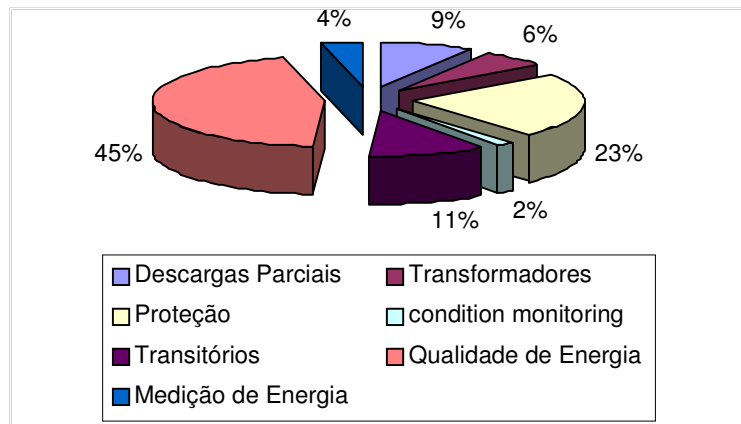


Fig. 3.9. Principais áreas de aplicação da TW em SEP [Kim & Aggarwal, 2001].

## Capítulo 4 – Técnicas de Reconstrução do Espaço-Fase

Este capítulo apresenta alguns conceitos sobre a reconstrução do espaço-fase de um dado sistema dinâmico a partir do método das coordenadas em atraso e discute também sua aplicação no problema de seleção de variáveis de entrada para previsores neurais.

### 4.1 Introdução

Considere um sistema puramente determinístico. Uma vez que seu estado corrente seja fixado, os estados em todos os instantes de tempo futuros também poderão ser conhecidos [Kantz & Schreiber, 1999]. Dessa forma, será importante estabelecer um espaço vetorial (também conhecido como *espaço de estados* ou *espaço-fase*) para o sistema em análise, de modo que, especificando-se um ponto nesse espaço, especifica-se o estado do sistema e vice-versa. A partir daí, pode-se estudar a dinâmica do sistema através do estudo da dinâmica dos pontos correspondentes no espaço-fase. Na teoria, sistemas dinâmicos são geralmente definidos por um conjunto de Equações Diferenciais Ordinárias (EDO) de primeira ordem atuando em um dado espaço-fase. A teoria matemática das EDO garante a existência e a unicidade de trajetórias caso certas condições sejam atendidas.

Matematicamente falando, considere que o espaço-fase é um espaço vetorial de dimensão finita  $m$ ,  $\mathfrak{R}^m$ . Um *estado* é especificado por um vetor  $\mathbf{x} \in \mathfrak{R}^m$ . Então, pode-se descrever a dinâmica através de um mapa  $m$ -dimensional ou por um sistema explícito de  $m$ -EDO de primeira ordem. No primeiro caso, o tempo é uma variável discreta:

$$\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n), \quad n \in \mathfrak{J}. \quad \text{Eq. 4.1}$$

Já no segundo exemplo, o tempo é uma variável contínua:

$$\frac{d}{dt} \mathbf{x}(t) = \mathbf{f}(t, \mathbf{x}(t)), \quad t \in \mathfrak{R}. \quad \text{Eq. 4.2}$$

Uma seqüência de pontos  $\mathbf{x}_n$  ou  $\mathbf{x}(t)$  que solucione as equações acima é chamada de *trajetória* do sistema dinâmico. Também,  $\mathbf{x}_0$  e  $\mathbf{x}(0)$  são as condições iniciais. Normalmente, trajetórias típicas se dirigem para o infinito com o passar do tempo ou ficam “presas” em uma

dada região do espaço-fase (que é o caso de maior interesse). Tais comportamentos dependem tanto da forma de  $\mathbf{F}$  (ou, respectivamente,  $\mathbf{f}$ ) e das condições iniciais. Ressalta-se que muitos sistemas permitem ambos tipos de solução [Kantz & Schreiber, 1999]. O conjunto de condições iniciais que conduz ao mesmo comportamento assintótico da trajetória é chamado de *bacia de atração*. Um subconjunto do espaço-fase que tem a propriedade de atrair (ou repelir) trajetórias é chamado de *atrator* (ou repulsor). Uma vantagem de se considerar atratores no espaço-fase como uma representação alternativa de uma série temporal é que diversos resultados geométricos e topológicos podem ser utilizados. A forma e a dimensão de um atrator no espaço-fase estão diretamente relacionadas à dinâmica da respectiva série temporal. Portanto, atratores simples (de baixa dimensionalidade) correspondem a dinâmicas simples no domínio do tempo. Alguns exemplos de atratores no plano, tais como, ponto fixo e ciclo limite são apresentados na Fig. 4.1.a e 4.1.b, respectivamente. Atratores estruturalmente mais complexos podem ocorrer para fluxos tridimensionais (ver Fig. 4.2).

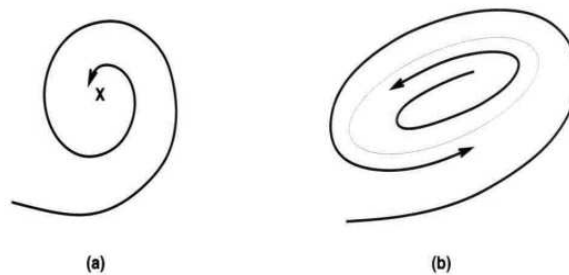


Fig. 4.1. Exemplo de atratores. (a) ponto fixo; (b) ciclo limite [Kantz & Schreiber, 1999].

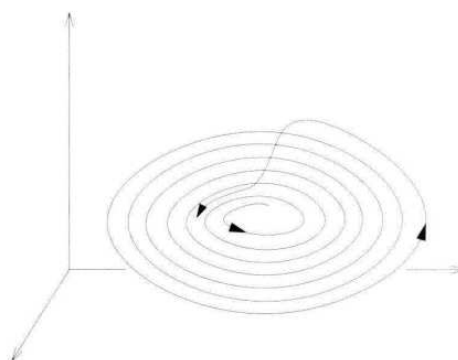


Fig. 4.2. Em um espaço tridimensional, uma trajetória pode ser reinjetada sem que o determinismo seja violado [Kantz & Schreiber, 1999].

Com relação a dimensão de objetos fractais e de certos atratores que se comportam de forma caótica, as dimensões Euclidianas<sup>1</sup> não se aplicam e outros tipos de dimensão devem ser considerados. A título de exemplo, a chamada *dimensão fracionária* é uma das mais difundidas e foi apresentada por F. Hausdorff em 1918 [Hausdorff, 1918]. Basicamente, para todo objeto de tamanho  $T$ , construído de unidades menores de tamanho  $t$ , em que  $N$  é o número de vezes que  $t$  cabe em  $T$ , a dimensão fracionária é dada por:

$$D = \frac{\log(N)}{\log\left(\frac{T}{t}\right)} \tag{Eq. 4.3}$$

As figuras 4.3 e 4.4 ilustram alguns exemplos da aplicação da equação acima. Note que a dimensão Euclidiana de alguns objetos (Casos A-C, Fig. 4.3) também pode ser determinada pela Eq. 4.3.


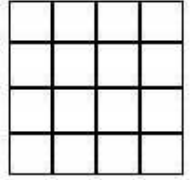
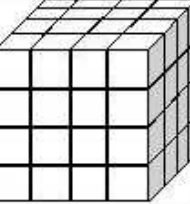
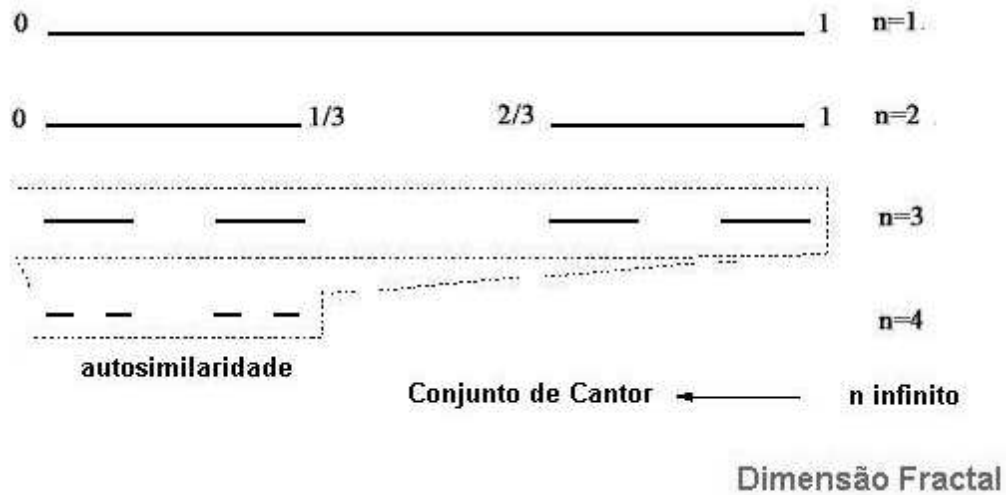
<b>Caso A - Retas</b>	<b>Gráfico</b>
$D = \frac{\log(N)}{\log\left(\frac{T}{t}\right)} = \frac{\log(4)}{\log\left(\frac{1}{\frac{1}{4}}\right)} = \frac{\log(4)}{\log(4)} = 1$	
<b>Caso B - Superfície</b>	<b>Gráfico</b>
$D = \frac{\log(N)}{\log\left(\frac{T}{t}\right)} = \frac{\log(16)}{\log\left(\frac{1}{\frac{1}{4}}\right)} = \frac{2\log(4)}{\log(4)} = 2$	
<b>Caso C - Volume</b>	<b>Gráfico</b>
$D = \frac{\log(N)}{\log\left(\frac{T}{t}\right)} = \frac{\log(64)}{\log\left(\frac{1}{\frac{1}{4}}\right)} = \frac{3\log(4)}{\log(4)} = 3$	

Fig. 4.3. Determinação da dimensão Euclidiana de alguns objetos.

<sup>1</sup> Um ponto tem dimensão 0, uma linha apresenta dimensão unitária, uma superfície (e.g., um quadrado) tem dimensão 2 e um volume (e.g., um cubo) possui 3 dimensões.





$$D = \frac{\log(N)}{\log\left(\frac{T}{t}\right)} = \frac{\log(2)}{\log\left(\frac{1}{\frac{1}{3}}\right)} = \frac{\log(2)}{\log(3)} \cong 0,63092$$

Fig. 4.4. Determinação da dimensão de Hausdorff de um objeto fractal.

## 4.2 Reconstrução do Espaço-Fase

Tendo mostrado a importância do espaço-fase para o estudo de sistemas com propriedades determinísticas, tem-se que enfrentar o primeiro problema: o que se consegue a partir de um experimento é uma série temporal e não um objeto de espaço-fase. Portanto, tem-se que converter esse conjunto de observações em vetores de estado. Esse problema de conversão é conhecido como Reconstrução do Espaço-Fase (REF).

Geralmente, uma série temporal é uma seqüência de medições escalares de alguma quantidade que depende do estado corrente do sistema e que são tomadas em intervalos de tempo igualmente espaçados, i.e.:

$$s_n = s(\mathbf{x}(n\Delta t)) + \eta_n,$$

em que  $s$  é uma função de medição e  $\eta$  o ruído de medição.

A principal idéia por trás da REF é que cada ponto que compõe a série temporal de um observável contém, em teoria, informação sobre variáveis de estado não observadas do sistema original. Se uma quantidade suficiente de pontos é amostrada, então a informação sobre variáveis de estado não observadas e suas interações deverão ser suficientes para definir um sistema dinâmico. O método de reconstrução deve garantir que as propriedades do atrator reconstruído serão iguais as do atrator original (e.g., mesmo espaço-fase, mesma dimensão do atrator, expoentes de Lyapunov idênticos, etc. [Kantz & Schreiber, 1999]).

A primeira idéia de como reconstruir um dado atrator num espaço de fases de dimensão conveniente foi sugerida por Packard et al. [Packard et al., 1980]. Os autores mostram que para o atrator do modelo de Rössler<sup>2</sup>, o espaço de fases  $x, \dot{x}, \ddot{x}$ , conduz às mesmas informações que o espaço real  $x, y, z$  das variáveis do modelo, i.e., as derivadas sucessivas de uma variável,  $dx^m/dt^m$ , contêm informações sobre a evolução temporal das outras variáveis do sistema. O cálculo das derivadas<sup>3</sup>, porém, amplifica erros experimentais e limita o uso do algoritmo, principalmente se o número de variáveis independentes envolvido for grande.

Uma outra idéia apresentada pelos autores sugere que seja possível reconstruir certas propriedades topológicas do atrator a partir de um conjunto de *coordenadas em atraso*. A idéia heurística por detrás do método de reconstrução é que para se especificar o estado de um sistema tridimensional em um dado instante, a medição de quaisquer três quantidades independentes deva ser suficiente. Suspeita-se que qualquer conjunto de três quantidades independentes que, unicamente e suavemente rotulam os estados do atrator, sejam difeomorficamente equivalentes.

As três quantidades mais usadas são os valores de cada coordenada,  $x(t)$ ,  $y(t)$  e  $z(t)$ . Descobriu-se que partindo de uma série temporal obtida através da amostragem de uma única coordenada do modelo de Rössler, pode-se obter uma variedade de três quantidades independentes que produzam uma boa representação da dinâmica no espaço original  $x, y, z$ . Um conjunto possível de tais três quantidades é o valor da coordenada  $x$  no instante de tempo  $t$ , mais seus dois valores anteriores (i.e.,  $x(t)$ ,  $x(t-1)$ ,  $x(t-2)$ ).

<sup>2</sup> O modelo de Rössler satisfaz o seguinte sistema de equações diferenciais:

$$\dot{x} = -y - z, \quad \dot{y} = x + a \cdot y, \quad \dot{z} = b + z \cdot (x - c).$$

<sup>3</sup> Na prática, as derivadas são estimadas a partir de dados experimentais por meio de  $\frac{dx_n}{dt} \sim \frac{x_{n+1} - x_n}{\Delta t}$ .

Em [Takens, 1981], o autor apresenta uma prova matemática formal para a REF via método das coordenadas em atraso. Uma reconstrução em atraso em  $d$  dimensões é então formada por vetores  $\varphi_n$  dada por:

$$\varphi_n = (s_{n-(d-1)\tau}, s_{n-(d-2)\tau}, \dots, s_{n-\tau}, s_n). \quad \text{Eq. 4.4}$$

A diferença temporal em número de amostras  $\tau$  entre componentes adjacentes dos vetores em atraso é referido como “time delay” ou tempo de atraso.

Embora o atrator reconstruído não seja idêntico ao original, pode-se demonstrar que as suas propriedades topológicas são preservadas. Da mesma forma, a dimensão  $d$  do espaço de fases reconstruído não é necessariamente idêntica à dimensão  $d_a$  do espaço de fases real dos vetores que representam a dinâmica do sistema físico de interesse. Em geral, é necessário reconstruir-se o atrator em espaços de fases com dimensão suficientemente elevada, a saber:

$$d \geq 2 \cdot d_a + 1, \quad \text{Eq. 4.5}$$

em que  $d_a$  é a dimensão de Hausdorff do atrator original.

Definida a equação para a REF, pode-se questionar sob quais circunstâncias e até que extensão o objeto geométrico formado pelos vetores  $\varphi_n$  é equivalente à trajetória original  $\mathbf{x}_n$ . Na verdade, sob circunstâncias bastante gerais, o atrator formado por  $\varphi_n$  é equivalente ao atrator “verdadeiro” se a dimensão de imersão  $d$  do espaço-fase reconstruído for suficientemente grande [Takens, 1981]. Existem vários algoritmos para a determinação desse parâmetro. O item 4.2.1 apresentará um deles que é baseado na localização de *falsos vizinhos* no espaço-fase [Kennel et al., 1992].

Por outro lado, a escolha do valor em atraso  $\tau$  não é assunto dos teoremas de imersão, pois esses consideram que os dados tenham precisão e duração infinitas. Isso significa dizer que *a priori* poderia utilizar-se qualquer valor de  $\tau$  na REF. No entanto, uma escolha razoável do valor em atraso ganha importância devido ao fato de que os dados utilizados na prática são finitos e contêm algum tipo de ruído. Especificamente, se  $\tau$  for muito pequeno em comparação ao tempo de recorrência de um sistema, que é definido como o valor recíproco da frequência dominante no espectro de potência do sistema, a dinâmica ficará confinada à hiper-diagonal do espaço-fase sugerindo que dois estados consecutivos sejam linearmente

dependentes. Definitivamente, esse não é o caso para observáveis de um sistema não linear. Por outro lado, um valor em atraso que seja muito maior do que o tempo de recorrência fará com que dois estados consecutivos fiquem completamente desconexos, descaracterizando o sistema original. Portanto, utilizar-se-á na identificação de um valor de  $\tau$  adequado, um algoritmo baseado no cálculo da Informação Mútua Média (IMM) de uma série [Fraser & Swinney, 1986]. Esse algoritmo será apresentado no item 4.2.2. Finalmente, o problema de REF adaptado a PCCP é assunto do item 4.3.

### 4.2.1 Falsos vizinhos no espaço-fase

Um método para determinar a dimensão de imersão mínima foi proposto por Kennel et al. [Kennel et al., 1992]. Esse método é chamado de ‘Método dos falsos vizinhos’. A idéia é bastante intuitiva. Suponha que a dimensão mínima de imersão para uma dada série  $\{S_i\}$  seja  $m_0$ . Isto significa que em um espaço-fase  $m_0$ -dimensional o atrator reconstruído é uma imagem unívoca do atrator no espaço-fase original. Particularmente, as propriedades topológicas são preservadas. Assim, os vizinhos de um dado ponto são mapeados nos vizinhos do espaço-fase reconstruído. Devido à suavidade da dinâmica assumida a priori, vizinhanças de pontos são mapeadas em vizinhanças novamente. Suponha agora que a série  $\{S_i\}$  tenha sido imersa em um espaço-fase de dimensão  $m < m_0$ . Nesse caso, a estrutura topológica não é mais preservada. Pontos são projetados na vizinhança de outros pontos para as quais eles não pertenceriam em dimensões mais elevadas. Essa ausência de localização única de todas as imagens em  $m_0-1$  dimensões é refletida através da identificação de falsos vizinhos, significando que o determinismo foi violado. Esses pontos são chamados falsos vizinhos. Se agora a dinâmica é aplicada, esses falsos vizinhos não são mapeados na imagem da vizinhança, mas em um outro lugar qualquer.

### 4.2.2 Informação Mútua Média

Conceitualmente, a informação mútua descreve a quantidade de informação conhecida sobre  $y$  dado  $x$ . Quando assume-se que  $x$  seja o estado presente e  $y$  o estado após um tempo  $\Delta t$ , suas informações mútuas descrevem quanto de informação é conhecida sobre um estado no futuro baseado no conhecimento do estado presente. O cálculo da Informação Mútua Média

(IMM) como uma ferramenta para determinar o valor em atraso “ótimo”<sup>4</sup> para REF foi sugerido por Fraser & Swinney [Fraser & Swinney, 1986]. Diferentemente da função de autocorrelação linear, a IMM também leva em conta correlações não lineares.

Com base na série de interesse, cria-se um histograma para a distribuição de probabilidade dos dados. Denota-se por  $p_i$  a probabilidade de que o sinal assuma um valor na  $i$ -ésima caixa<sup>5</sup> do histograma, e considera-se que  $p_{ij}(\tau)$  seja a probabilidade de que  $x$  esteja na caixa  $i$  e  $y$  na caixa  $j$ . Então, a informação mútua para o valor em atraso  $\tau$  pode ser computada a partir de:

$$I(\tau) = \sum_{i,j} p_{ij}(\tau) \ln p_{ij}(\tau) - 2 \sum_i p_i \ln p_i . \quad \text{Eq. 4.6}$$

O primeiro valor mínimo da IMM relativo a um certo valor em atraso é atribuído como sendo o valor “ótimo” de  $\tau$ . A Eq. 4.6 é calculada para valores em atraso variando de 1 a um dado valor em atraso máximo.

### 4.3 Reconstrução do Espaço-Fase aplicada à PCCP

A idéia de REF de uma série temporal, juntamente com a descrição matemática do procedimento de imersão apropriado para o problema de previsão, é mostrada na Fig. 4.5. Considera-se que a dinâmica original  $f$  gere uma série temporal obtida de uma sistema dinâmico através de uma função de medição  $h$ . No caso estudado nesta tese,  $h$  corresponde às leituras dos dispositivos de medição de potência ativa realizadas por uma dada concessionária de energia elétrica. Esse processo apresenta erros de medição, dados com ruído, além de tornar um processo originalmente contínuo em um processo discreto com tempo de amostragem  $T_s$ . O próximo passo no processo de previsão baseado no espaço-fase requer a REF. Essa reconstrução é feita através de um mapa  $\phi$ . A dinâmica  $F^n$  tem que ser estimada no espaço-fase reconstruído de modo que ela represente uma dinâmica equivalente à dinâmica original  $f$ . Para esse fim, RNA serão utilizadas. A previsão é feita no espaço-fase reconstruído usando a dinâmica  $F^n$  aprendida.

<sup>4</sup> Em [Kantz & Schreiber, 1999], os autores advertem que um espaço-fase “ótimo” somente pode ser definido relativamente a um propósito específico para o qual a imersão será utilizada.

<sup>5</sup> *Bin*, do inglês.

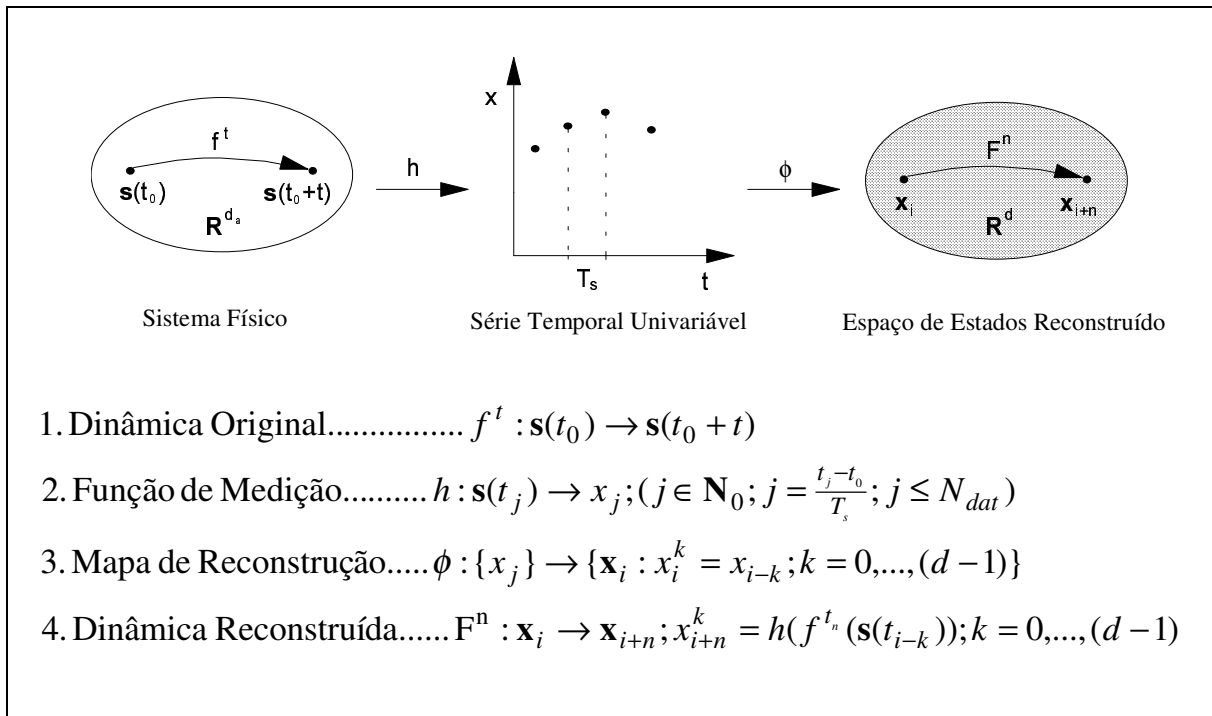


Fig.4.5. Diagrama esquemático da técnica de REF (adaptado de [Drezga, 1996]).

### 4.3.1 Implementação

A metodologia aplicada na reconstrução do espaço-fase foi dividida em 3 sub-itens. São eles:

1. Cálculo do valor em atraso através da IMM;
2. Cálculo da dimensão de imersão através do algoritmo de falsos vizinhos;
3. Reconstrução do espaço-fase via método das coordenadas em atraso.

Na seqüência, aplicar-se-ão os algoritmos supraditos a um segmento de uma série de carga de uma concessionária de EE norte-americana. O período de análise vai de 01/10/1990 a 31/10/1990 (ver Fig. 4.6). Dois esquemas de pré-processamento de dados de carga serão aplicados ao segmento considerado. O primeiro esquema faz a normalização dos dados no intervalo de 0 a 1. Os resultados obtidos com esse sinal servirão para auxiliar na identificação das variáveis de entrada do modelo de previsão discutido no item 5.9.5. Já no segundo esquema, faz-se a AMR via DWT da série de carga considerada. Para tanto, emprega-se a wavelet mãe Daubechies (daub2), considerando-se 3 níveis de resolução (i.e.,  $A_3$ ,  $D_3$ ,  $D_2$  e

D<sub>1</sub>). Como  $A_3$  é uma versão suavizada da série de carga, somente essa componente será considerada. A informação de alta frequência presente nos níveis de detalhes não será levada em conta aqui. A Fig. 4.7 mostra  $A_3$ . Os resultados obtidos com este esquema serão utilizados no previsor discutido no item 5.9.6. Finalizando, a Fig. 4.8 mostra as séries pré-processadas (painel superior). Uma ampliação de parte dessas séries pode ser encontrada no painel inferior da mesma.

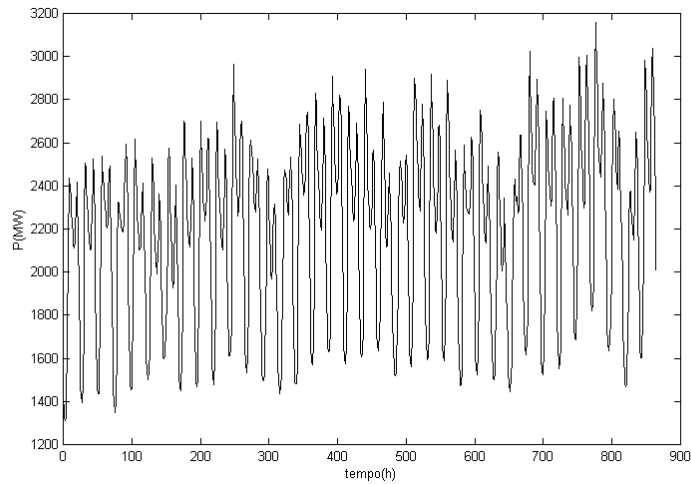


Fig. 4.6. Segmento de série de carga considerado.

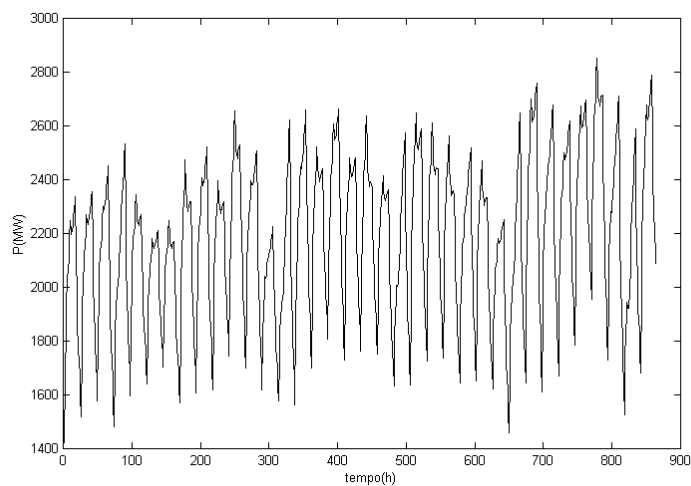


Fig. 4.7. Série de carga suavizada (componente  $A_3$ ).

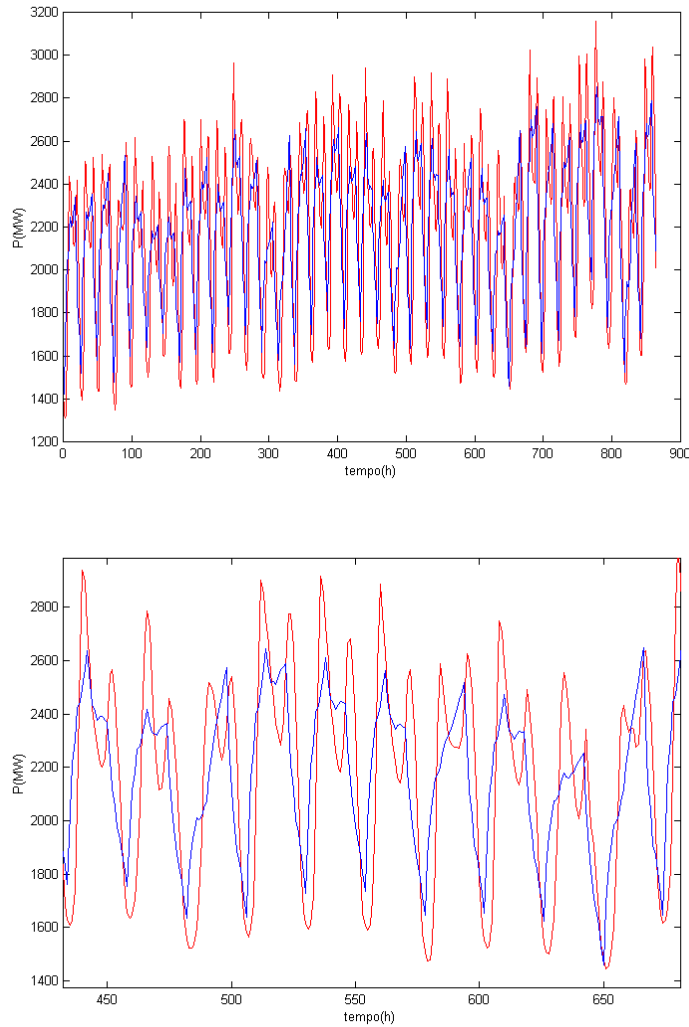


Fig. 4.8. Um comparativo entre as séries pré-processadas (painel superior). Uma ampliação de um segmento dos dois sinais pode ser encontrada no painel inferior (a série normalizada está em vermelho) .

#### 4.3.1.1 Cálculo do valor em atraso através da IMM

Para o cálculo da IMM nas séries apresentadas anteriormente, utiliza-se uma rotina implementada no pacote computacional conhecido como TiSeAn<sup>6</sup> [Helger et al., 1999]. O primeiro valor mínimo da IMM, relativo a um certo valor em atraso, é atribuído como sendo o valor em atraso “ótimo” mesmo quando o método apresente valores menores de IMM para valores em atraso maiores [Buzug & Pfister, 1992]. Além disso, alguns resultados experimentais apresentados por Buzug e Pfister sugerem que o valor em atraso “ótimo” deve ser menor que a metade do período da periodicidade dominante em um sistema (no caso, 12h).

<sup>6</sup> *Time Series Analysis*



Para a série normalizada, a aplicação da IMM apontou o valor em atraso  $\tau=2$  (Fig. 4.9). Esse resultado foi o mesmo obtido por I. Drezga em [Drezga, 1996]. Já para a aproximação  $A_3$ , um valor em atraso  $\tau=3$  foi obtido através do algoritmo. Uma inspeção visual na Fig. 4.12.c mostra uma boa utilização do espaço-fase com esse valor em atraso (i.e.,  $\tau=3$ ).

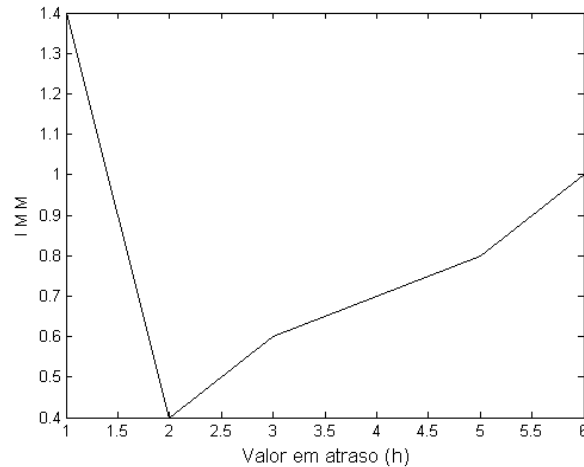


Fig. 4.9. Informação Mútua Média em função do valor em atraso para a série normalizada.

#### 4.3.1.2 Cálculo da dimensão de imersão via algoritmo de falsos vizinhos

Aplica-se o método de localização de falsos vizinhos nas duas séries. Novamente, o TiSeAn foi utilizado. Um valor de dimensão de imersão é encontrado quando o número de falsos vizinhos no espaço-fase é suficientemente pequeno ou igual a zero. Em ambos os casos, o algoritmo aponta que uma dimensão igual a 5 deva ser suficiente para garantir a reconstrução do espaço-fase (Fig. 4.10).

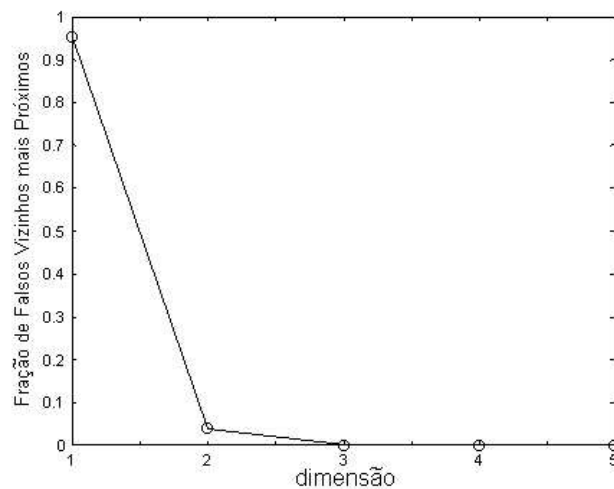


Fig. 4.10. Fração de falsos vizinhos mais próximos em função da dimensão para a série normalizada.

### 4.3.1.3 Reconstrução do espaço-fase via método das coordenadas em atraso – Um exemplo prático

Tendo determinado os parâmetros valor em atraso e dimensão de imersão, pode-se então reconstruir o espaço-fase de um dado sinal via Eq. 4.4. Nesta seção, dois exemplos serão considerados. No primeiro deles, utiliza-se novamente a série de carga normalizada. A Fig. 4.11.a-h ilustra várias representações de espaço-fase para diferentes valores em atraso dessa série. Esse conjunto de figuras mostra os dados imersos no espaço-fase.

No segundo exemplo, a série de carga utilizada foi aquela pré-processada via wavelets. A Fig. 4.12 apresenta várias representações de espaço-fase para diferentes valores em atraso dessa série. Na Fig. 4.12.a, nota-se que os dados se concentram próximos à diagonal. À medida que  $\tau$  cresce (Fig.4.11.b a 4.11.f), observa-se uma expansão do atrator reconstruído e, por conseguinte, uma melhor utilização do espaço-fase até  $\tau = 3$ . A partir daí, percebe-se que o atrator começa a se contrair até atingir representações desnecessariamente complicadas (Fig. 4.11.h). Nota-se ainda uma grande semelhança entre as duas figuras do ponto de vista de ocupação do espaço-fase. Já a característica “quadrada” presente nas representações da Fig. 4.12 está relacionada com a wavelet mãe utilizada na análise.

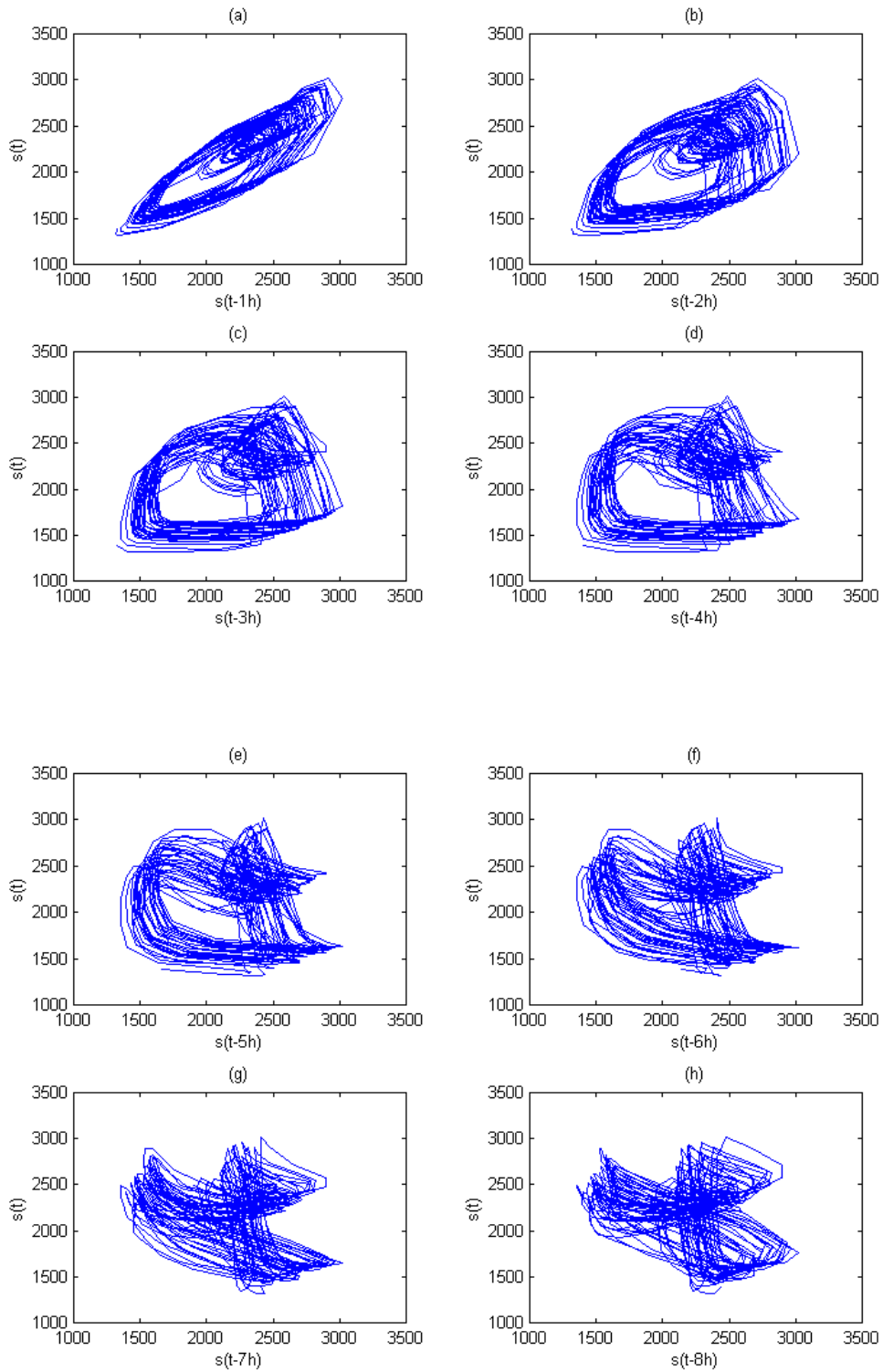


Fig. 4.11. Representação do espaço-fase para valores em atraso de 1 a 8h (série de carga normalizada).

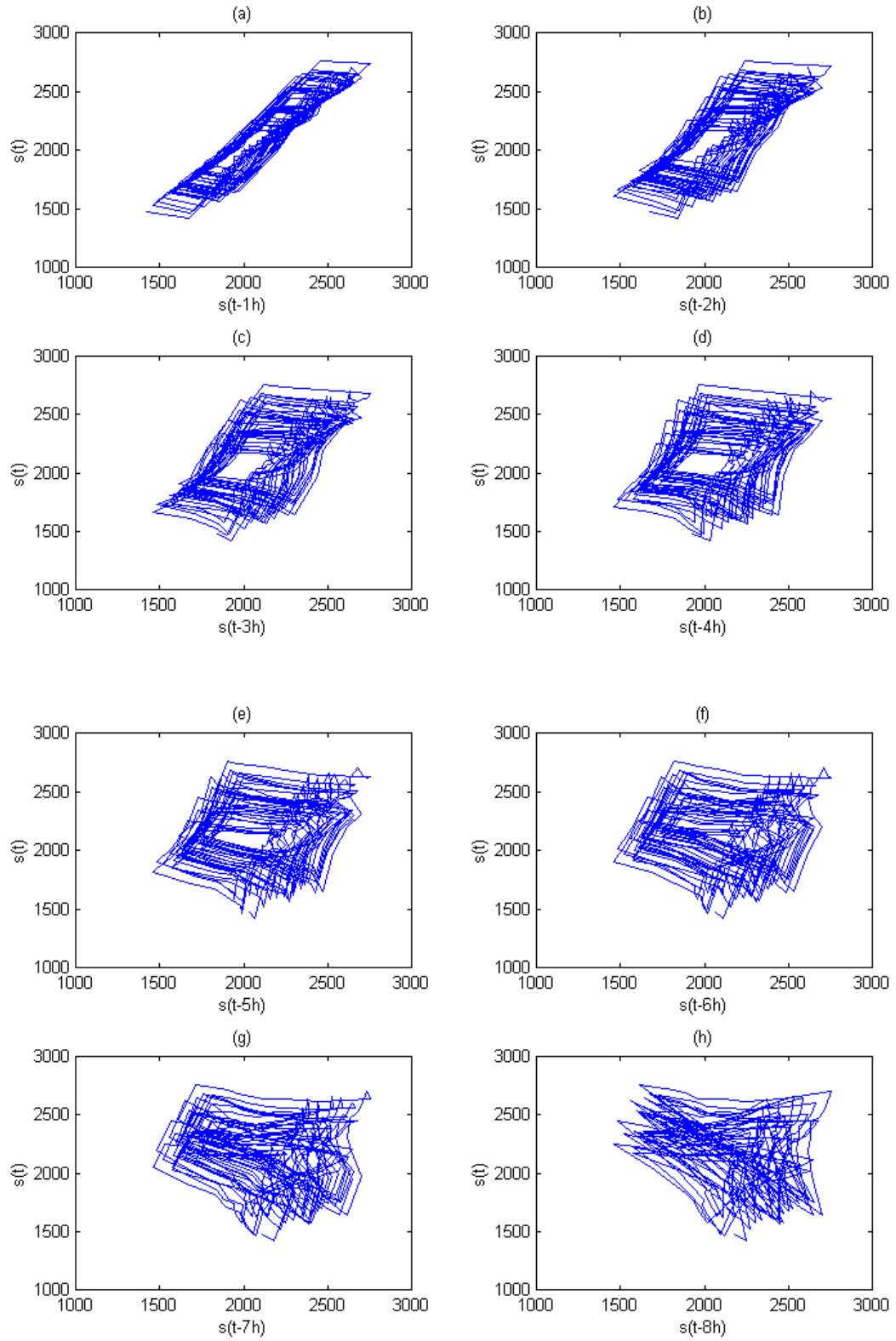


Fig.4.12. Representação do espaço-fase para valores em atraso de 1 a 8h (dados suavizados).

Na seqüência, o capítulo 5 tratará da implementação computacional das ferramentas até aqui apresentadas em previsores de carga elétrica via redes neurais artificiais. Vários testes foram realizados. Seus resultados serão apresentados e comentados.

## Capítulo 5 - Testes e Resultados

Neste capítulo serão apresentadas e detalhadas as etapas de desenvolvimento dos previsores implementados neste trabalho. Vários modelos de PCCP foram simulados em ambiente MATLAB 6.5 e seus resultados serão apresentados e analisados. A implementação realizada foi dividida nos treze passos seguintes. São eles:

1. Definição das séries históricas de carga e temperatura;
2. Agrupamento de dados de previsão;
3. Identificação das variáveis explicativas do processo gerador da série histórica;
4. Definição do número de pontos utilizados no processo de treinamento/validação;
5. Etapa de pré-processamento dos dados disponíveis;
6. Partição ótima dos dados;
7. Definição da estrutura do PMC;
8. Algoritmo de aprendizagem;
9. Modelos de PCCP via PMC;
10. Etapas de treinamento/validação;
11. Tipos de previsão;
12. Índices de desempenho;
13. Testes.

### 5.1 Definição das séries históricas de carga e temperatura

Dados horários de carga e temperatura de uma concessionária de energia elétrica norte-americana, compreendendo o período de 1º de janeiro de 1988 a 12 de outubro de 1992, foram usados neste trabalho. Alguns dados faltantes foram identificados e substituídos utilizando-se interpolação simples de pontos adjacentes. Nenhum outlier foi detectado. Algumas informações sobre os dados de carga e temperatura podem ser encontradas na Tab. 5.1.

Como em qualquer área de pesquisa, na previsão de carga também é importante permitir a repetição dos resultados publicados na literatura. O único modo de se fazer isso é através da utilização de uma base de dados de domínio público, tal qual aquela utilizada aqui, e que pode ser encontrada no seguinte *website*: <http://www.ee.washington.edu/class/559/2002spr>.

Tab. 5.1. Informações sobre os dados de carga e temperatura utilizados.

Ano	1988	1989	1990	1991	1992
Min. (MW)	1083	1110	1194	1231	1252
Máx. (MW)	3618	4282	4635	4094	3833
Média (MW)	1999,5	2104,4	2209,8	2228,7	2129,8
Min. (° F)	24	07	12	26	27
Máx. (° F)	98	91	94	98	91

## 5.2 Agrupamento de dados de previsão

O modelo de carga é um reflexo das atividades da população. Assim, tem-se características bem definidas para certos dias e certas épocas do ano. Curvas de carga de fins de semana e de dias de semana apresentam características bem diferentes.

Diante disso, optou-se por agrupar os dados disponíveis por dias semelhantes, treinando-se uma PMC para cada tipo de dia da semana.

## 5.3 Identificação das variáveis explicativas do processo gerador da série histórica

Muitos métodos têm sido usados na seleção de variáveis de entrada para uma RNA [Chen et al., 1992; Lu et al., 1993; Papalexopoulos et al., 1994; Piras et al., 1996]. A maioria deles utiliza funções de autocorrelação, acompanhadas ou não de heurísticas, e da experiência do operador. Outros se baseiam no espectro de potência da série. Em [Drezga & Rahman, 1998], os autores usaram um método baseado no cálculo da deformação local<sup>1</sup> de trajetórias no espaço-fase proposto por [Buzug & Pfister, 1992] para descobrir relações não lineares entre as variáveis. Drezga & Rahman fizeram uma abordagem univariável do problema de REF.

Nesta tese, duas metodologias distintas foram empregadas. Na primeira delas, utilizou-se a função de autocorrelação na identificação das variáveis de entrada referentes à carga para os diversos previsores testados. Um dos métodos mais populares para se calcular a autocorrelação é mostrado na equação 5.3. As equações 5.1 e 5.2 são, respectivamente, a autocovariância e a variância da série. Esse método foi empregado nos modelos descritos nos sub-itens 5.9.1 a 5.9.4.

<sup>1</sup> *Integral Local Deformation*, do inglês.

$$C_k = \frac{1}{n-k} \sum_{i=k+1}^n (z_i - \bar{z}) \cdot (z_{i-k} - \bar{z}) \quad \text{Eq.5.1}$$

$$C_0 = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2 \quad \text{Eq.5.2}$$

$$r_k = \frac{C_k}{C_0} \quad \text{Eq.5.3}$$

A segunda metodologia é baseada no procedimento de reconstrução do espaço-fase apresentado no item 4.3 do capítulo anterior. Em linhas gerais, escolhe-se um segmento da curva de carga (e.g., um mês de dados) e estimam-se os parâmetros valor em atraso ( $\tau$ ) e dimensão de imersão ( $d$ ) a partir dos algoritmos para o cálculo da IMM e identificação de falsos vizinhos no espaço-fase, respectivamente. De posse desses parâmetros, a reconstrução do espaço-fase do sistema de interesse pode ser obtida a partir do método das coordenadas em atraso (Eq. 4.4). Finalmente, baseado nos resultados da REF, um conjunto independente de variáveis de carga pode ser selecionado. Esse algoritmo foi empregado nos modelos de previsão 5 e 6. Tais modelos serão discutidos nos sub-itens 5.9.5 e 5.9.6.

## 5.4 Definição do número de pontos utilizados no processo de treinamento/validação

Concatenações de janelas de seis semanas para o ano corrente e para os períodos equivalentes – um ano antes e dois anos antes, foram usadas para fins de treinamento e validação. Ressalta-se que janelas de sete e oito semanas também foram testadas. No entanto, a utilização de janelas de seis semanas produziu resultados mais satisfatórios que as demais em termos de MAPE. Esse índice será discutido no item 5.12.



## 5.5 Etapa de pré-processamento dos dados disponíveis

Antes que um conjunto de dados esteja pronto para formar um conjunto de treinamento para RNA, é aconselhável (senão obrigatório) pré-processá-lo. Técnicas tais como normalização, padronização, diferenciação e análise de multiresolução via transformada wavelet discreta foram consideradas neste trabalho. Além da normalização usual, um procedimento complementar baseado na padronização das variáveis, i.e., transformar as variáveis de modo que elas tenham média zero e desvio padrão unitário pode ser muito útil. O processo de diferenciação subtrai valores de carga adjacentes, i.e., a nova série representa as variações da série no domínio do tempo. Uma boa razão para se aplicar a diferenciação reside no fato que, dependendo da variável, suas variações podem ser tão importantes quanto os seus valores originais. Essas técnicas são discutidas em detalhes em [Reis, 1999]. Finalmente, a AMR via TWD extrai as características importantes da dinâmica da carga.

## 5.6 Partição ótima dos dados

Uma teoria estatística do fenômeno de sobreparametrização (i.e., *overfitting*), apresentada em [Amari et al., 1996], analisa dois modos de comportamento de um MLP dependendo do tamanho do conjunto de treinamento. São eles:

### a) Modo não assintótico

O modo não assintótico é caracterizado por  $N < 30W$ , onde  $N$  é o número total de padrões de treinamento disponíveis e  $W$  é o número de parâmetros livres da rede. Para esse modo de comportamento, o Critério de Parada Prematura (CPP) discutido no item I.5.6, aumenta a capacidade de generalização da rede se comparado com o treinamento exaustivo (i.e., quando um conjunto de exemplos é usado para o treinamento e a sessão de treinamento não é interrompida). Esse resultado sugere que a sobreparametrização pode ocorrer para  $N < 30W$ , e existe mérito prático em se usar a validação cruzada para parar o treinamento. O valor ótimo do parâmetro  $p$  que determina a partição dos dados disponíveis em conjunto de treinamento e conjunto de teste é definido por:

$$p_{\text{ótimo}} = \frac{\sqrt{2W - 1} - 1}{2(W - 1)} \quad \text{Eq.5.4}$$

Consequentemente,

$$N_{\text{treinamento}} = p_{\text{ótimo}} \cdot N, \quad \text{Eq.5.5}$$

$$N_{\text{teste}} = N - N_{\text{treinamento}}, \quad \text{Eq.5.6}$$

em que,  $N$  é o número de padrões disponíveis,  $N_{\text{treinamento}}$  é o número de padrões para fins de treinamento e  $N_{\text{validação}}$  é o número de padrões para fins de validação.

### b) Modo assintótico

Caso  $N > 30W$ , o aumento da capacidade de generalização da rede produzido pelo CPP é pequeno se comparado ao treinamento exaustivo. Em outras palavras, a aprendizagem exaustiva é satisfatória quando o número de padrões de treinamentos disponível é suficientemente grande se comparado ao número de parâmetros livres da rede.

Em todas as situações analisadas neste trabalho, pôde-se constatar que todos os previsores implementados operavam no modo não assintótico. Portanto, com o intuito de sistematizar o cálculo de  $W$  e valendo-se da Fig. 5.1, deduziu-se a seguinte equação<sup>2</sup>:

$$W = [S_1 \cdot (1 + n_1 + S_2) + S_2], \quad \text{Eq. 5.7}$$

em que:

$W$  é o número de parâmetros livres da rede;

$S_1$  é o número de neurônios na camada oculta;

$S_2$  é o número de neurônios na camada de saída;

$n_1$  é o número de entradas da rede.

<sup>2</sup> A Eq. 5.7 é válida para uma única camada oculta

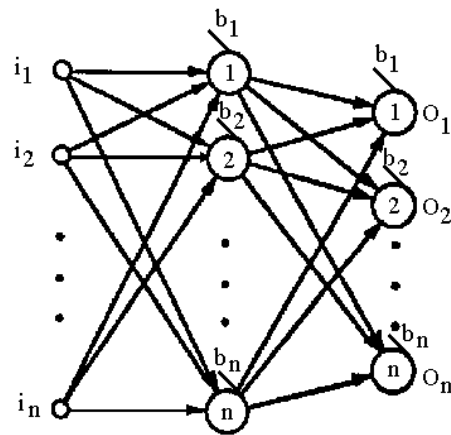


Fig. 5.1. Estrutura típica de um MLP

Esta seção termina com um exemplo. Suponha que uma estrutura MLP, com duas unidades na camada oculta e uma única unidade de saída, tenha seis entradas e 144 padrões de treinamento disponíveis. Qual seria a partição ótima entre os conjuntos de treinamento e validação?

$$S_1=2; S_2=1; n_1=6; N=144;$$

$$W=?$$

$$W = [2 \cdot (1 + 6 + 1) + 1] \Rightarrow W = 17$$

$$\therefore 30W = 510$$

Como  $N < 30W$ ,

$$p_{\text{ótimo}} = \frac{\sqrt{2 \cdot 17 - 1} - 1}{2 \cdot (17 - 1)} = 0,85, \text{ logo:}$$

$$N_{\text{treinamento}} = 0,85 \cdot 144 = 122,6 \Rightarrow N_{\text{treinamento}} = 123 \text{ padrões.}$$

$$N_{\text{validação}} = 144 - 123 \Rightarrow N_{\text{validação}} = 21 \text{ padrões.}$$

## 5.7 Definição da estrutura do PMC

O procedimento para se definir a estrutura de um PMC envolve basicamente os seguintes tópicos: seleção das variáveis de entrada, escolha do número de neurônios nas camadas oculta e de saída. O primeiro tópico foi assunto da seção 5.3 e não será tratado aqui novamente. Quanto à escolha do número de neurônios na camada oculta, via de regra, o *princípio da navalha de Occam*<sup>3</sup> deve ser seguido. Em linhas gerais, esse princípio implica em parcimônia. No caso, quanto mais simples a estrutura da rede, menor o número de pesos e bias a serem ajustados, e menor o risco da ocorrência da sobreparametrização durante a fase de treinamento.

Em [Zhang & Dong, 2001], os autores sugerem o uso de uma estrutura neural similar a uma “garrafa”, em que o número de neurônios na camada oculta é menor que o número de variáveis de entrada.

Para uma discussão mais elaborada sobre a seleção de estrutura para modelos não lineares (incluindo redes neurais), ver [Aguirre, 1994]. Em [Drezga, 1996], o autor propõe um algoritmo para a determinação do número de neurônios na camada oculta de um PMC.

Nesta tese, a seguinte metodologia foi adotada:

- 1º passo: Criar uma estrutura com um único neurônio na camada oculta;
- 2º passo: Realizar atividades de treinamento, validação e teste, armazenando os erros observados durante a etapa de testes da estrutura (um ano de dados foi utilizado para fins de testes);
- 3º passo: Aumentar o número de neurônios em uma unidade, mantendo todos os demais parâmetros inalterados (e.g., número de épocas, semente, etc.);
- 4º passo: Voltar para o 1º passo até que o número de neurônios na camada oculta seja igual a, e.g., 10.

## 5.8 Algoritmo de aprendizagem

O algoritmo utilizado no treinamento do PMC foi uma variante do algoritmo de Retropropagação de Erro (EBP) [Masters, 1995]. Via de regra, o EBP padrão é bastante lento porque requer baixas taxas de aprendizagem para que se consiga um treinamento estável. A

versão utilizada aqui conta com dois parâmetros adicionais, a saber: taxa de *momentum*  $\xi$  e taxa de aprendizagem adaptativa  $\eta$ .

O parâmetro  $\xi$  diminui a sensibilidade do EBP a pequenos detalhes na superfície de erro. Isso auxilia a RNA a não se prender a um mínimo local, permitindo que a rede possa encontrar uma solução melhor, ou até mesmo, atingir o mínimo global da função. Com  $\xi$ , a rede passa a responder não apenas ao gradiente local, mas também a tendências recentes na superfície de erro.

Por outro lado, o tempo de treinamento também pode ser diminuído usando-se uma taxa de aprendizagem adaptativa que mantenha o tamanho do passo de aprendizagem tão grande quanto possível, enquanto que mantendo um aprendizado estável. O parâmetro  $\eta$  é sensível à complexidade da superfície de erro local. O fluxograma abaixo ilustra a implementação da taxa de aprendizagem adaptativa no EBP padrão.

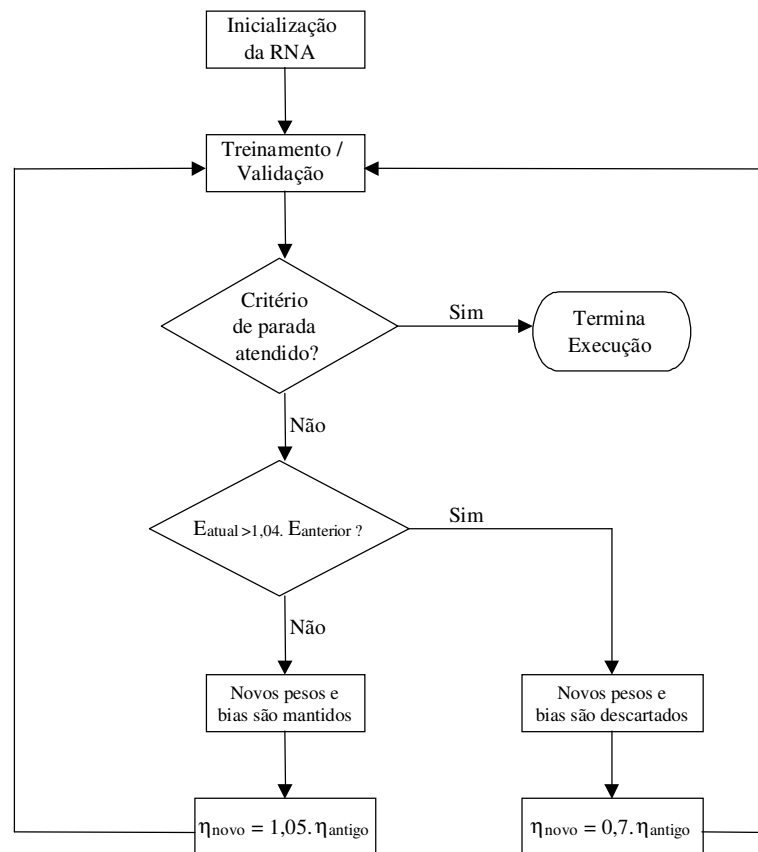


Fig. 5.2. Algoritmo para a implementação de  $\eta$  adaptativa.

<sup>3</sup> Occam's razor principle, do inglês.

## 5.9 Modelos de PCCP via PMC

### 5.9.1 Modelo 1 (M1)

O modelo 1 é, de certa forma, bastante conhecido. Após definir-se a “janela” de dados, faz-se a normalização e a padronização da mesma. Esse tipo de pré-processamento faz com que os dados fiquem no intervalo  $[-0,5; 0,5]$ .

A Fig. 5.3 mostra um autocorrelograma típico obtido para um dado segmento de carga (outubro/1990). Com base nele, os valores em atraso 1, 2, 24 e 168h foram selecionados como variáveis de entrada para o previsor (i.e.,  $P(t-1)$ ,  $P(t-2)$ ,  $P(t-24)$ ,  $P(t-168)$ ). Em paralelo, variáveis de temperatura foram adicionadas para cada um dos instantes de tempo considerados anteriormente, mais a temperatura prevista para a hora alvo. Finalmente, utilizou-se ainda duas entradas adicionais com a finalidade de codificar as 24 horas do dia. Essas entradas foram denominadas por “Hora\_Seno” e “Hora\_Cos” e foram calculadas a partir das equações 5.8 e 5.9.

$$Hora\_Seno(t) = \text{sen}\left(\frac{2\pi t}{24}\right), \quad \text{Eq.5.8}$$

$$Hora\_Cos(t) = \text{cos}\left(\frac{2\pi t}{24}\right), \quad \text{Eq.5.9}$$

em que  $t$  é a hora corrente.

Para fins de treinamento, a saída da RNA é a potência ativa no instante de tempo  $t$  (i.e.,  $P(t)$ ). A arquitetura da rede é composta de 11 entradas, 4 neurônios na camada oculta e 1 neurônio na camada de saída. As funções de ativação dos neurônios da camada oculta e de saída são do tipo tangente hiperbólica e linear, respectivamente. A Fig. 5.4 representa o modelo 1.

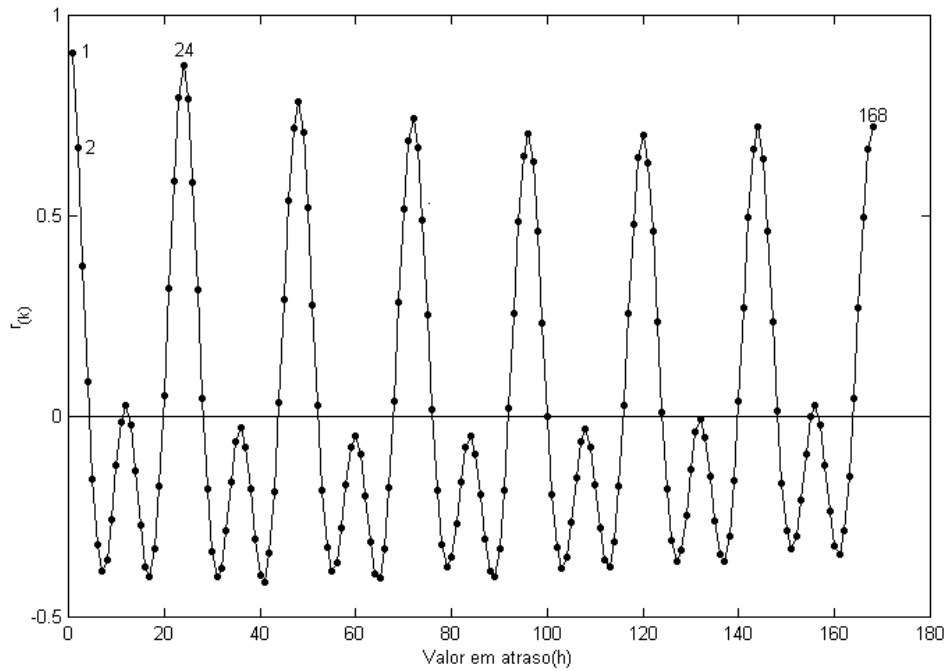


Fig. 5.3 Autocorrelograma típico de um segmento de série de carga (outubro/1990).

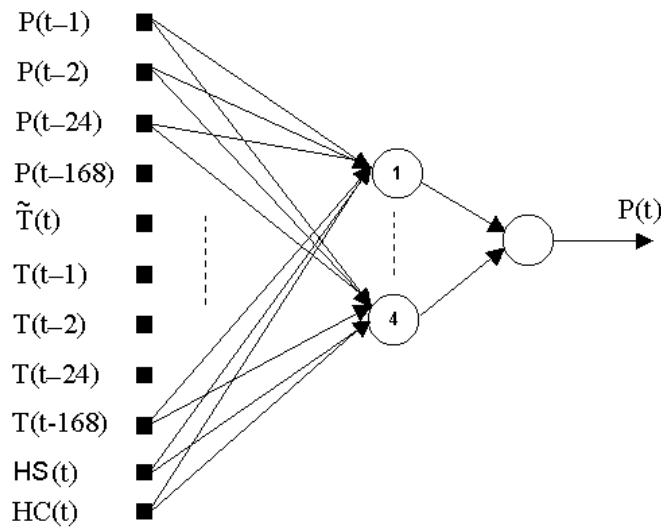


Fig. 5.4. Diagrama para M1.

### 5.9.2 Modelo 2 (M2)

O segundo previsor de carga adiciona variáveis diferenciadas ao conjunto de entradas de M1. Portanto, três séries são utilizadas: carga normalizada/padronizada, temperatura normalizada/padronizada e carga diferenciada. A diferenciação é aplicada à série de carga padronizada. Para M2, apenas as diferenças de primeira ordem são consideradas. As variações

de carga foram adicionadas para cada um dos instantes de tempo considerados para a variável carga. A Tab. 5.2 mostra o conjunto de entradas para M2.

Lembrando-se que a série diferenciada tem um ponto a menos que sua respectiva série no domínio do tempo, adotou-se o procedimento abaixo para que as séries apresentassem o mesmo comprimento.

1. Define-se a série de interesse com um número de pontos igual a 1033 (em M1, essa série contava com 1032 pontos), armazenando-a. Esse ponto extra é o ponto imediatamente anterior ao primeiro ponto da série de carga definida para M1.
2. Aplica-se a normalização, padronização e diferenciação simples à série. A série diferenciada contém 1032 pontos e é multiplicada por um fator de escala  $C$  (no caso  $C=2$ ).
3. Descarta-se o primeiro ponto da série definida no passo 1.

Realizado esse procedimento, as duas séries apresentam o mesmo comprimento e os padrões de entrada/saída podem ser montados sem que a referência temporal seja perdida (esse procedimento deve ser repetido para cada uma das três janelas de dados consideradas no item 5.4). A Fig. 5.5 apresenta o diagrama esquemático do modelo. A saída da RNA é a carga medida no instante de tempo  $t$  da qual é subtraída a carga medida uma hora antes (i.e.,  $D(t)=P(t)-P(t-1)$ ). A arquitetura neural consiste de 15 entradas, 4 neurônios na camada oculta e 1 neurônio na camada de saída.

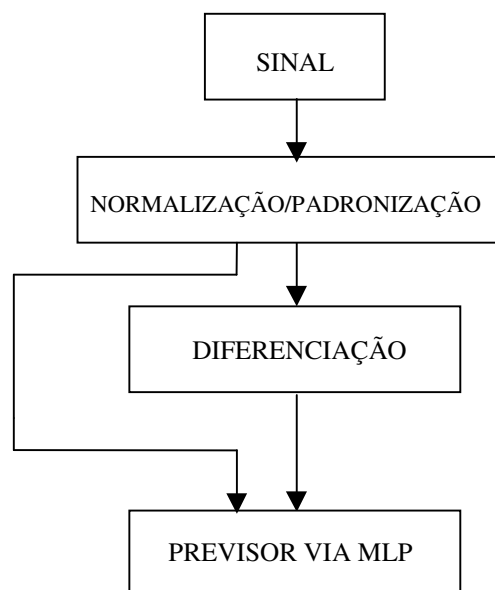


Fig. 5.5 Diagrama esquemático do modelo 2.



Tab. 5.2 Variáveis de entrada para M2.

N.º da entrada	Tipo de variável	Lags (t)	Representação
1–4	Carga (P)	1, 2, 24, 168	P(t-1), P(t-2), P(t-24), P(t-168)
5–8	Carga diferenciada (D)	1, 2, 24, 168	D(t-1), D(t-2), D(t-24), D(t-168),
9	Temperatura prevista	0*	$\tilde{T}(t)$
10–13	Temperatura (T)	1, 2, 24, 168	T(t-1), T(t-2), T(t-24), T(t-168)
14	Hora seno	0*	H_seno (t)
15	Hora coseno	0*	H_cos (t)

\* O lag 0 representa o instante alvo.

### 5.9.3 Modelo 3 (M3)

O modelo 3 (M3) combina normalização, padronização e análise de multiresolução via wavelets. As subfamílias da wavelet mãe Daubechies de ordem 2 (daub2), 3 (daub3) e 4 (daub4) foram avaliadas e tiveram os seus desempenhos comparados. Três níveis de decomposição foram considerados.

Há, basicamente, duas razões principais que justificam a incorporação das wavelets neste modelo. Primeiramente, as características importantes de uma série de carga podem ser enfatizadas. Em segundo lugar, a componente ruidosa da série carga pode ser separada. A Fig. 5.6 apresenta o esquema de M3.

Como uma tentativa inicial, a aproximação  $A_3$  e todos os três níveis de detalhes ( $D_1$ ,  $D_2$  e  $D_3$ ) foram adicionados ao conjunto de entradas de M2. Com essa configuração, M3 não produziu resultados satisfatórios. Isso já era esperado de alguma forma devido ao fato de que todas as fontes de informação (incluindo informação ruidosa) foram apresentadas a M3.

Analisando-se os níveis de resolução, constata-se que as subséries mais informativas são  $A_3$  e  $D_3$ . Já os níveis  $D_1$  e  $D_2$  contêm algumas irregularidades que são, via de regra, provenientes de erros de medição e de outras fontes aleatórias de ruído. Essa breve análise sugere que a informação fornecida pelas subséries  $D_1$  e  $D_2$  não deve ser incorporada a M3.

O conjunto de entradas para M3 pode ser encontrado na Tab. 5.3. A saída da RNA é a carga medida no instante de tempo  $t$  da qual é subtraída a carga medida uma hora antes (i.e.,

$D(t)=P(t)-P(t-1)$ ). A arquitetura neural consiste em 23 entradas, 4 neurônios na camada oculta e 1 neurônio na camada de saída.

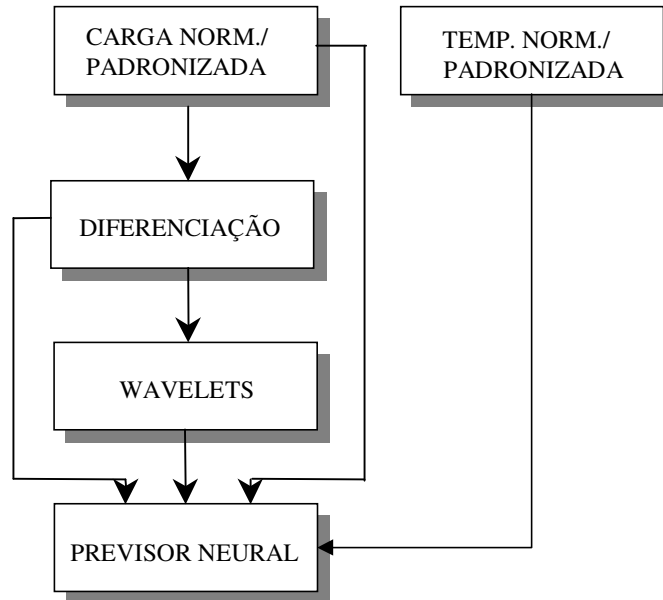


Fig. 5.6. Diagrama esquemático do modelo 3.

**Tab. 5.3** Variáveis de entrada para M3.

N.º da entrada	Tipo de variável	Lags (t)	Representação
1–4	Carga Diferenciada (CD)	1, 2, 24, 168	CD(t-1), CD(t-2), CD(t-24), CD(t-168),
5–8	Aproximação ( $A_3$ )	1, 2, 24, 168	$A_3(t-1)$ , $A_3(t-2)$ , $A_3(t-24)$ , $A_3(t-168)$
9–12	Carga (P)	1, 2, 24, 168	P(t-1), P(t-2), P(t-24), P(t-168)
13–16	Detalhe ( $D_3$ )	1, 2, 24, 168	$D_3(t-1)$ , $D_3(t-2)$ , $D_3(t-24)$ , $D_3(t-168)$
17	Temperatura Prevista	0	~ (t)
18–21	Temperatura (T)	1, 2, 24, 168	T(t-1), T(t-2), T(t-24), T(t-168)
22	Hora seno	0	H_seno (t)
23	Hora coseno	0	H_cos (t)

### 5.9.4 Modelo 4 (M4)

O modelo 4 apresenta uma abordagem completamente diferente. Aqui, a idéia é decompor a série de carga usando-se AMR via wavelets da família Daubechies e modelá-la através do ajuste individual de cada nível de resolução. Isto é, cada componente ( $A_3$ ,  $D_3$ ,  $D_2$  e  $D_1$ ) é modelada separadamente e a previsão final é obtida somando-se as quatro previsões realizadas. Note que para M4, a aproximação  $A_3$  e os três níveis de detalhes devem ser necessariamente levados em conta para que as previsões possam ser conhecidas. A Fig. 5.7 mostra o diagrama esquemático desse modelo. Cada um de seus submodelos será apresentado na seqüência.

#### a) Submodelo para $A_3$

Como a aproximação  $A_3$  é uma versão suavizada da série de carga, as mesmas variáveis de entrada de M1 foram selecionadas (Tab. 5.4). A única diferença é que esse submodelo consiste de 3 neurônios na camada oculta (ao invés de 4) e um neurônio na camada de saída.

#### b) Submodelos para $D_3$ e $D_2$

As funções de autocorrelação e correlação cruzada foram utilizadas para selecionar as variáveis de entrada para  $D_3$  e  $D_2$ . A Tab. 5.4 apresenta o conjunto de entradas e saídas para esses submodelos. Nota-se que ambos modelos apresentam a mesma estrutura. No entanto, eles são bastantes diferentes do submodelo anterior. As principais diferenças são:

- 1) 5 variáveis de entrada (ao invés de 4) foram selecionadas pela função de autocorrelação – note que o valor em atraso de 12 horas foi incluído;
- 2) a variável temperatura foi descartada – isso se deve ao fato de que as correlações cruzadas obtidas entre esses níveis e a temperatura não são significativas. Cada um desses submodelos consiste de 3 neurônios na camada oculta e um neurônio na camada de saída.

#### c) Submodelo para $D_1$

Como o nível de detalhe  $D_1$  está mais relacionado com a parte ruidosa da série de carga, as previsões para esse nível de resolução são baseadas exclusivamente em valores médios. A média é estimada sobre os dados das janelas de seis semanas para o ano corrente e para os períodos equivalentes – um ano antes e dois anos antes.

Tab. 5.4 Variáveis de entrada para M4.

RNA	Entrada	Variável	Lags (t)	Representação	Saída
1	1–4	Aproximação ( $A_3$ )	1, 2, 24, 168	$A_3(t-1)$ , $A_3(t-2)$ , $A_3(t-24)$ , $A_3(t-168)$	$A_3(t)$
	5–8	Série $A_3$ diferenciada ( $DA_3$ )	1, 2, 24, 168	$DA_3(t-1)$ , $DA_3(t-2)$ , $DA_3(t-24)$ , $DA_3(t-168)$	
	9	Temp. prevista	0	$\tilde{T}(t)$	
	10–13	Temperatura (T)	1, 2, 24, 168	$T(t-1)$ , $T(t-2)$ , $T(t-24)$ , $T(t-168)$	
	14	Hora seno	0	$H_{\text{seno}}(t)$	
	15	Hora coseno	0	$H_{\text{cos}}(t)$	
2	1–5	Detalhe ( $D_3$ )	1, 2, 12, 24, 168	$D_3(t-1)$ , $D_3(t-2)$ , $D_3(t-12)$ , $D_3(t-24)$ , $D_3(t-168)$	$D_3(t)$
	6–10	Série $D_3$ diferenciada ( $DD_3$ )	1, 2, 12, 24, 168	$DD_3(t-1)$ , $DD_3(t-2)$ , $DD_3(t-12)$ , $DD_3(t-24)$ , $DD_3(t-168)$	
	11	Hora seno	0	$H_{\text{seno}}(t)$	
	12	Hora coseno	0	$H_{\text{cos}}(t)$	
3	1–5	Detalhe ( $D_2$ )	1, 2, 12, 24, 168	$D_2(t-1)$ , $D_2(t-2)$ , $D_2(t-12)$ , $D_2(t-24)$ , $D_2(t-168)$	$D_2(t)$
	6–10	Série $D_2$ diferenciada ( $DD_2$ )	1, 2, 12, 24, 168	$DD_2(t-1)$ , $DD_2(t-2)$ , $DD_2(t-12)$ , $DD_2(t-24)$ , $DD_2(t-168)$	
	11	Hora seno	0	$H_{\text{seno}}(t)$	
	12	Hora coseno	0	$H_{\text{cos}}(t)$	

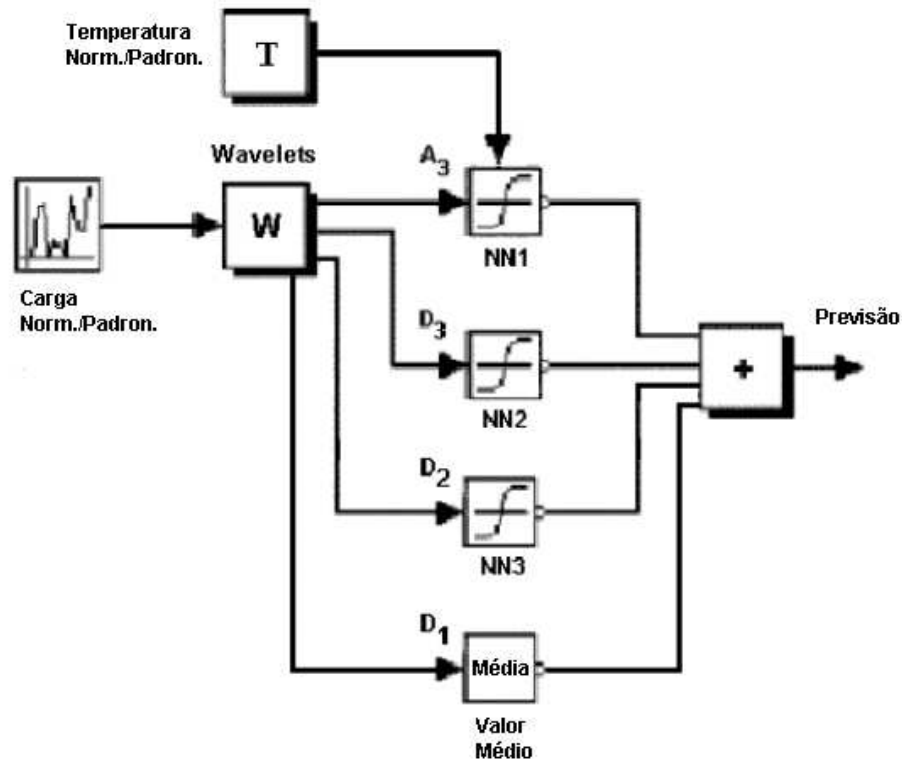


Fig. 5.7. Diagrama esquemático do modelo 4.

### 5.9.5 Modelo 5 (M5)

O modelo 5 é similar ao modelo 1. A principal diferença entre eles está relacionada com o método pelo qual as suas variáveis de entrada foram selecionadas. Para M1, utilizou-se a função de autocorrelação. Já para M5, considerou-se a metodologia para seleção de variáveis de entrada via reconstrução do espaço-fase apresentada em [Drezga e Rahman, 1998]. Em suma, a abordagem proposta pelos autores pode ser sintetizada em quatro passos, a saber:

1. Considera-se um segmento de série de carga (e.g., um mês de dados de carga);
2. Estimam-se os parâmetros  $\tau$  (valor em atraso) e  $d$  (dimensão de imersão);
3. De posse desses parâmetros, o método das coordenadas em atraso pode ser utilizado (ver Eq. 4.4) e a seleção das variáveis de entrada referentes à carga é realizada escolhendo-se cada coordenada como variável de entrada.
4. Finalmente, inclui-se no conjunto de entradas do previsor neural, as variáveis de entrada adicionais – temperatura, hora\_seno e hora\_cos.

Para M5, a mesma metodologia foi empregada com o objetivo de verificar sua aplicabilidade na série de carga disponível.

Definido o segmento de série de carga a ser analisado, aplica-se os algoritmos IMM e identificação de falsos vizinhos no espaço-fase para calcular-se os parâmetros  $\tau$  e  $d$ . Para o nosso caso,  $\tau$  vale 2 e  $d$  é igual a 5. Substituindo-se esses valores na Eq. 4.4, o seguinte mapa de reconstrução pode ser obtido:  $\varphi_{n-1} = (s_{n-9}, s_{n-7}, s_{n-5}, s_{n-3}, s_{n-1})$ .

Portanto, conclui-se que a metodologia aponta as entradas P(t-1), P(t-3), P(t-5), P(t-7), P(t-9), como sendo aquelas mais representativas da série de carga. A Tab. 5.5 apresenta o conjunto completo das variáveis utilizadas por M5.

Para fins de treinamento, a saída da RNA é a potência ativa no instante de tempo  $t$  (i.e., P(t)). A arquitetura da rede é composta de 13 entradas, 4 neurônios na camada oculta e 1 neurônio na camada de saída. As funções de ativação dos neurônios da camada oculta e de saída são do tipo tangente hiperbólica e linear, respectivamente.

Tab. 5.5 Variáveis de entrada selecionadas para M5.

N.º da entrada	Tipo de variável	Lags (t)	Representação
1–5	Carga (P)	1, 3, 5, 7, 9	P(t-1), P(t-3), P(t-5), P(t-7), P(t-9)
6	Temperatura prevista	0*	$\tilde{T}(t)$
7–11	Temperatura (T)	1, 3, 5, 7, 9	T(t-1), T(t-3), T(t-5), T(t-7), T(t-9)
12	Hora seno	0*	H_seno (t)
13	Hora coseno	0*	H_cos (t)

### 5.9.6 Modelo 6 (M6)

O modelo 6 é similar ao modelo 4. A principal diferença é que ao invés de se utilizar a função de autocorrelação na determinação das variáveis de entrada para o submodelo  $A_3$ , usou-se a metodologia apresentada no item anterior. Com essa abordagem, as entradas  $A_3(t-1)$ ,  $A_3(t-4)$ ,  $A_3(t-7)$ ,  $A_3(t-10)$ ,  $A_3(t-13)$ , foram identificadas como sendo aquelas mais representativas da série de carga. A Tab. 5.6 apresenta o conjunto completo das variáveis utilizadas por M6. Cada um dos previsores utilizados pelos submodelos foi apresentado no item 5.9.4 e permaneceram inalterados.

Tab. 5.6 Variáveis de entrada para M6.

RNA	Entrada	Variável	Lags (t)	Representação	Saída
1	1–5	Aproximação ( $A_3$ )	1, 3, 6, 9, 12	$A_3(t-1), A_3(t-3),$ $A_3(t-6), A_3(t-9), A_3(t-12)$	$A_3(t)$
	6–10	Série $A_3$ diferenciada ( $DA_3$ )	1, 3, 6, 9, 12	$DA_3(t-1), DA_3(t-3),$ $DA_3(t-6), DA_3(t-9),$ $DA_3(t-12)$	
	11	Temp. prevista	0	$\tilde{T}(t)$	
	12–15	Temperatura (T)	1, 3, 6, 9, 12	$T(t-1), T(t-3), T(t-6),$ $T(t-9), T(t-12)$	
	16	Hora seno	0	$H_{\text{seno}}(t)$	
	17	Hora coseno	0	$H_{\text{cos}}(t)$	
2	1–5	Detalhe ( $D_3$ )	1, 2, 12, 24, 168	$D_3(t-1), D_3(t-2), D_3(t-12),$ $D_3(t-24), D_3(t-168)$	$D_3(t)$
	6–10	Série $D_3$ diferenciada ( $DD_3$ )	1, 2, 12, 24, 168	$DD_3(t-1), DD_3(t-2),$ $DD_3(t-12), DD_3(t-24),$ $DD_3(t-168)$	
	11	Hora seno	0	$H_{\text{seno}}(t)$	
	12	Hora coseno	0	$H_{\text{cos}}(t)$	
3	1–5	Detalhe ( $D_2$ )	1, 2, 12, 24, 168	$D_2(t-1), D_2(t-2), D_2(t-12),$ $D_2(t-24), D_2(t-168)$	$D_2(t)$
	6–10	Série $D_2$ diferenciada ( $DD_2$ )	1, 2, 12, 24, 168	$DD_2(t-1), DD_2(t-2),$ $DD_2(t-12), DD_2(t-24),$ $DD_2(t-168)$	
	11	Hora seno	0	$H_{\text{seno}}(t)$	
	12	Hora coseno	0	$H_{\text{cos}}(t)$	

## 5.10 Etapas de treinamento/teste

As atividades de treinamento e teste dos modelos apresentados foram feitas baseadas em validação cruzada. A seguir, apresenta-se o procedimento adotado.

1º passo: Define-se a ‘janela’ de interesse.

2º passo: Faz-se o pré-processamento dos dados adequado para cada um dos modelos apresentados.

3º passo: Montam-se os padrões de entrada/saída (432 padrões de entrada/saída estarão disponíveis; o que equivale a 18 dias semelhantes, sendo 6 dias semelhantes por ano).

4º passo: Uma RNA é inicializada com pesos aleatórios e de pequena magnitude (entre -0,5 e 0,5).

5º passo: Define-se um valor inicial de “semente” para ser utilizado na geração de números pseudo-aleatórios (e.g.,  $seed=1$ ). Define-se também um valor para o erro médio quadrático de teste inicial (e.g.,  $e_{teste}=1000$ ).

6º passo: Geram-se 432 números aleatórios distribuídos uniformemente. Esses números serão associados aos padrões de entrada e variarão entre 1 e 432. Eles terão a função de rotular cada um dos padrões.

7º passo: Particionam-se os padrões de entrada disponíveis em conjunto de treinamento e validação, valendo-se das Eqs. 5.4, 5.5 e 5.6.

8º passo: Treina-se a rede por 500 épocas.

9º passo: Avalia-se a rede, apresentando-lhe os padrões que compõem o conjunto de validação.

10º passo: Calcula-se o desvio médio quadrático total para o conjunto de validação. Caso o erro atual seja menor que o valor anterior ( $e_{teste}$  anterior), os pesos e bias são armazenados. Senão, são descartados.

11º passo: soma-se 1 ao valor da semente inicial e volta-se ao 6º passo.

Esse procedimento é repetido 50 vezes.



Uma vez treinada a rede, desloca-se a janela de dados do ano corrente 24 pontos à frente, descartando-se os primeiros 24 pontos da janela anterior. Desse modo, a nova janela permanecerá com 1032 pontos e uma nova rede poderá ser treinada. O mesmo procedimento deve ser repetido para as mesmas janelas dos dois anos anteriores. Finalmente, ressalta-se que o enchimento deve ser considerado para os modelos M3, M4 e M6 (aqueles que utilizam wavelets).

## 5.11 Tipos de previsão

Para problemas de previsão de carga, via de regra, considera-se dois tipos de previsão:

- previsão um passo à frente;
- previsão recursiva de 1 a 24 passos à frente.

Na previsão de um passo à frente, o que se objetiva é prever, a partir de um dado instante, o consumo para a próxima hora. A rede é alimentada exclusivamente com dados observados. Já na previsão recursiva, objetiva-se prever o consumo de EE para as 24 horas seguintes. Nesse tipo de previsão, somente a previsão do primeiro passo é realizada com as entradas observadas. As demais previsões são realizadas baseadas nas previsões anteriores.

Existem também dois modos de como essas previsões podem ser realizadas: modos off-line ou on-line. No modo off-line, o operador somente pode obter a previsão para as próximas 24 horas, a partir de uma dada hora fixa (e.g., zero hora). No modo on-line, previsões de 1 a 24 passos podem ser obtidas a partir de qualquer instante.

## 5.12 Índices de desempenho

Durante a fase de desenvolvimento dos previsores sabe-se de antemão os valores verdadeiros para cada uma das previsões realizadas. O problema é expressar o desempenho de cada um dos modelos de maneira conclusiva.

Nesta tese, utilizaram-se quatro índices para essa tarefa. O primeiro deles é o *erro percentual absoluto médio* (MAPE) (ver Eq. 5.10). Esse índice é tido pelas concessionárias de EE como um índice padrão para a avaliação do desempenho de previsores de carga. Em geral,

MAPEs abaixo de 3% para previsões de 1-24 passos à frente são considerados bastante satisfatórios pelas concessionárias de EE [Khotanzad et al., 1998].

$$MAPE = \frac{1}{24} \sum_{h=1}^{24} \frac{|previsto_{(t)} - real_{(t)}|}{real_{(t)}} \times 100\%, \quad \text{Eq.5.10}$$

em que,  $real_{(t)}$  é o valor observado no instante  $t$  e  $previsto_{(t)}$  é o valor previsto no instante  $t$ .

O segundo índice é o *erro médio quadrático* (MSE) (Eq. 5.11).

$$MSE = \frac{1}{24} \sum_{h=1}^{24} (previsto_{(t)} - real_{(t)})^2 \quad \text{Eq. 5.11}$$

O terceiro é o *erro médio* (ME). Sua equação é dada por:

$$ME = \frac{1}{24} \sum_{h=1}^{24} (previsto_{(t)} - real_{(t)}) \quad \text{Eq. 5.12}$$

O quarto e último índice quantificam em termos percentuais, a melhoria de desempenho obtida de um modelo em relação a outro com respeito ao índice MAPE (Eq. 5.13).

$$ganho(\%) = \left( \frac{MAPE_{anterior} - MAPE_{atual}}{MAPE_{anterior}} \right) \cdot 100\% \quad \text{Eq. 5.13}$$

Finalizando, deve-se ressaltar que cada um dos três primeiros índices fornece informações diferentes a respeito dos modelos. Em linhas gerais, o MAPE permite avaliar, em média, o comportamento de cada um dos modelos. O MSE dá informações a respeito de qual dos modelos apresenta os maiores erros horários. Já o ME informa o quão tendencioso é um dado modelo.

## 5.13 Testes

Na ausência de dados referentes à previsão de temperatura, dois estudos foram considerados. No primeiro deles (Caso A), seguindo a prática usual, dados referentes à temperatura medida foram empregados. Já no segundo estudo (Caso B), investigou-se o efeito da adição de ruído Gaussiano nos dados de temperatura medidos para simular erros de previsão. Finalmente, uma comparação entre o desempenho dos modelos M1 e M4 com outros dois modelos neurais apresentados em [Brace et al., 1991] foi realizada (Caso C).

### 5.13.1 Caso A

Uma comparação entre os seis modelos propostos foi realizada. Para fins de treinamento e validação dos modelos, janelas de seis semanas do ano corrente, mais janelas de seis semanas do mesmo período dos dois anos anteriores foram utilizadas (e.g., out./nov. 1988, out./nov. 1989 e out./nov. 1990). O agrupamento de dados foi feito de acordo com o dia da semana. Para cada dia da semana, um PMC foi treinado com uma variante do algoritmo EBP com validação cruzada [Masters, 1995]. Diferentes partições para os conjuntos de treinamento e validação foram criadas aleatoriamente a cada 500 épocas (em média, 85% dos padrões disponíveis foram utilizados para treinamento). Durante o treinamento, nenhum tratamento especial para feriados foi aplicado.

Após o treinamento para previsões um passo à frente, previsões de carga para a hora seguinte (i.e., um passo à frente) e recursivas de 1–24 passos à frente são computadas. Elas se iniciam à meia-noite. Os previsores são retreinados no fim do dia a fim de incorporar a informação de carga mais recente. Move-se a janela de treinamento um dia à frente e as previsões para as próximas 24 horas são realizadas. Esse procedimento de teste é repetido para um período de dois anos completos. Os índices MAPE, MSE, ME e Erro Máximo (MAX) foram utilizados para avaliar os modelos de previsão de carga.

## a) Avaliando-se dias típicos

As tabelas 5.7 a 5.10 sintetizam os resultados obtidos exclusivamente para dias de semana regulares. Feriados e dias atípicos são tratados à parte.

### 1) Análise da previsão recursiva

Com base na Tab. 5.8, nota-se que M1 (nosso ‘benchmark’) teve um desempenho razoável. No entanto, a incorporação de variáveis de carga diferenciadas no modelo, melhorou ligeiramente o seu desempenho (M2). Com M3, que explora a propriedade de multiresolução, uma melhoria considerável pôde ser verificada. Já M4, em média, é o modelo mais acurado. Para múltiplos passos à frente, M4 apresenta erros de previsão significativamente menores que os demais (veja Tab. 5.7). Utilizando-se M4, ‘ganhos’ médios de desempenho de 26,25% com relação a M1 e 22,58% com respeito a M2 foram obtidos. Já os modelos M5 e M6, que utilizam técnicas de reconstrução do espaço-fase na identificação de suas variáveis de entrada, não apresentaram bons resultados. MAPEs de 5,62% e 4,65% para M5 e M6, respectivamente, foram obtidos. Comparando-se M1 com M5 (que são estruturalmente os modelos mais simples) verificam-se MAPEs de 3,58% e 5,62%, respectivamente. Nota-se que a sofisticada metodologia empregada na seleção de variáveis de entrada para M5 foi superada pela função de autocorrelação usada em M1. O fato de M5 não considerar o valor em atraso 24h como variável de entrada pode ter comprometido o seu desempenho. Por outro lado, comparando-se M4 e M6 (que são estruturalmente os modelos mais complexos) verificam-se MAPEs de 2,64% e 4,65%, respectivamente.

É importante ressaltar que um horizonte de previsão maior não implica necessariamente em um erro de previsão maior. Isso depende da variabilidade dos dados para os diferentes períodos do dia e do modelo de previsão adotado [Alves da Silva e Moulin, 2000]. Notar também que o índice MSE, na Tab. 5.8, aponta um número menor de erros elevados para M4 em relação aos demais.

Exemplos de previsões de carga para um dia à frente são apresentados nas Fig. 5.8 a 5.11. Finalmente, a Fig. 5.12 mostra a carga observada para o dia 15 de julho de 1990 decomposta em três níveis de resolução e as suas previsões recursivas correspondentes obtidas com M4.

## 2) Análise da previsão um passo à frente

Com base nas tabelas 5.9 e 5.10, nota-se que M2 e M3 apresentaram os melhores resultados e desempenhos comparáveis. O fato de M4 não ter apresentado um bom desempenho deve-se a própria ‘natureza’ do modelo. Portanto, a aplicação de M4 para previsões do tipo um passo à frente é altamente dependente da qualidade do enchimento.

### **b) Avaliando-se feriados**

Embora a previsão de feriados e dias atípicos não seja o objeto principal desse estudo, alguma informação sobre esse assunto será dada. Apenas a previsão do tipo recursiva será considerada.

Feriados e dias atípicos são uma tarefa muito difícil para qualquer ferramenta de previsão de carga. Isso se deve principalmente àquelas condições de carga anômalas que são raras e bem distintas daquelas observadas nos dias regulares. Por exemplo, existe apenas um dia do trabalhador por ano, e o comportamento da carga associada varia dependendo do dia da semana.

Neste trabalho, 22 dias atípicos foram considerados. Nenhuma estratégia especial de previsão foi adotada. MAPEs médios de 11,04%, 11,01%, 9,97%, 7,77%, 11,03% e 9,18%, para M1, M2, M3, M4, M5 e M6, respectivamente, foram verificados. A título de comparação, o previsor de carga chamado ANNSTLF-Generation III [Khotanzad et al., 1998] emprega um esquema dedicado para dias atípicos e apresenta MAPEs variando de 4,40 a 9,68% para cinco concessionárias de energia elétrica.

Mais informação sobre a previsão de carga para períodos anômalos pode ser encontrada em [Lamedica et al., 1991].

Tab. 5.7. MAPE horário para os 6 modelos (1-24 passos à frente).

M1	M2	M3	M4	M5	M6
2,09	1,00	0,90	1,18	2,45	2,39
4,15	1,84	1,44	1,91	5,64	6,12
4,34	2,20	2,02	2,22	7,04	5,08
3,93	2,61	2,57	2,24	6,70	5,76
3,94	2,94	3,07	2,91	6,59	4,87
4,19	3,36	3,50	3,05	7,35	5,49
5,42	4,07	4,22	3,99	9,44	6,78
4,44	3,96	4,20	3,20	8,72	4,98
3,53	3,33	3,48	2,53	5,79	4,89
2,90	3,10	3,00	2,36	4,24	4,82
2,81	3,38	3,08	2,21	3,81	3,79
2,87	3,52	3,29	2,39	3,71	3,91
2,91	3,66	3,40	2,51	4,07	3,72
3,05	3,79	3,44	2,61	4,75	3,96
3,18	3,95	3,58	2,80	5,04	4,10
3,23	4,12	3,71	2,93	5,13	4,30
3,39	4,12	3,63	2,93	5,09	4,79
3,87	3,98	3,47	2,84	5,51	7,74
3,51	3,77	3,24	2,84	5,27	4,33
3,46	3,83	3,20	2,67	5,10	3,83
3,51	3,78	3,23	2,68	5,63	3,79
3,44	3,77	3,31	2,60	6,58	3,69
3,69	3,76	3,31	2,72	5,64	3,84
4,18	3,93	3,47	2,94	5,54	4,56

Tab. 5.8. Avaliação global média (1-24 passos à frente).

Índice	M1	M2	M3	M4	M5	M6
MAPE	3,58	3,41	3,16	2,64	5,62	4,65
MSE	11754	12011	10522	6599	28364	17950
ME	-12,42	-6,73	-9,48	-6,39	-47,12	-11,85
MAX	34,57	40,22	30,78	29,44	48,83	45,23

Tab. 5.9. MAPE horário para os 6 modelos (1 passo à frente).

M1	M2	M3	M4	M5	M6
2,09	1,00	0,90	1,18	2,45	2,39
2,59	0,97	0,87	1,66	3,67	4,34
1,96	1,02	0,97	1,32	1,92	7,02
2,10	0,97	1,01	1,55	2,87	3,02
2,07	1,06	1,14	2,46	3,19	3,14
3,02	1,53	1,60	2,59	3,15	3,12
3,32	1,81	1,98	3,12	4,87	4,19
2,16	1,15	1,46	2,71	1,89	2,92
2,10	1,21	1,20	1,54	2,36	2,25
1,67	1,31	1,33	2,17	1,62	2,72
1,86	0,99	1,04	1,84	2,13	1,89
1,59	0,87	0,89	2,19	1,50	2,35
1,68	0,83	0,88	1,61	1,61	1,79
1,71	0,82	0,87	2,41	2,17	2,45
1,77	0,85	0,89	1,73	1,77	1,73
1,79	0,90	0,90	2,48	1,82	2,66
2,07	0,99	1,02	1,95	2,35	3,09
2,24	1,06	1,10	2,33	2,38	5,81
1,92	1,17	1,06	1,74	1,92	4,04
2,08	1,13	1,09	2,13	1,92	2,64
2,05	1,20	1,15	1,85	3,01	2,35
2,11	1,16	1,14	2,11	3,04	2,37
2,49	1,40	1,22	1,47	3,35	1,99
1,97	1,03	1,06	1,63	3,26	2,36

Tab. 5.10. Avaliação global média (um passo à frente).

Índice	M1	M2	M3	M4	M5	M6
MAPE	2,10	1,10	1,12	1,99	2,51	3,03
MSE	3708	1166	1240	3974	4879	7218
ME	-4,66	-0,18	-1,6	-3,99	-5,56	-6,34
MAX	20,22	15,92	18,03	25,35	16,13	25,2

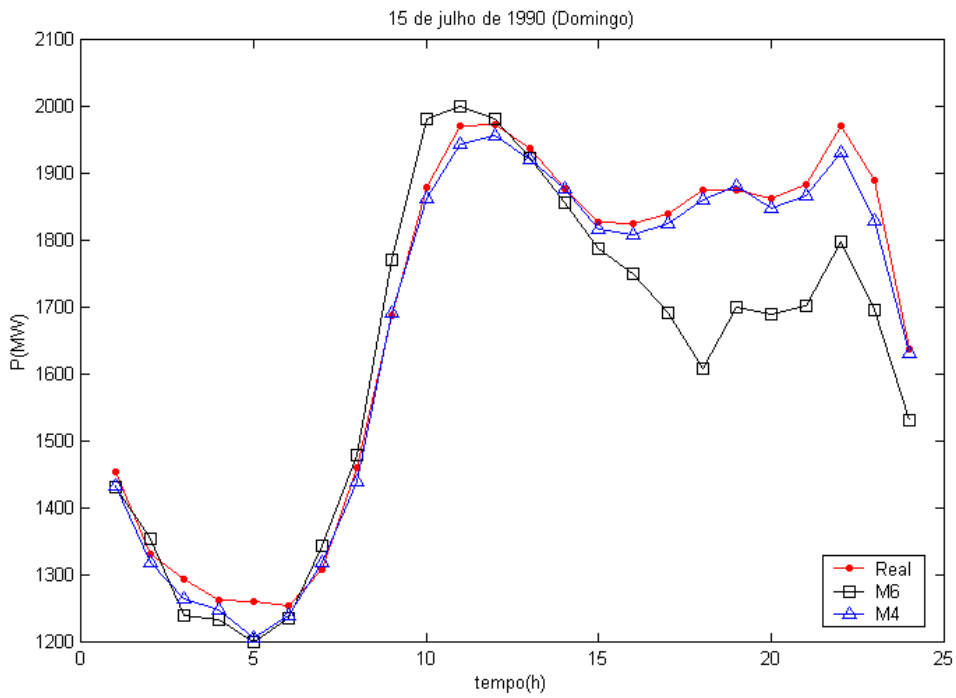


Fig. 5.8. Desempenho dos modelos M4 e M6 para previsões recursivas (domingo).

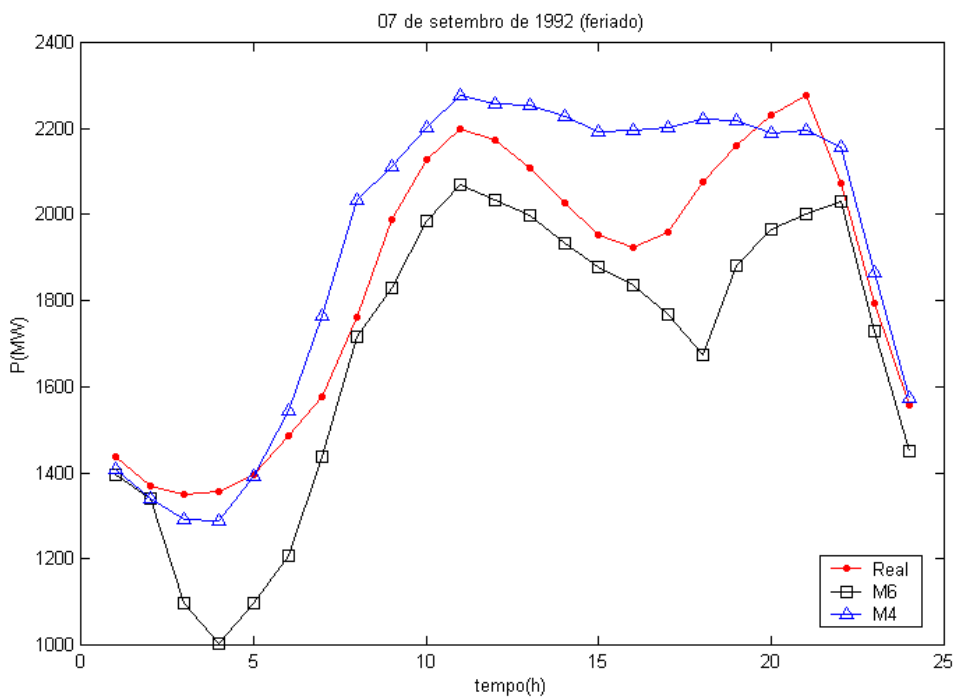


Fig. 5.9. Desempenho dos modelos M4 e M6 para previsões recursivas (feriado).

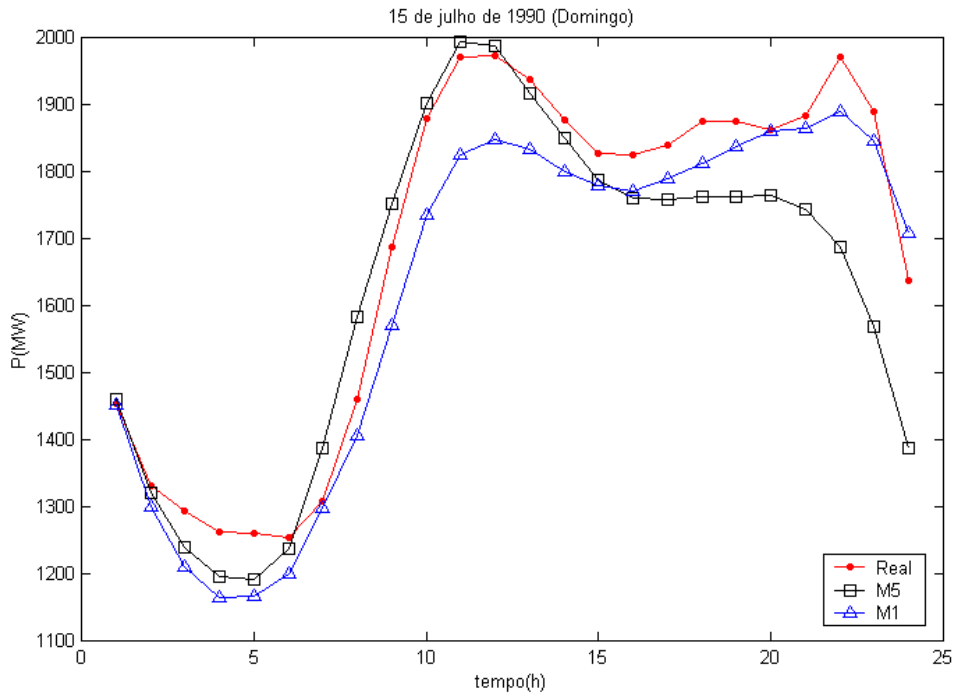


Fig. 5.10. Desempenho dos modelos M1 e M5 para previsões recursivas (domingo).

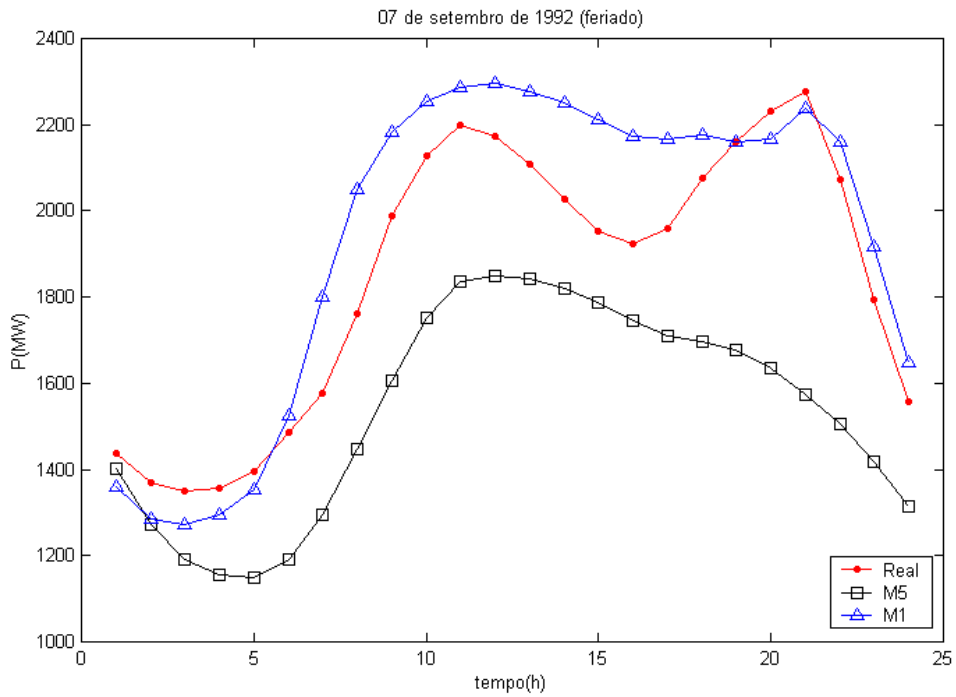


Fig. 5.11. Desempenho dos modelos M1 e M5 para previsões recursivas (feriado).



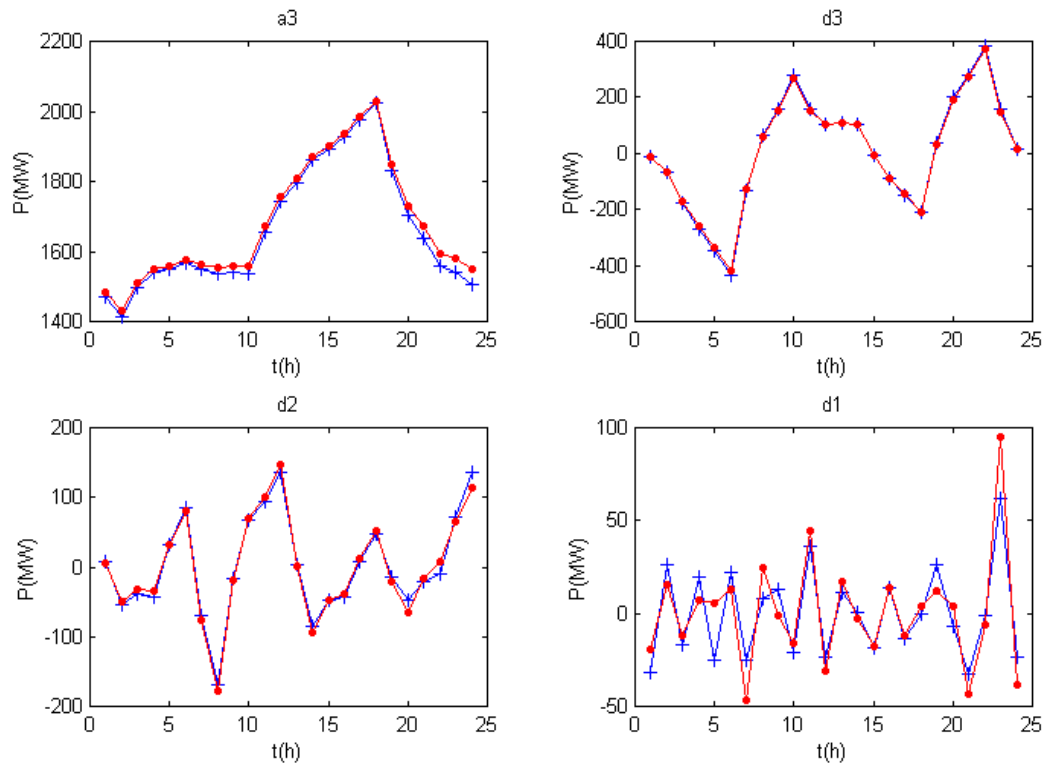


Fig. 5.12. Carga real e as respectivas previsões (+) obtidas com M4.

### 5.13.2 Caso B

Neste caso, o efeito de se adicionar ruído Gaussiano de média zero e desvio padrão  $\sigma=1^\circ\text{F}$  à série de temperatura para simular erros de previsão de temperatura foi investigada.

#### a) Avaliando-se dias de semana típicos

As Tabelas 5.11 e 5.12 resumem os resultados dos testes realizados. Comparando-se a Tab. 5.8 com a Tab. 5.11, nota-se que a exatidão de M1 degradou moderadamente com as temperaturas “previstas” ( $\sim 0,88\%$ ). Por outro lado, M2 apresentou resultados bem melhores que M1. Com relação a M3, pode-se dizer que os erros de previsão da temperatura não o impactaram de maneira muito severa. Um MAPE de 3,16% para o caso sem ruído, aumentou para 3,38% para o caso com ruído. M4 também não foi muito influenciado devido à adição de ruído à série de temperatura. Seu MAPE geral mudou de 2,64% para 2,82%, indicando um grau de degradação menor. A robustez de M4 é provavelmente devida à sua própria natureza. Lembre-se que somente o submodelo A<sub>3</sub> depende da informação de temperatura. Os outros

submodelos (para  $D_1$ ,  $D_2$  e  $D_3$ ) dependem exclusivamente da informação de carga. Um ganho percentual de desempenho de 36,77% com respeito a M1 e 22,52% com respeito a M2 pôde ser verificado. Já os modelos M5 e M6 foram severamente impactados pela presença do ruído. Os MAPES passaram de 5,62% e 4,65% (caso sem ruído) para 6,92% e 5,23%.

A Tab. 5.12 apresenta uma avaliação global média dos seis modelos para previsões de 1 passo à frente. Novamente, como no Caso A, M2 e M3 foram os modelos que apresentarem uma melhor adaptabilidade. As diferenças encontradas nos índices MAPE e MSE para os modelos 2 e 3 não têm significância estatística. Já com relação ao viés<sup>4</sup> presente nas previsões, i.e., aqueles apontados pelos índices ME, não se constata quaisquer tendências acentuadas (para mais ou para menos) nas previsões. Finalmente, com relação ao índice MAPE, verifica-se um ganho percentual de desempenho de M3 com respeito a M1 de 46,98% (i.e.,  $=1-1,14 \div 2,15$ ).

## b) Avaliando-se feriados

MAPEs médios de 11,47%, 10,39%, 9,62%, 7,46%, 10,49% e 9,67% para M1, M2, M3, M4, M5 e M6, respectivamente, são verificados para dias atípicos e feriados. Note, mais uma vez, que a presença de ruído Gaussiano na série de temperatura não impactou severamente M4.

### 5.13.3 Caso C

Uma comparação dos desempenhos de M1 e M4 com outros dois modelos neurais apresentados em [Brace et al., 1991] foi realizada. Apenas o índice MAPE foi calculado. Dados de carga e temperatura da concessionária *Puget Sound Power & Light Company* foram utilizados. O período de teste teve início em 7 de novembro de 1990, terminando em 31 de março de 1992 (excetuando-se os principais feriados observados), de modo que cada modelo foi testado com base em 137 dias.

O previsor de carga baseado em uma rede neural recorrente conhecido como ANN2, faz modelagem não linear com realimentação e produziu um MAPE médio de 5,4%. Já OH2, uma rede neural adaptativa, produziu um MAPE médio de 6,1%. Esses modelos são discutidos em [Brace et al., 1991]. Por outro lado, M1 e M4 produziram MAPEs de 4,04% e 3,67%, respectivamente.

Tab. 5.11. Avaliação global média (recursiva).

Índice	M1	M2	M3	M4	M5	M6
MAPE	4,46	3,64	3,38	2,82	6,92	5,23
MSE	18084	13127	11750	7548	41269	22109
ME	-6,27	-0,69	-6,58	-6,13	-47,16	-17,12
MAX	46,33	33,57	31,17	29,67	59,88	41,55

Tab. 5.12. Avaliação global média (1 passo à frente).

Índice	M1	M2	M3	M4	M5	M6
MAPE	2,15	1,11	1,14	2,04	2,56	3,49
MSE	3841	1207	1291	4116	5074	9483
ME	-4,52	0,13	-1,5	-3,99	-4,97	-7,55
MAX	18,95	15,86	18,04	25,04	15,46	24,81

<sup>4</sup> *Bias*, do inglês.

# Capítulo 6 – Conclusões e trabalhos futuros

## 6.1 Conclusões

Com o crescimento dos sistemas de potência e o aumento na sua complexidade, muitos fatores têm influenciado significativamente a geração e o consumo da energia elétrica. Portanto, o processo de previsão tem se tornado mais complexo, o que requer previsores de carga mais precisos. A previsão de carga a curto prazo é essencial para alimentar os métodos analíticos usados para determinar o preço da energia a curto prazo. A variabilidade e a não estacionariedade das cargas elétricas têm crescido devido à dinâmica das tarifas de energia. Previsores de carga mais robustos e autônomos são necessários no novo ambiente competitivo.

Muitos métodos baseados em estatística clássica têm sido utilizados no desenvolvimento de previsores de carga a curto prazo. Basicamente, esses métodos podem ser divididos em dois grupos:

- modelos univariados de séries temporais;
- modelos causais.

O primeiro grupo é aquele no qual a carga é modelada como uma função dos seus valores observados. Modelos autoregressivos, filtro de Kalman e modelos dinâmicos lineares são alguns exemplos deste grupo. Já nos modelos causais, a carga é modelada como uma função de alguns fatores exógenos, como e.g., variáveis sociais e climáticas. Funções de transferência de Box & Jenkins, regressão não paramétrica e modelos ARMAX são alguns exemplos de modelos causais.

Apesar do grande número de alternativas, os modelos causais mais populares ainda são os regressores lineares e os modelos que decompõem a carga em carga base e parcelas dependentes de variáveis climáticas. Esses modelos são atrativos porque alguma interpretação física do processo pode ser conseguida, permitindo aos operadores de sistemas elétricos entenderem melhor o seu comportamento. No entanto, eles são dispositivos essencialmente lineares, e as séries de carga que eles tentam explicar são funções não lineares de variáveis exógenas.

Por outro lado, muita pesquisa tem sido realizada na aplicação de redes neurais na previsão de carga a curto prazo. Isso se deve ao fato delas terem demonstrado que são capazes

de encontrar representações internas de interdependências dentro de um histórico de dados que não sejam explicitamente dadas ou mesmo conhecidas por um especialista humano [Niebur, 1993]. Muitos trabalhos apresentam bem sucedidas aplicações de RNA na PCCP ([Park et al., 1991], [Lee et al., 1992], [Bakirtzis et al., 1996], [Papalexopoulos et al., 1994]). Várias concessionárias de energia elétrica ao redor do mundo utilizam previsores baseados em redes neurais em caráter de testes ou em operação [Mori, 1996; Papalexopoulos, 1996; Khotanzad et al., 1998].

Nesta tese, novas abordagens para o problema de previsão de carga a curto prazo via redes neurais através da utilização de wavelets e técnicas de reconstrução do espaço-fase foram discutidas. Wavelets podem ser descritas como um sinal de curta duração com energia finita e média zero. Elas são localizadas no tempo (ou espaço), diferentemente das funções trigonométricas. Essa característica torna-as adequadas para a análise de vários sinais não estacionários [Strang & Nguyen, 1996]. Com relação às técnicas de reconstrução do espaço-fase de um sistema dinâmico, no caso da PCCP, está-se interessado em estabelecer um mapeamento entre um dado vetor de carga na hora corrente  $t$  e a carga na hora seguinte ( $t+1$ ). Aqui, seis modelos de previsão foram apresentados e tiveram seus desempenhos comparados. Previsões para um passo à frente e recursiva de 1 a 24 passos à frente foram computadas por perceptrons de múltiplas camadas treinados com uma versão modificada do algoritmo de retropropagação de erros. Dois anos completos de dados de carga e temperatura reais de uma concessionária norte americana foram utilizados para testar as metodologias propostas.

A modelagem de uma série de carga (no domínio do tempo) através do ajuste individual de cada componente no domínio de wavelet, M4, não apresentou a mesma adaptabilidade que M3 (modelo que combina normalização, padronização e Análise de Multiresolução (AMR) via wavelets da família Daubechies) para o horizonte de curtíssimo prazo. Os erros de previsão de M4 para previsões de um passo à frente são mais sensíveis à distorção de borda. Por outro lado, em média, o modelo 4 foi o mais acurado para o horizonte de 1-24 h à frente embora M3 tenha sido bastante competitivo. Esses resultados sugerem que, considerando os dois horizontes apresentados, o modelo 3 tenha apresentado o melhor desempenho global. Apesar de existirem mais entradas para serem manipuladas por M3, a combinação de características da série de carga original, série diferenciada, análise de multiresolução e da série de temperatura fornece um conteúdo de informação mais significativo para o modelo. De maneira geral, os resultados obtidos com M3 e M4 sugerem que a AMR via wavelets pode melhorar a previsão de carga a curto prazo via redes neurais.

Por outro lado, os modelos que utilizaram técnicas de reconstrução do espaço-fase na identificação de suas variáveis de entrada não apresentaram bons resultados (M5 e M6). O fato desses modelos não considerarem, e.g., o valor em atraso 24h como variável de entrada pode ter comprometido os seus desempenhos.

Um conjunto de decisões importantes com relação a estimação dos modelos 3 e 4 (i.e., aqueles baseados em AMR) está relacionado com os seguintes tópicos: seleção da wavelet mãe, definição do número de níveis de decomposição e tipo de enchimento. Valendo-se dos apelos teórico e prático da wavelet Daubechies, decidiu-se investigar sua aplicabilidade no contexto deste trabalho. Wavelets Daubechies de ordens 2, 3 e 4 foram avaliadas. Com relação ao número de níveis de decomposição, com base na dinâmica da carga apenas, três estratégias de análise via wavelets com 3, 4 e 5 níveis foram consideradas. A representação com 3 níveis de decomposição se mostrou como a escolha mais promissora porque ela descreve a série de carga de maneira mais significativa que as demais. Essa conclusão tem a ver principalmente com o conteúdo de baixa frequência (aproximação) que é a componente mais importante do sinal. Nota-se que o esquema de 3 níveis de decomposição enfatiza o comportamento regular da série de carga e revela padrões ocultos que não podem ser vistos com os demais esquemas de decomposição. Finalizando, com respeito ao enchimento, quatro esquemas foram comparados. Anexar dados previamente medidos no começo da série de carga e valores previstos no final da mesma foi a melhor maneira de minimizar a distorção que aparece nas extremidades da série devida a convolução em sinais de comprimento finito. O melhor predictor de carga que não fez uso da transformada wavelet (M2) foi usado para enchimento. Na verdade, ele é também bastante competitivo já que o mesmo é uma versão simplificada de M3. O modelo 2 combina características da série de carga original com componentes de alta frequência da própria série obtidos via diferenciação de pontos adjacentes. Porém, os seus desvios médios quadráticos (i.e., o índice MSE) para previsões de um dia à frente indicam uma incidência maior de erros de previsão elevados.

É verdade que o esforço computacional aumenta com a aplicação da AMR (desde que M2 também se torna necessário). Contudo, considerando-se os ganhos percentuais de desempenho, a PCCP via RNA e wavelets é bastante atrativa.

## 6.2 Trabalhos futuros

A evolução natural deste trabalho seria partir para a implementação de um previsor operando em modo on-line. Seria desejável também que os testes fossem executados para outras concessionárias de energia elétrica situadas em pontos geograficamente diferentes.

Em [Corrêa et al., 1999], os autores apresentam várias abordagens que podem ser utilizadas juntamente com a minimização do erro de previsão para fins de validação de modelos neurais. Tais abordagens também poderão ser consideradas.

Com relação a definição da estrutura de um PMC, a principal limitação do tipo de abordagem empregada neste trabalho (conhecida como “Um-elemento-de-cada-vez”) se deve ao fato da mesma não permitir a interação entre fatores (e.g., número de neurônios, número de épocas de treinamento, semente). Na prática, a interação entre fatores é muito comum, e se elas ocorrerem, a estratégia Um-elemento-de-cada-vez poderá não produzir bons resultados. Portanto, com o intuito de contornar esse problema, averiguar-se-á a aplicabilidade de técnicas de Projeto e Análise de Experimentos (DOE) [Montgomery, 2000] na determinação da melhor estrutura para um previsor neural.

Outras famílias de wavelets também deverão ser testadas e intervalos de confiança para as previsões deverão ser estimados [Alves da Silva e Moulin, 2000]. Finalmente, mais investigação será necessária para averiguar melhor a aplicabilidade das técnicas de reconstrução do espaço-fase na seleção de variáveis de entrada para a previsão de carga.

## 6.3 Contribuições

A principal contribuição deste trabalho está relacionada com melhorias na capacidade de generalização de uma rede neural através da extração das características mais relevantes de uma série de carga via análise de multiresolução. A metodologia prática desenvolvida aqui poderá ser aplicada a vários outros problemas envolvendo a previsão de séries temporais via redes neurais artificiais. Também, a inclusão da variável temperatura na PCCP via RNA e wavelets deve ser destacada, já que a grande maioria dos artigos do gênero não consideram essa informação. Finalmente, o esquema prático de enchimento proposto para lidar com o problema de distorção de bordas deve ser mencionado. Tal assunto tem sido sistematicamente ignorado na literatura especializada.

## Referências

---

[Abraham & Ledolter, 1983] B. Abraham & J. Ledolter. *Statistical Methods for Forecasting*. John Wiley & Sons, 1983.

[Aguirre, 1994] Luis Antonio Aguirre, “Some remarks on structure selection for nonlinear models”, *International Journal of Bifurcation and Chaos*, Vol. 4, N° .6, pp. 1707-1714, 1994.

[Alves da Silva & Moulin, 2000] A.P. Alves da Silva & L.S. Moulin, “Confidence intervals for neural network based short-term load forecasting”, *IEEE Trans. Power Systems*, Vol. 15, No. 4, pp. 1191-1196, Nov. 2000.

[Amari et al., 1996] S. Amari, N. Murata, K.-R. Müller, M. Finke, H. Yang, “Statistical Theory of Overtraining – Is cross-validation asymptotically effective?”. *Advances in Neural Information Processing Systems*, Vol. 8, pp.176-182, Cambridge, MA:MIT Press, 1996.

[Barakat et al., 1990] E. H. Barakat, M.A. Qayyum, M.N. Hamed, S.A. Al Rashed, “Short Term Peak Demand Forecasting in fast Developing Utility with Inherent Dynamic Load Characteristics – Part I and II”. *IEEE Transactions on Power Systems*, Vol.5, pp. 813-824, 1990.

[Bakirtzis et al., 1996] A.G. Bakirtzis, V. Petridis, S.J. Klartzis, M.C. Alexiadis, and A.H. Maissis, “A neural network short-term load forecasting model for the Greek power system”. *IEEE Trans. Power Systems*, vol. 11, no. 2, Maio 1996, pp. 858-863.

[Benítez et al., 1997] J. M. Benítez, J.L. Castro, I. Requena, “Are Artificial Neural Networks Black Boxes?”. *IEEE Transactions on Neural Networks*, Vol.8, n° .5, pp.1156-1164, Sep.1997.

[Box & Jenkins, 1970] G.E.P. BOX & G. M. JENKINS. *Time Series Analysis: Forecasting and Control*. Holden Day, 1970.

[Brace et al., 1991] M.C. Brace, J. Schmidt, M. Hadlin, “Comparison of the forecasting accuracy of neural networks with other established techniques”, in *Proc. of the First*



International Forum on Applications of Neural Networks to Power Systems, pp. 31-35, Julho 1991.

[Brockwell & Davis, 1996] P. J. BROCKWELL & R.A. DAVIS. *Introduction to Time Series and Forecasting*. Springer-Verlag, NY, 1996.

[Bunn, 2000] D.W. Bunn, 'Forecasting loads and prices in competitive power markets', *Proceedings of the IEEE*, Vol. 88, No. 2, pp. 163-169, Fev. 2000.

[Buzug & Pfister, 1992] Th. Buzug & G. Pfister, "Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global static and local dynamical behaviour of strange attractors". *Physical Review A*, vol. 45, n° 10, pp.7073-7084, Maio 1992.

[Chen et al., 1992] S.-T. Chen, D.C. Yu, and A.R. Moghaddamjo, "Weather Sensitive Short-Term Load Forecasting Using Nonfully Connected Artificial Neural Network", *IEEE Trans. on Power Systems*, Vol. 7, No. 3, Ago. 1992, pp. 1098-1105.

[Coifman & Wickerhauser, 1992] R.R. Coifman & M.V. Wickerhauser, 'Entropy-based algorithms for best basis selection', *IEEE Trans. on Inf. Theory*, Vol. 38, No. 2, pp.713-718, Março 1992.

[Corrêa et al., 1999] M. V. Corrêa, L. A. Aguirre, A. P. Braga, 'Validação de Modelos Neurais Identificados a partir de um Sistema Caótico', *Anais do IV Congresso Brasileiro de Redes Neurais*, São José dos Campos, SP, julho 1999, pp. 152-157.

[Daubechies, 1992] I. Daubechies. *Ten lectures on wavelets*. CBMS-NSF Conference Series in Applied Mathematics. SIAM Ed., 1992.

[Drezga, 1996] Irislav Drezga, *A generalized ANN-based model for short-term load forecasting*. Tese de doutorado, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, EUA, Dez. 1996.

[Drezga & Rahman, 1998] I. Drezga & S. Rahman, 'Input Variable Selection for ANN-Based Short-term Load Forecasting'. *IEEE Trans. on Power Systems*, vol.13, n° .4, pp.1238-1244, Nov. 1998.

[Fraser & Swinney, 1986] A. M. Fraser & H. L. Swinney, ‘Independent Coordinates for Strange Attractors from Mutual Information’. *Phys. Review A.*, vol. 33, 1986, pp. 1134.

[Goh et al., 1986] T.N. Goh et al. ‘A New Approach to Statistical Forecasting of Daily Peak Power Demand’. *Electrical Power Systems Research*, Vol.10, n° 2, pp.145-148, Mar. 1986.

[Hausdorff, 1918] Hausdorff, F. Dimension und äußeres Maß. *Math. Annalen*, 79, 157.

[Haykin, 1999] S. Haykin. *Neural networks – A Comprehensive Foundation*. Prentice Hall, NJ, 2nd ed., 1999.

[Heinemann et al., 1966] G.T. Heinemann et al. ‘The Relationship between Summer Weather and Summer Loads – A Regression Analysis’. *IEEE Transaction on Power App. Systems*, Vol. PAS-85, pp. 1144-1154, Nov. 1966.

[Hippert et al., 2001] H. S. Hippert, Carlos E. Pedreira, Reinaldo C. Souza, ‘Neural Networks for Short-term Load Forecasting: A Review and Evaluation’, *IEEE Transactions on Power Systems*, Vol. 16, N° 1, Feb. 2001, pp.44-55.

[Hornik et al., 1989] K. Hornik, M. Stinchcombe, and H. White, ‘Multilayer feedforward networks are universal approximators’, *Neural Networks*, Vol. 2, pp. 359-366, 1989.

[Huang & Yang, 2001] C.-M. Huang & H.-T. Yang, ‘Evolving wavelet-based networks for short-term load forecasting’, *IEE Proc. -Gener. Transm. Distrib.*, Vol. 148, No. 3, pp. 222-228, Maio 2001.

[IEEE Committee report, 1981] IEEE Committee report , ‘Load Forecast Bibliography Phase II’. *IEEE Transactions on Power App. Systems*, Vol.PAS -100, n°7, pp.3217-3220, Jul. 1981.

[Kantz & Schreiber, 1999] H. Kantz & T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, UK, 1999.

[Kennel et al., 1992] M. B. Kennel, R. Brown, H. D. I. Abarbanel, ‘Determining Embedding Dimension for Phase-Space Reconstruction using a Geometrical Construction’. *Physical Review A*, vol. 45, 1992, pp. 3403.

[Kim & Aggarwal, 2001] C. W. Kim & R. Aggarwal, "Wavelets Transform in Power Systems - Part 2 - Examples of Application to Power Systems Transients". Power Engineering Journal, Agosto 2001, pp.193-202.

[Kim et al., 2002] C.-i. Kim, I.-k. Yu, Y. H. Song, "Kohonen neural network and wavelet transform based approach to short-term load forecasting", Electric Power Systems Research, vol. 63, pp. 169-176, 2002.

[Khotanzad et al., 1998] A. Khotanzad, R. Afkhami-Rohani, D. Maratukulam, "ANNSTLF – artificial neural network short-term load forecaster – generation three". IEEE Trans. Power Systems, vol. 13, no. 4, Nov. 1998, pp. 1413-1422.

[Lamedica et al., 1991] R. Lamedica, A. Prudenzi, M. Sforna, M. Caciotta, V. O. Cencelli, "A neural network based technique for short-term forecasting of anomalous load periods", IEEE Trans. Power Systems, Vol. 11, No. 4, pp. 1749-1756, Maio 1991.

[Lee et al., 1992] K.Y. Lee, Y.T. Cha, and J.H. Park, "Short-term load forecasting using an artificial neural network", IEEE Trans. Power Systems, Vol. 7, No. 1, pp. 124-132, Fev. 1992.

[Liu et al., 1996] K. Liu, S. Subbarayan, R.R. Shoults, M.T. Manry, C. Kwan, F.L. Lewis, and J. Naccarino, "Comparison of very short-term load forecasting techniques". IEEE Trans. Power Systems, vol. 11, no. 2, Maio 1996, pp. 877-882.

[Lu et al., 1993] C.N. Lu, H.T. Wu, and S. Vemuri, "Neural Network Based Short Term Load Forecasting", IEEE Transactions on Power Systems, Vol. 8, No. 1, Fev. 1993, pp. 336-342.

[Makridakis et al., 1998] S. Makridakis, S. Wheelwright, R. J. Hyndman. *Forecasting Methods and Applications*. 3. ed. New York: John Wiley & Sons, 1998.

[Mallat, 1989] S. Mallat, "A Theory for Multiresolution Signal Decomposition - The Wavelet Representation". IEEE on Pattern Analysis and Machine Intelligence, vol. 11, N° 7, Julho 1989, pp.674-693.

[Marín et al., 2002] F.J. Marín, F.Garcia-Lagos, G. Joya, F. Sandoval, “Global model for short-term load forecasting using artificial neural networks”, IEE Proc.-Gener. Transm. Distrib. Vol. 149, nº 2, Março 2002.

[Masters, 1995] T. Masters. *Neural, Novel & Hybrid Algorithms for Time Series Prediction*. New York: Wiley, 1995, p. 514.

[Misiti et al., 1993] M. Misiti, Y. Misiti, G. Oppenheim, J.-M. Poggi, “Ondelettes en Statistique et Traitement du Signal”. *Revue de Statistique Appliquée*, vol. XLI, nº. 4, 1993, pp.33-43.

[Misiti et al., 1994] M. Misiti, Y. Misiti, G. Oppenheim, J.-M. Poggi, “Décomposition Par Ondelettes et Méthodes comparatives: Étude d’une Courbe de Charge Électrique”. *Revue Statistique Appliquée*, vol. XLII, nº 2, 1994, pp. 57-77.

[Misiti et al., 1996] M. Misiti, Y. Misiti, G. Oppenheim, J.-M. Poggi. *Wavelet Toolbox Manual – User’s guide*. The Math Works Inc., USA, 1996, p. 626.

[Mohammed et al., 1995] O. Mohammed, D. Park, R. Merchant, T. Dinh, C. Tong, A. Azeem, J. Farah, and C. Drake, “Practical experiences with an adaptive neural network short-term load forecasting system”. *IEEE Trans. Power S systems*, vol. 10, no. 1, Fev. 1995, pp. 254-265.

[Montgomery, 2000] D. C. Montgomery. *Design and Analysis of Experiments*. Willey, 5<sup>a</sup> ed., 2000.

[Morettin, 1997] Morettin, P.A. *Ondaletas e seus usos na Estatística*. EDUSP, São Paulo, 1997, pp.1-86.

[Morettin & Toloí, 1987] P.A. Morettin & C.M.C. Toloí. *Previsão de Séries Temporais*. Atual Editora Ltda., São Paulo, 2<sup>a</sup> edição, 1987.

[Mori, 1996] H. MORI, “State-of-the-art overview on artificial neural networks in power systems”. *A Tutorial Course on Artificial Neural Networks with Applications to Power Systems*, IEEE Catalog no. 96TP112-0, Edited by M.A. El-Sharkawi and D. Niebur, 1996, pp. 51-70.

[Niebur, 1993] D. Niebur, "Artificial Neural Networks for Power Systems: A Literature Survey". International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications, Vol.1, nº .3, pp.133-158, Dez. 1993.

[Ning e Yunping, 1998] M. Ning & C. Yunping, "An ANN and Wavelet transformation based method for short term load forecast", Anais da International Conference on Energy Management and Power Delivery, Singapore, Vol. 2, pp. 405-410, Março de 1998.

[Oonsivilai e El-Hawary, 1999] A. Oonsivilai & M. El-Hawary, "Wavelet neural network based short term load forecasting of electric power system commercial load". Anais do IEEE Canadian Conference on Electrical and Computer Engineering, Edmonton, Vol. 3, pp. 1223-1228, Maio de 1999.

[Packard et al., 1980] N. H. Packard, J. P. Crutchfield, J. D. Farmer, R. S. Shaw, "Geometry from a time series", Physical Review Letters, vol. 45, N.º 9, Set. 1980, pp. 712-716.

[Papalexopoulos et al., 1994] A. Papalexopoulos, S. Hao, T. M. Peng, "An implementation of a Neural Network based Load Forecasting Model for the EMS", IEEE Transactions on Power Systems, Vol. 9, N.º.4, Nov. 1994, pp.1956-1962.

[Papalexopoulos, 1996] A. Papalexopoulos, "System Load Forecasting: the US perspective, A Tutorial Course on Artificial Neural Networks with Applications to Power Systems". IEEE Power Engineering Society, edited by M. El-Sharkawi and D. Niebur, Ch. 7, 1996, pp. 71-89.

[Park et al., 1991] D.C. Park, M.A. El-Sharkawi, R.J. Marks II, L.E. Atlas, and M.J. Damborg, "Electric load forecasting using an artificial neural network". IEEE Trans. Power Systems, vol. 6, no. 2, Maio 1991, pp. 442-449.

[Piras et al., 1996] A. Piras, A. Germond, B. Buchenel, K. Imhof and Y. Jaccard, "Heterogeneous Artificial Neural Network for Short Term Electrical Load Forecasting", IEEE Transactions on Power Systems, Vol. 11, No. 1, Fev. 1996, pp. 397-402.

[Reis, 1999] A. J. Rocha Reis. *Pré-processamento de Séries Temporais para um Previsor de Carga Elétrica a Curto Prazo via Redes Neurais Artificiais*. Dissertação de mestrado. Escola Federal de Engenharia de Itajubá, Itajubá, MG, Setembro de 1999.

[Reis e Alves da Silva, 2000] A. J. Rocha Reis & A.P. Alves da Silva, ‘Pré-processamento de séries temporais para previsão de carga a curto prazo via redes neurais artificiais.’ Anais do XIII Congresso Brasileiro de Automática – CBA 2000, Florianópolis, Set. de 2000, pp. 1138-1143.

[Rich & Knight, 1994] E. Rich & K. Knight. *Inteligência Artificial*. McGraw-Hill, São Paulo, 2ª edição, 1994.

[Rioul & Vetterli, 1991] O. Rioul & M. Vetterli, ‘Wavelets and Signal Processing’. IEEE Signal Processing Magazine, Out. 1991, pp.14-38.

[Russel & Norvig, 1995] S. J. Russel & P. Norvig. *Artificial Intelligence – A Modern Approach*. Prentice Hall, NJ, 1995.

[Sauer et al., 1991] T. Sauer, J. Yorke, M. Casdagli, ‘Embedology’, Journal of Statistical Physics, vol. 65, N.ºs 3 e 4, 1991, pp. 579-616.

[Schweppe, 1973] F.C. Schweppe. *Uncertain Dynamic Systems*. Englewood Cliffs, NJ, 1973.

[Strang, 1986] G. Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Wellesley, 1986, pp.263-290 / pp.382-385.

[Strang & Nguyen, 1996] G. Strang & T. Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, 1996.

[Takens, 1981] F. Takens, ‘Detecting Strange Attractors in Turbulence’. Em D.A. Rand e L. - S. Young (Eds.), *Dynamical Systems and Turbulence*, Warwick 1980, Springer Verlag, 1981, pp. 366-381.

[Taylor & Buizza, 2002] James W. Taylor & Roberto Buizza, ‘Neural Network load forecasting with weather ensemble predictions’, IEEE Trans. Power Systems, vol. 17, no. 3, Agosto 2002, pp. 626-632.

[Werbos, 1994] P.J. Werbos. *The Roots of Backpropagation – From Ordered Derivatives to Neural Networks and Political Forecasting*. John Wiley & Sons, NY, 1994.

[Wildberger, 1994] M. Wildberger, “Wavelets: A Tool for Enhanced Information Processing”. EPRI Technical Brief, RP8018, set.1994.

[Zhang & Dong, 2001] B.-L. Zhang, Z.-Y. Dong, “An adaptive neural -wavelet model for short term load forecasting”, *Electric Power Systems Research*, vol. 59, pp. 121 -129, 2001.

# Anexo I – Redes Neurais Artificiais

## I.1 Introdução

Redes Neurais Artificiais são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através de experiência. É uma estrutura de processamento não algorítmica na forma de um grafo orientado, onde cada nó, denominado **neurônio**, possui muitas entradas e um único sinal de saída. O processamento em cada nó depende exclusivamente das entradas deste nó e dos valores acumulados em uma pequena memória local do mesmo. A seguir, uma breve descrição do funcionamento de um neurônio biológico.

## I.2 O Neurônio Biológico

Sabe-se que o **neurônio**, ou célula nervosa, é a unidade funcional fundamental de todo o tecido nervoso, incluindo o cérebro. Eles têm um papel essencial na determinação do funcionamento do corpo humano e do raciocínio.

Os neurônios são formados pelo corpo celular ou **soma**, onde se encontra o núcleo da célula, pelos **dendritos**, que representam um conjunto de ramificações de entrada, e pelos **axônios**, que são longos terminais de saída. Os dendritos se concentram nas proximidades do soma. Já o axônio se estende por uma distância de cerca de cem vezes o diâmetro da célula. Na Figura I.1 está representado, de maneira simplificada, um neurônio biológico típico dos seres humanos.



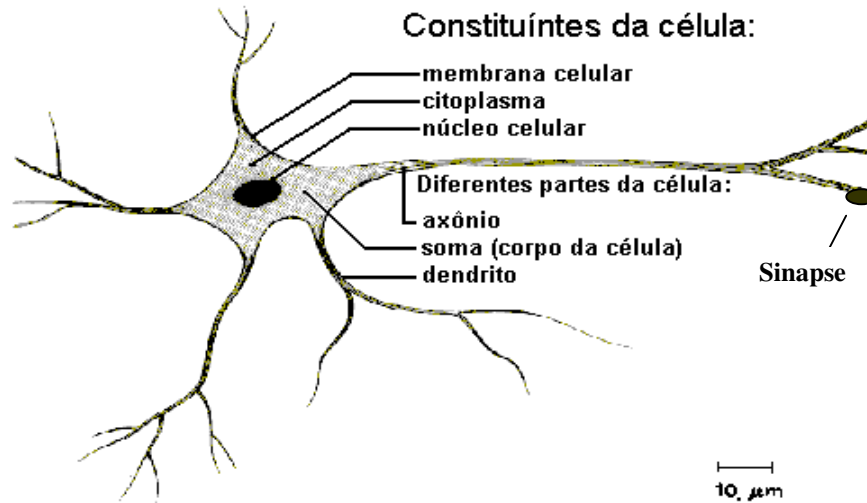


Fig. I.1. Partes constituintes de um neurônio biológico.

Eventualmente, o axônio também se ramifica e se conecta aos dendritos de outros neurônios. O ponto onde ocorre a junção entre um dado axônio **A** e um dendrito **B** é conhecido como **sinapse**. Em média, cada neurônio forma entre mil e dez mil sinapses. O cérebro humano possui cerca de  $10^{11}$  neurônios, e o número de sinapses é superior a  $10^{14}$ , possibilitando a formação de redes muito complexas. Uma das propriedades mais importantes das sinapses é a **plasticidade**, i.e., mudanças a longo prazo na força das conexões em resposta a padrões de estímulo.

### I.2.1 O Processo de Comunicação entre Neurônios

Sinais são propagados de neurônio a neurônio por uma complicada reação eletroquímica. Substâncias transmissoras químicas são liberadas das sinapses e entram nos dendritos, aumentando ou diminuindo o potencial elétrico do soma. Quando esse potencial atinge um limite, um pulso elétrico é enviado via axônio, alcançando outras sinapses e liberando transmissores no soma de outros neurônios. As sinapses que aumentam o potencial da célula são chamadas **excitatórias**. Em contra partida, aquelas que diminuem o potencial são chamadas **inibitórias**. Este processo depende de vários fatores, como a geometria da sinapse e o tipo de neurotransmissor.

### I.3 Representação Matemática de um Neurônio

Uma Rede Neural Artificial (RNA) é composta de um número de **neurônios** ( $a_j$ ) ou unidades, interligados por **conexões** (“*Links*”) distribuídas em **camadas**. Cada conexão está associada a um **peso** ( $W_{j,i}$ ) numérico. A Figura I.2 apresenta um neurônio artificial típico.

Cada unidade realiza a seguinte computação: as unidades fazem operações apenas sobre seus dados locais, i.e., as entradas recebidas multiplicadas pelos pesos de suas respectivas conexões. A seguir, um novo nível de ativação é calculado (via uma dada função de ativação  $g$ ) e enviado para as conexões de saída.

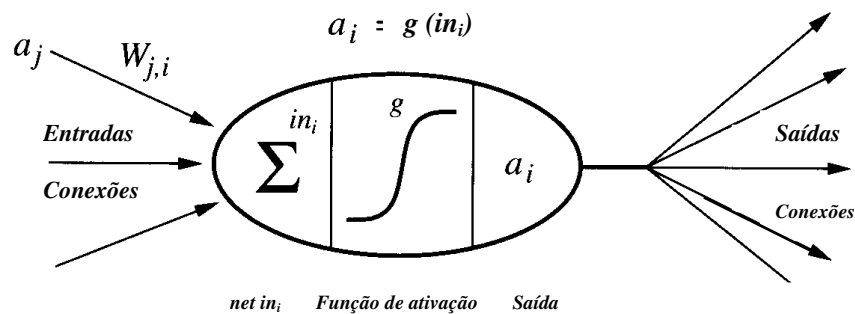


Fig. I.2. Representação Matemática de uma Unidade [Russel & Norvig, 1995].

Matematicamente falando,

$$in_i = \sum_j W_{j,i} \cdot a_j = W_i \bullet a_i$$

$$a_i \leftarrow g(in_i) = g\left(\sum_j W_{j,i} \cdot a_j\right)$$

### I.3.1 Funções de Ativação

Diferentes modelos podem ser obtidos através do uso de diferentes funções matemáticas para  $g$ . As três funções mais usadas são as do tipo Passo, Degrau e Sigmoidal; a saber:

$$Passo_t(x) = \begin{cases} 1, se & x \geq t \\ 0, se & x < t \end{cases}$$

$$Degrau(x) = \begin{cases} +1, se & x \geq 0 \\ -1, se & x < 0 \end{cases}$$

$$Sigmoidal(x) = \frac{1}{1 + e^{-x}}$$

A Figura I.3 ilustra essas funções.

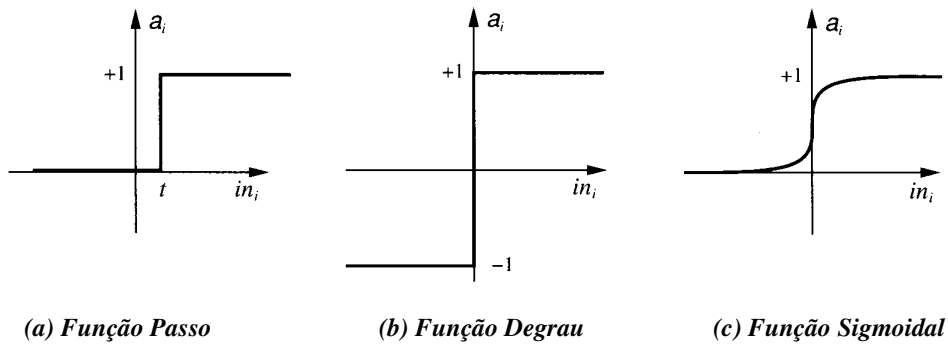


Fig. I.3. Três diferentes funções de ativação.

## I.4 Tipos de Redes Neurais Artificiais

Existe uma grande variedade de estruturas de redes, e.g., perceptrons de múltiplas camadas (PMC), redes de Hopfield, Máquinas de Boltzman, Mapas de características de Kohonen, entre outras. Cada uma delas apresenta suas características próprias e devem ser escolhidas de acordo com o problema a ser solucionado. No entanto, a principal distinção entre elas se deve ao fato de as mesmas serem do tipo **unidirecionais** (i.e., grafos acíclicos direcionados - DAG) ou do tipo **recorrentes** (i.e., as conexões podem assumir topologias arbitrárias). A título de exemplo, a Figura I.4 ilustra uma estrutura unidirecional do tipo Perceptron. Já a Figura I.5 mostra uma estrutura recorrente conhecida como Rede de Hopfield.

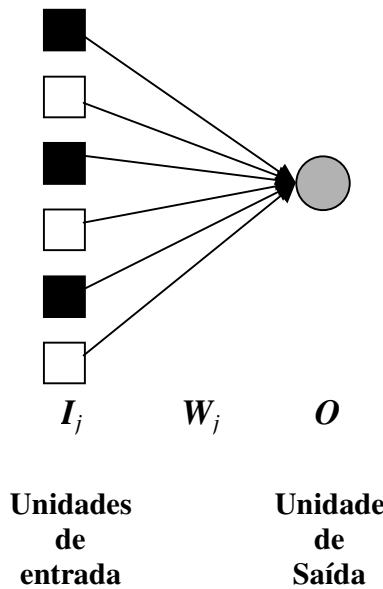


Fig. I.4 Estrutura de Perceptron simples.

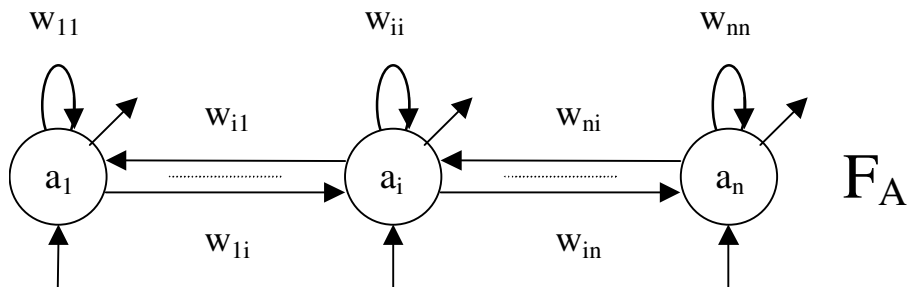


Fig. I.5. Rede de Hopfield e Máquinas de Boltzman.

Para maiores informações sobre as diversas estruturas de RNA, ver [Haykin, 1999]. A seção seguinte tem como objetivo detalhar a estrutura de uma PMC e explicar como se dá o seu processo de aprendizagem.

## I.5 Perceptrons de Múltiplas Camadas

### I.5.1 Estrutura

Arquiteturas neurais do tipo PMC são tipicamente organizadas em camadas. Cada camada tem uma matriz de pesos  $W$ , um vetor de *bias*  $b$  e um vetor de saída  $O$ . A Figura I.6 mostra uma PMC típica.

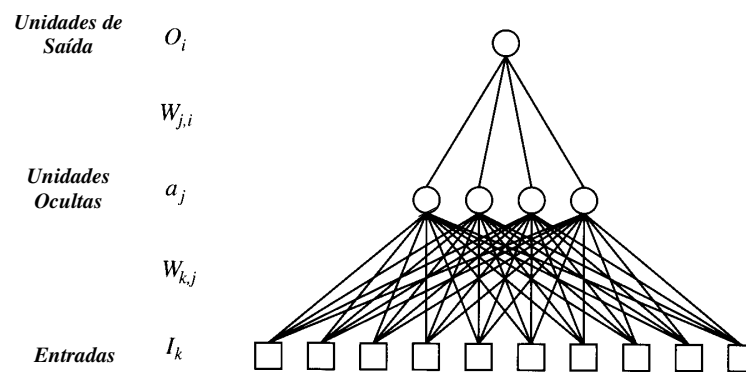


Fig. I.6. Perceptron de Múltiplas Camadas [Russel & Norvig, 1995].

As camadas de uma PMC desempenham funções diferentes. A camada que produz a saída da rede é chamada de *camada de saída*. Todas as demais camadas são denominadas de *camadas ocultas* (ou intermediárias) cuja função é a de extração de características complexas dos dados ou de representações internas. A rede de duas camadas mostrada acima tem uma camada de saída e uma oculta. Alguns autores referem-se às entradas como uma terceira camada (i.e., como uma camada de entrada). No entanto, para se evitar ambigüidade, essa convenção não será adotada aqui. A seguir, falar-se-á do processo de aprendizagem de uma PMC.

## I.5.2 Processo de Aprendizagem de uma PMC

Todo o “conhecimento” adquirido pela RNA está armazenado nos pesos relativos às conexões entre os seus neurônios [Rich & Knight, 1994]. Conseqüentemente, o processo de aprendizagem consiste em apresentar um conjunto de dados representativos do problema a ser tratado pela rede e ajustar-se os pesos de suas conexões de maneira iterativa (via algoritmo de treinamento), objetivando-se capturar o conhecimento contido nos dados. Em última análise, o que se busca é desenvolver um processo de treinamento que maximize a capacidade de generalização *da RNA* para a dada tarefa.

## I.5.3 Procedimentos de Aprendizado

Dependendo da maneira pela qual uma rede neural se relaciona com o ambiente externo, três tipos principais de treinamento podem ser verificados:

- **Treinamento Supervisionado** (entrada/saída) - Quando se utiliza um agente externo que indica à rede a resposta desejada para o padrão de entrada (i.e., informa-se para a rede qual é o conjunto de entrada e qual a saída que está associada a ele).
- **Treinamento Não Supervisionado** (entrada) - Quando não existe uma agente externo indicando a resposta desejada para os padrões de entrada. Dado um conjunto de dados de entrada, a rede tem permissão de avaliá-los e tentar descobrir regularidades e relacionamentos entre partes diferentes da entrada.
- **Treinamento por Reforço** (entrada/certo ou errado) - Quando um crítico externo avalia a resposta fornecida pela rede, recompensando-a caso sua resposta seja satisfatória e punindo-a caso contrário. Portanto, a rede sempre tenta encontrar um conjunto de pesos que impeça avaliações negativas.

## I.5.4 Algoritmo de Aprendizado

Denomina-se algoritmo de aprendizado a um conjunto de regras bem definidas para a solução de um problema de aprendizagem. Existem muitos tipos de algoritmos específicos para determinados modelos de redes neurais. Estes algoritmos diferem entre si, principalmente pelo modo como os pesos são modificados. Um algoritmo de treinamento que tem sido largamente utilizado é conhecido como “*Backpropagation*” ou algoritmo de retropropagação de erros [Werbos, 1994]. Esse algoritmo é uma generalização da regra delta e será apresentado a seguir.

### Regra delta padrão

O procedimento de aprendizado implica no uso de um conjunto de pares de padrões de entrada e saída. Para cada padrão de entrada, uma saída é calculada pela rede e esta é comparada com o padrão de saída correspondente. Não havendo diferença entre as saídas, conseqüentemente não haverá mudança nos pesos. De outro modo, os pesos deverão sofrer mudanças a fim de minimizar a diferença.

A regra delta padrão é utilizada para perceptrons que não possuem camadas internas. A equação que gera a mudança dos pesos para cada padrão  $p$  de entrada-saída, para funções de ativação lineares, é dada por

$$\Delta_p W_{ji} \propto -\frac{\partial E_p}{\partial W_{ji}} = \eta(t_{pj} - o_{pj})i_{pi} = \eta\delta_{pj}i_{pi} \quad \text{Eq.I.1}$$

onde  $j$  e  $i$  representam os índices dos neurônios ligados através do peso  $W_{ji}$ . Já o desvio médio quadrático é dado por:

$$E_p = \frac{1}{2} \sum_j (t_{pj} - o_{pj})^2 \quad \text{Eq.I.2}$$

em que,

$t_{pj}$  é a saída desejada para o neurônio de saída  $j$ ;

$o_{pj}$  é a saída calculada, do mesmo neurônio, correspondente ao padrão de entrada;

$i_{pi}$  é o valor da padrão de entrada;

$\eta$  é a taxa de aprendizagem.

$\Delta_p W_{ji}$  é a variação do valor do peso  $W_{ji}$ , correspondente ao padrão  $p$  do conjunto de treinamento.

A camada de entrada atua apenas como repetidora dos sinais de entrada, distribuindo-os às entradas dos neurônios da camada de saída.

### Regra delta generalizada

Para generalizar a regra delta padrão, primeiro define-se o erro correspondente a um padrão  $p$  de entrada-saída como sendo o erro quadrático parcial dado pela Eq.I.2. O erro quadrático global é dado por  $E = \sum E_p$  (que é calculado após a apresentação de todos os padrões utilizados no treinamento) e será minimizado pelo algoritmo.

Tomando-se como base a camada  $i$ , anterior à camada  $j$ , define-se:

$$S_{pj} = \sum_i W_{ji} o_{pi} \quad \text{Eq.I.3}$$

A saída calculada  $o_{pj}$  usa a função sigmoideal e é dada por:

$$o_{pj} = f_j(S_{pj}) \quad \text{Eq.I.4}$$

Para se obter a generalização correta da regra delta,

$$\Delta_p W_{ji} \propto - \frac{\partial E_p}{\partial W_{ji}} \quad \text{Eq.I.5}$$



A derivada da Eq.I.5 é calculada da seguinte maneira,

$$\frac{\partial E_p}{\partial W_{ji}} = \frac{\partial E_p}{\partial S_{pj}} \frac{\partial S_{pj}}{\partial W_{ji}} \quad \text{Eq.I.6}$$

Da Eq.I.3 tem-se que:

$$\frac{\partial S_{pj}}{\partial W_{ji}} = \frac{\partial}{\partial W_{ji}} \sum_k W_{jk} o_{pk} = o_{pi} \quad \text{Eq.I.7}$$

Definindo-se  $\delta_{pj}$  como sendo:

$$\delta_{pj} = -\frac{\partial E_p}{\partial S_{pj}} \quad \text{Eq.I.8}$$

e substituindo-se I.7 e I.8 em I.6 tem-se,

$$-\frac{\partial E_p}{\partial W_{ji}} = \delta_{pj} o_{pi} \quad \text{Eq.I.9}$$

Então, para a implementação do gradiente de descida em  $E$ , a variação do peso é dada por:

$$\boxed{\Delta_p W_{ji} = \eta \delta_{pj} o_{pi}} \quad \text{Eq.I.10}$$

O objetivo é fazer a propagação reversa dos  $\delta$ 's através da rede, para isso, a Eq.I.8 é calculada da seguinte maneira,

$$\delta_{pj} = -\frac{\partial E_p}{\partial S_{pj}} = -\frac{\partial E_p}{\partial o_{pj}} \frac{\partial o_{pj}}{\partial S_{pj}} \quad \text{Eq.I.11}$$

Da Eq.I.4 tem-se que,

$$\frac{\partial o_{pj}}{\partial S_{pj}} = f'_j(S_{pj}) \quad \text{Eq.I.12}$$

Resta agora calcular o segundo fator. Para tanto existem dois casos a se considerar. No primeiro assume-se que a unidade  $j$  é uma unidade de saída da rede. Logo, a partir da definição de  $E_p$ , tem-se que:

$$\frac{\partial E_p}{\partial o_{pj}} = -(t_{pj} - o_{pj}) \quad \text{Eq.I.13}$$

Substituindo as duas últimas equações em I.11, tem-se

$$\boxed{\delta_{pj} = (t_{pj} - o_{pj}) f'_j(S_{pj})} \quad \text{Eq.I.14}$$

para qualquer unidade de saída  $u_j$ . Se  $u_j$  não é uma unidade de saída, então:

$$\sum_k \frac{\partial E_p}{\partial S_{pk}} \frac{\partial S_{pk}}{\partial o_{pj}} = \sum_k \frac{\partial E_p}{\partial S_{pk}} \frac{\partial}{\partial o_{pj}} \sum_i W_{ki} o_{pi} = \sum_k \frac{\partial E_p}{\partial S_{pk}} W_{kj} = -\sum_k \delta_{pk} W_{kj} \quad \text{Eq.I.15}$$

onde  $k$  representa o somatório sobre os neurônios da camada posterior à camada do neurônio  $j$ .

Substituindo I.15 e I.12 em I.11, resulta em:

$$\boxed{\delta_{pj} = f'_j(S_{pj}) \sum_k \delta_{pk} W_{kj}} \quad \text{Eq.I.16}$$

A regra delta generalizada fornece, através das equações I.14, I.16 e I.10, o procedimento computacional necessário para o ajuste de todos os pesos de uma rede do tipo PMC.

Sintetizando, o EBP pode ser descrito através dos seguintes passos de execução:

- 1) Inicializar pesos e *bias* com valores aleatórios pequenos (entre -0,5 e 0,5);
- 2) Propagar um padrão de entrada pela rede;
- 3) Comparar o sinal de saída obtido com o valor desejado;
- 4) Calcular e retropropagar a medida de erro (começando pela camada de saída) através da rede;
- 5) Minimizar o erro ajustando-se os pesos das conexões através de:

$$W_{ji}(K+1) = W_{ji}(K) + \Delta_p W_{ji}(K) \quad \text{Eq.I.17}$$

em que,

$$\Delta_p W_{ji} = \eta \delta_{pj} o_{pi};$$

$K$  refere-se a iteração corrente.

Nesta equação deve ser observado que se a unidade  $j$  é uma unidade de saída, então a equação I.14 deve ser usada no cálculo de  $\delta_{pj}$ . De outro modo, se a unidade  $j$  é uma unidade interna, então, a equação I.16 deverá ser usada. Os bias podem ser encarados como pesos e ajustados de maneira similar. Um termo de *momentum* ( $\xi$ ) pode ser adicionado na equação I.17 com o objetivo de evitar que a rede se prenda a algum mínimo local da superfície de erro. A saber,

$$W_{ji}(K+1) = W_{ji}(K) + \Delta_p W_{ji}(K) + \xi \Delta_p W_{ji}(K-1) \quad \text{Eq.I.18}$$

6) Apresentar o próximo padrão de entrada-saída e retornar ao passo 2. Os padrões de entrada-saída são apresentados, de forma cíclica e contínua, até que o critério de parada definido a priori seja atendido.

### I.5.5 “Back-propagation” como busca em Gradiente Descendente

A superfície que descreve o erro em cada exemplo (ou no conjunto de treinamento como um todo) como uma função de todos os pesos na rede, é conhecida como *superfície de erro*. A Figura I.7 apresenta uma superfície de erro característica de problemas linearmente separáveis. O conjunto atual de pesos define um ponto nessa superfície. Nesse ponto, nota-se uma inclinação da superfície ao longo do eixo formado por cada peso. Isso é conhecido como *derivada parcial* da superfície com respeito a cada peso - quanto o erro mudaria se fosse feita uma pequena mudança no peso. Portanto, altera-se os pesos em uma quantidade proporcional à inclinação naquela direção. Isso leva a rede como um todo na direção mais íngreme da superfície de erro (i.e.,  $w_1=a$  e  $w_2=b$ ).

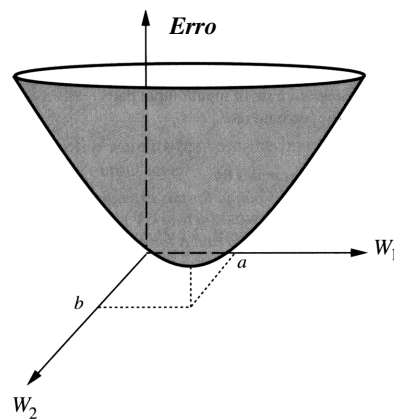


Fig. I.7. Superfície de erro característica de problemas linearmente separáveis [Russel & Norvig, 1995].

A superfície de erro acima tem a forma de um parabolóide e por conseguinte, apresenta apenas um mínimo. Entretanto, a superfície de erro pode não ser tão bem comportada e suas derivadas mais difíceis de serem calculadas. A título de exemplo, a Figura I.8 ilustra uma superfície de erro do tipo multimodal. Superfícies dessa natureza são caracterizadas por diversos mínimos locais, o que se torna um problema adicional. No

entanto, há vários métodos para superar esse problema. O fator de *momentum*, que tende a fazer com que as mudanças dos pesos mantenham sempre a mesma direção, permite que o algoritmo pule os mínimos menos “profundos”. O recozimento simulado<sup>1</sup>, assim como outras técnicas de otimização não convexa, também são úteis. Finalmente, ajustar a forma da função de ativação de uma unidade pode ter efeito sobre a susceptibilidade da rede a mínimos locais.

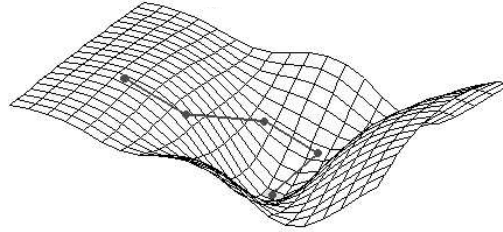


Fig. I.8. Superfície de erro multimodal

Em suma, a aprendizagem via “back-propagation” é um procedimento para dividir o cálculo do gradiente entre as unidades, de modo que a mudança em cada peso possa ser calculada para a unidade a qual o peso está ligado, usando apenas informação local.

### I.5.6 Generalização

Se todas as entradas e saídas possíveis forem apresentadas a uma rede de retropropagação, a mesma provavelmente encontrará um conjunto de pesos que faça o mapeamento das entradas para as saídas [Knight & Rich, 1994]. No entanto, na prática, se torna impossível fornecer todas essas entradas possíveis, mesmo porque, se dispuséssemos de um banco de dados completo e irrestrito, não faria muito sentido desenvolver-se uma ferramenta para desempenhar atividades de classificação ou regressão (por exemplo); bastaria criar algumas tabelas e consultá-las sempre que necessário. A título de exemplo, considere o problema de reconhecimento de rostos. Existe um número infinito de expressões e orientações em um rosto. Ainda assim, os seres humanos aprendem a classificá-los facilmente com alguns poucos exemplos. Seria de se esperar que as redes fizessem o mesmo. Felizmente, a retropropagação se mostra promissora como mecanismo de *generalização*. Se trabalharmos em um domínio (e.g.,

<sup>1</sup> *Simulated annealing*, do inglês.

reconhecimento de padrões) em que entradas semelhantes são mapeadas para saídas semelhantes, a retropropagação interpolará quando receber entradas que nunca tenha visto antes. A generalização também ajudará a superar ruídos indesejáveis nas entradas. É justamente isso que se deseja; desenvolver uma ferramenta que após ter sido treinada, classifique (no caso) de maneira satisfatória padrões de entrada desconhecidos.

Todavia, há algumas armadilhas durante o processo de aprendizagem de uma RNA. Antes de falar-se a respeito delas, torna-se propício elucidar como o banco de dados disponível é particionado. Basicamente, o mesmo será dividido em dois conjuntos distintos, são eles:

- conjunto de treinamento;
- conjunto de teste.

O conjunto de treinamento será formado por  $2/3$  do conjunto total<sup>2</sup> de dados e será usado efetivamente na tarefa de aprendizagem da rede conforme apresentado em seções anteriores. Por conseguinte, o conjunto de teste contará com o restante dos dados. A particularidade desse último conjunto é que os exemplos que o constituem, não serão apresentados à rede. Esse conjunto tem a função de testar o desempenho da rede durante o processo de treinamento, o que justifica a sua não inclusão.

A Figura I.9 mostra o desempenho de uma rede em função do tempo de treinamento. Durante a primeira parte do treinamento, o desempenho para o conjunto de treinamento melhora quando a rede ajusta seus pesos através da retropropagação. O desempenho para o conjunto de teste também melhora, embora não seja tão bom quanto o do conjunto de treinamento (ver Região 1). Depois de um certo tempo, o desempenho da rede atinge um platô, à procura de um caminho que leve a uma melhoria ainda melhor. Finalmente, esse percurso é encontrado e o desempenho para o conjunto de treinamento melhora de novo. Em contrapartida, o desempenho para o conjunto de teste piora. Por quê? O que ocorreu é que a rede começou a *memorizar* os pares entrada-saída ao invés de procurar pesos que descrevam de modo geral o mapeamento de todos os exemplos. Em outras palavras, caso se continue o processo de treinamento, a partir desse ponto<sup>3</sup> a rede começa a perder gradativamente a sua capacidade de generalização e começa a se *super-ajustar* (Figura I.12) aos exemplos de treinamento. É claro que tanto poder assim é indesejável e alguma

---

<sup>2</sup> Esse valor é de caráter orientativo.

<sup>3</sup> Passagem da região 1 para a 2.

providência deve ser tomada para impedir que a retropropagação recorra a um esquema de consulta de tabelas. Uma delas seria detectar a fronteira entre as regiões 1 e 2, e interromper o treinamento. Esse procedimento é conhecido como *Critério de Parada Prematura*<sup>4</sup> e tem como base a suposição de que qualquer melhoria virá através de “trapaça”. Uma outra maneira seria acrescentar pequenas quantidade de ruído ao conjunto de treinamento. O ruído deve ser suficiente para evitar a memorização (ou “overfitting”), mas não deve ser muito, pois uma quantidade excessiva poderia confundir o classificador.

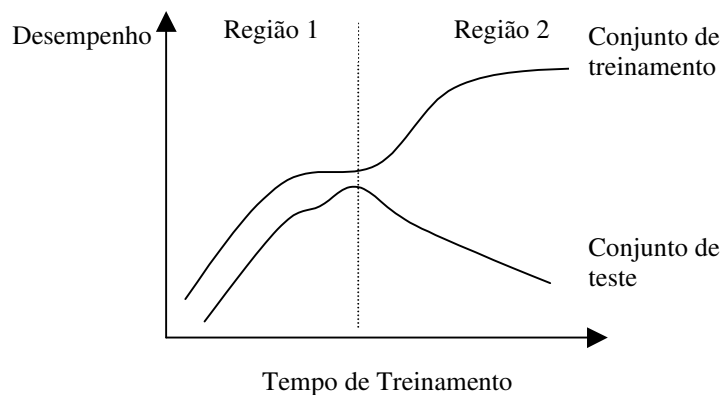


Fig. I.9 Generalização em aprendizagem de RNA.

### I.5.7 Estrutura de Rede Ótima

Até então, do ponto de vista prático, muito pouco foi dito a respeito da estrutura de RNA do tipo PMC. A Figura I.6 apresentou uma estrutura fixa, determinada por algum agente externo, mas nenhuma justificativa em particular foi dada a respeito da escolha do número de camadas e do número de unidades presentes em cada uma dessas camadas. Logo, pode-se questionar:

- a) Quantas camadas comporão uma determinada rede? Uma, duas ou três?
- b) Qual o número de unidades que cada camada conterà?

<sup>4</sup> *Early Stopping Method*, do inglês.

A resposta para essas perguntas é no mínimo inusitada; depende do problema. Atualmente, não se dispõe de uma metodologia sistemática para a determinação da estrutura de uma PMC. O que existe, são certos procedimentos que nos auxiliam na determinação de uma estrutura de rede ótima. A seguir, discutir-se-á algumas conseqüências ocasionadas pela escolha de estrutura de rede inadequada e o que fazer para contornar esse problema.

A escolha da estrutura de uma PMC, por si só, é um ponto fraco em potencial, pois uma arquitetura mal dimensionada pode comprometer o desempenho da rede [Russel & Norvig, 1995]. Se escolhe-se uma rede muito pequena (i.e., com número de camadas e/ou unidades reduzidos), então o modelo, via de regra, será incapaz de representar a função desejada; isso é conhecido como *sub-ajuste* (ou “underfitting”). Por outro lado, se escolhe-se uma rede muito grande (i.e., com um número muito grande de parâmetros livres), ela poderá ao invés de generalizar, fazer uma memorização excessiva dos dados de treinamento.

A título de exemplo, considere um problema de ajuste funcional. A Figura I.10 ilustra o problema proposto. A Figura I.11 apresenta os resultados gerados por uma rede de estrutura muito pequena (camada oculta: duas unidades). Pode-se notar que a rede foi incapaz de gerar um conjunto de pesos e *bias* que se ajustem bem ao problema de aproximação funcional proposto. A Figura I.12 ilustra a ocorrência do “overfitting” (camada oculta: 40 unidades). Finalmente, a Figura I.13 apresenta a resposta de uma rede com arquitetura adequada para o problema em questão (camada oculta: 5 unidades).

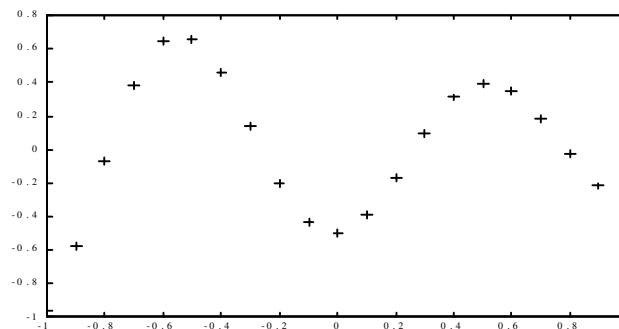


Fig. I.10. Problema proposto.



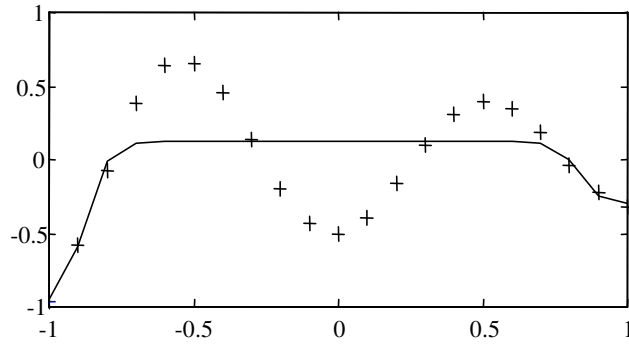


Fig. I.11. 'Underfitting'.

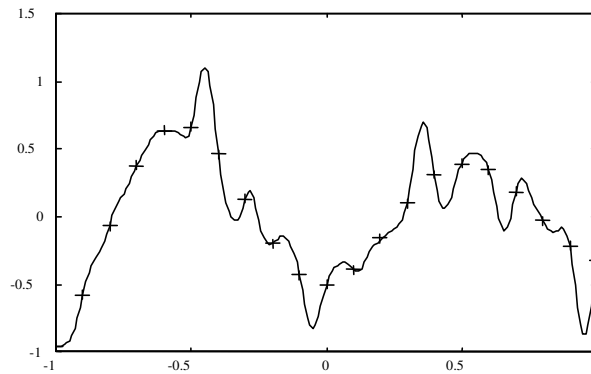


Fig. I.12. 'Overfitting'.

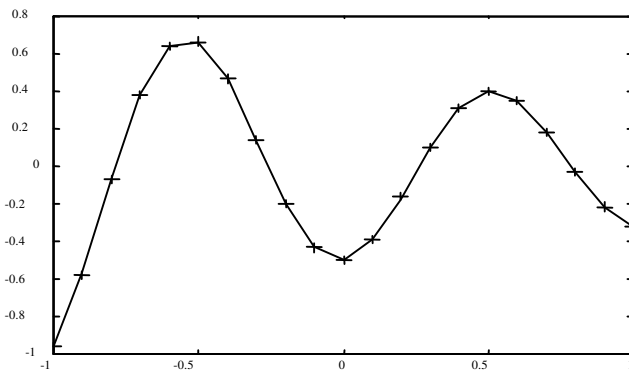


Fig. I.13. Generalização adequada.

Com respeito ao problemas citados anteriormente, pode-se pensar no problema de encontrar a estrutura de rede adequada, como um problema de busca. Recentemente, uma abordagem que tem sido adotada é usar algoritmos genéticos (GA) para buscar a rede

ideal no espaço de estruturas de rede. Um inconveniente desta técnica é o elevado esforço computacional requerido devido ao tamanho do espaço de busca. Uma outra solução mais tradicional é fazer-se buscas do tipo “*hill-climbing*” que seletivamente modificam a estrutura de uma rede. Existem duas variantes: iniciar com uma rede maior e torná-la menor, ou iniciar com uma rede pequena e torná-la maior. Técnicas de validação cruzada também são úteis para decidir quando a estrutura de rede ideal foi encontrada para um determinado problema.

### I.5.8 Utilização

Depois que a rede estiver treinada, ela poderá ser utilizada como uma ferramenta para, e.g., classificação de novos dados. Para isto, a rede deverá ser utilizada apenas no modo progressivo, i.e., novas entradas são apresentadas à rede, são processadas nas camadas ocultas e os resultados são obtidos na camada de saída, como no treinamento, mas sem a retropropagação do erro. O resultado apresentado pela rede, refletirá o que ela aprendeu durante a fase de treinamento.

## I.6 Discussão

A seguir, examinar-se-á um conjunto de tópicos que se destacaram neste anexo, são eles:

- **Expressividade:** RNA são claramente uma representação baseada em atributos, e não têm poder expressivo de gerar representações lógicas. São indicadas para tratar com entradas e saídas contínuas. Em particular, uma PMC pode representar qualquer função desejada de um conjunto de atributos. Na prática, grande parte dos problemas podem ser resolvidos com modelos parcimoniosos.
- **Transparência:** As redes neurais que utilizam EBP, assim como muitos outros tipos de RNA, podem ser vistas como “caixas pretas” (os modelos não apresentam justificativas para suas respostas). Neste sentido, muitas pesquisas vêm sendo realizadas visando a extração de conhecimento de redes neurais artificiais, e na

criação de procedimentos explicativos, onde se tenta justificar o comportamento da rede em determinadas situações [Benítez et al., 1997].

- **Esforço computacional:** Em se tratando de limitações, uma delas refere-se ao tempo de treinamento de redes neurais utilizando “back-propagation”, que tende a ser muito longo. Algumas vezes, são necessárias milhares de épocas para se chegar a resultados aceitáveis. Para minimizar esse problema, já existem algumas variações do back-propagation que fazem com o que o treinamento seja mais rápido e eficaz (ver capítulo 4). Paralelamente, muitos estudos estão sendo realizados para a implementação de redes neurais em computadores paralelos, além da construção de *chips* neurais (e.g., o Intel 80170NX Electronically Trainable ANN).
- **Generalização:** RNA têm a capacidade de generalizar bem, desde que se determine uma estrutura de rede adequada para o problema a ser tratado. Nenhum teorema foi demonstrado aqui provando isso. No entanto, na literatura especializada, pode-se encontrar várias publicações que reportam aplicações bem sucedidas com RNA em um grande número de problemas do mundo real.

## I.7 Aplicações de RNA em Sistemas Elétricos de Potência

Nesta seção serão apresentados alguns exemplos das aplicações mais significativas de RNA em SEP. Os resultados de pesquisas feitas por [Niebur, 1993] e [Mori,1996] também serão apresentados com o intuito de se verificar a evolução da aplicação dos diversos tipos de RNA em SEP. Para finalizar, serão apresentados alguns números relativos ao artigos recomendados pelo comitê técnico do ISAP99 (*Intelligent Systems Application to Power Systems*), realizado no Rio de Janeiro, em abril de 1999. Dos setenta e dois artigos publicados, vinte e dois envolviam RNA.

Algumas áreas típicas de aplicação de RNA são:

- Previsão de Carga;
- Planejamento;
- Análise de segurança;
- Detecção e Diagnóstico de Falta;
- Controle;
- Análise;
- Proteção.

A Fig. I.14 mostra a proporção dos tipos de aplicações de RNA em SEP de acordo com [Niebur, 1993]. Já os modelos de RNA mais empregados são apresentados na Fig. I.15.

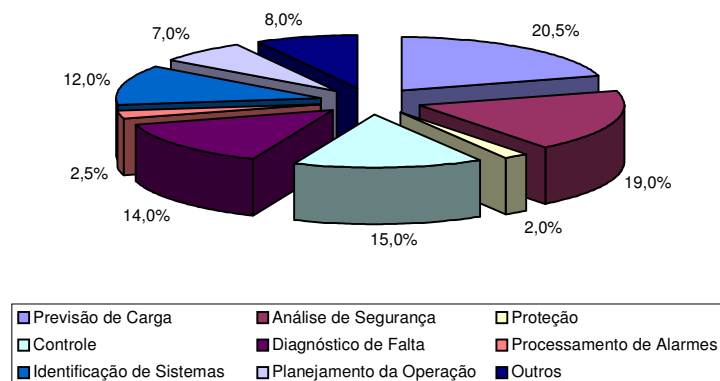


Fig. I.14. Principais áreas de aplicação em abril de 1993.

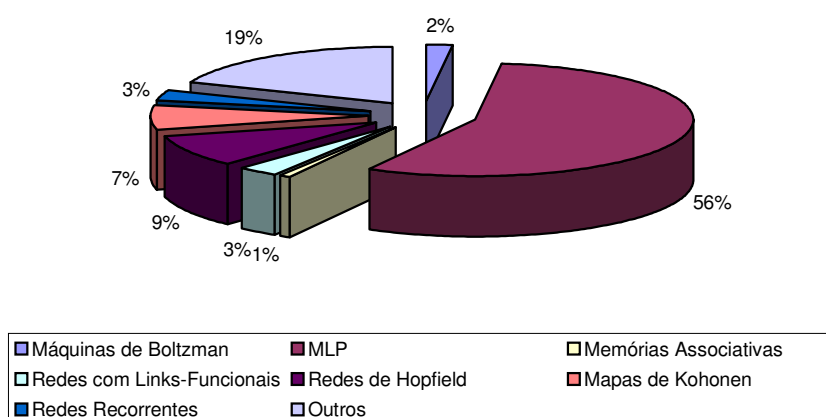


Fig. I.15 Modelos de RNA usados em problemas de SEP (1993).

De acordo com [Mori, 1996], a proporção dos tipos de aplicações de RNA em SEP encontrava-se da seguinte forma:

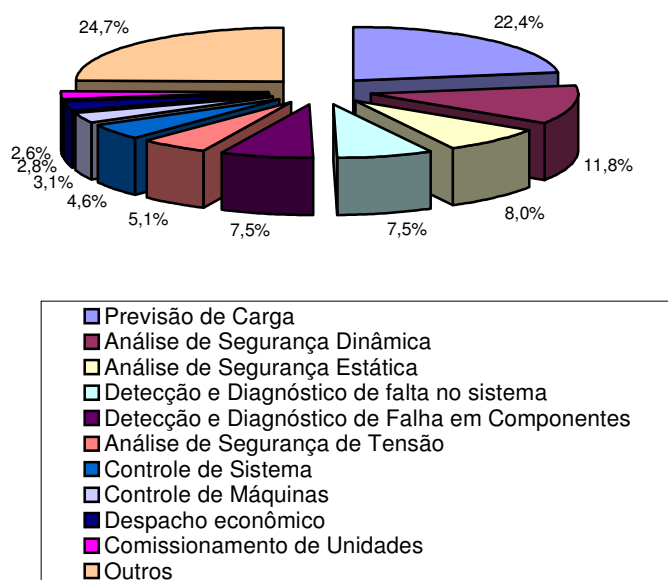


Fig. I.16. Principais áreas de aplicação em 1995.

Já os modelos de RNA mais empregados são apresentados na Figura I.17. São eles:

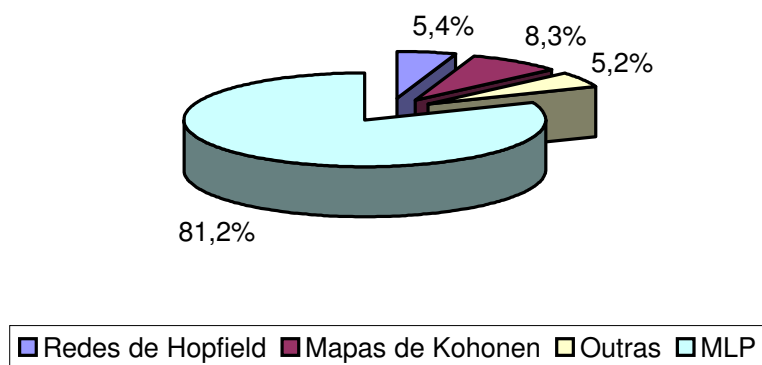


Fig. I.17. Modelos de RNA usados em problemas de SEP (1995).

A Tab. I.1 mostra a área de aplicação das RNA em SEP e o respectivo número de artigos apresentados no ISAP 99, a saber:

Tab. I.1 RNA no ISAP 99.

Área de Aplicação	Número de Artigos publicados
Previsão	5
Proteção	3
Identificação e Controle	2
Diagnóstico de faltas no sistema	1
Diagnóstico de faltas em componentes	3
Avaliação da segurança	3
Análise de segurança de tensão	1
Estabilização de sistemas de potência	1
Qualidade de energia	3

Analisando-se a Tab. I.1, nota-se que as RNA estão sendo aplicadas em outras áreas dos SEP (e.g., Qualidade de Energia).

## I.8 Principais Referências de Aplicações de RNA em SEP

### *IEEE Transactions / Journals / Proceedings:*

- PES Transactions on Power Systems
- PES Transactions on Power Delivery
- Computer Application in Power
- Transactions on Neural Networks
- Proceedings of IEEE
- PES Winter/Summer Meeting
- Power Industry Computer Application Conference (PICA)
- Intelligent Symposium on Circuits and Systems (ISCAS)
- Intelligent Conference on SMC
- Intelligent Conference on Neural Networks (ICNN)

***Outros Journals/Conference/Proceedings:***

- AANNM (NSF Workshop on Applications of Artificial Neural Network Methodology in Power System Engineering; Clemson, SC, 1990)
- ANNPS (International Forum on Application of Neural Networks to Power Systems; Seattle, WA, 1991, and Yokohama, Japão, 1993)
- Bulk Power System Voltage Phenomena II, and III-Voltage Stability and Security; Deep Creek Lake, MD, 1991; Davos, Suíça, 1994)
- Electrical Power Systems Research (EPRI)
- ESAP (International Symposium. on Expert Systems Application to Power Systems; Seattle 1989; Tokyo/Kobe 1991, and Melbourne 1993)
- EPRI/NSF Workshop on Bulk Power System Voltage Phenomena II, Voltage Stability and Security
- ICARCV (International Conference on Automation, Robotics, and Computer Vision)
- ICANN (International Conference on Artificial Neural Networks)
- IKE Proceedings Pt. C
- IJCNN(IEEE/INNS Intelligent Joint Conference on Neural Networks)
- IFAC/IFIP/IMACS (Intelligent Symposium on Artificial Intelligence in Real time; Intelligent Journal of Electrical Power and Energy Systems; Intelligent Conference on Fuzzy Logic, Neural Nets and Soft Computing)
- ISAP (International Conference on Intelligent System Application Systems; Montpellier, França, 1994)
- Journees d' Electronique
- Neurocomputing
- NNCEPI (INNS/EPRI Summer Workshop on Neural Network Computing in Electric Power Industry; Stanford, CA, 1992)
- NEURONET ' 90
- NOLTA (International Symposium on Nonlinear Theory and Its Applications)
- PSCC(Power System Computation Conference)
- ISAP (International Conference on Intelligent System Application to Power Systems)

## I.9 Resumo

Este anexo apresentou alguns conceitos e idéias básicas a respeito das RNA. Os principais pontos discutidos serão enumerados a seguir:

1. Uma rede neural artificial é um modelo computacional que compartilha algumas propriedades do cérebro humano: ela consiste de muitas unidades simples trabalhando em paralelo, sem uma unidade de controle central. As conexões entre as unidades tem pesos numéricos que vão sendo modificados durante o processo de aprendizagem. Todo o “conhecimento” adquirido pela rede está codificado nesses pesos.
2. O comportamento de uma rede neural é determinado por sua topologia e pela natureza de suas unidades individuais.
3. Redes do tipo PMC podem representar qualquer função (desde que possua uma topologia adequada para tanto).
4. O algoritmo de *Back-propagation* é aplicado em redes do tipo PMC. Se utiliza o método de descida em gradiente, no espaço de pesos, para minimizar o erro de saída. É um algoritmo bastante eficaz, porém sua convergência é lenta (o tempo de treinamento é proporcional ao cubo do número de interconexões).
5. Vários modelos de redes neurais têm sido aplicados a problemas de SEP e resultados expressivos têm sido publicados na literatura especializada. Áreas como previsão de carga, análise de segurança e diagnóstico de faltas estão consolidadas.



## Anexo II – Domínio da Freqüência

### Introdução ao Domínio da Freqüência

Uma série temporal é uma coleção de amostras de uma variável medida em diferentes instantes. Usualmente, essas medições são tomadas em intervalos de tempo equiespaçados. Desde que os dados sejam tomados em função do tempo, diz-se que a série se encontra no *domínio do tempo*.

Existe um grande número de domínios que podem ser usados para expressar exatamente a mesma informação contida em uma série temporal. Um dos mais úteis é o chamado *domínio da freqüência*. A seguir, será apresentada uma ferramenta matemática que é responsável pela transformação de uma série no domínio do tempo para uma série no domínio da freqüência.

### II.1 Análise de Fourier para sinais discretos

Considere que  $x_0, x_1, \dots, x_{n-1}$  seja uma série temporal consistindo de  $n$  igualmente espaçadas amostras de uma variável (apesar de ser perfeitamente legal que uma variável seja complexa). A Transformada de Fourier Discreta (DFT) é uma série de  $n$  números complexos que serão denotados por  $w_0, w_1, \dots, w_{n-1}$ . Essa representação no domínio da freqüência pode ser computada com a Eq.II.1, a saber:

$$w_j = \sum_{k=0}^{n-1} [x_k \cos\left(\frac{2\pi jk}{n}\right) + x_k \sin\left(\frac{2\pi jk}{n}\right)i] \quad \text{Eq.II.1}$$

A Eq.II.1 consiste de dois termos quase idênticos. O primeiro deles é a multiplicação da série no domínio do tempo com uma série definida pela função coseno. O segundo é a multiplicação da série, com uma série definida pela função seno. Esse segundo termo é ainda multiplicado pelo número imaginário  $i$ .

Em suma, tem-se que multiplicar cada um dos  $n$  pontos da série temporal por um par de funções trigonométricas e então somar. Isso deve ser feito para cada um dos  $n$  termos na

DFT. Desse modo, o número total de operações aritméticas para a computação completa de uma DFT é da ordem de  $n^2$ .

Caso a série de interesse apresente poucos pontos, não há nenhum problema em calcular-se a DFT<sup>1</sup>. No entanto, quando a série apresenta muitos pontos, o tempo computacional para o cálculo de uma DFT completa pode ser proibitivo. Felizmente, várias propriedades trigonométricas podem ser utilizadas para se reduzir drasticamente o esforço computacional exigido.

Um algoritmo chamado *Transformada Rápida de Fourier* (FFT) foi inventado há várias décadas atrás. Ele, recursivamente, divide a série de interesse pela metade, transformando cada metade separadamente em um quarto do tempo necessário para transformar a série inteira e rapidamente agrupando os resultados. O número de operações nesse algoritmo é da ordem de  $n \log n$ . Uma grande melhoria em relação à DFT padrão. A única exigência desse algoritmo é que a série tenha um número de pontos que seja potência de 2. Via de regra, na prática, esse problema pode ser facilmente contornado.

A título de exemplo, a Tabela II.1 mostra uma série com seis pontos na primeira coluna, a parte real de sua DFT na segunda coluna, e a parte imaginária na terceira coluna.

Tab. II.1 Aplicação da DFT.

<b>Ponto</b>	<b>Série</b>	<b>Parte real da DFT</b>	<b>Parte Imag. da DFT</b>
0	3	10	0
1	1	0,5	2,6
2	5	-3,5	0,9
3 ( $n/2$ )	-2	14	0
4	4	-3,5	-0,9
5	-1	0,5	-2,6

A partir de [Strang, 1986], e após uma análise minuciosa na Tabela II.1, pode-se concluir que:

<sup>1</sup> Do ponto de vista de esforço computacional

- $W_i$  representa uma componente periódica que se repete exatamente  $i$  vezes durante o período de tempo definido pela série que vai de  $x_0$  a  $x_{n-1}$ ;
- Se uma componente periódica se repete  $i$  vezes a cada  $n$  amostras, o período de repetição é  $n/i$ ;
- O menor período resolvido pela DFT é de duas amostras e o mais longo de  $n$  amostras;
- A componente de frequência mais alta está em  $n/2$ ;
- Apenas os termos de zero até  $n/2$  da parte real da DFT são **únicos**;
- A parte imaginária do termo zero e do termo  $n/2$  são sempre **zero**;
- Existem exatamente  $n$  números **únicos** definidos pela DFT.

Para maiores detalhes sobre a análise de Fourier, ver [Strang,1986].

### II.1.1 Frequência de Nyquist

Uma das conclusões da seção anterior foi que a componente de frequência mais alta está associada ao ponto  $n/2$ . Essa é uma frequência especial e é conhecida como a *Frequência de Nyquist*. Existe um famoso teorema chamado “Teorema de amostragem de Nyquist” que declara, de forma sucinta, que: “Se uma forma de onda contínua não contém quaisquer frequências superiores a um limite  $f$ , pode-se capturar toda a informação contida nesse sinal, amostrando-o numa taxa de no mínimo  $2f$  amostras por unidade de tempo”.

A tecnologia empregada nos CDs de áudio faz uso desse teorema. Com ela, amostra-se o sinal gravado 44.000 vezes por segundo. Isso permite capturar perfeitamente toda a informação até 22.000 ciclos por segundo. Como o limite da audição humano é de aproximadamente 20.000 ciclos por segundo, esse procedimento é suficiente já que todas as frequências audíveis são capturadas.

## II.2 O Espectro de Potência

Cada termo  $W_i$  da DFT é um número complexo. Analisar as partes reais e imaginárias de cada  $W_i$  pode ser algumas vezes difícil. Em situações onde o que realmente se quer saber sobre a série temporal é quanto de sua energia é devido às diferentes componentes de

frequência de interesse, existe uma maneira mais conveniente – calcula-se o espectro de potência (PS) da série. A energia de uma componente da DFT é igual ao quadrado de seu módulo. Isso permite definir o espectro de potência discreto como sendo:

$$\begin{aligned}
 P_0 &= \frac{|w_0|^2}{n^2} \\
 P_i &= 2 \frac{|w_i|^2}{n^2} \quad 0 < i < n/2 \\
 P_{n/2} &= \frac{|w_{n/2}|^2}{n^2}
 \end{aligned}
 \tag{Eq.II.2}$$

A Eq.II.2 é extremamente útil. Ela mostra como usar a DFT para estimar a energia contida em cada faixa de frequência discreta. Abaixo estão listados alguns possíveis usos para essa ferramenta, são eles:

- Avaliação da necessidade de desazonalização;
- Identificação de séries não estacionárias;
- Auxílio na sintonia de filtros digitais.

## II.3 Transformada Inversa de Fourier

Conforme foi visto no item II.1, a DFT pode ser usada para decompor uma série temporal em componentes de frequência individual. A série original pode ser reconstruída, a partir do domínio da frequência, através da Eq.II.3.

$$x_k = \frac{1}{n} \sum_{j=0}^{n-1} [w_j \cos(\frac{2\pi jk}{n}) - w_j \operatorname{sen}(\frac{2\pi jk}{n})i]
 \tag{Eq.II.3}$$

## Anexo III – Fundamentos da Wavelet de Haar

Por analogia com a análise de Fourier, considere o espaço  $L^2(\mathfrak{R})$  de todas as funções mensuráveis de quadrado integrável<sup>1</sup> sobre  $\mathfrak{R}$ . Aqui as funções  $f(x)$  devem decair rapidamente a zero, quando  $|x| \rightarrow \infty$ . A idéia é considerar dilatações (compressões) e translações de uma única função  $\psi$ , de modo a cobrir  $\mathfrak{R}$ , i.e.,

$$\left\{ \psi\left(\frac{x-b}{a}\right), (a,b) \in \mathfrak{R}^+ \times \mathfrak{R} \right\}.$$

A princípio,  $a$  e  $b$  poderiam assumir quaisquer valores. No entanto, é conveniente tomar valores especiais para eles, a saber :  $a = 2^{-j}$  e  $b = k \cdot 2^j$ , em que  $k$  e  $j$  são inteiros. Essa escolha produzirá uma *base esparsa* permitindo uma reconstrução precisa do sinal com poucos vetores ([Morettin, 1997], [Strang, 1986]).

### III.1 Wavelet de Haar

Considere a mais antiga e mais simples wavelets de todas, a chamada Wavelet de Haar,  $\psi(x)$ . Ela é uma função degrau tomando valores de 1 e -1, nos intervalos  $[0,1/2)$  e  $[1/2,1)$ , respectivamente (ver Fig.3.1, pág.31). A wavelet de Haar é conhecida por mais de oitenta anos e tem sido usada em vários campos da matemática.

É sabido que qualquer função contínua pode ser aproximada uniformemente por funções de Haar. Dilatações e translações da função  $\psi$ , do tipo:

$$\psi_{jk}(x) = \text{const} \cdot \psi(2^j \cdot x - k),$$

definem uma base ortogonal em  $L^2(\mathfrak{R})$ . Isso significa que qualquer elemento em  $L^2(\mathfrak{R})$  pode ser representado como uma combinação linear dessa função base.

Para se checar a ortogonalidade de  $\psi_{jk}$ , verifica-se se  $\langle \psi_{jk}, \psi_{j'k'} \rangle = 0$  (exceto quando  $j = j'$  e  $k = k'$ ), i.e.,

---

<sup>1</sup> Uma função  $f$  está em  $L^2(S)$  se  $\int_S f^2$  é finita.

$$\int \psi_{jk} \cdot \psi_{j'k'} = 0 . \quad \text{Eq.III.1}$$

Se  $j \neq j'$  (com  $j' < j$ ), então valores não nulos da Wavelet  $\psi_{j'k'}$  estão contidos no conjunto onde a Wavelet  $\psi_{jk}$  é constante. Isso faz com que a integral (Eq.III.1) assumam valor zero. Por outro lado, se  $j = j'$ , mas  $k \neq k'$ , então, no mínimo, um fator no produto  $\psi_{j'k'} \cdot \psi_{jk}$  é zero. Daí, pode-se afirmar que as funções  $\psi_{jk}$  são ortogonais.

Para que  $\psi_{jk}$ , além de ser ortogonal, seja também ortonormal, é necessário que

$$\| \psi_{jk} \| =_{\text{def}} \langle \psi_{jk}, \psi_{jk} \rangle = \int \psi_{jk}^2 = 1.$$

Daí, calcularemos a constante *const* como sendo:

$$1 = (\text{const})^2 \cdot \int \psi^2(2^j \cdot x - k) \cdot dx = (\text{const})^2 \cdot 2^{-j} \cdot \int \psi^2(t) \cdot dt = (\text{const})^2 \cdot 2^{-j} .$$

$$\boxed{\therefore \text{const} = 2^{j/2}}$$

Logo, o conjunto  $\{\psi_{jk}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$  define uma base ortonormal em  $L^2$ . Como uma alternativa, consideraremos bases ortonormais da forma  $\{\phi_{j_0,k}, \psi_{jk}, j \geq j_0, k \in \mathbb{Z}\}$ , onde  $\phi_{00}$  é chamado de *função escala* associada com a base da wavelet  $\psi_{jk}$ . O conjunto  $\{\phi_{j_0,k}, k \in \mathbb{Z}\}$  atravessa o mesmo subespaço que  $\{\psi_{jk}, j < j_0, k \in \mathbb{Z}\}$ . Para a base da wavelet de Haar, a função escala é muito simples. Ela é igual a unidade no intervalo  $[0,1)$ , i.e.

$$\phi(x) = 1(0 \leq x < 1) .$$

E ainda, devemos salientar que:

$$\psi_{j,k}^{(\text{Haar})}(x) = \begin{cases} 2^{j/2}, & 2^{-j} \cdot k \leq x < 2^{-j} \cdot (k + \frac{1}{2}) \\ -2^{j/2}, & 2^{-j} \cdot (k + \frac{1}{2}) \leq x < 2^{-j} \cdot (k + 1) \\ 0, & \text{caso contrario} \end{cases}$$

## III.2 Representação em Wavelets de funções geradas por um conjunto de dados

Considere que  $y = (y_0, y_1, \dots, y_{2^n-1})'$  seja um vetor de dados de dimensão  $2^n$ . Esse vetor pode ser associado com uma função constante discreta  $f$  em  $[0,1)$  gerado por  $y$  da seguinte forma:

$$f(x) = \sum_{k=0}^{2^n-1} y_k \cdot 1(k \cdot 2^{-n} \leq x < (k+1) \cdot 2^{-n}).$$

Como a função  $f$  está no espaço  $L^2[0,1)$ , a decomposição em wavelet de  $f$  será:

$$f(x) = c_{00}\phi(x) + \sum_{j=0}^{n-1} \sum_{k=0}^{2^j-1} d_{jk}\psi_{jk}(x). \quad \text{Eq.III.2}$$

Daí, é importante salientar que:

- o somatório em relação a  $j$  é finito, pois  $f$  é uma função degrau, e tudo pode ser exatamente descrito pelos primeiros  $(n-1)$  níveis de resolução;
- cada nível com relação a  $k$  também é finito pois o domínio de  $f$  é finito;
- no caso, nenhuma translação da função escala  $\phi_{00}$  é necessária.

## III.3 Um exemplo numérico

- a) Suponha que o vetor de dados  $y = (1, 0, -3, 2, 1, 0, 1, 2)$ , represente observações de um processo estocástico. Suponha que  $T = 2^M$ , com  $M > 0 \in \mathbb{Z}$ . Encontrar a decomposição em wavelets explicitada na Eq.III.2. A função correspondente  $f$  é dada na Fig.III.1.

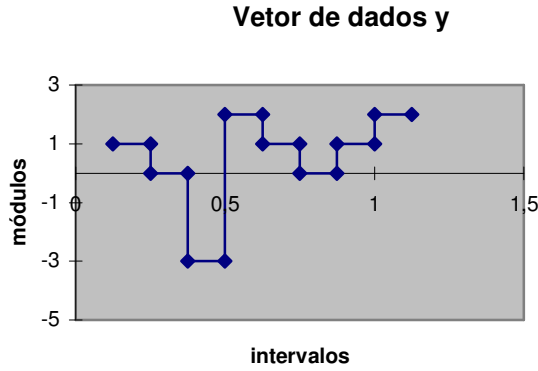


Fig. III.1. Representação gráfica dos dados.

Define-se a transformada wavelet discreta  $d$  de  $y$ , com respeito à wavelet mãe  $\psi$ , como sendo:

$$d_{j,k}^{(\psi)} = \sum_{t=0}^{T-1} Y_t \psi_{j,k}(t/T).$$

Na forma matricial, teremos que:  $d = Wy$

Como condições de fronteira apropriadas foram utilizadas, pode-se afirmar que a transformada é ortogonal e podemos invertê-la na forma:  $y = W'd$ .

Assim, teremos que<sup>2</sup>:

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \sqrt{2} & 0 & 2 & 0 & 0 & 0 \\ 1 & 1 & \sqrt{2} & 0 & -2 & 0 & 0 & 0 \\ 1 & 1 & -\sqrt{2} & 0 & 0 & 2 & 0 & 0 \\ 1 & 1 & -\sqrt{2} & 0 & 0 & -2 & 0 & 0 \\ 1 & -1 & 0 & \sqrt{2} & 0 & 0 & 2 & 0 \\ 1 & -1 & 0 & \sqrt{2} & 0 & 0 & -2 & 0 \\ 1 & -1 & 0 & -\sqrt{2} & 0 & 0 & 0 & 2 \\ 1 & -1 & 0 & -\sqrt{2} & 0 & 0 & 0 & -2 \end{pmatrix} \cdot \begin{pmatrix} c_{00} \\ d_{00} \\ d_{10} \\ d_{11} \\ d_{20} \\ d_{21} \\ d_{22} \\ d_{23} \end{pmatrix}$$

Daí, a solução é:

<sup>2</sup> Para se calcular as matrizes  $W'$  e  $d$ , foi implementada uma subrotina em ambiente *Matlab*. A listagem completa se encontra no item III.4.



$$\begin{pmatrix} c_{00} \\ d_{00} \\ d_{10} \\ d_{11} \\ d_{20} \\ d_{21} \\ d_{22} \\ d_{23} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2\sqrt{2}} \\ -\frac{1}{2\sqrt{2}} \\ \frac{1}{4} \\ \frac{5}{4} \\ -\frac{1}{4} \\ \frac{1}{4} \end{pmatrix}.$$

Assim,

$$f = \frac{1}{2}\phi - \frac{1}{2}\psi_{00} + \frac{1}{2\sqrt{2}}\psi_{10} - \frac{1}{2\sqrt{2}}\psi_{11} + \frac{1}{4}\psi_{20} - \frac{5}{4}\psi_{21} + \frac{1}{4}\psi_{22} - \frac{1}{4}\psi_{23}. \quad \text{Eq.III.3}$$

b) Reconstruir o sinal a partir de Eq.III.3, quando  $x \in \left[0, \frac{1}{8}\right)$ .

$$f(x) = \frac{1}{2} - \frac{1}{2} \cdot 1 + \frac{1}{2\sqrt{2}} \cdot \sqrt{2} + \frac{1}{4} \cdot 2 = 1,$$

que nada mais é, que a multiplicação da primeira linha de  $W'$  pela matriz coluna  $d$ . O resultado é igual a  $y_0$ .

Terminado este exemplo, e em se tratando de vetores  $y$  de dimensões elevadas, a resolução através de equações matriciais se torna uma tarefa um tanto quanto árdua. Daí, a análise de multiresolução de Mallat desempenha um papel fundamental na resolução de tais problemas.

### III.4 Subrotina implementada

```
% Wavelets de Haar
%=====
%A. J. Rocha Reis, 07/10/1997 - EFEI
%=====
clc
disp(' Wavelet de Haar' )
```

```

disp(' Esta subrotina decompõe o vetor y apresentado na seção III.3, através da TWD.' )
input(' Entre com o valor de M (níveis de resolução): ');
M=ans;
psi=0;
T=2^3;
teste=0;
lna=1;
cla=2;
for t=1:T
    for jota=0:(M-1)
        for k=0:(2^(jota)-1)

            % cálculo de psi
            % =====
            a=2^(-jota)*k;
            b=2^(-jota)*(k+.5);
            c=b;
            d=2^(-jota)*(k+1);

            if ((teste >= a) & (teste<b))
                psi(lna,cla)=2^(jota/2);
            elseif ((teste >= c) & (teste < d))
                psi(lna,cla)=-2^(jota/2);
            else
                psi(lna,cla)=0;
            end
            saida=psi;
            cla=cla+1;
        end
    end
    teste=t/T;
    lna=lna+1;
    cla=2;
end
clc
for x=1:T
    saida(x,1)=1;
end
clc
disp(' Matriz W transposta ' )
W=saida
D=[1;0;-3;2;1;0;1;2];
coeficientes=inv(W)*D;
pause (5)
disp(' ' )
disp(' Matriz coluna dos coeficientes ' )
coeficientes

```

## Anexo IV - Publicações

---

### *Periódicos:*

#### *Periódicos Nacionais: 1*

1. **Agnaldo J.R. Reis** & Alexandre P. Alves da Silva: “Aplicação da Transformada Wavelet Discreta na Previsão de Carga a Curto Prazo via Redes Neurais”, Aceito para publicação na revista da SBA Controle & Automação.

#### *Periódicos Internacionais: 1*

2. **Agnaldo J.R. Reis** & Alexandre P. Alves da Silva, “Multiresolution Short-Term Forecasting for Electrical Load Using Wavelet decompositions and Neural Networks”, WSEAS Transactions on Systems, Issue 3, Vol. 2, July 2003.

### *Anais de Congressos:*

#### *Conferências Internacionais: 4*

#### *Conferências Nacionais: 7*

3. **Agnaldo J.R. Reis**, Alexandre P. Alves da Silva, “Seleção de Variáveis de Entrada para Previsores Neurais – Um Estudo Comparativo”, aceito para publicação nos anais do *IX Symposium of Specialists in Electric Operational and Expansion Planning* (SEPOPE 2004).
4. Alexandre P. Alves da Silva, Ubiratan P. Rodrigues, **Agnaldo J.R. Reis**, Luciano S. Moulin, Paulo C. do Nascimento, “Artificial Neural Network-Based Short-Term Demand Forecaster”, aceito para publicação nos anais do *5<sup>th</sup> Latin-American Congress: Electricity Generation and Transmission* (5<sup>th</sup> CLAGTEE – São Pedro-SP, Novembro/2003)
5. **Agnaldo J.R. Reis** & Alexandre P. Alves da Silva, “Enhancing Neural Network Based Load Forecasting Via Discrete Wavelet Transform”, anais da *International Conference on Intelligent System Application to Power Systems* (ISAP – Lemnos, Grécia, Agosto/ 2003).

6. **Agnaldo J.R. Reis** & Alexandre P. Alves da Silva, ‘Previsor Neural de Carga Elétrica baseado na Transformada Wavelet Discreta’, anais do VI Congresso Brasileiro de Redes Neurais (VI CBRN), São Paulo, Junho de 2003.
7. **Agnaldo J.R. Reis** & Alexandre P. Alves da Silva: ‘Aplicação da Transformada Wavelet Discreta na Previsão de Carga a Curto Prazo via Redes Neurais’, Anais do XIV Congresso Brasileiro de Automática (XIV CBA), Natal, Setembro de 2002.
8. **Agnaldo J.R. Reis** & Alexandre P. Alves da Silva: ‘Previsor Neural de Carga Elétrica Baseado em Análise de Multiresolução via Wavelets’, anais do VIII Simpósio de Especialistas em Planejamento de Operação e Expansão Elétrica (VII SEPOPE), Brasília, Maio de 2002, ref. IP-052.
9. Alexandre P. Alves da Silva, Ubiratan P. Rodrigues, **Agnaldo J.R. Reis**, Luciano S. Moulin, & Carlos A. Silva Neto: ‘ORÁCULO – Uma Ferramenta para Previsão de Carga’, anais do XVI Seminário Nacional de Produção e Transmissão de Energia Elétrica (SNPTEE), Campinas, Outubro 2001, Grupo IX, 6 páginas.
10. Alexandre P. Alves da Silva, Ubiratan P. Rodrigues, **Agnaldo J.R. Reis**, & Luciano S. Moulin: ‘NeuroDem – A Neural Network Based Short Term Demand Forecaster’, anais do IEEE Power Tech, Porto, Portugal, Setembro de 2001.
11. Alexandre P. Alves da Silva, **Agnaldo J.R. Reis**, Mohamed A. El-Sharkawi, & Robert J. Marks II: ‘Enhancing Neural Network Based Load Forecasting Via Preprocessing’, anais da International Conference on Intelligent System Application to Power Systems, Budapest, Hungary, June 2001, pp. 118-123.
12. Otávio A.S. Carpinteiro, Alexandre P. Alves da Silva, & **Agnaldo J.R. Reis**: ‘Short Term Load Forecasting via a Hierarchical Neural Model’, anais do V Congresso Brasileiro de Redes Neurais, Rio de Janeiro, Abril de 2001, pp. 55-59.
13. **Agnaldo J.R. Reis** & Alexandre P. Alves da Silva: ‘Pré-processamento de Séries Temporais para Previsão de Carga a Curto Prazo Via Redes Neurais’, anais do XIII Congresso Brasileiro de Automática, Florianópolis, Setembro de 2000, pp. 1138-1143.

*Artigos aguardando julgamento*

- Otávio A.S. Carpinteiro, Alexandre P. Alves da Silva, & **Agnaldo J.R. Reis**, “A hierarchical self-organizing map model in short-term load forecasting”, submetido ao corpo editorial do periódico *Neural Processing Letters* em maio de 2002.
- **A.J. Rocha Reis** & A.P. Alves da Silva, “Feature Extraction via Multiresolution Analysis for Short-term Load Forecasting”, submetido ao corpo editorial do IEEE Trans. on Power Systems em outubro de 2003.

*Síntese*

Artigo PI (Publicação Internacional) – 1 artigo publicado e 2 artigos aguardando julgamento.

Artigos PN (Publicação Nacional) – 1 artigo aceito para publicação.

Artigos em conferências internacionais – 4 artigos publicados (ou aceitos).

Artigos em conferências nacionais – 7 artigos publicados (aceitos).

## **Currículo Resumido**

Nasceu em Belo Horizonte aos 14 de dezembro de 1970. Engenheiro Eletricista formado pela Pontifícia Universidade Católica de Minas Gerais (PUC-MG) em 1996. Mestre em Engenharia Elétrica pela Escola Federal de Engenharia de Itajubá em 1999. Doutor em Engenharia Elétrica pela Universidade Federal de Itajubá em 2003. Membro do *Institute of Electrical and Electronic Engineers* (IEEE), da Sociedade Brasileira de Automática (SBA) e da Sociedade Brasileira de Redes Neurais (SBRN). Suas principais áreas de interesse são: Reconhecimento de Padrões, Previsão de Séries Temporais (e.g., previsão de carga elétrica), Inteligência Artificial (redes neurais, principalmente), Wavelets e Sistemas Dinâmicos Não Lineares (incluindo caos). Agnaldo J. R. Reis tem 13 trabalhos publicados em revistas e anais de congressos nacionais e internacionais. Atuou como revisor de um artigo para a revista científica *NeuroComputing* e foi membro de Comitês Técnicos e/ou Executivos de alguns congressos nacionais (XIII CBA e V CBRN) e internacionais (ISAP 1999 e PSCC 2002) na área de sistemas inteligentes e sistemas de potência. E-mail: <[agnaldo\\_reis@yahoo.com](mailto:agnaldo_reis@yahoo.com)>