

**TESE ESCOLA FEDERAL DE ENGENHARIA DE ITAJUBÁ**  
**1016**

**DISSERTAÇÃO DE MESTRADO**

**PRÉ - PROCESSAMENTO DE SÉRIES TEMPORAIS PARA UM  
PREVISOR DE CARGA ELÉTRICA A CURTO PRAZO VIA  
REDES NEURASIS ARTIFICIAIS**

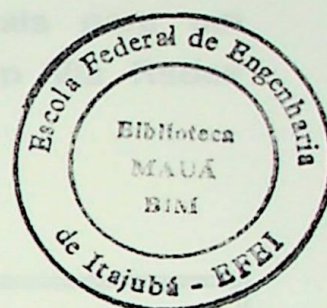
**AGNALDO JOSÉ DA ROCHA REIS**

**Itajubá, Setembro de 1999**



**ESCOLA FEDERAL DE ENGENHARIA DE ITAJUBÁ**

**DISSERTAÇÃO DE MESTRADO**



**PRÉ-PROCESSAMENTO DE SÉRIES TEMPORAIS PARA UM  
PREVISOR DE CARGA ELÉTRICA A CURTO PRAZO VIA  
REDES NEURAS ARTIFICIAIS**

**AGNALDO JOSÉ DA ROCHA REIS**

Dissertação submetida ao programa de pós-graduação em Engenharia Elétrica da Escola Federal de Engenharia de Itajubá, como requisito parcial para a obtenção do grau de Mestre em Ciências em Engenharia Elétrica

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

Itajubá, setembro de 1999



## ANEXO II

### FOLHA DE JULGAMENTO DA BANCA EXAMINADORA

Título da Dissertação: "Pré-Procesamento de Séries Temporais para um Previsor de Carga Elétrica a Curto Prazo Via Redes Neurais Artificiais"

Autor: Agnaldo José da Rocha Reis

#### JULGAMENTO

Examinadores	Conceito	Rubrica
1º	A+	Cairo Lúcio Nascimento Júnior
2º	A+	Luiz Eduardo Borges da Silva
3º	A+	Alexandre Pinto Alves da Silva

Resultado Médio: Conceito A+, ou seja, Aprovação com distinção

Observações: \_\_\_\_\_

Itajubá 02 de setembro de 1999-09-02

*Cairo Lúcio Nascimento Júnior*  
Prof. Cairo Lúcio Nascimento Júnior

1 Examinador- ITA

*Luiz Eduardo Borges da Silva*  
Prof. Luiz Eduardo Borges da Silva

2 Examinador- EFEI

*Alexandre Pinto Alves da Silva*  
Prof. Alexandre Pinto Alves da Silva

3 Examinador- EFEI - Orientador

## Agradecimentos

Agradeço ao prof. Alexandre Silva pela toda sua ajuda, confiança e orientação sempre.

A família e amigos pela ajuda.

A todos os professores e amigos que, direta ou indiretamente, contribuíram para o meu crescimento profissional.

Agradecimentos ao Conselho CAPES e CNPq pelo auxílio financeiro.

Aos meus pais e a Srta. Lucilene pelo incentivo e apoio sempre.

“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.

# Agradecimentos

Agradeço ao prof. Alexandre Silva por todo seu apoio, confiança e orientação segura.

À família GESis pela acolhida.

A todos os professores e amigos que, direta ou indiretamente, contribuíram para o meu crescimento profissional.

Às agências de fomento CAPES e CNPq pelo suporte financeiro.

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

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

# Sumário

<b>Resumo</b>	vi
<b>Abstract</b>	vii
<b>Lista de Figuras</b>	viii
<b>Lista de Tabelas</b>	x
<b>Lista de Abreviaturas</b>	xi
<b>Capítulo 1 - Previsão de Carga Elétrica a Curto Prazo.....</b>	<b>1</b>
1.1 Visão geral do problema .....	1
1.2 A Importância da PCCP .....	3
1.3 Características gerais das cargas elétricas.....	5
1.4 Classificação quanto ao tipo de modelo de carga .....	7
1.5 Técnicas de previsão .....	9
<b>Capítulo 2 - Métodos Utilizados na Previsão de Carga a Curto Prazo .....</b>	<b>11</b>
2.1 Análise de Séries Temporais.....	11
2.2 Métodos Convencionais.....	14
2.3 Métodos Não Convencionais .....	19
2.4 Redes Neurais Artificiais como solução para o problema de previsão de carga .....	20
<b>Capítulo 3 - Pré-Processamento de Séries Temporais .....</b>	<b>22</b>
Introdução .....	22
3.1 Normalização dos dados .....	22
3.2 Transformações.....	31
3.3 Centralização.....	34
3.4 Diferenciação .....	34
3.5 Desazonalização.....	38
3.6 Filtros Digitais.....	41
<b>Capítulo 4 - Testes e Resultados.....</b>	<b>58</b>
4.1 Definição da série histórica de carga .....	58
4.2 Agrupamento de dados de previsão .....	59
4.3 Identificação das variáveis explicativas do processo gerador da série histórica.....	59
4.4 Definição do número de pontos utilizados no processo de treinamento/teste .....	62
4.5 Etapa de normalização dos dados disponíveis .....	62
4.6 Partição ótima dos dados.....	62
4.7 “Expurga” de padrões atípicos.....	65
4.8 Definição da estrutura do MLP.....	65
4.9 Algoritmo de aprendizagem.....	66
4.10 Modelos de PCCP via MLP .....	67
4.11 Etapas de treinamento/teste.....	77
4.12 Tipos de previsão .....	79
4.13 Validações.....	80
4.15 Exemplo de previsões .....	95
4.16 Esforço computacional.....	98
<b>Capítulo 5 - Conclusões e Sugestões para Trabalhos Futuros .....</b>	<b>99</b>
5.1 Conclusões .....	99
5.2 Sugestões para trabalhos futuros.....	100

<b>Referências .....</b>	<b>101</b>
<b>Anexo I – Teoria sobre Redes Neurais Artificiais .....</b>	<b>105</b>
I.1 Introdução .....	105
I.2 O Neurônio Biológico.....	105
I.3 Cérebro versus Computador .....	107
I.4 Um Breve Histórico.....	108
I.5 Representação Matemática de um Neurônio .....	109
I.6 Tipos de Redes Neurais Artificiais.....	111
I.7 Perceptrons de Múltiplas Camadas.....	112
I.8 Discussão .....	126
I.9 Aplicações de RNA em Sistemas Elétricos de Potência .....	127
I.10 Principais Referências de Aplicações de RNA em SEP .....	130
I.11 Resumo .....	132
<b>Anexo II – Domínio da Freqüência.....</b>	<b>133</b>
Introdução ao Domínio da Freqüência.....	133
II.1 Análise de Fourier para sinais discretos.....	133
II.2 O Espectro de Potência.....	136
II.3 Transformada Inversa de Fourier .....	137

resposta é menor sensível a dados com ruídos do que os modelos usualmente mencionados.

Previsões baseadas em arquiteturas neurais podem fornecer boas previsões. Porém, o que se tem observado é que algumas etapas do processo de previsão não tem recebido atenção de maneira adequada. Esta dissertação apresenta uma metodologia para o pré-processamento de ST. Combinando-se ferramentas que normalmente tem o seu uso restrito (e.g., diferenciação) com filtros digitais, mostra-se á como elas podem dar uma valiosa assistência na previsão de movimentos futuros em uma série temporal.

Para testar a validade de duas séries de carga reais de uma concessionária de São de Janeiro foram utilizadas. As previsões realizadas foram do tipo *off-line*, recursiva e de 1 a 24 passos à frente. Os resultados indicam que a exatidão das previsões pode ser aumentada significativamente com a metodologia proposta. Conseqüentemente, este trabalho representa um passo à frente na determinação de um procedimento geral para o pré-processamento de ST para previsores de carga a curto prazo via RNA.

## Resumo

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 à frente 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. Portanto, fica evidente que a previsão de séries temporais se torna vital na busca por decisões corretas.

Com o passar dos anos muitos métodos para previsão de valores futuros de uma Série Temporal (ST) têm sido desenvolvidos. Muitos deles têm sólidos fundamentos teóricos e provas de competência (e.g., ARIMA, alisamento exponencial, etc.). Por outro lado, existem também aqueles baseados em Redes Neurais Artificiais (RNA) que são bastante robustos, versáteis e menos sensíveis a dados com ruídos do que os modelos anteriormente mencionados.

Previsores baseados em arquiteturas neurais podem fornecer boas previsões. Porém, o que se tem observado é que algumas etapas do processo de previsão não tem recebido atenção de maneira adequada. Essa dissertação apresenta uma metodologia para o pré-processamento de ST. Combinando-se ferramentas que normalmente tem o seu uso restrito (e.g., diferenciação) com filtros digitais, mostrar-se-á como elas podem dar uma valiosa assistência na previsão de movimentos futuros em uma série temporal.

Para testar a viabilidade desse estudo, séries de carga reais de uma concessionária do Rio de Janeiro foram utilizadas. As previsões realizadas foram do tipo *off-line*, recursiva e de 1 a 24 passos à frente. Os resultados indicam que a exatidão das previsões pode ser aumentada significativamente com a metodologia proposta. Conseqüentemente, esse trabalho representa um passo à frente na determinação de um procedimento geral para o pré-processamento de ST para previsores de carga a curto prazo via RNA.

# ***Abstract***

Some of the decisions related to the power system operation require knowledge of future load behavior. For instance, in security assessment, future knowledge of loads for the next 1-24 h can be used to check for those periods of vulnerability and to plan maintenance outages for lines, transformers and generators. Hence, it is evident that time series forecasting tasks become vital in the search for correct decisions.

Over the years, many methods for forecasting future values of Time Series (TS) have been developed. Most of them having solid theoretical foundation and proof of competence (e.g., ARIMA, exponential smoothing, etc.). On the other hand, there are those based on Artificial Neural Networks (ANN) which are quite robust, versatile and less sensitive to bad data than the traditional models previously presented.

Forecasters based on neural architectures could come up with good predictions. However, it can be noticed that some stages of the forecasting process have not received suitable attention. This dissertation describes a methodology for the preprocessing of TS. Combining many common tools that traditionally have an excessively narrow range of use (e.g., differencing) with digital filters, it will be shown how these tools can give a valuable assistance in foreseeing future movements in time series.

To test the viability of this work, real load data for one Brazilian electric utility have been used and 24-hour ahead load forecastings have been carried out. The results indicate that the accuracy can be improved with the proposed methodology. Consequently, this dissertation represents a step ahead in determining a general procedure for the preprocessing stage of TS for an ANN-based Short Term Load Forecasting.

## Lista de Figuras

Figura 2.1 Série de carga .....	12
Figura 2.2 Tendência de crescimento de carga no período de 1992-1995 .....	14
Figura 3.1 Tipos de variáveis e sua hierarquia.....	25
Figura 3.2 Probabilidade dos valores de Z excederem o intervalo $[-2; 2]$ .....	29
Figura 3.3 Exemplo de aplicação do método escala .....	31
Figura 3.4 Histogramas.....	33
Figura 3.5 Processo de diferenciação.....	35
Figura 3.6 Exemplo de séries originais e diferenciadas .....	35
Figura 3.7 Dois tipos de não estacionariedade homogênea .....	37
Figura 3.8 Processo de sazonalização .....	39
Figura 3.9 Autocorrelograma de uma série típica do outono .....	40
Figura 3.10 Autocorrelograma da série diferenciada .....	40
Figura 3.11 Autocorrelograma da série diferenciada e sazonalizada de 24 “lags” 41	
Figura 3.12 Autocorrelograma de uma série dif. e saz. de 24 e 168 “lags” .....	41
Figura 3.13 Exemplo de aplicação de filtros digitais.....	43
Figura 3.14 Três diferentes valores de $l$ .....	46
Figura 3.15 O efeito “wraparound” .....	49
Figura 3.16 FPB oriundo de um FPF .....	51
Figura 3.17 Séries filtradas de referência .....	52
Figura 3.18 Séries filtradas sem “padding” .....	53
Figura 3.19 Séries filtradas utilizando zeros como “padding” .....	54
Figura 3.20 Séries filtradas utilizando previsões como “padding” .....	55
Figura 3.21 Comparação entre as diversas séries filtradas.....	56
Figura 3.22 Filtros sintonizados.....	57
Figura 4.1 Autocorrelograma da série analisada .....	61
Figura 4.2 Estrutura típica de um MLP.....	64
Figura 4.3 Algoritmo para a implementação de $\eta$ adaptativa .....	67
Figura 4.4 Diagrama esquemático do Modelo 1 .....	68
Figura 4.5 Arquitetura neural do Modelo 1 .....	69
Figura 4.6 Diagrama esquemático do Modelo 2 .....	69
Figura 4.7 Diagrama esquemático do Modelo 3 .....	70
Figura 4.8 Arquitetura neural do Modelo 3 .....	71
Figura 4.9 Diagrama esquemático do Modelo 4 .....	72

Figura 4.10 Arquitetura neural do Modelo 4 .....	73
Figura 4.11 Diagrama esquemático do Modelo 5 .....	74
Figura 4.12 Arquitetura neural do Modelo 5 .....	74
Figura 4.13 Diagrama esquemático do Modelo 6 .....	76
Figura 4.14 Arquitetura neural do Modelo 6 .....	76
Figura 4.15 Representação da curva de carga para o período do verão .....	78
Figura 4.16 Previsão recursiva.....	79
Figura 4.17 Previsão de carga de 1 a 24 passos à frente para o dia 31/01/1995 .....	95
Figura 4.18 Erros de treinamento e teste para os Modelos 3 e 6 .....	96
Figura 4.19 Previsão de carga de 1 a 24 passos à frente para o dia 29/08/1995 .....	97
Figura 4.20 Erros de treinamento e teste para os Modelos 3 e 5 .....	98
Figura I.1 Partes constituintes de um neurônio biológico .....	106
Figura I.2 Representação Matemática de uma Unidade .....	109
Figura I.3 Três diferentes funções de ativação .....	110
Figura I.4 Estrutura de Perceptron simples .....	111
Figura I.5 Rede de Hopfield e Máquinas de Boltzman.....	112
Figura I.6 Perceptron de Múltiplas Camadas .....	112
Figura I.7 Superfície de erro característica de problemas linearmente separáveis. ....	119
Figura I.8 Superfície de erro multimodal .....	120
Figura I.9 Generalização em aprendizagem de RNA .....	122
Figura I.10 Problema proposto .....	124
Figura I.11 “Underfitting” .....	124
Figura I.12 “Overfitting” .....	124
Figura I.13 Generalização adequada.....	125
Figura I.14 Principais áreas de aplicação em abril de 1993 .....	128
Figura I.15 Modelos de RNA usados em problemas de SEP (1993) .....	128
Figura I.16 Principais áreas de aplicação em 1995 .....	129
Figura I.17 Modelos de RNA usados em problemas de SEP (1995) .....	129

## Lista de Tabelas

Tabela 2.1	Previsores de carga baseados em redes neurais.....	21
Tabela 3.1	Prob. dos valores das variáveis reduzidas excederem alguns intervalos.....	29
Tabela 4.1	$r_k$ da série analisada.....	60
Tabela 4.2	MAPE horário obtido com o Modelo 1.....	82
Tabela 4.3	MAPE diário obtido com o Modelo 1.....	82
Tabela 4.4	Índices de avaliação do Modelo 1 .....	83
Tabela 4.5	MAPE horário obtido com o Modelo 2 .....	84
Tabela 4.6	MAPE diário obtido com o Modelo 2 .....	85
Tabela 4.7	Índices de avaliação do Modelo 2.....	85
Tabela 4.8	MAPE horário obtido com o Modelo 3.....	86
Tabela 4.9	MAPE diário obtido com o Modelo 3 .....	87
Tabela 4.10	Índices de avaliação do Modelo 3.....	87
Tabela 4.11	MAPE horário obtido com o Modelo 4.....	88
Tabela 4.12	MAPE diário obtido com o Modelo 4.....	89
Tabela 4.13	Índices de avaliação do Modelo 4 .....	89
Tabela 4.14	MAPE horário obtido com o Modelo 5.....	90
Tabela 4.15	MAPE diário obtido com o Modelo 5.....	91
Tabela 4.16	Índices de avaliação do Modelo 5 .....	91
Tabela 4.17	MAPE horário obtido com o Modelo 6.....	92
Tabela 4.18	MAPE diário obtido com o Modelo 6.....	93
Tabela 4.19	Índices de avaliação do Modelo 6 .....	93
Tabela 4.20	Índices de avaliação dos modelos para o período 1 (verão).....	94
Tabela 4.21	Índices de avaliação dos modelos para o período 2 (inverno).....	94
Tabela 4.22	Índices de avaliação dos modelos para o período 3 (inverno).....	94
Tabela 4.23	Índices de avaliação dos Modelos 3 e 6 para o dia 31/01/1995.....	95
Tabela 4.24	Índices de avaliação dos modelos 3 e 5 para o dia 29/08/1995.....	97
Tabela I.1	Comparação entre o cérebro e o computador (para realização de tarefas)..	107
Tabela I.2	Comparação entre o cérebro e o computador (recursos disponíveis).....	107
Tabela I.3	RNA no ISAP'99 .....	130
Tabela II.1	Aplicação da DFT .....	135

## Lista de abreviaturas

<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>EE</b>	Energia Elétrica
<b>EESP</b>	Estimador de Estado do Sistema de Potência
<b>e.g.</b>	Por exemplo
<b>FFT</b>	Transformada Rápida de Fourier
<b>FPA</b>	Filtro Passa Alta
<b>FPB</b>	Filtro Passa Baixa
<b>FPF</b>	Filtro Passa Faixa
<b>GA</b>	Algoritmos Genéticos
<b>i.e.</b>	Isto é
<b>MAPE</b>	Erro Percentual Médio Absoluto
<b>MÁX.</b>	Erro Máximo Médio
<b>ME</b>	Erro Médio
<b>MÍN.</b>	Erro Mínimo Médio
<b>MLP</b>	Perceptrons de Múltiplas Camadas
<b>MSE</b>	Erro Médio Quadrático
<b>PCCP</b>	Previsão de Carga a Curto Prazo
<b><math>r</math></b>	Função de Autocorrelação
<b>RNA</b>	Redes Neurais Artificiais
<b>SE</b>	Sistemas Especialistas
<b>SEP</b>	Sistemas Elétricos de Potência

# Capítulo 1 - Previsão de Carga Elétrica a Curto Prazo

## 1.1 Visão geral do problema

É bem conhecido que a energia elétrica (EE) não pode ser armazenada 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 a frente. Já a previsão de carga a curto prazo (PCCP) refere-se à previsão horária da carga, i.e., das próximas 24 horas a uma semana a 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 apresentados têm um caráter orientativo, já que os mesmos variam de concessionária para concessionária, segundo sua conveniência.

Os critérios que são relevantes para a previsão em cada um dos intervalos supra mencionados variam. Por exemplo, para uma PCCP, normalmente desconsidera-se tendências a longo prazo da carga que podem ocorrer nos meses e/ou anos subsequentes. Em geral, o comportamento da carga é influenciado por fatores típicos tais como:

- influências climáticas devido a mudanças na 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, lutas, decisões de campeonatos, especiais de TV;
- fatores aleatórios desconhecidos.

Para uma previsão a médio e a longo prazo, aspectos econômicos e decisões políticas devem ser levadas em consideração. 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 à 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 telemidas 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.

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 (compatibilização da geração

com a carga, sendo que um desequilíbrio pode causar problemas de elevação ou queda de tensão). 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 (no caso de uma usina termelétrica), avaliação da segurança do sistema, planejamento para transações de energia, entre outras.

A importância da exatidão da PCCP aumentará num futuro próximo por causa das diversas mudanças na estrutura do setor eletro-energético brasileiro (desregulamentação e competição). Isso exigirá que melhores previsores sejam desenvolvidos.

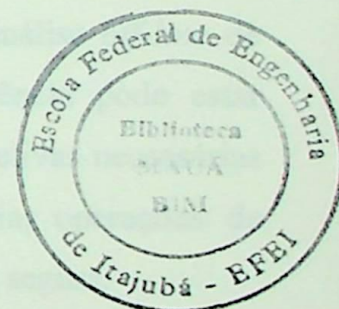
Este trabalho diz respeito a área de PCCP na operação de Sistemas Elétricos de Potência (SEP). A grandeza básica a ser prevista será a carga total do sistema, usualmente integralizada em base horária. A partir desse ponto, os outros horizontes de previsão não serão mais citados.

## 1.2 A Importância da PCCP

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

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

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



A aplicação primária da função PCCP é fornecer dados para as funções de programação que determinam o comissionamento 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 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 de forma mais segura.

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 análise da segurança tanto *on-line* quanto *off-line*.

### 1.3 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 certas tendências de consumo que podem ser estatisticamente previstas. Tais tendências refletem a influência de uma série de fatores cujo entendimento é essencial para se conseguir modelos que representem as cargas do sistema com 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.

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.

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 horazonais, celebração de contratos de suprimento com previsão de condições de interrupção, hierarquização geográfica da confiabilidade, apelo ao público, etc.

## 1.4 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.

### 1.4.1 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));*

onde 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.

### 1.4.2 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 são também 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. Atualmente, 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:

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

onde 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 cosenos 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  $\beta_i$  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 Estado. Estes modelos serão abordados no capítulo 2, seção 2.2.

## 1.5 Técnicas de previsão

Basicamente, as técnicas de previsão de carga podem ser divididas em: convencionais e não convencionais.

### 1.5.1 Técnicas convencionais de previsão

Por técnicas convencionais de previsão, entende-se aquelas técnicas baseadas em modelos estatísticos. 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 extrapolar-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).

## 1.5.2 Técnicas não convencionais de previsão

Técnicas não convencionais de previsão se apresentam como uma alternativa àquelas baseadas em modelos estatísticos. 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 (RNA) 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. Maiores detalhes sobre essas técnicas serão dados na seção 2.3.

Uma visão geral sobre alguns métodos empregados no problema de previsão de carga pode ser encontrada no capítulo 2. Veja também [Gross & Galiana, 1987].

### 2.1.1 Considerações Gerais

Uma série temporal é um conjunto de observações sequenciais no tempo acerca de um sistema. Essas observações são geralmente tomadas em intervalos de tempo regulares, o que não é obrigatório.

Uma série temporal definida pelos valores  $Y_1, Y_2, \dots, Y_n$  de uma variável  $Y$  (temperatura, valor de fechamento de uma ação, carga elétrica), nos instantes  $t_1, t_2, \dots, t_n$  será representada por  $Y = Y(t)$ .

A título de exemplo, a Figura 2.1 representa graficamente uma série temporal onde a variável  $Y$  se refere a potência real integralizada diária, partindo-se de 1<sup>o</sup> de janeiro de 1992 a 31 de dezembro de 1993. Essas medições foram realizadas por uma concessionária do Rio de Janeiro.

## Capítulo 2 - Métodos Utilizados na Previsão de Carga a Curto Prazo

Neste capítulo serão abordados alguns métodos utilizados em previsão de carga. Inicialmente, serão apresentados os métodos estatísticos, aqui denominados de métodos convencionais; em seguida, serão enfocados os métodos denominados não convencionais (e.g., sistemas especialistas, redes neurais artificiais, sistemas híbridos, etc.).

Antes de abordar tais métodos será dada uma visão geral sobre séries temporais.

### 2.1 Análise de Séries Temporais

#### 2.1.1 Considerações Gerais

Uma série temporal é um conjunto de observações seqüenciais no tempo acerca de um fenômeno. Essas observações são geralmente tomadas em intervalos de tempos equiespaçados, o que não é obrigatório.

Uma série temporal definida pelos valores  $Y_1, Y_2, \dots, Y_n$  de uma variável  $Y$  (temperatura, valor de fechamento de uma ação, carga elétrica), nos instantes  $t_1, t_2, \dots, t_n$  será representada por:  $Y = F(t)$ .

A título de exemplo, a Figura 2.1 representa graficamente uma série temporal onde a variável  $Y$  se refere a potência real integralizada diária, partindo-se de 1º de janeiro de 1992 a 31 de dezembro de 1995. Essas medições foram realizadas por uma concessionária do Rio de Janeiro.

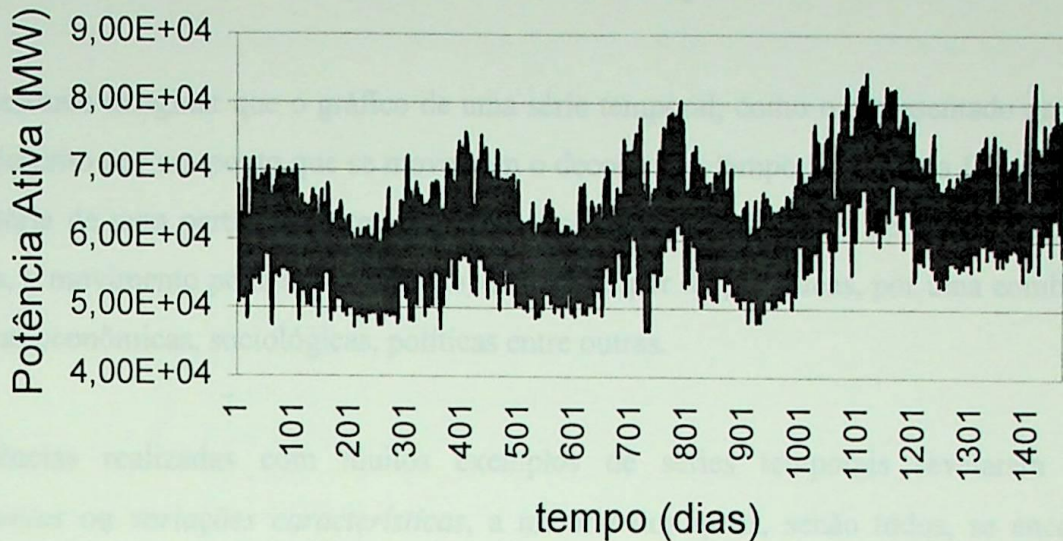


Figura 2.1 Série de carga

### 2.1.2 Séries temporais usadas em PCCP

As séries temporais usadas em previsão de carga são do tipo discretas (pois  $t$  é um conjunto finito de pontos), suas variáveis (valor medido da carga elétrica num dado instante  $t$ ) são contínuas e as observações são tomadas em intervalos de tempos equiespaçados. Na realidade, uma série temporal é a trajetória de um processo estocástico.

### 2.1.3 O que se objetiva com o estudo de séries temporais?

Via de regra, existem dois aspectos principais a serem considerados no estudo de séries temporais: a modelagem e a análise.

Com a modelagem de uma série temporal objetiva-se:

- entender o mecanismo do sistema gerador da série temporal;
- prever valores futuros.

Já a análise da série temporal visa sintetizar suas propriedades e caracterizar seu comportamento, identificando ou sugerindo um modelo adequado.

## 2.1.4 Movimentos característicos das séries temporais

É interessante imaginar que o gráfico de uma série temporal, como o representado na Figura 2.1, é descrito por um ponto que se move com o decorrer do tempo, de alguma forma análogo à trajetória de uma partícula material que se desloca sob a influência de forças físicas. No entanto, o movimento pode ser provocado, em vez de por forças físicas, por uma combinação de forças econômicas, sociológicas, políticas entre outras.

Experiências realizadas com muitos exemplos de séries temporais revelaram certos *movimentos* ou *variações características*, a maioria dos quais, senão todos, se encontram presentes nas séries temporais em graus diversos. A análise desses *movimentos* é de grande valor quando o que se deseja é a *previsão de movimentos futuros* em uma dada série temporal.

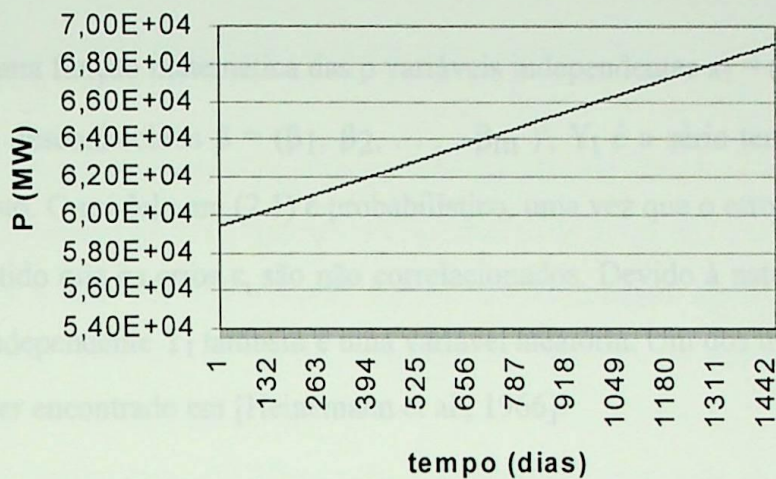
## 2.1.5 Classificação dos movimentos das séries temporais

Os movimentos característicos das séries temporais podem ser classificados em quatro tipos principais, são eles:

1. *movimentos a longo prazo, seculares ou tendência* - referem-se à direção geral, segundo a qual, parece que o gráfico da série temporal se desenvolve, em um longo intervalo de tempo. A Figura 2.2 ilustra a tendência de crescimento de carga no referido período.
2. *variações cíclicas* - referem-se às oscilações a longo prazo ou aos desvios em torno da reta ou da curva de tendência. Esses *ciclos*, podem ser ou não periódicos, i.e., podem ou não seguir exatamente padrões análogos, depois de intervalos de tempo iguais;
3. *variações por estações ou sazonalidade* - referem-se a padrões idênticos, ou quase, que uma série temporal parece obedecer durante os mesmos meses de anos sucessivos. Esses movimentos são resultantes de eventos periódicos que ocorrem anualmente<sup>1</sup>, como por exemplo, as estações do ano.

<sup>1</sup> Esses eventos periódicos, além de anuais, podem ser diários, horários, semanais, etc.

4. *movimentos irregulares ou aleatórios* - referem-se aos deslocamentos esporádicos das séries temporais, provocados por eventos casuais como enchentes, greves, eleições, etc. Esse tipo de movimento requer uma atenção especial, pois, dependendo de sua intensidade, eles podem acarretar novos movimentos cíclicos ou de outra natureza.



**Figura 2.2** Tendência de crescimento de carga no período de 1992-1995

A seguir, apresentar-se-á diversos métodos utilizados na PCCP. No entanto, será dada uma ênfase maior no método baseado em modelos conexionistas.

## 2.2 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.

### 2.2.1 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  $\beta = (\beta_1, \beta_2, \dots, \beta_m)^t$ ,  $Y_t$  é a série temporal e " $^t$ " é o operador transposto. 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 requer um número muito grande de dados para a estimação dos parâmetros do modelo. Esta 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].

## 2.2.2 Modelos de Alisamento Exponencial (MAE)

Uma grande classe de métodos de previsão é a do tipo Alisamento Exponencial (*exponential smoothing*). 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. Esta 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 ajustar a uma 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. Esta metodologia não permite considerar mais de uma série temporal.

### 2.2.3 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>2</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

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 ( $a_{t-1}, a_{t-2}, \dots$ ). 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:

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

onde:

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

é o polinômio AR simples de grau “p”;

<sup>2</sup> Representação adequada com poucos parâmetros

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

é o polinômio **MA** simples de grau “q” e  $a_t$  é ruído branco.

Uma série temporal definida conforme um processo *ARMA* é chamada de um processo estacionário. Isto significa que a média da série de qualquer um destes processos e que as covariâncias entre as suas observações não mudam com o tempo. Se o processo for não estacionário (raramente os processos encontrados na prática são estacionários), a série terá de ser transformada, idealmente, em um processo estacionário (ou mais estacionário). Isto pode ser alcançado, sob certas circunstâncias, por diferenciação (sucessiva ou não) e desazonalização. Com a diferenciação destendenciam-se o processo. Já com a desazonalização remove-se o período sazonal de interesse<sup>3</sup>.

Introduzindo-se o operador  $\nabla$ , uma série temporal diferenciada de ordem 1 pode ser escrita como  $\nabla Y_t = Y_t - Y_{t-1} = (1 - B) Y_t$ , usando-se a definição do operador de atraso **B**.

Consequentemente, uma série temporal diferenciada de ordem “d” é escrita como  $\nabla^d Y_t = (1 - B)^d Y_t$ . Assim, o modelo de Box & Jenkins toma a forma ARIMA (p, d, q) e pode ser escrito da seguinte maneira:

$$\phi(B)\nabla^d Y_t = \theta(B)a_t \quad \text{Eq.2.7}$$

sendo “d” a ordem da série diferenciada.

Para maiores informações sobre a teoria que sustenta este modelo, ver [Box & Jenkins, 1970; Abraham & Ledolter, 1983; Brockwell & Davis, 1996].

<sup>3</sup> Para maiores detalhes, ver as seções 3.4 e 3.5.

## Considerações Finais

A identificação dos parâmetros de um modelo ARMA é uma tarefa computacionalmente intensiva; entretanto, este esforço extra é necessário com a finalidade de 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 entrada 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.

## 2.3 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 chamados *Sistemas Especialistas*.

Outra metodologia que tem-se destacado bastante ultimamente é aquela baseada em Redes Neurais Artificiais (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 robustas no tratamento das correlações existentes entre as variáveis e as interdependências empíricas entre as diferentes seqüências temporais.

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

As Redes Neurais Artificiais têm demonstrado serem capazes de encontrarem representações internas de interdependências dentro de um histórico de dados disponível que não sejam explicitamente dadas ou mesmo conhecidas por um especialista humano [Cigré,1993]. Elas apresentam duas características bastante atrativas e que incentivaram vários pesquisadores a aplicarem RNA à tarefa de previsão de carga.

São elas:

- as RNA são tão versáteis que não requerem a escolha de um modelo a priori. Ao invés de termos que escolher entre um modelo 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*) ou dados que apresentem componentes caóticas do que a maioria dos métodos clássicos de previsão.

O único perigo inerente na maioria das técnicas de RNA é conhecido como *overfitting* (super ajuste ou super especialização). A elucidação e a definição precisa desse termo se encontra na seção I.7.6.

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 (T) ou em estado operativo (O) [Mori, 1996; Papalexopoulos, 1996; Khotanzad et al., 1998]. A Tabela 2.1 apresenta alguns exemplos.

**Tabela 2.1** Previsores de carga baseados em redes neurais

PREVISORES	ESTADO
Pacific Gas & Electric, USA	O
Florida International University / Florida Power Light, USA	O
Companhia Western Bohemia, República Tcheca	O
Southern Methodist University / EPRI*, USA	O
University of Washington / Puget Sound Power & Light, USA	T
Tractebel, Bélgica	T
Siemens / Austria / Byernwerke AG	T
KEPCO, Korea	T
ABB / EOS, Suíça	T
CEPCO / Fuji Electric, Japão	T
EFEI / Furnas, Brasil**	D

\* em 40 concessionárias norte-americanas e canadenses

\*\* em desenvolvimento

Para maiores informações sobre aplicações de RNA em sistemas elétricos de potência ver seção I.9. Na seqüência, apresentar-se-á as técnicas de pré-processamento de dados utilizadas nessa dissertação.

# Capítulo 3 - Pré-Processamento de Séries Temporais

## Introdução

Pré-processamento é a chave do sucesso. Qualquer modelo de previsão pode ter o seu desempenho melhorado se alguma etapa de pré-processamento for aplicada às suas variáveis de entrada, revelando claramente a informação mais importante contida nos dados [Masters,1995]. O melhor previsor é incapaz de fazer um bom trabalho caso os seus padrões de entrada se apresentem de forma complexa ou confusa. Nesse capítulo, o tema acima será discutido. Algumas abordagens clássicas e outras que tem seu uso restrito serão apresentadas.

### 3.1 Normalização dos dados

Apesar de as redes neurais artificiais serem mais robustas que a maioria das técnicas estatísticas padrões, não significa que se possa abusar delas com impunidade. Deve-se ter sempre em mente que no processo de preparação dos dados que formarão o conjunto de treinamento e teste, uma preparação adequada desses dados pode ser o diferencial entre uma rede que treina em cinco dias e apresenta bons resultados, e uma segunda que treina em cinco minutos e apresenta resultados excelentes.

Portanto, antes de iniciar o processo de treinamento de uma RNA, as seguintes questões devem ser levantadas:

- Foi fornecido ao previsor informação correta e suficiente para que ele possa realizar um bom trabalho?
- Parte da informação considerada significativa foi de alguma maneira deixada de fora, na esperança de que a rede possa inferir sua existência?
- A informação irrelevante e/ou confusa, que possa dificultar o trabalho da RNA, foi removida tanto quanto possível?

Todos esses questionamentos devem ser feitos antes de se iniciar um processo de treinamento, i.e., antes de se pagar pelo tempo de treinamento de uma rede. Por outro lado, um

conhecimento do tipo das variáveis manipuladas é fundamental para que as decisões mais acertadas sejam tomadas. A seguir, falar-se-á de quatro tipos existentes.

### 3.1.1 Tipos de variáveis

As variáveis que serão apresentadas à rede podem ser categorizadas de acordo com o tipo e o grau de informação que elas contém. Pensar sobre o tipo de cada variável em um dado experimento pode ajudar ao pesquisador a apresentá-la da maneira mais correta à rede. A seguir, as principais categorias serão definidas. Uma atenção maior será dada às ditas *variáveis de razão*, já que as séries temporais em estudo são melhor representadas por elas. Para maiores detalhes, ver [Masters, 1993; Freund, 1988].

#### Variáveis lingüísticas

A principal característica de uma variável lingüística é que ela não possui qualquer relação de ordem. Valores medidos de uma variável desta natureza nunca apresentam relações como “maior que” ou “menor que”. As únicas relações matemáticas possíveis são as do tipo igualdade e desigualdade.

Uma variável chamada “tipo de fruta”, que pode assumir os valores “maçã”, “pêra” e “banana”, é uma variável do tipo lingüística. O sexo também é um exemplo desse tipo de variável.

Por outro lado, dados numéricos também podem ser lingüísticos. O código de endereçamento postal (CEP) e o código telefônico de área para cada cidade são exemplos clássicos.

#### Variáveis de ordem

Variáveis medidas em uma escala ordinal apresentam relações de ordem. Os valores que podem ser assumidos por uma variável de ordem só poderão ser ordenados de forma única. As relações “maior que” e “menor que” têm algum significado, mas os valores numéricos reais de uma variável de ordem não exprimem qualquer outra informação que seja além da sua própria ordem.

A título de exemplo, suponha que uma variável de ordem tenha sido medida para três assuntos. Os valores medidos foram 3, 4 e 100 respectivamente. Daí, pode-se afirmar que o segundo assunto medido é maior que o primeiro nesta variável, e que o terceiro assunto é maior que os demais. Porém, não se pode dizer que o terceiro assunto é maior que o segundo por um grau maior que o segundo é maior que o primeiro. O ato de se fazer isso implicaria na necessidade de se usar outro tipo de variável que não a de ordem, provavelmente uma de nível superior.

### Variáveis de intervalo

As variáveis de intervalo são aquelas cuja informação é a segunda de mais alto nível. Essas variáveis compartilham a propriedade de relação de ordem com as variáveis de ordem.

Como exemplo, considere três variáveis de intervalo  $A$ ,  $B$  e  $C$ , e suponha que elas obedeçam a seguinte relação de ordem  $A < B < C$ . No entanto, considerando-se esse tipo de variável, pode-se dar um passo a mais e, e.g., dizer que  $(C - B) > (B - A)$ . E ainda, se  $A = 3$ ,  $B = 4$  e  $C = 100$ , pode-se afirmar com toda segurança que  $C$  é muito maior que  $B$ , enquanto que  $B$  é ligeiramente maior que  $A$ , o que não é possível de se dizer quando as variáveis são do tipo de ordem.

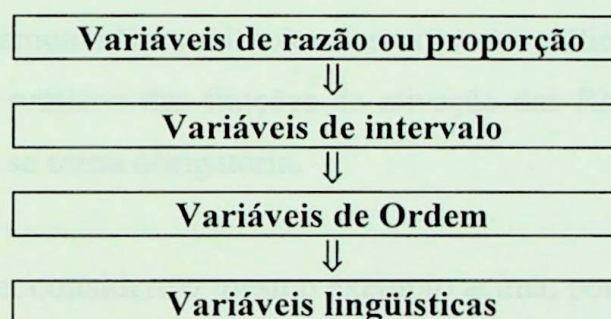
### Variáveis de razão ou proporção

Variáveis desse tipo são ditas “*as de mais alto nível ou de nível superior*”, devido ao tipo de informação que representam. São idênticas às variáveis de intervalo, porém o valor *zero* de uma variável de intervalo é *arbitrário*. Já o valor *zero* de uma variável de razão é *informativo*. Para explicar melhor essa diferença, considere os exemplos a seguir.

As escalas de temperatura Celsius e Fahrenheit são do tipo intervalo, pois o zero é arbitrariamente definido. Por outro lado, a escala Kelvin é do tipo razão. Zero graus Kelvin é o *zero absoluto*; não existe movimento molecular naquela temperatura. Um outro exemplo é a medida da quantidade de potência ativa injetada em um determinado barramento. Zero Watts de potência injetada, realmente significa *zero*, i.e., inexistência de potência útil.

Portanto, só tem sentido falar em proporção quando as variáveis são do tipo razão. Um objeto cuja massa é de 100 gramas, apresenta duas vezes mais massa que um outro objeto de massa 50 gramas. Da mesma forma, um barramento que apresente uma potência de 10 MW, apresenta a metade da potência de um outro de 20 MW. Entretanto, não se pode dizer que a sensação provocada por uma temperatura de 40° C é duas vezes maior que uma outra provocada por uma temperatura de 20° C; isso não procede.

A Figura 3.1 permite uma visualização dos tipos de variáveis anteriormente descritos, seguindo sua ordem hierárquica, i.e., partindo-se da variável de mais alto nível ou superior (variáveis de razão) até a de mais baixo nível (variáveis lingüísticas). É conveniente frisar que essa hierarquia é determinada de acordo com o tipo de informação que cada variável representa.



**Figura 3.1** Tipos de variáveis e sua hierarquia

### 3.1.2 Método Escala

#### Considerações gerais

Como toda a preparação dos dados está sendo feita para ser usada em uma arquitetura neural treinada para realizar previsões, há a necessidade de que os dados que formarão os conjuntos de teste e de treinamento sejam de alguma forma normalizados para serem **compatíveis** com os limites da função de ativação do modelo neural empregado.

A seguir, discutir-se-á se o “Método Escala” é realmente necessário e quais seriam os reais ganhos ao utilizá-lo.

### O processo de adaptação dos dados originais é sempre necessário?

Uma breve resposta para esta pergunta seria **não**. No entanto, ele é bastante recomendável, senão obrigatório em muitos casos.

Em uma primeira análise, a título de exemplo, considere que uma variável está sendo usada para treinar um MLP e que ela seja a medida da quantidade de potência real consumida numa dada hora do dia. Considere ainda que a camada de saída dessa rede seja composta por neurônios do tipo não-linear. Caso esses neurônios sejam do tipo **logsig**<sup>1</sup>, i.e., suas funções de ativação sejam do tipo sigmoidal<sup>2</sup>, suas saídas podem variar de 0 a 1, teoricamente. Na prática, esses valores variam entre 0,1 e 0,9.

Como via de regra, a ordem de grandeza das variáveis de entrada é de  $10^6$ , deve-se, no mínimo, fazer um mapeamento linear simples dos extremos práticos das variáveis de entrada e saída para os extremos práticos das funções de ativação das RNA. Nesse caso, não se tem escolha, a normalização se torna obrigatória.

Em uma segunda análise, considere o mesmo exemplo acima, porém com camada de saída do tipo **linear**. Tais redes não apresentam limites de saída, de modo que, teoricamente, não é necessário a normalização das variáveis de entrada. Mesmo assim, a normalização ainda é recomendada, já que as RNA treinam de maneira mais eficiente se suas entradas e saídas estão compreendidas em uma faixa “razoável”.

### Justificando a normalização

Uma das razões para usar-se dados normalizados é para inicialmente igualar a importância das variáveis. Teoricamente, deve ser possível para uma rede aprender a importância relativa das variáveis através do ajuste dos pesos das entradas. Se uma variável tem uma ordem de grandeza de  $10^6$  enquanto que outra  $10^{-6}$ , a rede deve ser capaz de aprender a usar minúsculos pesos de entrada para a primeira variável e imensos pesos para a segunda. Mas

<sup>1</sup> Notação MATLAB para um neurônio com função de ativação do tipo sigmoidal

<sup>2</sup>  $Sigmoidal(x) = \frac{1}{1 + e^{-x}}$  (ver item I.5.1)

isso é pedir demais para qualquer algoritmo de treinamento. Na verdade, muitos deles impõem limites fixos nos seus pesos para evitar que o processo de aprendizagem da rede fique comprometido. Esses algoritmos simplesmente não são capazes de aprender tais valores extremos. Logo, pode-se tornar a vida da rede mais fácil, fornecendo-lhe dados normalizados segundo algum critério, de modo que todos os pesos possam ficar em faixas pequenas, previsíveis.

Uma outra razão, tão ou mais importante que a primeira, se dá quando os dados estão sendo aprendidos pelos neurônios de saída. A maioria dos algoritmos de treinamento minimizam o erro total de todas as saídas. Se as variáveis de saída estão desigualmente normalizadas, aquelas com maior variabilidade serão favorecidas, de modo que elas dominarão o somatório dos erros. Isso pode ter sérias conseqüências e é um problema que deve ser sempre considerado.

Uma terceira razão é que com o processo de normalização pode-se melhorar a interpretabilidade dos pesos da rede. Após os processos de treinamento e validação, pode-se querer visualizar todos os pesos aprendidos<sup>3</sup> na esperança de que eles tenham algum significado, i.e., de que eles dêem uma indicação de como a rede toma suas decisões; o que sob as melhores condições ainda é uma tarefa muito difícil.

Após ter-se justificado o uso do processo de normalização dos dados originais, apresentar-se-á duas metodologias utilizadas para tanto.

### **Teoria sobre o Método Escala**

Foi mencionado anteriormente que o processo de normalização dos dados seria um mapeamento linear simples dos extremos práticos das variáveis que formam a base de dados para os extremos práticos das funções de ativação de um MLP.

Um método de normalização bastante popular é baseado nos valores máximos e mínimos da série de carga de interesse. Esse método é matematicamente representado por:

<sup>3</sup> e.g. diagrama de Hinton

$$V_{Normalizado(k)} = C \cdot \left( \frac{V_{MW(k)} - V_{min}}{V_{max} - V_{min}} \right) \quad \text{Eq.3.1}$$

onde,

$V_{MW(k)}$  é o valor da potência ativa em MW no instante k;

$C$  é o fator escala;

$V_{min}$  é o valor mínimo da série de carga;

$V_{max}$  é o valor máximo da série de carga.

No entanto, se a base de dados tem uma distribuição aproximadamente normal, unimodal, quase simetricamente distribuída sobre sua média aritmética e nunca apresenta variáveis com valores extremamente distantes de sua média, pode-se querer empregar uma normalização mais sofisticada baseada na sua média e no seu desvio padrão amostrais. A normalização baseada nessa metodologia tem um apelo mais teórico e mesmo mais prático.

Para tanto, considere a variável:  $Z = \frac{X - \mu}{\sigma}$  Eq.3.2

onde:

$X$  é a variável a ser normalizada;

$\mu$  é a média aritmética ( $\mu = \frac{1}{n} \sum_{i=1}^n x_n$ );

$\sigma$  é o desvio padrão ( $\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2}$ ).



A variável  $Z$  que mede o desvio em relação à média, em unidades de desvio padrão, é denominada *variável reduzida*<sup>4</sup> e é uma quantidade abstrata, pois independe das unidades usadas.

Isso pode ser visualizado com o seguinte exemplo. Dois pesquisadores trabalhando em países distintos conduzem essencialmente o mesmo experimento. Um deles mede a quantidade de massa de um determinado objeto em quilogramas. O outro faz a mesma medição, porém usa libra como unidade. Imediatamente, pode-se concluir que essas medições não são diretamente

<sup>4</sup> A variável  $Z$  também é conhecida como *unidades reduzidas* ou *escores reduzidos*

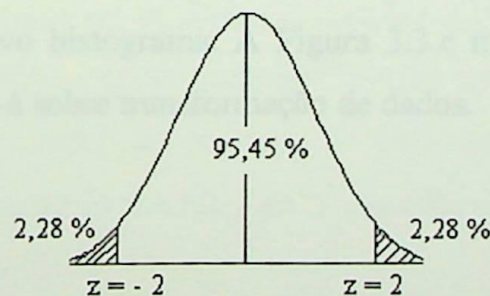
comparáveis. Usando-se um processo de normalização via variável reduzida, essas medições se tornam comparáveis.

Simplesmente normalizar os dados originais via variáveis reduzidas geralmente não é suficiente. Esses novos valores normalizados ainda poderiam exceder os limites práticos de um MLP. Para solucionar esse problema, considere que os valores máximos e mínimos<sup>5</sup> esperados da variável  $Z$  sejam designados por  $V_{\max}$  e  $V_{\min}$ , e que os limites práticos das funções de ativação do MLP sejam  $A_{\max}$  e  $A_{\min}$ . Para uma rede unidirecional, com função de ativação do tipo sigmoideal, esses limites de ativação de saída seriam 0,1 e 0,9, respectivamente.

A Tabela 3.1 mostra a probabilidade  $P$  dos valores das variáveis reduzidas excederem alguns intervalos. A Figura 3.2 é uma ilustração da 2ª linha dessa tabela.

**Tabela 3.1** Probabilidade dos valores das variáveis reduzidas excederem alguns intervalos

Intervalo de $Z$	Prob.exceder (%)
$-1 < Z < 1$	31,73
$-2 < Z < 2$	4,55
$-3 < Z < 3$	0,27



**Figura 3.2** Probabilidade dos valores de  $Z$  excederem o intervalo  $[-2; 2]$

<sup>5</sup> Esses valores nada têm a ver com aqueles apresentados na Eq.3.1

Para a determinação de  $V_{\max}$  e  $V_{\min}$ , o que se tem a fazer é escolher um valor de probabilidade que seja adequado ao problema tratado e então fixar  $V_{\max}$  como sendo o limite superior do intervalo de  $Z$  correspondente e  $V_{\min}$  como o limite inferior.

Portanto, considere que:

$$r = \frac{A_{\max} - A_{\min}}{V_{\max} - V_{\min}} \quad \text{Eq.3.3}$$

$$\begin{aligned} V_{\text{Normalizado}(k)} &= r \cdot \left( \frac{X(k) - \mu}{\sigma} - V_{\min} \right) + A_{\min} \\ &= \frac{r}{\sigma} X(k) + \left( A_{\min} - r \left( \frac{\mu}{\sigma} + V_{\min} \right) \right) \end{aligned} \quad \text{Eq.3.4}$$

Invertendo-se a Eq.3.4, obtém-se a variável na sua escala original.

$$X(k) = \frac{\sigma}{r} V_{\text{Normalizado}(k)} + \left( \mu + \sigma \left( V_{\min} - \frac{A_{\min}}{r} \right) \right) \quad \text{Eq.3.5}$$

Essa seção termina com um exemplo. As figuras 3.3.a e 3.3.b mostram uma série em sua escala original e seu respectivo histograma. A Figura 3.3.c mostra a série normalizada via Eq.3.4. Na seqüência, falar-se-á sobre transformação de dados.

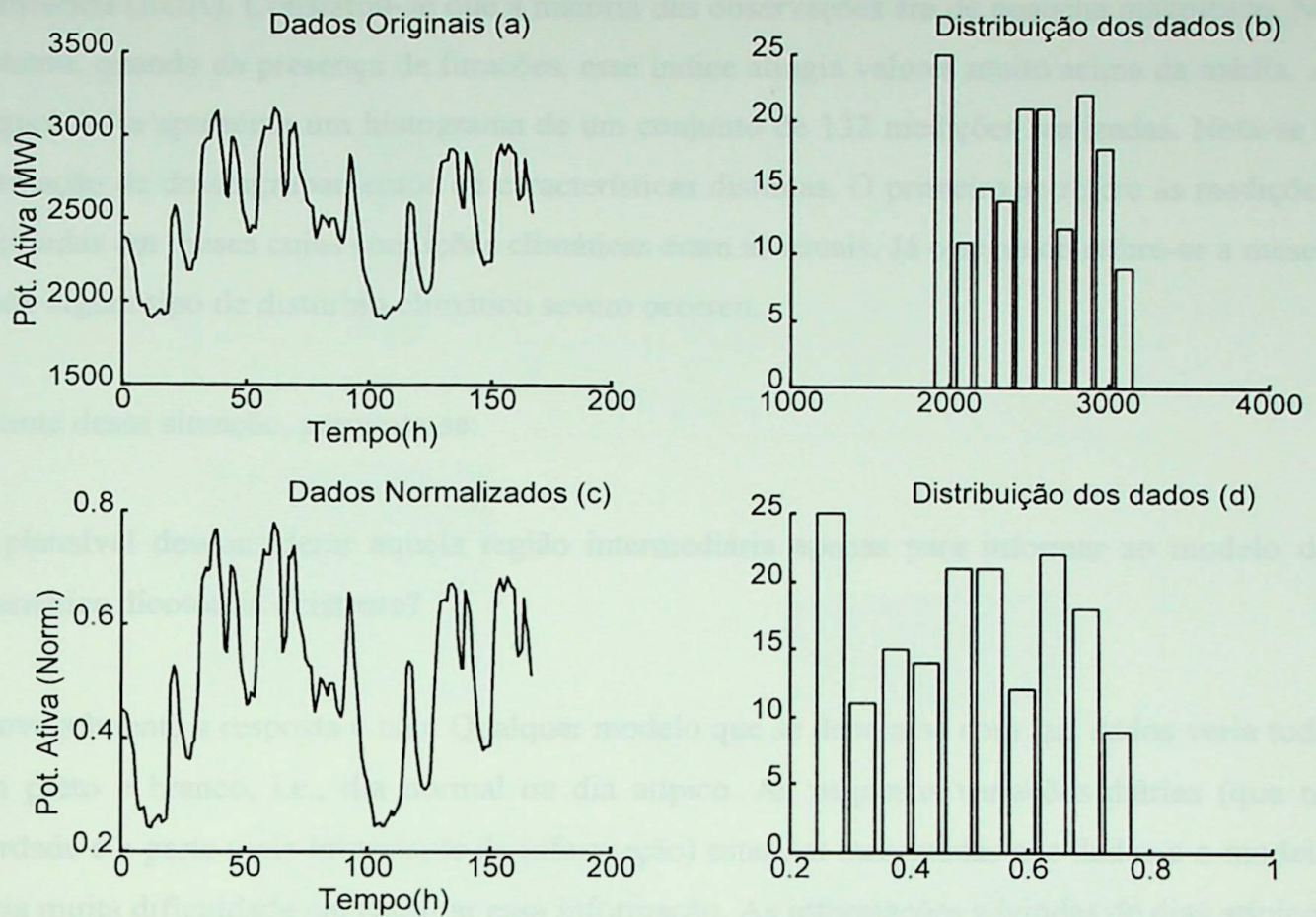


Figura 3.3 Exemplo de aplicação do método escala

## 3.2 Transformações

Em qualquer tarefa de processamento de dados, os números que apresentam as informações sobre o estado de um processo devem assumir valores que sejam *compatíveis* com o significado prático do estado atual do mesmo. O conceito de uma variável refletir a informação sobre o seu estado, de forma clara para o modelo de previsão adotado, é crucial para se obter bons resultados.

Com o intuito de esclarecer a declaração anterior, considere o exemplo hipotético a seguir. Suponha que um grupo de pesquisa esteja desenvolvendo um previsor de carga para uma dada concessionária de energia elétrica norte americana. Durante a etapa de levantamento de quais variáveis seriam as mais relevantes para o previsor, verificou-se que o índice pluviométrico seria uma delas. Com isso, foram realizadas medições mensais de precipitação no estado da

Califórnia (EUA). Constatou-se que a maioria das observações era de pequena magnitude. No entanto, quando da presença de furacões, esse índice atingia valores muito acima da média. A Figura 3.4.a apresenta um histograma de um conjunto de 132 medições realizadas. Nota-se a formação de dois agrupamentos de características distintas. O primeiro se refere às medições efetuadas em meses cujas condições climáticas eram as usuais. Já o segundo refere-se a meses onde algum tipo de distúrbio climático severo ocorreu.

Diante desse situação, pergunta-se:

É plausível desconsiderar aquela região intermediária apenas para informar ao modelo da dramática dicotomia existente?

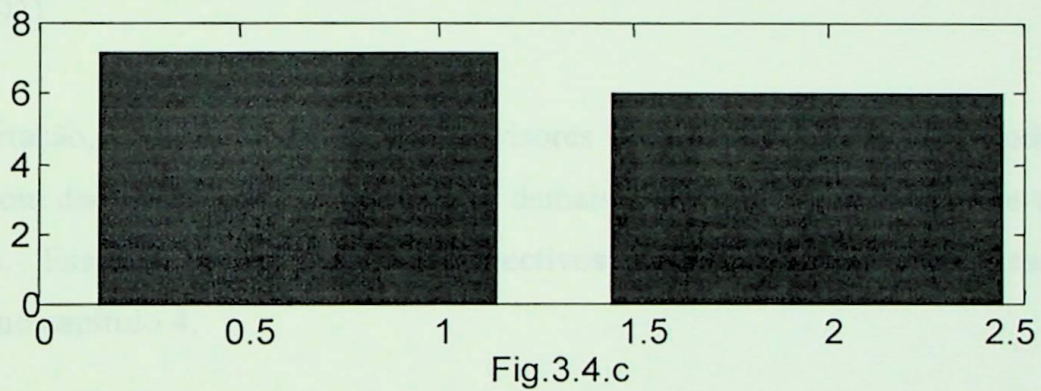
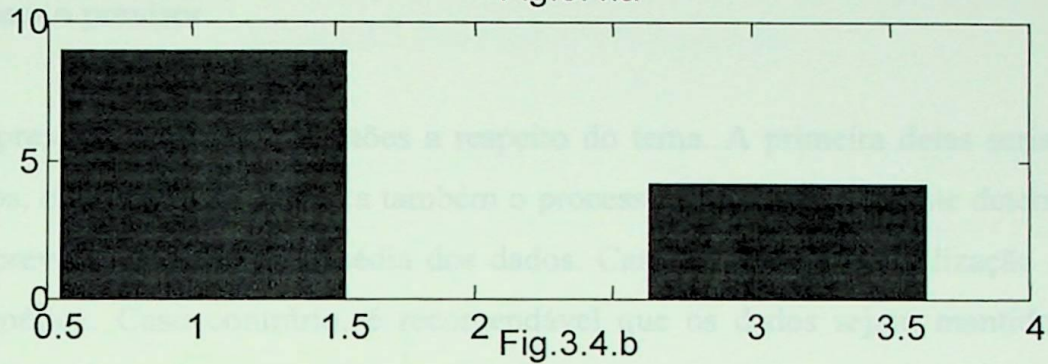
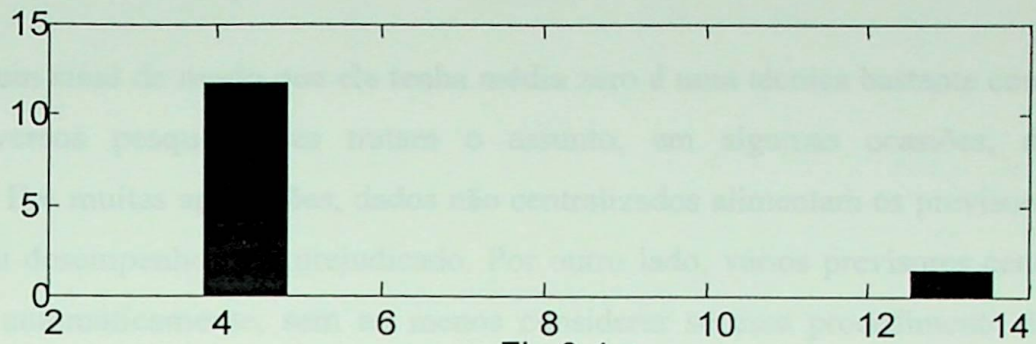
Provavelmente a resposta é não. Qualquer modelo que se deparasse com tais dados veria tudo em preto e branco, i.e., dia normal ou dia atípico. As pequenas variações diárias (que na verdade é a parte mais importante da informação) estariam mascaradas nos dados e o modelo teria muita dificuldade em capturar essa informação. As informações advindas de dias atípicos são importantes. No entanto, elas não podem se destacar tanto quanto as primeiras.

Em tais situações é sempre aconselhável usar uma função de mapeamento do tipo compressora em todos os dados disponíveis para mover as observações extremas em direção à maioria. Isso pode ser visualizado nas Figura 3.4.b e 3.4.c. Essas figuras mostram histogramas cujos dados foram mapeados via uma função do tipo raiz quadrada<sup>6</sup> e raiz cúbica<sup>7</sup>, respectivamente.

---

<sup>6</sup>  $V_{Transformado(k)} = X_{(k)}^{1/2}$

<sup>7</sup>  $V_{Transformado(k)} = X_{(k)}^{1/3}$



**Figura 3.4** Histogramas

Em linhas gerais, o primeiro princípio de transformação é que algumas decisões são necessariamente arbitrárias em diversas situações. Isso é devido ao fato de que, ocasionalmente, baseado em algum conhecimento prévio acerca do processo, um dado tipo de transformação pode apresentar uma propriedade matemática desejada e portanto justificar a sua escolha.

Por outro lado, o que geralmente se faz na prática é examinar o histograma dos dados e tomar uma decisão intuitiva baseada nessa examinação visual. Caso seja conveniente aplicar algum tipo de transformação, um novo exame nos dados transformados é obrigatório. Maiores detalhes sobre o assunto ver [ Acton,1959; Kendal, 1976].

A seguir, falar-se-á do processo de centralização dos dados.

### 3.3 Centralização

Centralizar um sinal de modo que ele tenha média zero é uma técnica bastante conhecida. No entanto, diversos pesquisadores tratam o assunto, em algumas ocasiões, de maneira equivocada. Em muitas aplicações, dados não centralizados alimentam os previsores, fazendo com que seu desempenho seja prejudicado. Por outro lado, vários previsores centralizam os seus dados automaticamente, sem ao menos considerar se esse procedimento é realmente necessário para o previsor.

Esse item apresenta algumas sugestões a respeito do tema. A primeira delas seria: conheça bem os dados, e se possível, conheça também o processo que os gerou. Tente determinar se o modelo de previsão é sensível à média dos dados. Caso ele seja, a centralização dos dados pode ser benéfica. Caso contrário, é recomendável que os dados sejam mantidos intactos [Masters, 1993].

Nessa dissertação, vários modelos de previsores foram concebidos. O modelo 1 foi alimentado com dados não centralizados. Já os demais modelos foram alimentados com dados centralizados. Esses modelos e seus respectivos desempenhos serão apresentados e comparados no capítulo 4.

### 3.4 Diferenciação

Em algumas ocasiões, pode-se querer trabalhar com os dados essencialmente do modo que são coletados. Em outras, não. Além das técnicas de pré-processamento de séries temporais anteriormente discutidas, uma outra conceitualmente simples e útil, é conhecida por *diferenciação de pontos adjacentes* ou simplesmente, *diferenciação*.

O processo de se diferenciar uma série consiste em se computar a diferença entre sucessivos pontos. Assim, ao invés de se trabalhar com os dados originais, trabalha-se com suas “variações”. A Figura 3.5 ilustra essa técnica, tomando como exemplo a série mostrada na Figura 3.6.a (série típica de verão). A Figura 3.6.b é a série anterior, diferenciada. A Figura 3.6.c é uma série típica de outono. Suas “variações” são exibidas na Figura 3.6.d.

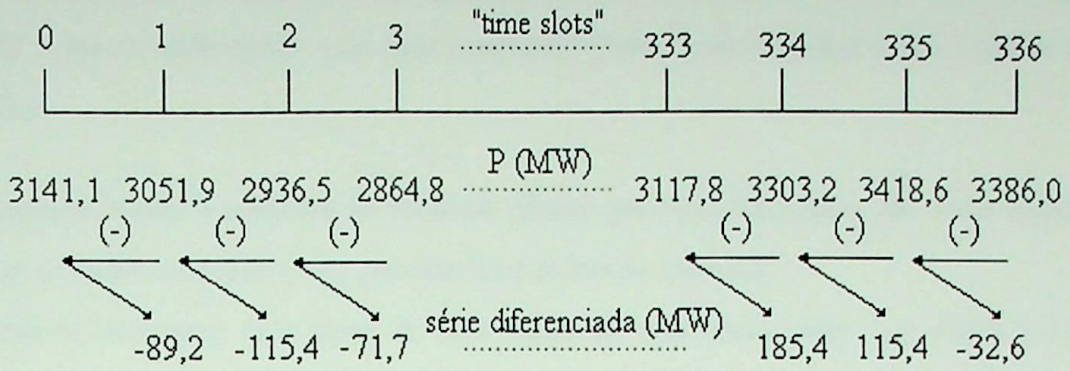


Figura 3.5 Processo de diferenciação

Para se computar a primeira diferença, subtrai-se o valor do “slot n°1” pelo valor do “slot n°0”, e assim, sucessivamente. Notar que a série diferenciada tem um ponto a menos que a série original.

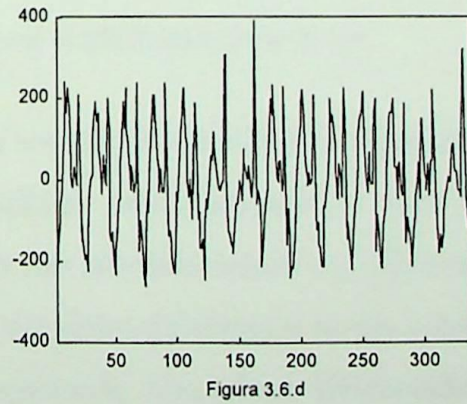
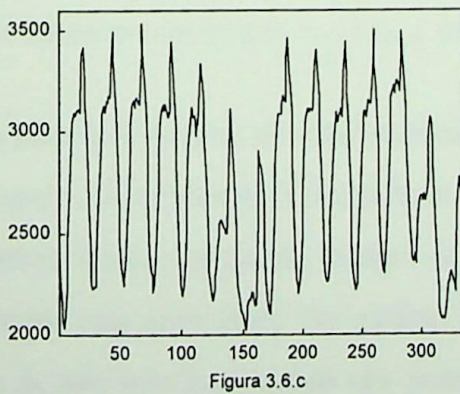
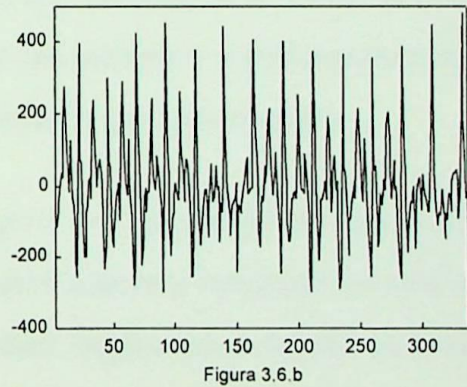
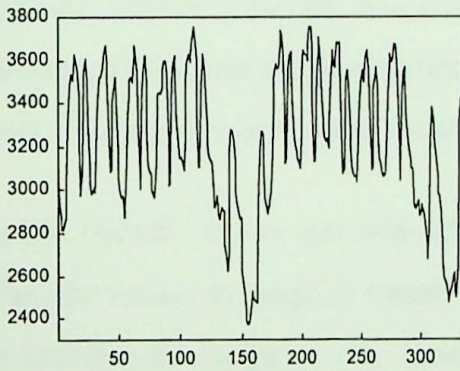


Figura 3.6 Exemplo de séries originais e diferenciadas

Após apresentar-se o tópico diferenciação, pode-se questionar: Quais motivos levariam alguém a querer diferenciar uma série temporal? Umas poucas razões estão listadas abaixo. São elas:

- Algumas vezes a natureza do processo afirma que suas diferenças são mais importantes que os dados originais (e.g., previsão para bolsas de valores);
- Pode-se conseguir dois tipos de informação de uma única série. Um deles relativo às recentes mudanças ocorridas; e outro, devido aos próprios valores da série original;
- O desempenho de uma RNA pode ser melhorado, caso ela seja alimentada com ambas séries: original e diferenciada<sup>8</sup>;
- Pode-se acreditar que informações de alta frequência em uma série sejam tão importantes quanto informações de baixa frequência. A diferenciação é uma maneira eficaz e facilmente reversível de se aplicar um filtro passa alta a uma série temporal.

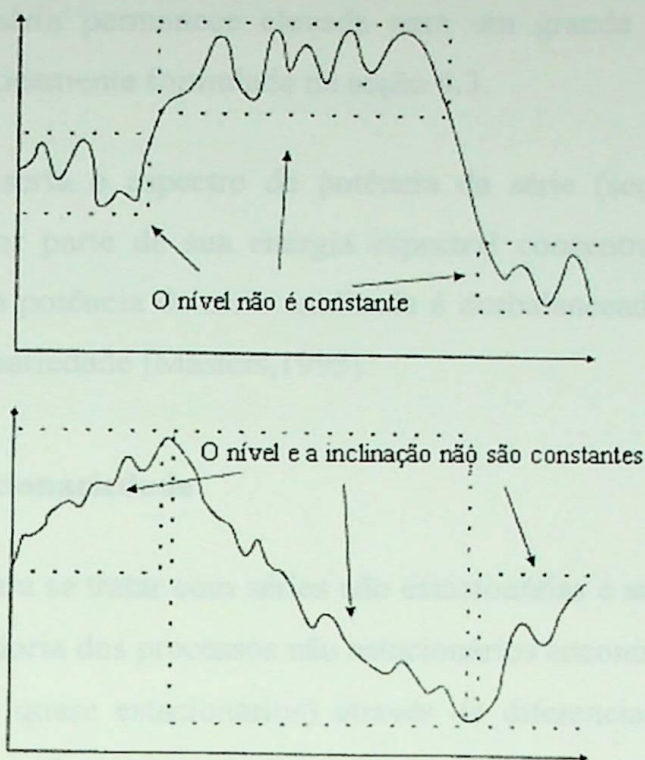
## Estacionariedade

Intuitivamente, pode-se dizer que uma série temporal é estacionária se o seu comportamento não muda com o passar do tempo. Isso implica que não somente o comportamento dos pontos individuais é analisado, mas também o comportamento de um conjunto deles.

Estritamente falando, diz-se que uma série temporal é *estacionária* se suas distribuições estatísticas não variam ao longo do tempo. Mais especificamente, considere um conjunto de  $m$  amostras tomadas nos tempos  $t_1$  a  $t_m$ . Estacionariedade implica que a função de densidade de probabilidade composta daquelas  $m$  amostras é idêntica à distribuição conjunta de outras  $m$  amostras tomadas nos tempos  $t_{1+k}$  a  $t_{m+k}$ . Isso deve ser verdade para todo  $m$  e  $k$ .

A razão pela qual trata-se de estacionariedade em uma seção sobre diferenciação deve-se ao fato de que a diferenciação é um método de indução de estacionariedade em uma série não estacionária. Existe um grande número de tipos de não estacionariedade e a diferenciação só funciona em uma parte deles. No entanto, muitas das séries encontradas no dia a dia exibem um tipo de não estacionariedade que pode ser considerada *homogênea*. Desconsiderando-se mudanças ocasionais em nível (ou até mesmo em nível e inclinação), essas séries exibem geralmente um comportamento padrão ao longo do tempo (Figura 3.7).

<sup>8</sup> Esse resultado foi obtido de forma empírica



**Figura 3.7** Dois tipos de não estacionariedade homogênea

### Identificando séries não estacionárias

Não existe uma fórmula única para decidir se uma série é não estacionária, e se sua não estacionariedade é do tipo homogênea. No entanto, os procedimentos listados abaixo podem servir como linhas gerais na tarefa de identificação de séries não estacionárias.

O primeiro passo é examinar visualmente a série de interesse. Pode ser difícil para o olho humano detectar mudanças sutis em correlações seriais. Por outro lado, mudanças de nível e inclinação são óbvias. Mudanças nítidas de nível e/ou inclinação (como aquelas apresentadas na Figura 3.7) sugerem que a diferenciação deve ser usada.

Uma ferramenta útil para detectar-se comportamento não estacionário de séries temporais é a *função de autocorrelação*. A autocorrelação ( $r$ ) é, por si só, uma série cujos elementos são as correlações entre o valor corrente e valores passados da série original. O primeiro ponto no autocorrelograma é a correlação entre o ponto corrente e o ponto anterior. O segundo é a correlação entre o ponto corrente e dois pontos anteriores e assim, por diante. A autocorrelação de uma série quase não estacionária cai a zero rapidamente, enquanto que a de

uma série não estacionária permanece elevada para um grande número de “lags”. Essa ferramenta está matematicamente formulada na seção 4.3.

Uma outra ferramenta seria o espectro de potência da série (seção II.2). Uma série não estacionária terá a maior parte de sua energia espectral concentrada em frequências mais baixas. Se o espectro de potência da série analisada é desbalanceado dessa maneira, deve-se suspeitar de não estacionariedade [Masters,1995].

### **Removendo a não estacionariedade**

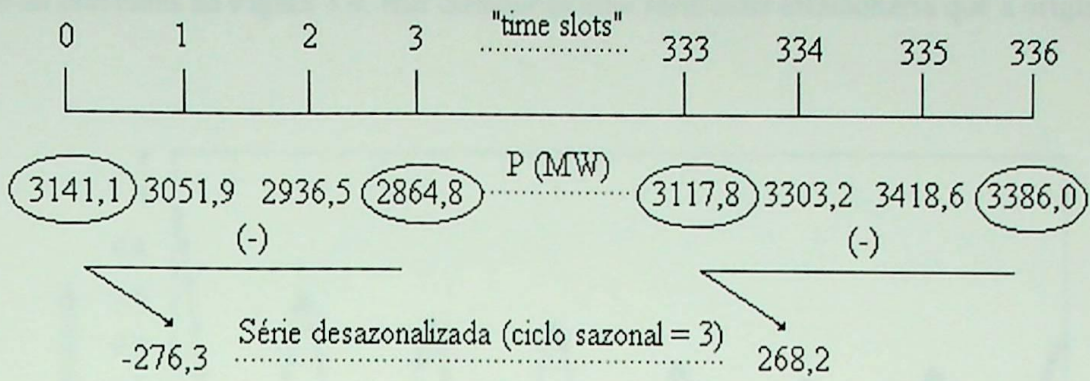
O método mais direto para se tratar com séries não estacionárias é através da diferenciação de pontos adjacentes. A maioria dos processos não estacionários encontrados na prática podem se tornar estacionários (ou quase estacionários) através de diferenciação simples e única. No entanto, algumas séries podem requerer dois estágios de diferenciação, o que não é muito comum.

### **Considerações Finais sobre diferenciação e estacionariedade**

Concluir-se-á esse tópico com uma advertência. Nunca inadvertidamente descarte informação. Isso é especialmente válido quando RNA estão envolvidas. Séries não estacionárias podem impedir a habilidade de uma RNA generalizar bem. Isso implica que tais sinais não devem ser usados como as únicas entradas da rede. Diferenças, mesmo para séries estacionárias, podem ser imensamente úteis.

## **3.5 Desazonalização**

Um segundo tipo de diferenciação é a *diferenciação por pontos não adjacentes* ou *diferenciação sazonal*. Nesse tipo, uma nova série é computada como a mudança de uma amostra em relação a uma distância fixa no passado. A Figura 3.8 ilustra essa técnica. Nesse exemplo, inicialmente calcula-se a diferença entre o “slot” n° 3 e o n° 0. Em seguida, desloca-se para o “slot” imediatamente posterior (n.º 4) e uma nova diferença é calculada (em relação ao n.º1). E assim, sucessivamente. Notar que a série diferenciada tem um número de pontos igual ao número de pontos da série original subtraído do tamanho do ciclo sazonal (no caso, 3 pontos).



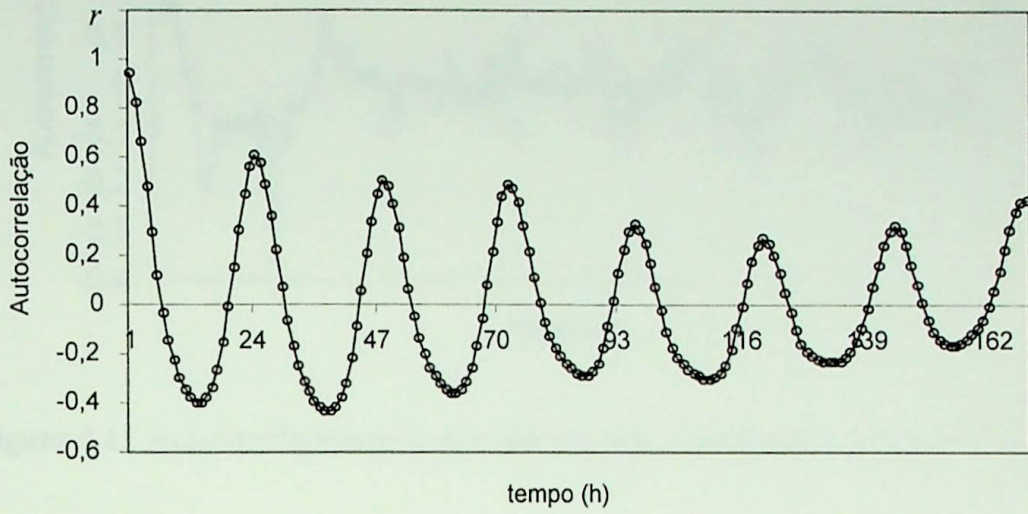
**Figura 3.8** Processo de desazonalização

Apresentado o conceito de desazonalização, uma questão similar àquela formulada para a diferenciação se faz presente: Por que desazonalizar uma série? A resposta está relacionada com o nome da técnica: *sazonalidade*. Esse tipo de diferenciação é útil quando a série de interesse exibe algum tipo de comportamento periódico, i.e., quando pode-se constatar padrões que se repetem em intervalos de tempos iguais. A título de exemplo, considere as figuras 3.6.a e 3.6.c. Em ambas, pode-se constatar padrões característicos que se repetem a cada 24 horas. Daí, diz-se que essa série exibe uma sazonalidade diária. Caso se faça um exame mais cuidadoso, ainda constatar-se-á uma sazonalidade semanal.

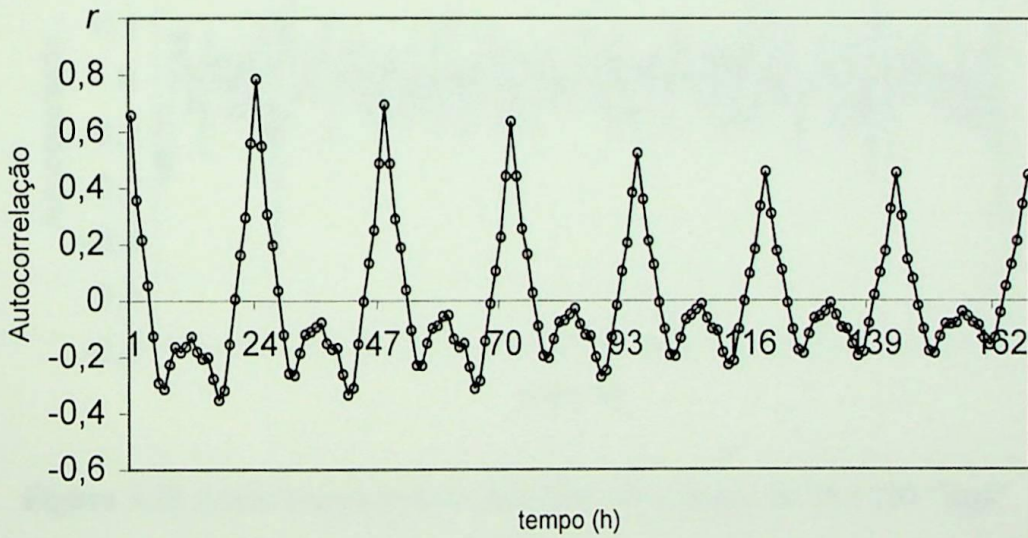
Em muitos casos, com uma simples inspeção visual, é possível identificar-se a presença de ciclos sazonais em uma série. Em outros, não. Novamente, utilizar-se-á o autocorrelograma da série, só que desta vez, para detectar-se ciclos sazonais. Um componente sazonal apresentará valores de autocorrelação elevados em múltiplos inteiros do comprimento do ciclo sazonal.

A Figura 3.9 mostra o autocorrelograma da série exibida na Figura 3.6.c. Duas características são marcantes: a série é não estacionária e contém, no mínimo, uma forte componente periódica. A Figura 3.10 apresenta o autocorrelograma da série diferenciada exibida na Figura 3.6.d. Constata-se elevadas autocorrelações nos instantes 24, 48, 72, ..., 168, ratificando o que foi obtido com uma inspeção visual dos dados. Já a Figura 3.11 exibe o autocorrelograma da série diferenciada e desazonalizada (ciclo sazonal = 24). Apesar de o ciclo diário ter sido removido, o ciclo semanal ainda está presente (autocorrelação do "lag" 168 elevada). A Figura 3.12 exibe o autocorrelograma da série diferenciada e desazonalizada para ambos ciclos

sazonais: diário e semanal. Nessa figura, nota-se a ausência de autocorrelações significativas como as presentes na Figura 3.9. Isso caracteriza uma série mais estacionária que a original.



**Figura 3.9** Autocorrelograma de uma série típica do outono



**Figura 3.10** Autocorrelograma da série diferenciada

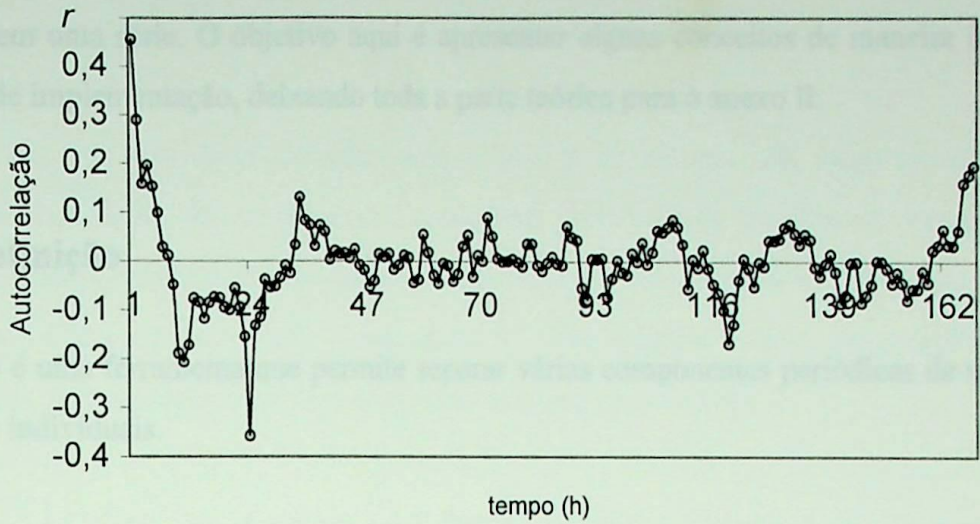


Figura 3.11 Autocorrelograma da série diferenciada e desazonalizada de 24 “lags”

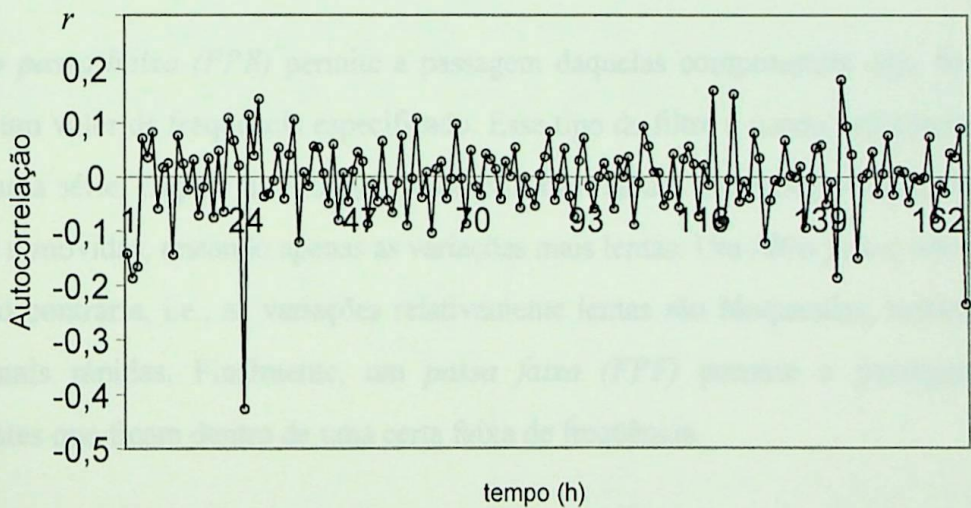


Figura 3.12 Autocorrelograma de uma série dif. e desaz. de 24 e 168 “lags”

### 3.6 Filtros Digitais

As séries temporais que se observa na prática são um amálgama de diferentes componentes individuais de informação. Parte dessa informação pode ser útil para responder certas questões, enquanto que outras não.

Essa seção discute brevemente o uso de filtros digitais como separadores de informações contidas em uma série. O objetivo aqui é apresentar alguns conceitos de maneira intuitiva e detalhes de implementação, deixando toda a parte teórica para o anexo II.

### 3.6.1 Definição

Um filtro é uma ferramenta que permite separar várias componentes periódicas de uma série, em séries individuais.

Existem três tipos básicos de filtros, são eles:

- Filtros passa baixa;
- Filtros passa alta;
- Filtros passa faixa.

Um *filtro passa baixa (FPB)* permite a passagem daquelas componentes cuja frequência é inferior a um valor de frequência especificado. Esse tipo de filtro é usado principalmente para suavizar uma série. Depois da filtragem, as oscilações rápidas presentes na série de interesse terão sido removidas, restando apenas as variações mais lentas. Um *filtro passa alta (FPA)* faz a operação contrária, i.e., as variações relativamente lentas são bloqueadas, restando apenas aquelas mais rápidas. Finalmente, um *passa faixa (FPF)* permite a passagem aquelas componentes que ficam dentro de uma certa faixa de frequência.

Existem basicamente três razões para se filtrar uma série antes de se treinar uma RNA. São elas:

- Fontes de ruído podem ser *reduzidas*;
- Importantes fontes de informação podem ser *ênfatizadas*;
- Uma série, que é um aglomerado de fontes de informação, pode ser particionada em duas ou mais partes, facilitando o trabalho da rede.

### Exemplo

Muitas vezes a informação em uma série surge de diferentes processos, sendo que cada uma se apresenta como uma componente individual que varia em uma taxa característica ou em

uma faixa de taxas. Quando essas fontes de informação separadas são combinadas em uma única série, suas contribuições individuais se perdem no somatório.

Suponha que em um dado instante, uma componente assuma um valor excessivamente elevado e, além disso, não usual; expressando uma parte vital da informação. Por outro lado, uma outra componente no mesmo instante, apresenta um valor extremamente baixo e também não usual. Caso se considere unicamente a série original, essas informações ficarão mascaradas, pois naquele instante o somatório será próximo de zero.

A Figura 3.13 ilustra um exemplo de aplicação de filtros digitais. A série original não é particularmente interessante. No entanto, a sua versão filtrada por um FPB revela claramente uma lenta variação que pode ser uma informação bastante útil para a rede (a série filtrada exprime o nível geral da série). Da mesma forma, a versão filtrada por um FPA exibe as oscilações rápidas; a ausência de tendências é uma característica marcante dessa série. Fornecendo à rede as três séries, seu trabalho será bastante simplificado. Apesar de existirem mais entradas para o modelo manusear, o conteúdo da informação de entrada está mais bem definido.

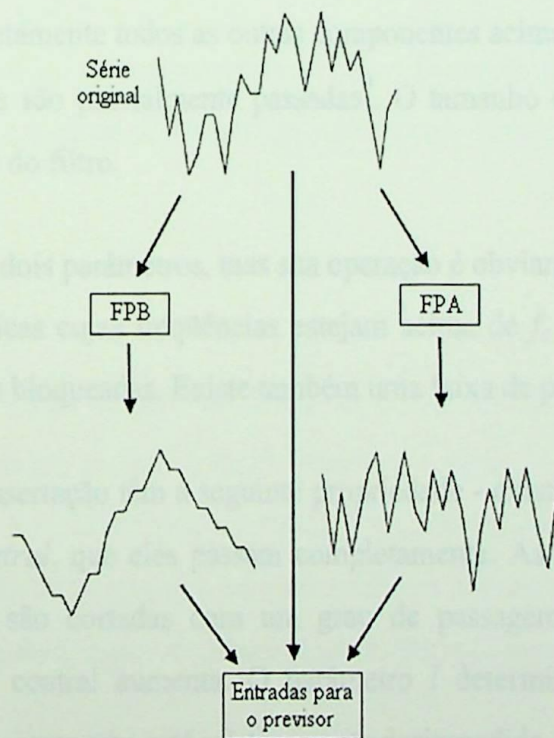


Figura 3.13 Exemplo de aplicação de filtros digitais

### 3.6.2 Filtrando no domínio da frequência

Em se tratando de filtros, o que se objetiva é enfatizar ou bloquear certas frequências. Então, porque não passar a série para o domínio da frequência, aplicar a função filtro nesse domínio, e reconstruir a série temporal a partir da série modificada? Essa abordagem faz sentido, já que na grande maioria das vezes se define um filtro em termos de frequência.

Antes de se apresentar o algoritmo utilizado para a construção dos filtros, definir-se-á os parâmetros frequência e largura.

### 3.6.3 Parâmetros dos Filtros

Independentemente do tipo de filtro (FPB, FPA ou FPF), os filtros são caracterizados por dois parâmetros: *frequência de corte* ( $f_c$ ) e *largura* ( $l$ ). Para um FPB, as componentes periódicas cujas frequências forem inferiores a  $f_c$  terão sua passagem desimpedida. A *largura* é a faixa de transição sobre a qual a resposta do filtro vai de um extremo (passagem permitida) ao outro (bloqueio). Na prática, é impossível projetar-se um filtro que tenha uma transição brusca, i.e., um filtro que passe completamente todos as componentes periódicas abaixo de um certo limiar e bloqueie completamente todos as outras componentes acima dele. Existe sempre uma faixa de frequências que são parcialmente passadas<sup>9</sup>. O tamanho dessa faixa é crucial para avaliar-se o desempenho do filtro.

Os FPA têm os mesmos dois parâmetros, mas sua operação é obviamente contrária a dos FPB. As componentes periódicas cujas frequências estejam acima de  $f_c$  serão passadas, enquanto que aquelas abaixo serão bloqueadas. Existe também uma faixa de passagem parcial.

Os FPF usados nessa dissertação têm a seguinte propriedade - existe uma frequência simples, chamada *frequência central*, que eles passam completamente. As frequências em ambos os lados dessa frequência são cortadas com um grau de passagem que decresce quando a distância da frequência central aumenta. O parâmetro  $l$  determina a largura da faixa de frequência sobre a qual a passagem está relativamente desimpedida.

Essa seria a abordagem tradicional. Contudo, existe uma abordagem alternativa que é frequentemente mais apropriada. A inspiração para tal metodologia será discutida a seguir.

Em muitas situações é desejável assumir que todo tipo de informação na série original é importante. Nenhuma parte dela deve ser inadvertidamente descartada. No entanto, a aplicação de um FPB, um FPA, e um ou mais FPF (sintonizados em diferentes frequências), podem descartar parte da informação. Na melhor das hipóteses, a informação total contida na saída dos filtros representará de maneira desequilibrada a informação original; enfatizando algumas frequências e mascarando outras.

A solução para esse problema é utilizar-se filtros *lineares*. O exemplo a seguir apresenta os benefícios dessa propriedade.

Suponha que aplicou-se um FPB e um FPA a uma série. Para garantir que toda a informação na série original seja capturada, necessita-se de aplicar um FPF que seja exatamente o complemento dos dois primeiros filtros. A computação direta e aplicação desse FPF seria relativamente difícil do ponto de vista prático. Porém, a linearidade permite que esse objetivo seja alcançado de uma maneira bastante simples e eficaz. Para tanto, *subtrai-se as séries filtradas pelos filtros passa baixa e passa alta da série original*. O resultado desse somatório é exatamente o complemento dos filtros FPB e FPA. Com isso, o particionamento da informação é perfeito.

Como uma regra geral, essa é a abordagem preferida. A filtragem direta através de um FPF é apropriada em situações nas quais se conhece de antemão que uma certa faixa de frequências é de interesse especial.

### Frequência

O parâmetro *frequência* está sempre na faixa de 0,0 a 0,5. No entanto, o ponto mais importante é que o parâmetro recíproco da frequência é o *período da componente periódica*. Por exemplo, considere a série exibida na Figura 3.6.c. As medições de potência ativa são tomadas de hora em hora. Suponha que se queira manter apenas aquelas componentes cujo período **exceda** 24 horas. Para tanto, se utilizaria um FPB cuja frequência de corte seria:

$$f_c = \frac{1}{24} = 0,04167 \text{ ciclos por amostra.}$$

<sup>9</sup> Essa faixa de passagem é conhecida como “gray zone”

Por outro lado, caso se quisesse manter apenas aquelas componentes cujo período fosse inferior a 4 horas, se usaria um FPA cuja frequência de corte seria:

$$f_c = \frac{1}{4} = 0,2500 \text{ ciclos por amostra.}$$

Finalmente, caso se quisesse enfatizar uma componente de período 12 horas, eliminando-se tendências mais lentas e oscilações mais rápidas, se utilizaria um FPF tendo uma frequência central igual a:

$$f_{central} = \frac{1}{12} = 0,08333 \text{ ciclos por amostra.}$$

### Largura

O parâmetro  $l$  define a largura da faixa de passagem e, via de regra, é mais difícil de se especificar. Ele é uma escolha arbitrária. No entanto, a Figura 3.14 pode ser bastante útil para sua determinação. Ela mostra o efeito de três valores diferentes de  $l$  no comportamento de um FPF cuja frequência central vale 0,25 ciclos por amostra.

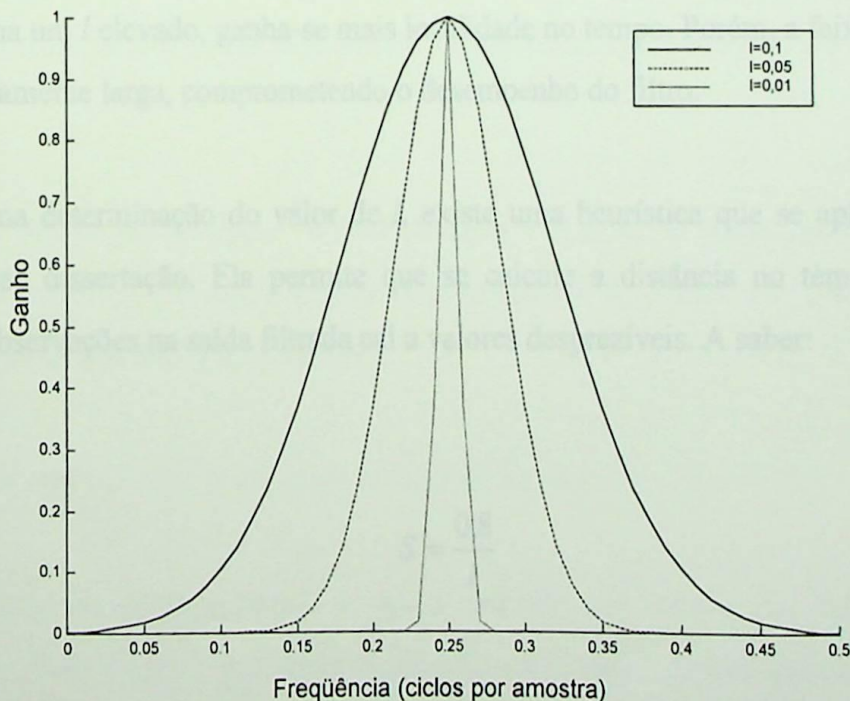


Figura 3.14 Três diferentes valores de  $l$

A menor largura, i.e.,  $l = 0,01$ , produz uma rápida transição. Se essa largura fosse usada para um FPB ou FPA, a faixa de passagem parcial seria bem estreita. A transição de passagem completa para atenuação completa seria quase instantânea. Já a maior largura, quando  $l = 0,1$ , produz uma região de transição bastante larga. A transição da condição de passagem total para bloqueio total seria lenta e gradual. Existirá um grande número de componentes periódicas na região de passagem parcial.

Em uma primeira análise se escolheria o menor  $l$  possível, pois uma transição rápida da passagem completa das componentes de interesse para o bloqueio completo seria desejável. Contudo, há um elevado preço a pagar por transições rápidas – *transições mais bruscas implicam em menos localização no tempo*. O procedimento a seguir explica melhor essa afirmação.

Para filtrar uma série temporal, grupos de observações são examinados. Para calcular, e.g., a 150ª saída filtrada, o filtro pode utilizar os valores originais da série que vão do instante 100 ao 200. Isso é um fato inevitável no processo de filtragem e, nesse caso, é bastante inconveniente. Se as observações fossem tomadas mensalmente, um valor observado para algum mês teria um impacto nas saídas filtradas 50 meses (mais de 4 anos) antes e depois daquela observação. Um pouco mais de localidade no tempo é aconselhável.

Caso se escolha um  $l$  elevado, ganha-se mais localidade no tempo. Porém, a faixa de transição será demasiadamente larga, comprometendo o desempenho do filtro.

Para auxiliar na determinação do valor de  $l$ , existe uma heurística que se aplica aos filtros utilizados nessa dissertação. Ela permite que se calcule a distância no tempo, na qual o impacto das observações na saída filtrada cai a valores desprezíveis. A saber:

$$S = \frac{0,8}{l} \quad \text{Eq.3.6}$$

onde:

$l$  é a largura do filtro;

$S$  é o número da observação mais distante que afeta cada saída filtrada.

Considerando-se  $l = 0,01$  e aplicando-se na Eq.3.6, tem-se que:

$$S = \frac{0,8}{0,01} = 80 \text{ observações, i.e., cada saída filtrada é afetada por valores da série original até}$$

80 observações anteriores e posteriores.

Caso se considere  $l = 0,1$ , tem-se que:

$$S = \frac{0,8}{0,1} = 8 \text{ observações.}$$



Em suma, a escolha de  $l$  pode ser feita a partir da Figura 3.14. Essa figura mostra, para cada uma das três larguras, o grau de atenuação em qualquer distância de uma freqüência específica. Ressalta-se que  $l$ , tipicamente, pode assumir qualquer valor na faixa de 0,01 a 0,5. Definido  $l$ , verifica-se a questão da localidade no tempo através da Eq.3.6.

### 3.6.4 Algoritmo para filtragem digital

O método de filtragem digital anteriormente mencionado, será adotado na implementação de cada um dos filtros usados nessa dissertação. O algoritmo apresentado a seguir dá maiores detalhes.

1. *Adiciona-se pontos no começo e no final da série temporal, de modo que a mesma tenha um número de amostras que seja potência de 2.* O número de pontos e a “qualidade” desses pontos extras serão discutidos no item 3.6.5;

2. *Calcula-se a DFT da série modificada através da Eq.II.1;*

3. *Multiplica-se cada termo  $w_i$  obtido no passo anterior, por valores que definam o tipo de filtro desejado.* Isso implica em fazer alguns iguais a zero, deixar outros inalterados, e suavizar os demais;

4. Utiliza-se a Eq.II.3 para transformar para o domínio do tempo, a série filtrada no domínio da frequência;

5. Descarta-se os pontos extras que foram anexados no primeiro passo.

Esse algoritmo é rápido e eficaz. O único aspecto que varia de acordo com o tipo de filtro a ser implementado está no passo 3. A escolha dos multiplicadores (que definem o filtro) para os termos  $w_i$  serão apresentados no item 3.6.6. A seguir, falar-se-á da operação de anexação de pontos extras<sup>10</sup> citada no passo 1.

### 3.6.5 Anexando pontos extras

O motivo pelo qual se torna imprescindível anexar dados extras no começo e no final da série é devido ao fato de que ao se computar os termos da DFT, a convolução circular é inerente ao processo. Seu efeito é que ao se calcular os últimos pontos filtrados, os dados do começo da série são incluídos nesse cálculo, o que gera uma certa distorção na série filtrada. Um efeito similar também ocorre quando se calcula os primeiros pontos filtrados. A única diferença é que dessa vez os pontos do final da série é que são incluídos no cálculo. Esse efeito é conhecido como “wraparound” e é ilustrado na Figura 3.15.

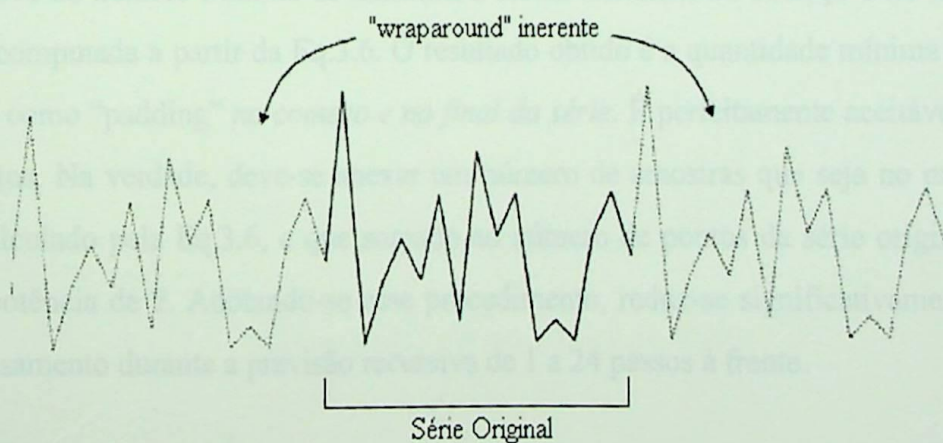


Figura 3.15 O efeito “wraparound”

O “wraparound” inerente no processo de filtragem é inevitável. Em se tratando de previsão de carga, e caso o grau da distorção seja elevado, alguma medida deve ser tomada para minimizar

o problema de distorção. A solução mais plausível seria anexar dados extras no começo e no final da série. Esses dados devem ser o mais inócuos possíveis.

Uma primeira abordagem seria anexar um certo número de zeros no final da série. Uma segunda seria calcular uma “série média” e anexá-la no fim da série. Essa série sintética “média” seria composta, e.g., pelos valores médios horários das últimas duas semanas. A segunda abordagem é superior à primeira, mas não se mostrou adequada para os propósitos dessa dissertação.

Uma terceira abordagem, que foi proposta e utilizada nesse trabalho, se mostrou superior às anteriores. No começo da série anexou-se medições imediatamente anteriores ao primeiro ponto da série de interesse. Esses dados são conhecidos, pois compõem o histórico de carga da concessionária de EE. Com isso, consegue-se eliminar a distorção que estaria presente nos primeiros pontos filtrados, caso nenhum tipo de “padding” fosse utilizado. Já no final da mesma, anexou-se valores previstos por um previsor neural, minimizando-se a distorção presente nos últimos pontos filtrados da série de carga.

O “padding” é sempre necessário. No entanto, até o momento nada se falou a respeito do comprimento desse “padding”, i.e., de quantos pontos devem ser anexados no começo e no final da série de carga. Para aqueles filtros que são empregados nesse trabalho, uma estimativa conservativa do número mínimo de amostras a serem anexadas no começo e no final da série pode ser computada a partir da Eq.3.6. O resultado obtido é a quantidade mínima de pontos a ser usado como “padding” *no começo e no final da série*. É perfeitamente aceitável utilizar-se mais pontos. Na verdade, deve-se anexar um número de amostras que seja no mínimo igual àquele calculado pela Eq.3.6, e que somado ao número de pontos da série original, seja um número potência de 2. Adotando-se esse procedimento, reduz-se significativamente o tempo de processamento durante a previsão recursiva de 1 a 24 passos à frente.

### 3.6.6 Definindo as funções filtro

As funções que multiplicarão a DFT da série de carga no domínio da frequência são mostradas na Eq.3.7.

---

<sup>10</sup> Essa operação é conhecida como “padding”.

Exemplo 1

$$H(f) = e^{-\frac{(f-f_0)^2}{l}} \tag{Eq.3.7}$$

Essa função atinge seu valor máximo (unitário) em  $f=f_0$ , e cai em direção a zero em ambos os lados da frequência central. A taxa de queda é determinada por  $l$ . Grandes valores de  $l$  implicam em uma queda mais lenta. O efeito de três diferentes valores de  $l$ , com  $f_0$  fixado em 0,25, foi mostrado na Figura 3.14.

Porém, a Eq.3.7 é válida somente quando se tratar de filtros do tipo FPF. Quando se tratar de FPBs ou FPAs, uma modificação deve ser feita. Para se criar um FPB a partir de um FPF, descarta-se toda a “cauda” à esquerda da frequência central, substituindo-se todos os pontos descartados pelo valor 1. De maneira similar, para se criar um FPA, substitui-se todos os pontos à direita da frequência central por 1. A Figura 3.16 mostra essa técnica para um FPB. A linha pontilhada representa os pontos descartados. A linha cheia é a resposta do FPB resultante.

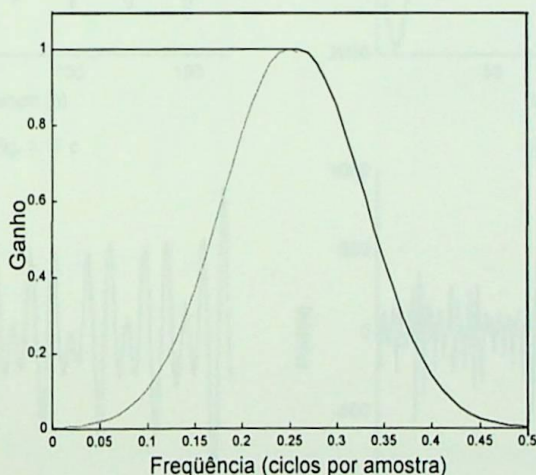


Figura 3.16 FPB oriundo de um FPF

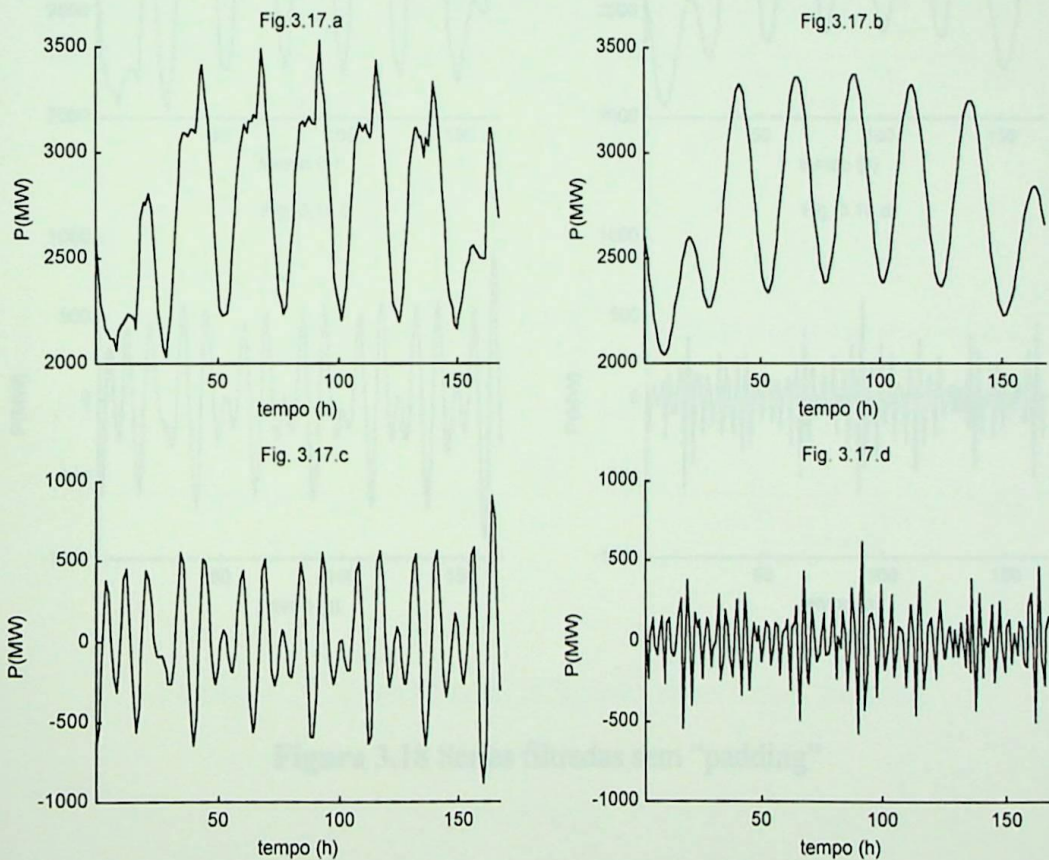
Exemplo de Filtragem

Esse capítulo será terminado com um exemplo de aplicação dos filtros estudados. Os exemplos a seguir têm como objetivo mostrar a forma típica de saída de cada um dos filtros apresentados e o efeito do “padding” nas séries filtradas.

### Exemplo 1

A série de interesse é exibida na Figura 3.17.a. As Figuras 3.17.b, 3.17.c, 3.17.d, são as séries filtradas por um FPB ( $f=1/24$ ;  $l=0,025$ ), um FPF ( $f=1/8$ ;  $l=0,03$ ) e um FPA ( $f=1/4$ ;  $l=0,03$ ), respectivamente (Figura 3.22). Anexou-se os dados verdadeiros no começo e no final da série.

Esse primeiro exemplo é apenas didático. Ele mostra as saídas filtradas que se gostaria de obter. Nenhum tipo de distorção está presente nas séries filtradas. Consequentemente, essas séries servirão de referência para se avaliar a qualidade dos dados que estão sendo anexados.



**Figura 3.17** Séries filtradas de referência

## Exemplo 2

No segundo exemplo (Figura 3.18) manteve-se a mesma série de carga e os parâmetros anteriores. No entanto, nenhum tipo de “padding” foi utilizado. Uma análise cuidadosa das séries mostra a distorção em relação às séries de referência.

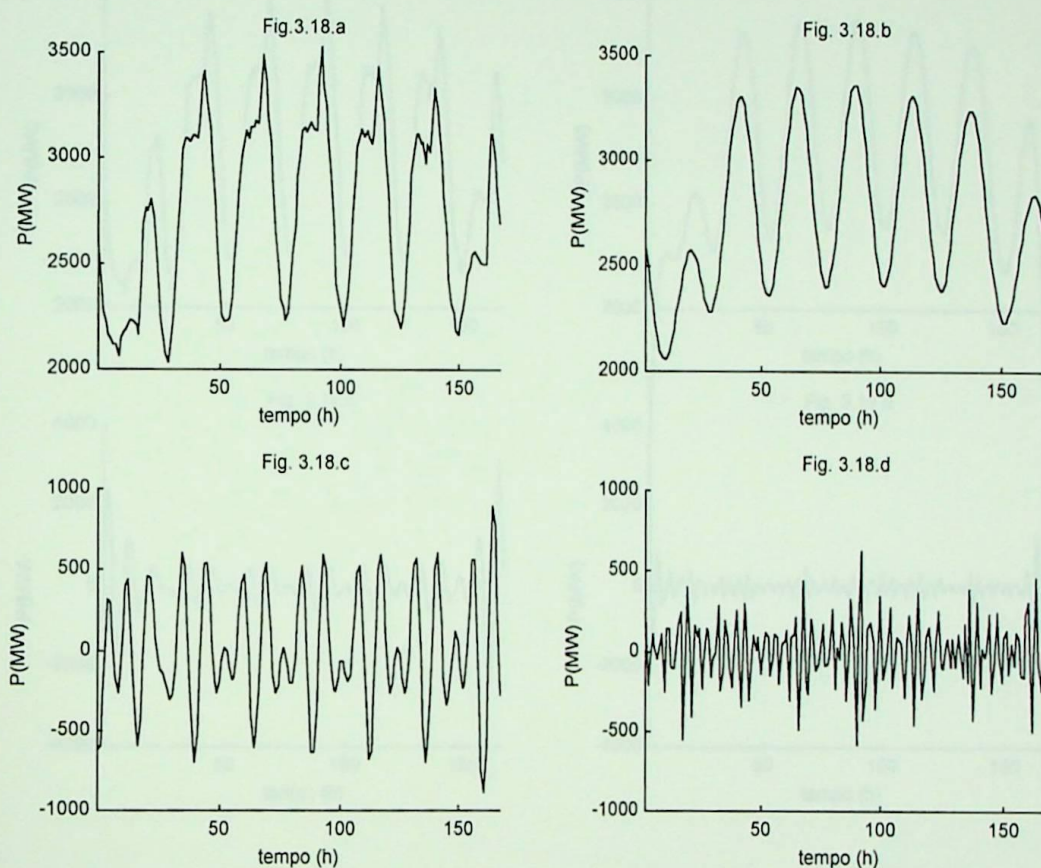


Figura 3.18 Séries filtradas sem “padding”

### Exemplo 3

No terceiro exemplo (Figura 3.19) substituiu-se os dados verdadeiros anexados no primeiro exemplo por zeros no começo e no final da série de interesse. Nota-se que o nível de distorção aumentou significativamente nos extremos.

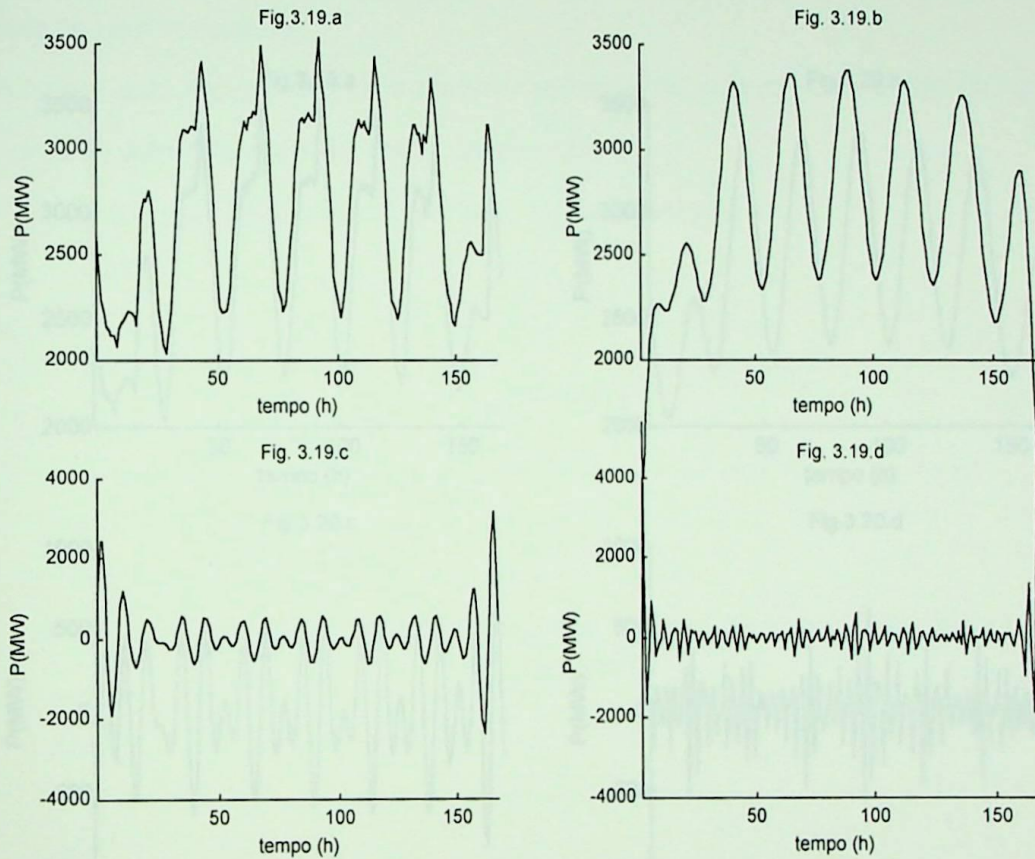
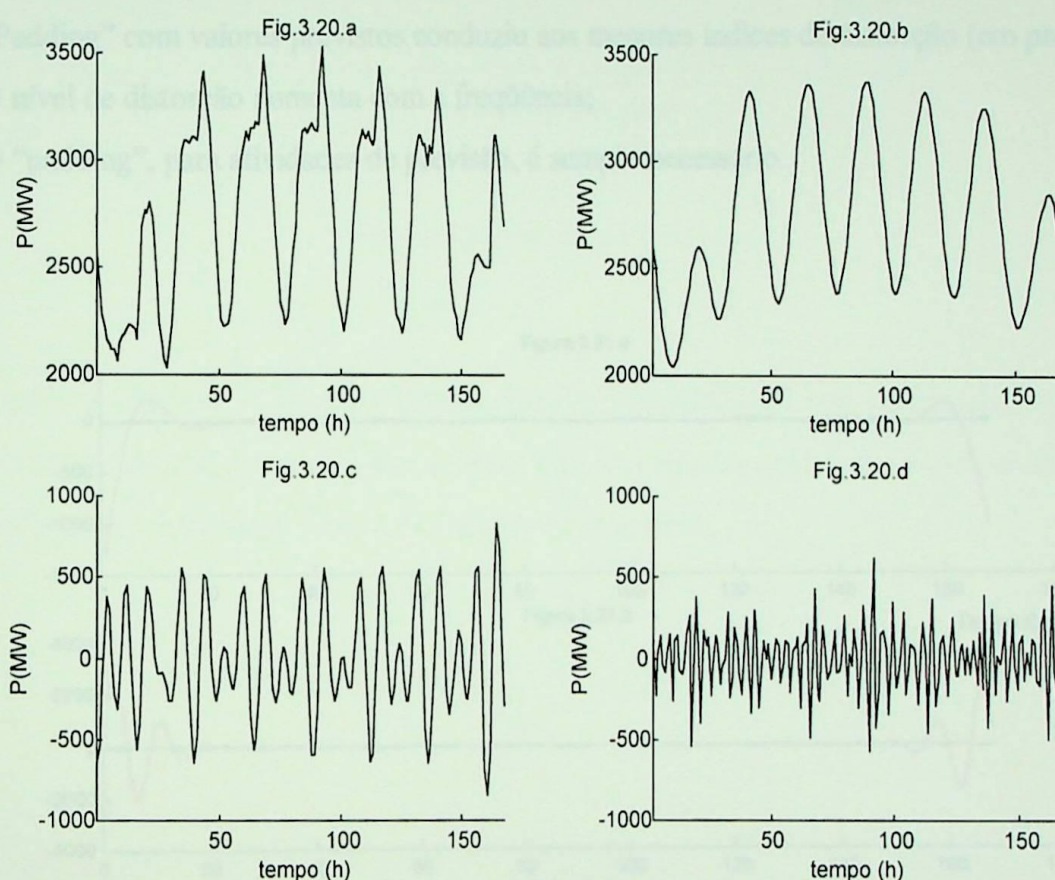


Figura 3.19 Séries filtradas utilizando zeros como “padding”

### Exemplo 4

No quarto e último exemplo (Figura 3.20), utilizou-se o “padding” proposto nessa dissertação. Como no exemplo 1, anexou-se os dados verdadeiros no começo da série. Já no final da mesma, valores previstos por um previsor (ver Modelo 3, item 4.10) foram adicionados. Nota-se agora que os níveis de distorção foram minimizados.



**Figura 3.20** Séries filtradas utilizando previsões como “padding”

Para finalizar esse capítulo, a Figura 3.21 mostra a diferença entre as saídas filtradas de referência com as demais, para os diversos tipos de filtros (Figura 3.21.a, FPB; Figura 3.21.b, FPF; Figura 3.21.c, FPA). A partir dela, pode-se concluir que:

- As séries filtradas sem nenhum tipo de “padding” apresentam distorções em todos os pontos (em azul);
- “Padding” com zeros produz níveis de distorção inaceitáveis nos extremos das séries filtradas (em vermelho);
- “Padding” com valores previstos conduziu aos menores índices de distorção (em preto);
- O nível de distorção aumenta com a frequência;
- O “padding”, para atividades de previsão, é sempre necessário.

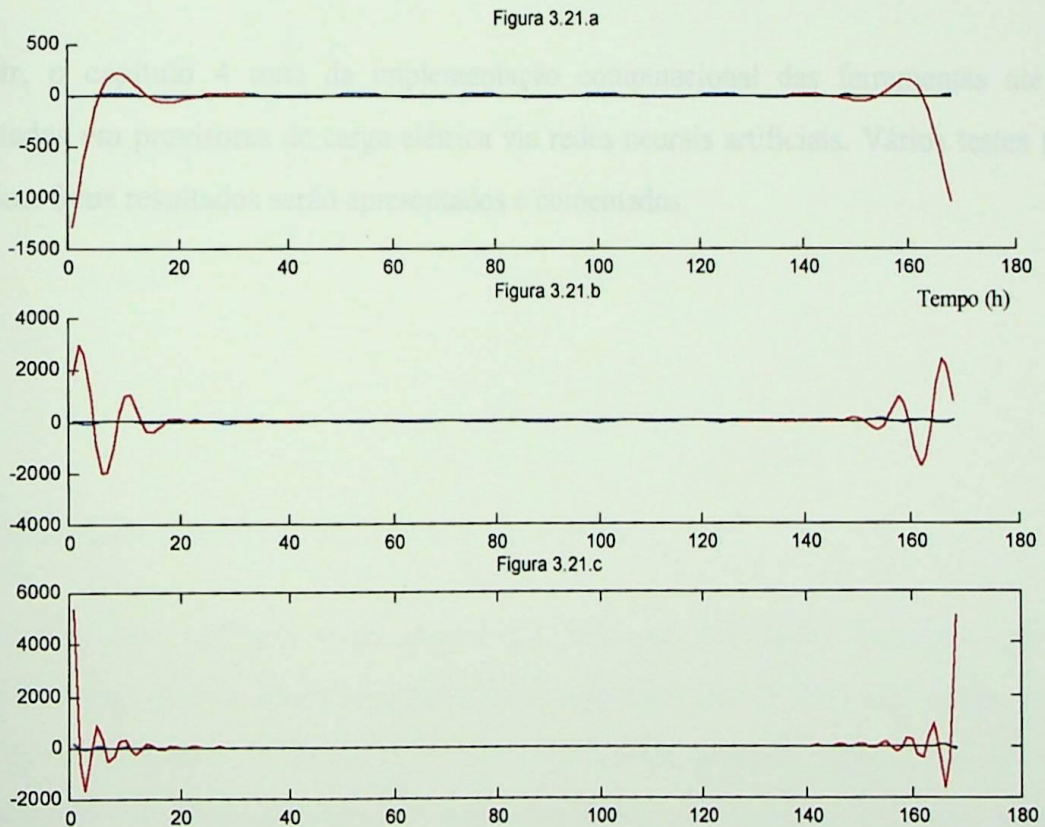


Figura 3.21 Comparação entre as diversas séries filtradas

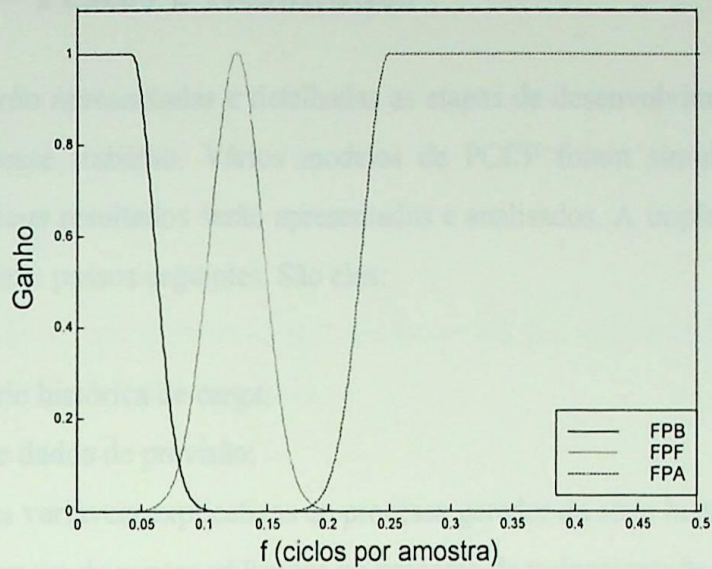


Figura 3.22 Filtros sintonizados

A seguir, o capítulo 4 trata 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.

#### 4.1 Definição da série histórica de carga

A série de carga utilizada nesta dissertação refere-se aos dados coletados por uma concessionária do Rio de Janeiro nos períodos de 8 de dezembro de 1994 a 23 de fevereiro de 1995, e de 14 de maio a 31 de agosto de 1995, abrangendo, portanto, os períodos de final de primavera/verão e final do outono/inverno. Os dados são tomados em 1500 horários e correspondem às medições de potência ativa em MW, totalizando 4.500 medições.

## Capítulo 4 - Testes e Resultados

Nesse capítulo serão apresentadas e detalhadas as etapas de desenvolvimento dos previsores implementados nesse trabalho. Vários modelos de PCCP foram simulados em ambiente MATLAB 5.2 e seus resultados serão apresentados e analisados. A implementação realizada foi dividida nos treze passos seguintes. São eles:

1. Definição da série histórica de carga;
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/teste;
5. Etapa de normalização dos dados disponíveis;
6. Partição ótima dos dados;
7. “Expurgo” de padrões atípicos;
8. Definição da estrutura do MLP;
9. Algoritmo de aprendizagem;
10. Modelos de PCCP via MLP;
11. Etapas de treinamento/teste;
12. Tipos de previsão;
13. Validação.

### 4.1 Definição da série histórica de carga

A série de carga utilizada nessa dissertação refere-se aos dados coletados por uma concessionária do Rio de Janeiro nos períodos de 8 de dezembro de 1994 a 23 de fevereiro de 1995, e de 14 de maio a 31 de agosto de 1995, abrangendo, portanto, os períodos de final da primavera/verão e final do outono/inverno. Os dados são tomados em base horária e correspondem às medições de potência ativa em MW, totalizando 4.512 medições.

## 4.2 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 MLP para cada tipo de dia da semana.

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

Uma RNA prevê dados futuros a partir de dados passados conhecidos. Portanto, pergunta-se: Qual ou quais “lag(s)” serão utilizados no processo de aprendizagem da RNA? E ainda, como identificá-los?

Muitos métodos para identificação das variáveis de entrada de uma RNA têm sido usados. A maioria deles utilizam funções de autocorrelação, acompanhadas ou não de heurísticas e da experiência do operador. Outros baseiam-se no espectro de potência da série. Mais recentemente, em [Drezga & Rahman, 1998], os autores usam um método baseado em PSE (*Phase-Space Embedding*) para descobrir relações não lineares entre as variáveis.

Nessa dissertação utilizou-se a função de autocorrelação na identificação desses “lags” explicativos. Tem-se a consciência de que o problema de previsão de carga, via de regra, é muito não linear e que tal função tenta identificar quase que exclusivamente relações lineares entre conjuntos de variáveis. No entanto, essa função se mostrou adequada para os propósitos desse trabalho.

Um dos métodos mais populares de se calcular a autocorrelação é mostrado na equação 4.3. As equações 4.1 e 4.2 são, respectivamente, a autocovariância e a variância da série. A Tabela 4.1 mostra os  $r_k$  para a série exibida na Figura 3.6.a. Já a Figura 4.1 exhibe o autocorrelograma da mesma.

$$C_k = \frac{1}{n} \sum_{i=k+1}^n (z_i - \bar{z}) \cdot (z_{i-k} - \bar{z}) \quad \text{Eq.4.1}$$

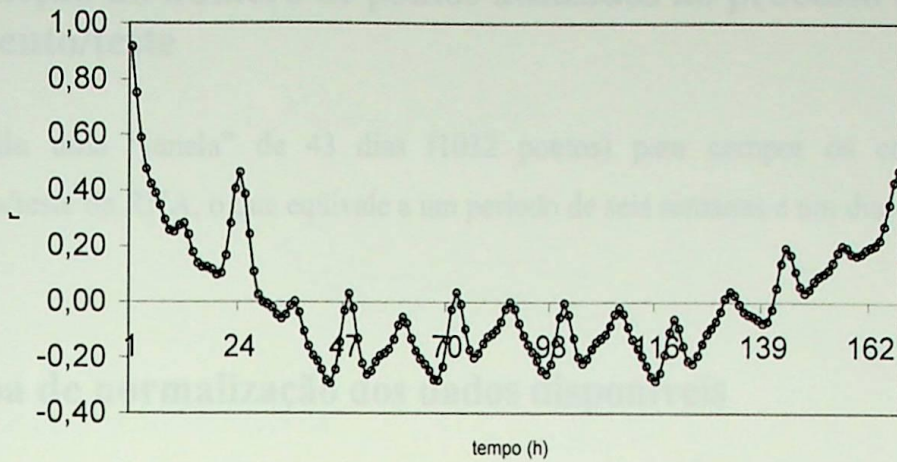
$$C_0 = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2 \quad \text{Eq.4.2}$$

$$r_k = \frac{C_k}{C_0} \quad \text{Eq.4.3}$$

Tabela 4.1  $r_k$  da série analisada

$r_1$ a $r_{24}$	$R_{25}$ a $r_{48}$	$r_{49}$ a $r_{72}$	$r_{73}$ a $r_{96}$	$r_{97}$ a $r_{120}$	$r_{121}$ a $r_{144}$	$r_{145}$ a $r_{168}$
<b>0,92</b>	0,39	-0,02	-0,01	-0,05	-0,09	0,17
<b>0,75</b>	0,24	-0,13	-0,10	-0,13	-0,16	0,11
0,59	0,11	-0,22	-0,17	-0,20	-0,21	0,06
0,48	0,03	-0,26	-0,20	-0,22	-0,22	0,03
0,42	0,00	-0,25	-0,18	-0,20	-0,19	0,05
0,39	-0,01	-0,21	-0,15	-0,17	-0,15	0,07
0,35	-0,02	-0,19	-0,12	-0,14	-0,11	0,09
0,29	-0,05	-0,18	-0,11	-0,13	-0,09	0,10
0,26	-0,06	-0,17	-0,10	-0,12	-0,07	0,12
0,25	-0,05	-0,13	-0,07	-0,09	-0,03	0,14
0,28	-0,01	-0,08	-0,02	-0,04	0,01	0,18
0,28	0,00	-0,05	0,00	-0,02	0,04	0,20
0,24	-0,04	-0,08	-0,03	-0,04	0,03	0,20
0,18	-0,11	-0,13	-0,08	-0,09	-0,01	0,18
0,14	-0,16	-0,18	-0,12	-0,13	-0,03	0,17
0,13	-0,19	-0,20	-0,15	-0,17	-0,05	0,18
0,13	-0,21	-0,22	-0,18	-0,19	-0,05	0,19
0,12	-0,24	-0,25	-0,21	-0,23	-0,06	0,20
0,10	-0,28	-0,28	-0,25	-0,27	-0,07	0,21
0,10	-0,29	-0,28	-0,26	-0,28	-0,07	0,23
0,17	-0,24	-0,23	-0,22	-0,25	-0,03	0,28
0,28	-0,15	-0,14	-0,15	-0,19	0,05	0,35
0,41	-0,03	-0,03	-0,06	-0,11	0,14	0,44
<b>0,46</b>	0,03	0,03	-0,01	-0,06	0,19	<b>0,48</b>





**Figura 4.1** Autocorrelograma da série analisada

A partir da Figura 4.1 e com o auxílio da Tabela 4.1, identificou-se os “lags” 1, 2, 24 e 168 como sendo os “lags” explicativos dessa série.

Além desses dados de carga, utilizou-se ainda duas entradas adicionais com a função de codificar as 24 horas do dia e de facilitar a captura de ciclos presentes na série de interesse [Drezga & Rahman, 1998]. Essas entradas foram denominadas por “Hora\_Sen” e “Hora\_Cos” e foram calculadas a partir das equações 4.4 e 4.5.

$$Hora\_Sen(k) = \frac{\text{sen}(2\pi k)}{24} \quad \text{Eq.4.4}$$

$$Hora\_Cos(k) = \frac{\text{cos}(2\pi k)}{24} \quad \text{Eq.4.5}$$

onde  $k$  é a hora corrente.

As entradas ora definidas serão utilizadas em cada um dos modelos de PCCP que serão apresentados no item 4.10.

## 4.4 Definição do número de pontos utilizados no processo de treinamento/teste

Foi utilizada uma “janela” de 43 dias (1032 pontos) para compor os conjuntos de treinamento/teste da RNA, o que equivale a um período de seis semanas e um dia.

## 4.5 Etapa de normalização dos dados disponíveis

Dois tipos de normalização foram testados e seus desempenhos comparados no item 4.13; a saber:

- Normalização baseada nos valores máximos e mínimos da série de carga (ver Eq.3.1);
- Normalização baseada em escores reduzidos (Eq.3.4).

## 4.6 Partição ótima dos dados

Uma teoria estatística do fenômeno de “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 (nessa dissertação,  $N$  pode variar de 96 a 144) 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.7.6, 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 o “overfitting” pode ocorrer para  $N < 30W$ , e existe mérito prático em usar-se 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:

$$P_{\text{ótimo}} = \frac{\sqrt{2W - 1} - 1}{2(W - 1)} \quad \text{Eq.4.6}$$

Consequentemente,

$$N_{\text{treinamento}} = p_{\text{ótimo}} \cdot N \quad \text{Eq.4.7}$$

$$N_{\text{teste}} = N - N_{\text{treinamento}} \quad \text{Eq.4.8}$$

onde,

$N$  é o número de padrões disponíveis;

$N_{\text{treinamento}}$  é o número de padrões para fins de treinamento;

$N_{\text{teste}}$  é o número de padrões para fins de teste.

### 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, 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 figura 4.2, deduziu-se a seguinte equação<sup>1</sup>:

$$W = [S_1 \cdot (1 + n_1 + S_2) + S_2] \quad \text{Eq.4.9}$$

onde:

$W$  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>1</sup> A Eq.4.9 é válida para uma única camada oculta

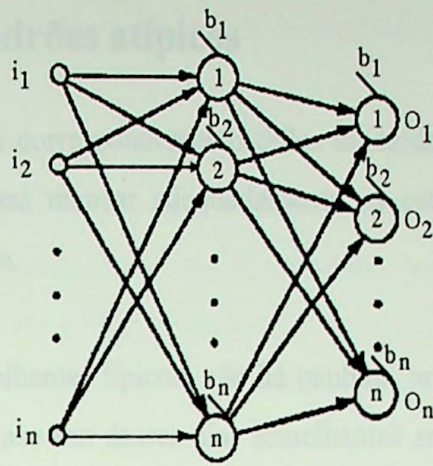


Figura 4.2 Estrutura típica de um MLP

Esse 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 teste?

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

$$W=?$$

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

$$\therefore 30W = 510$$

Como  $N < 30W$ ,

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

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

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

## 4.7 “Expurga” de padrões atípicos

Padrões de entrada e os seus correspondentes padrões de saída são usados para treinar uma RNA até que a mesma possa mapear adequadamente os estímulos da entrada para seus respectivos estímulos de saída.

Em se tratando de dias semelhantes típicos, não há nenhum problema com essa abordagem. No entanto, o que ocorrerá caso um desses dias semelhantes seja um feriado? Ou um dia de jogo da seleção brasileira em uma copa do mundo? Certamente, esse dia exibirá um comportamento atípico e poderá comprometer a capacidade de generalização da rede. Em tais situações, o que fazer?

O procedimento adotado nessa dissertação foi *excluir os padrões de entrada associados aos padrões de saída atípicos* (isso justifica o fato de  $N$  variar de 96 a 144).

Apesar desse processo de expurga implicar em uma redução do número de padrões disponíveis para treinamento e teste, ele evita que a RNA seja alimentada com informação não regular. Ressalta-se que previsões de dias atípicos fogem do escopo deste trabalho.

## 4.8 Definição da estrutura do MLP

Na ausência de procedimentos sistemáticos para a determinação da melhor estrutura de uma RNA (i.e., o número de unidades na camada oculta), diferentes arquiteturas de redes foram criadas e seus desempenhos comparados. Essa metodologia consiste em:

1º passo: Criar uma estrutura com um único neurônio na camada oculta;

2º passo: Realizar atividades de treinamento, teste e validação, armazenando os erros observados durante a etapa de validação da estrutura;

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 8.

De posse dos resultados obtidos no 2º passo, analisa-se todos os resultados (i.e., os erros de previsão). A estrutura que apresentou o melhor desempenho define a melhor estrutura de rede para aquele problema em questão. Mas, como avaliar o desempenho de uma dada estrutura? Quais índices deve-se utilizar? Esse assunto será tratado no item 4.13.

Todas as estruturas testadas contavam com um única camada oculta e um número de neurônios variando de 1 a 8. No entanto, essa metodologia pode ser estendida para estruturas com um número maior de neurônios e/ou duas camadas ocultas.

## 4.9 Algoritmo de aprendizagem

O algoritmo utilizado no treinamento do MLP foi uma variante do popular algoritmo de retropropagação de erro (EBP) apresentado no item I.7.4. Via de regra, o EBP padrão é bastante lento porque requer baixas taxas de aprendizagem para que se consiga um treinamento estável. A versão modificada conta com dois parâmetros adicionais, a saber:

- taxa de *momentum*  $\xi$ ;
- 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.

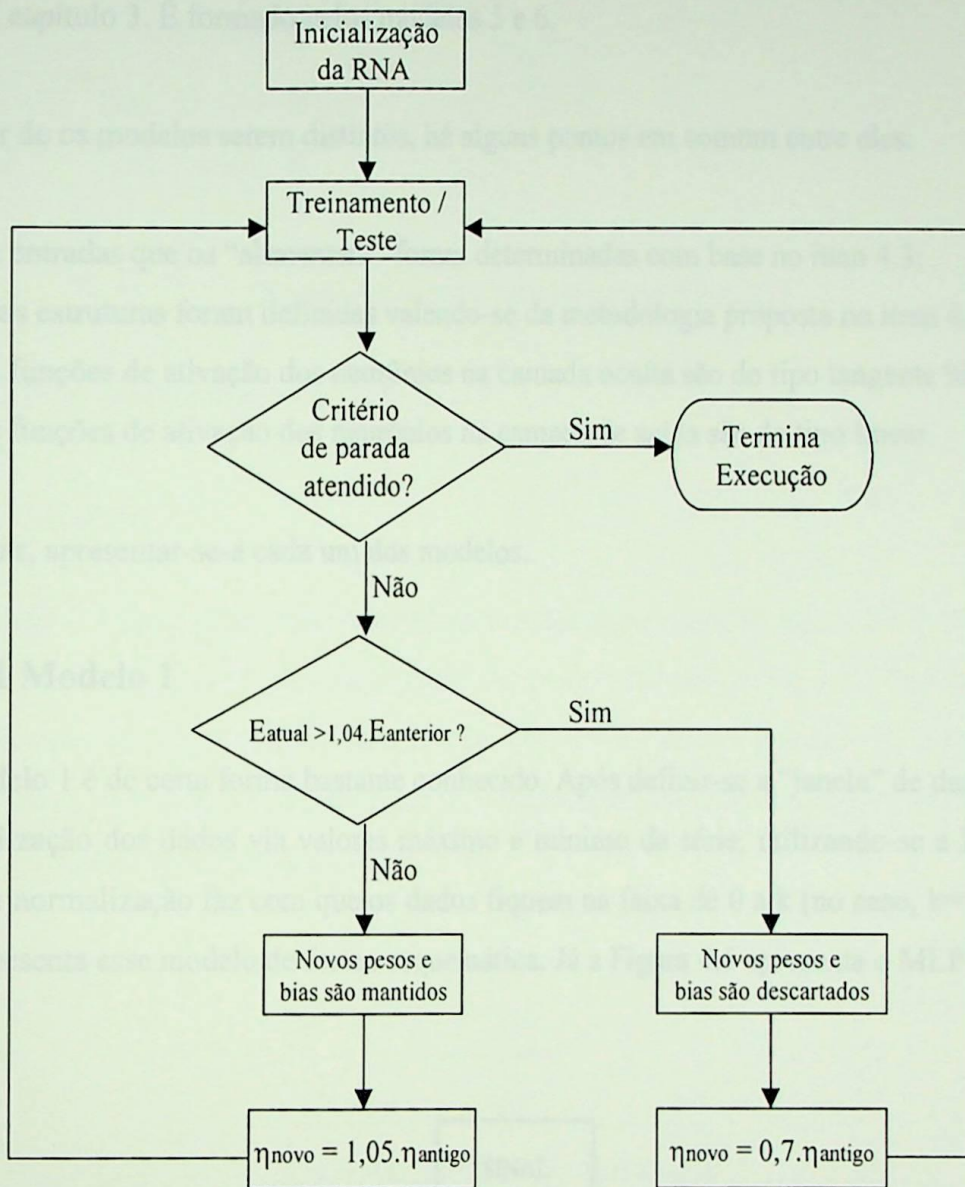


Figura 4.3 Algoritmo para a implementação de  $\eta$  adaptativa

#### 4.10 Modelos de PCCP via MLP

Durante o desenvolvimento dessa dissertação, cerca de 25 modelos de PCCP foram testados. Apesar de todos eles se basearem em MLP, o tipo e o número de entradas, a arquitetura das redes, entre outros parâmetros, variaram significativamente. Dentre eles, escolheu-se seis que exibissem características bastante peculiares. Os 6 modelos protótipos, por assim dizer, podem ser divididos em 2 grupos, a saber:

- **Grupo 1:** composto pelos modelos 1, 2, 3 e 4 não faz uso de qualquer recurso de filtragem digital;
- **Grupo 2:** uma combinação dos modelos anteriormente citados com os filtros apresentados no capítulo 3. É formado pelos modelos 5 e 6.

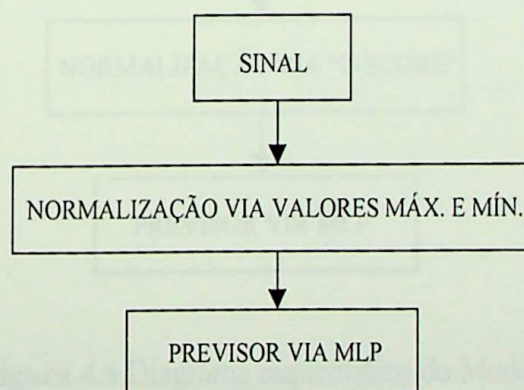
Apesar de os modelos serem distintos, há alguns pontos em comum entre eles:

1. As entradas que os “alimentam” foram determinadas com base no item 4.3;
2. Suas estruturas foram definidas valendo-se da metodologia proposta no item 4.8;
3. As funções de ativação dos neurônios na camada oculta são do tipo tangente hiperbólica;
4. As funções de ativação dos neurônios na camada de saída são do tipo linear.

A seguir, apresentar-se-á cada um dos modelos.

#### 4.10.1 Modelo 1

O Modelo 1 é de certa forma bastante conhecido. Após definir-se a “janela” de dados, faz-se a normalização dos dados via valores máximo e mínimo da série, utilizando-se a Eq.3.1. Esse tipo de normalização faz com que os dados fiquem na faixa de 0 a  $k$  (no caso,  $k=1$ ). A Figura 4.4 apresenta esse modelo de forma esquemática. Já a Figura 4.5 apresenta o MLP utilizado.



**Figura 4.4** Diagrama esquemático do Modelo 1

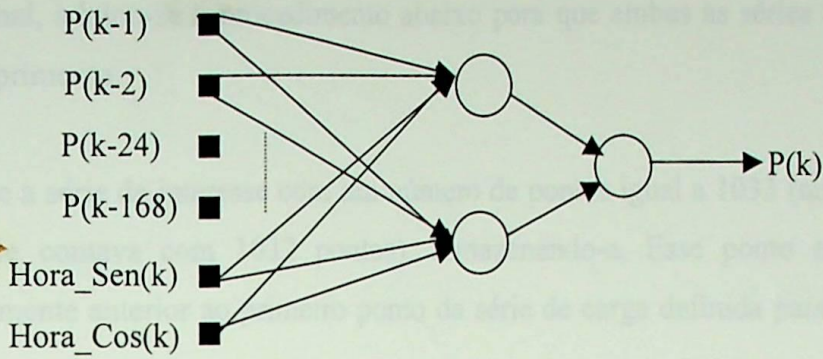


Figura 4.5 Arquitetura neural do Modelo 1

Onde,  $P(k)$  é a potência ativa no instante  $k$ .

### 4.10.2 Modelo 2

O Modelo 2 apresenta a mesma estrutura e as mesmas entradas que o Modelo 1. A única alteração é que os dados são normalizados via escores reduzidos (Eq.3.4). Esse procedimento faz com que os dados fiquem na faixa de -0,9 e 0,9 (limites práticos da função tangente hiperbólica). A Figura 4.6 mostra o diagrama esquemático desse modelo. A estrutura da rede é idêntica àquela apresentada na Figura 4.5.

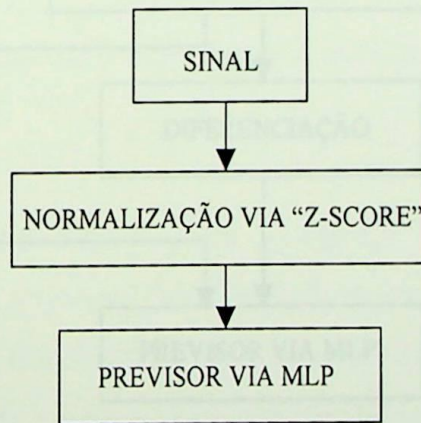


Figura 4.6 Diagrama esquemático do Modelo 2

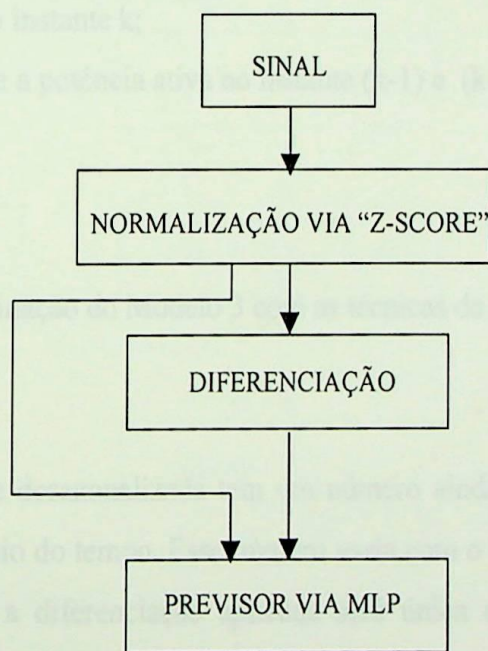
### 4.10.3 Modelo 3

O Modelo 3 é uma combinação do Modelo 2 com as técnicas de diferenciação apresentadas na seção 3.4. Nesse modelo, dois tipos de séries são empregadas: a original e outra no domínio das “diferenças”. Lembrando-se que a série diferenciada tem um ponto a menos que

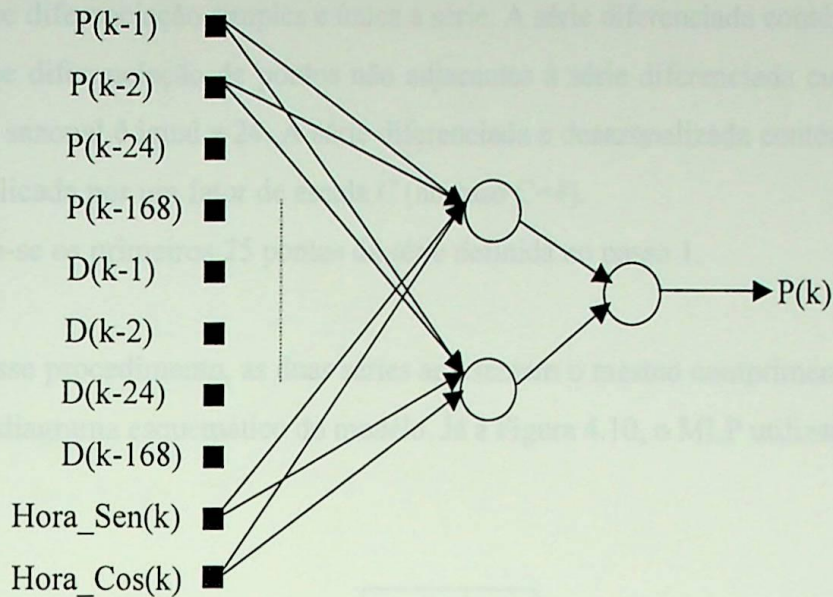
a série original, adotou-se o procedimento abaixo para que ambas as séries apresentassem o mesmo comprimento.

1. Define-se a série de interesse com um número de pontos igual a 1033 (nos modelos 1 e 2 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 os modelos 1 e 2.
2. Aplica-se diferenciação simples e única 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. A Figura 4.7 apresenta o diagrama esquemático do modelo. Já a Figura 4.8, o MLP utilizado.



**Figura 4.7** Diagrama esquemático do Modelo 3



**Figura 4.8** Arquitetura neural do Modelo 3

Onde:

$P(k)$  é a potência ativa no instante  $k$ ;

$D(k-1)$  é a diferença entre a potência ativa no instante  $(k-1)$  e  $(k-2)$ .

#### 4.10.4 Modelo 4

O Modelo 4 é uma combinação do Modelo 3 com as técnicas de desazonalização apresentadas na seção 3.5.

Uma série diferenciada e desazonalizada tem um número ainda menor de pontos que a sua correspondente no domínio do tempo. Esse número varia com o tamanho do ciclo sazonal que se deseja remover e se a diferenciação aplicada será única ou não. No caso, utilizou-se diferenciação única e o ciclo a ser removido continha 24 pontos. Logo, após a aplicação dos dois tipos de diferenciação à série original, a série diferenciada e desazonalizada terá 25 pontos a menos. Da mesma forma que para o Modelo 3, torna-se necessário adotar algum procedimento para que ambas as séries apresentem o mesmo comprimento. O procedimento utilizado é listado abaixo.

1. Define-se a série de interesse com um número de pontos igual a 1057, armazenando-a. Os pontos extras são os 25 pontos imediatamente anteriores ao primeiro ponto da série de carga definida para os modelos 1 e 2.

2. Aplica-se diferenciação simples e única à série. A série diferenciada contém 1056 pontos.
3. Aplica-se diferenciação de pontos não adjacentes à série diferenciada cujo comprimento do ciclo sazonal é igual a 24. A série diferenciada e desazonalizada contém 1032 pontos e é multiplicada por um fator de escala  $C$  (no caso  $C=4$ ).
4. Descarta-se os primeiros 25 pontos da série definida no passo 1.

Realizado esse procedimento, as duas séries apresentam o mesmo comprimento. A Figura 4.9 apresenta o diagrama esquemático do modelo. Já a Figura 4.10, o MLP utilizado.



**Figura 4.9** Diagrama esquemático do Modelo 4

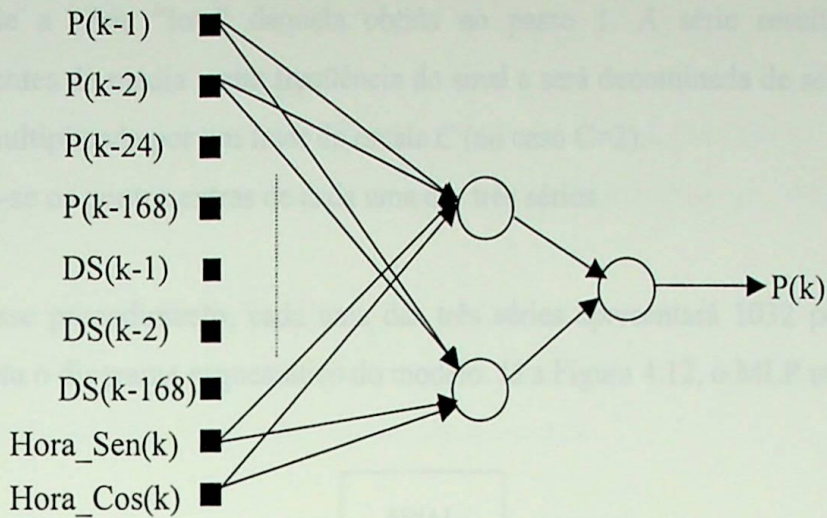


Figura 4.10 Arquitetura neural do Modelo 4

Onde:

$P(k)$  é a potência ativa no instante  $k$ ;

$DS(k-1) = P(k-1) - P(k-2) - P(k-24)$ .

#### 4.10.5 Modelo 5

O Modelo 5 é uma combinação do Modelo 2 com as técnicas de filtragem digital apresentadas na seção 3.6. A metodologia utilizada para a obtenção das séries que compõem o predictor será apresentada a seguir.

1. Define-se a série de interesse com um número de pontos igual a 1080, armazenando-a. Os pontos extras são os 48 pontos imediatamente anteriores ao primeiro ponto da série de carga definida para os modelos 1 e 2.
2. Anexa-se no final da série, 48 pontos previstos por um outro predictor (no caso, utilizou-se as previsões realizadas pelo Modelo 3). A série de carga passa a conter 1128 pontos.
3. Aplica-se um filtro passa baixas à série ( $f_c = \frac{1}{24}$  ciclos por amostra,  $l = 0,025$ ), armazenando-a. Essa série contém as oscilações mais lentas da série original de carga e será denominada de série “low”.

4. Subtrai-se a série “low” daquela obtida no passo 1. A série resultante exibirá as componentes de média e alta frequência do sinal e será denominada de série “band”. Essa série é multiplicada por um fator de escala  $C$  (no caso  $C=2$ ).
5. Descarta-se os pontos extras de cada uma das três séries.

Realizado esse procedimento, cada uma das três séries apresentará 1032 pontos. A Figura 4.11 apresenta o diagrama esquemático do modelo. Já a Figura 4.12, o MLP utilizado.

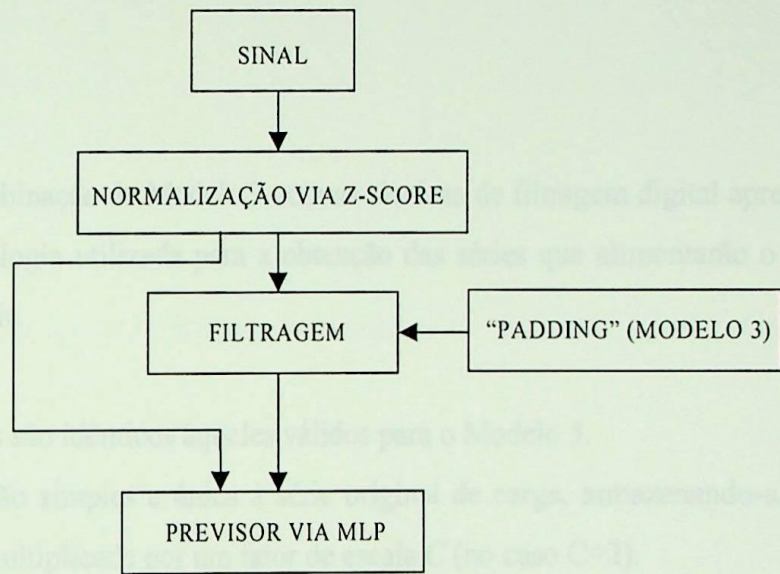


Figura 4.11 Diagrama esquemático do Modelo 5

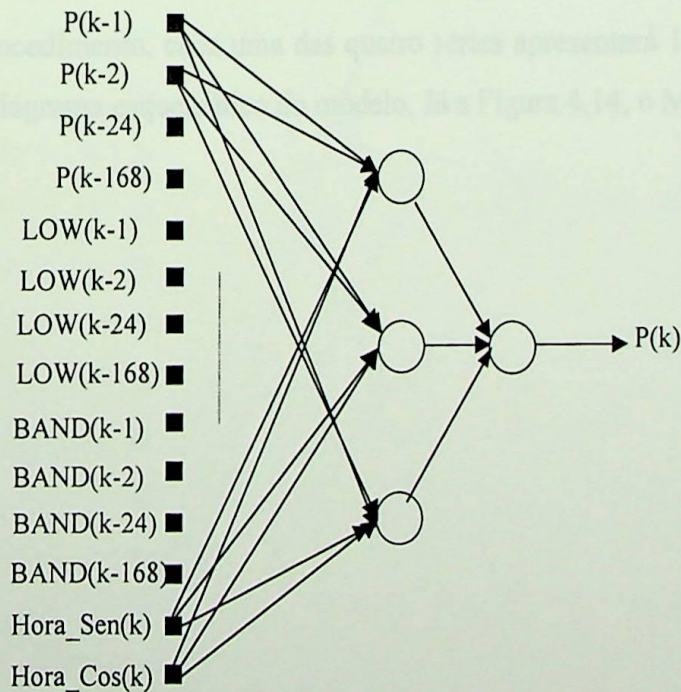


Figura 4.12 Arquitetura neural do Modelo 5

Onde,

$P(k)$  é a potência ativa no instante  $k$ ;

$LOW(k)$  é a componente de baixa frequência da potência ativa no instante  $k$ ;

$BAND(k)$  é a componente de média/alta frequência da potência ativa no instante  $k$ .

A arquitetura neural mostrada na Figura 4.12 é válida para o período do verão. A arquitetura para o período de inverno, que apresentou melhor capacidade de generalização, contava com as mesmas entradas e um único neurônio na camada oculta.

#### 4.10.6 Modelo 6

O Modelo 6 é uma combinação do Modelo 3 com as técnicas de filtragem digital apresentadas no item 3.6. A metodologia utilizada para a obtenção das séries que alimentarão o previsor será apresentada a seguir.

1. Os 4 primeiros passos são idênticos àqueles válidos para o Modelo 5.
2. Aplica-se diferenciação simples e única à série original de carga, armazenando-a. A série diferenciada deve ser multiplicada por um fator de escala  $C$  (no caso  $C=2$ ).
3. Descarta-se os pontos extras de cada uma das quatro séries (um cuidado adicional deve ser tomado quando se descartar os pontos extras da série diferenciada).

Realizado esse procedimento, cada uma das quatro séries apresentará 1032 pontos. A Figura 4.13 apresenta o diagrama esquemático do modelo. Já a Figura 4.14, o MLP utilizado.



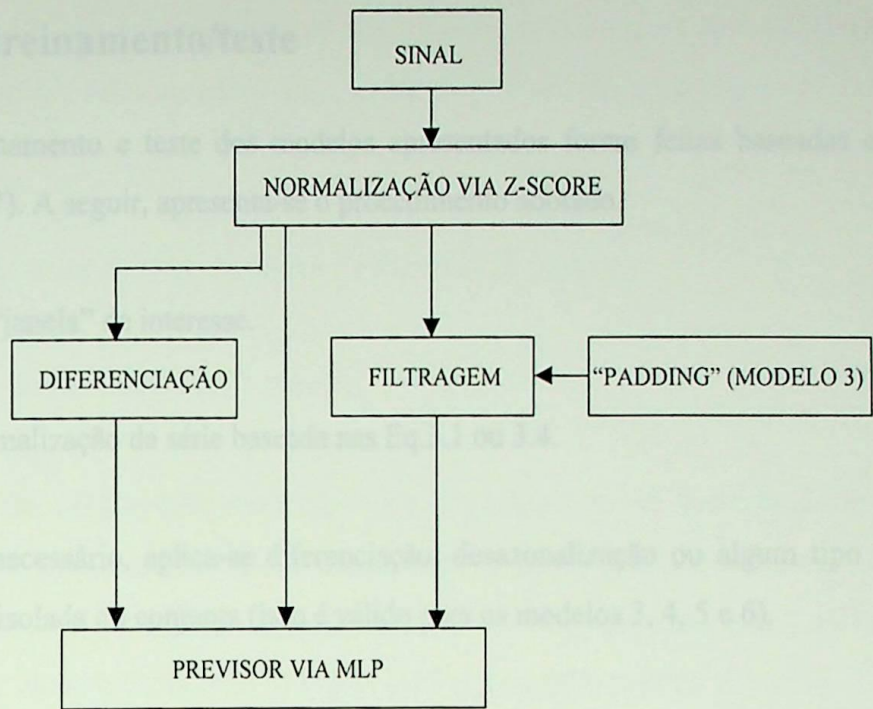


Figura 4.13 Diagrama esquemático do Modelo 6

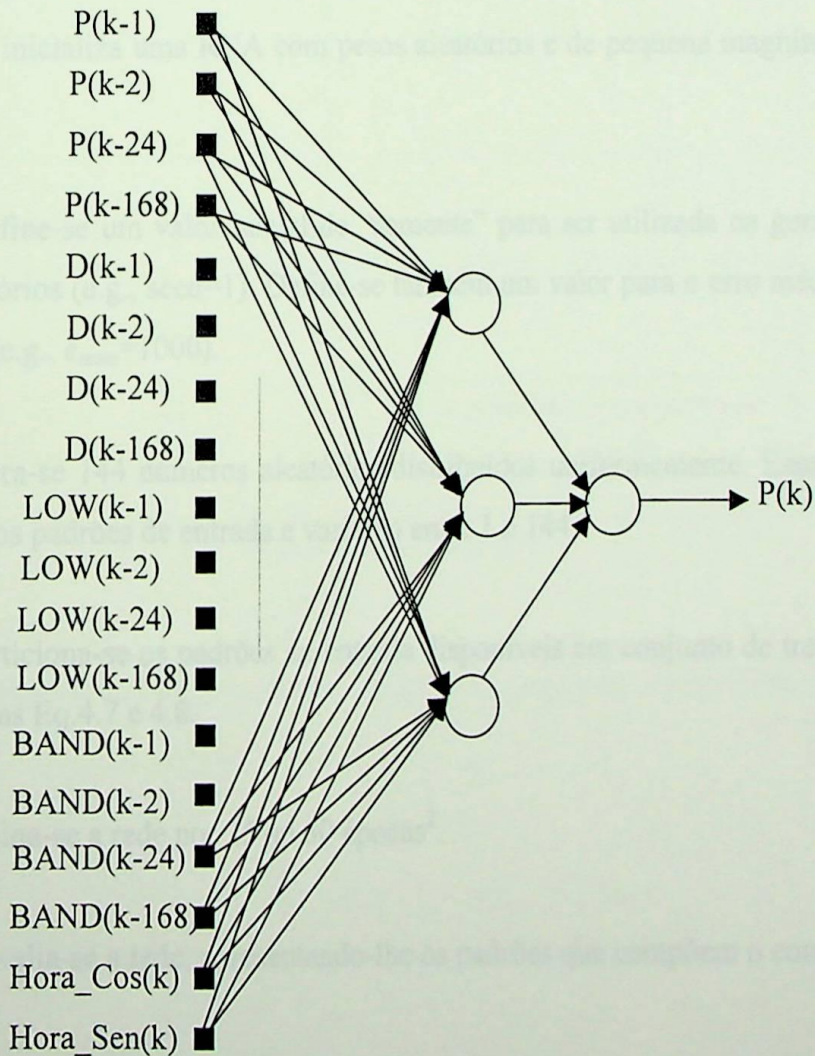


Figura 4.14 Arquitetura neural do Modelo 6

## 4.11 Etapas de treinamento/teste

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

1º passo: Define-se a “janela” de interesse.

2º passo: Faz-se a normalização da série baseada nas Eq.3.1 ou 3.4.

3º passo: Caso seja necessário, aplica-se diferenciação, sazonalização ou algum tipo de filtragem, de maneira isolada ou conjunta (isso é válido para os modelos 3, 4, 5 e 6).

4º passo: Monta-se os padrões de entrada/saída, expurgando-se alguns deles caso seja necessário. Supondo que nenhum padrão tenha sido expurgado, 144 padrões de entrada/saída estarão disponíveis; o que equivale a 6 dias semelhantes.

5º passo: Se inicializa uma RNA com pesos aleatórios e de pequena magnitude (entre  $-0,5$  e  $0,5$ ).

6º passo: Define-se um valor inicial de “semente” para ser utilizada 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.,  $e_{teste}=1000$ ).

7º passo: Gera-se 144 números aleatórios distribuídos uniformemente. Esses números serão associados aos padrões de entrada e variarão entre 1 e 144.

8º passo: Particiona-se os padrões de entrada disponíveis em conjunto de treinamento e teste, valendo-se das Eq.4.7 e 4.8.

9º passo: Treina-se a rede por 25 ou 50 épocas<sup>2</sup>.

10º passo: Avalia-se a rede, apresentando-lhe os padrões que compõem o conjunto de teste.

<sup>2</sup> Esse número varia de acordo com o modelo de previsão adotado.

11º passo: Calcula-se o desvio médio quadrático total (SSE) para o conjunto de teste. Caso o erro atual seja menor que o valor anterior ( $e_{teste}$ ), os pesos e “bias” são armazenados. Senão, são descartados.

12º passo: soma-se 1 ao valor da semente inicial e volta-se ao 7º passo.

Esse procedimento é repetido de 250 a 300 vezes<sup>3</sup>.

Uma vez treinada a rede, desloca-se a janela de dados 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 pode ser treinada. Esse procedimento foi adotado tanto para o período do verão quanto para o período do inverno. A Figura 4.15 representa o período do verão.

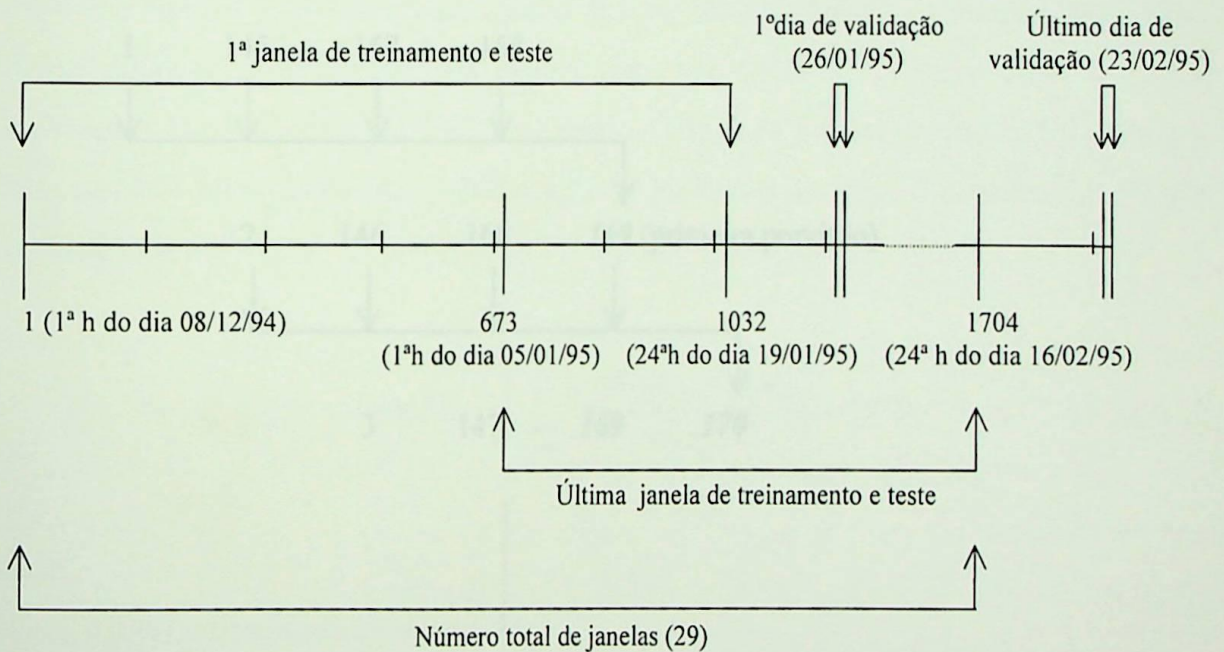


Figura 4.15 Representação da curva de carga para o período do verão

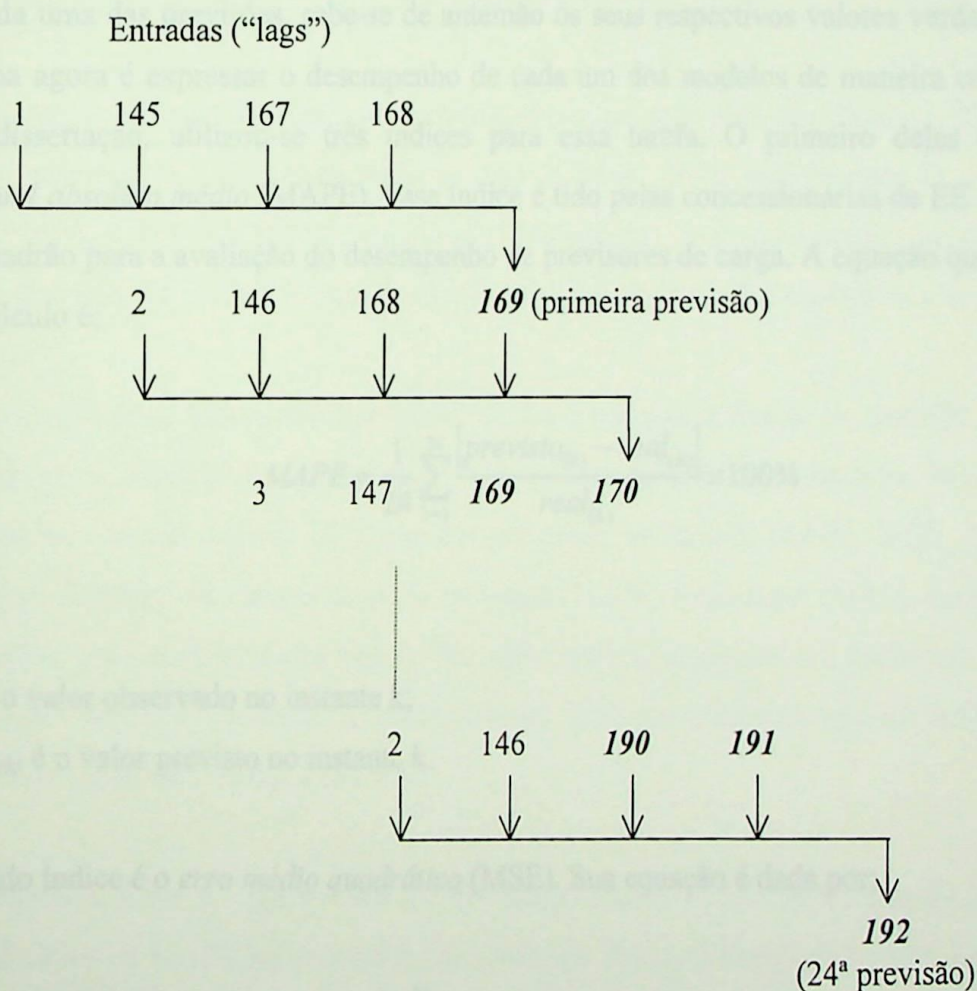
<sup>3</sup> Esse número também varia de acordo com o modelo de previsão adotado.

## 4.12 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 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. A Figura 4.16 mostra o processo de recursão.



**Figura 4.16** Previsão recursiva

Existem também dois modos de como essas previsões podem ser realizadas: de forma *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*, as mesmas 24 previsões podem ser obtidas a partir de qualquer instante.

### 4.13 Validações

Para avaliar-se o desempenho dos seis modelos apresentados, previsões recursivas de 24 passos à frente, do tipo *off-line*, são geradas para um período de 90 dias do histórico de dados da concessionária em questão. Os períodos de validação vão de 26/01 a 23/02/1995 (verão) e 02/07 a 31/08/1995 (inverno). As validações realizadas são conhecidas como “testes cegos”, pois utilizam dados completamente independentes daqueles usados nos processos de treinamento/teste das redes.

Para cada uma das previsões, sabe-se de antemão os seus respectivos valores verdadeiros. O problema agora é expressar o desempenho de cada um dos modelos de maneira conclusiva. Nessa dissertação, utilizou-se três índices para essa tarefa. O primeiro deles é o *erro percentual absoluto médio* (MAPE). Esse índice é tido pelas concessionárias de EE como um índice padrão para a avaliação do desempenho de previsores de carga. A equação que permite o seu cálculo é:

$$MAPE = \frac{1}{24} \sum_{k=1}^{24} \frac{|previsto_{(k)} - real_{(k)}|}{real_{(k)}} \times 100\% \quad \text{Eq.4.11}$$

onde,

$real_{(k)}$  é o valor observado no instante  $k$ ;

$previsto_{(k)}$  é o valor previsto no instante  $k$ .

O segundo índice é o *erro médio quadrático* (MSE). Sua equação é dada por:

$$MSE = \frac{1}{24} \sum_{k=1}^{24} (previsto_{(k)} - real_{(k)})^2 \quad \text{Eq.4.12}$$

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

$$ME = \frac{1}{24} \sum_{k=1}^{24} (previsto_{(k)} - real_{(k)}) \quad \text{Eq.4.13}$$

É óbvio que cada um dos índices fornecem informações diferentes a respeito dos modelos. Em linhas gerais, o MAPE permite avaliar, na média, o comportamento de cada um dos modelos. O MSE dá informações a respeito de qual dos modelos apresenta os maiores erros horários. Finalmente, o ME informa o quão tendencioso é um dado modelo.

Tendo discutido os modelos e os índices utilizados para avaliar seus desempenhos, apresentar-se-á os resultados obtidos.

#### 4.13.1 Teste 1

Inicialmente, o período de validação foi segmentado em três períodos distintos. O primeiro período, compreende o intervalo entre os dias 26/01 e 23/02/1995 (verão). O segundo, o intervalo entre 02/07 e 31/07/1995 (inverno). E finalmente, o terceiro compreende o período entre os dias 01/08 e 31/08/1995 (inverno). A soma dos três períodos totaliza noventa dias.

Nesse primeiro teste foram feitas previsões de 1 a 24 passos à frente, de maneira recursiva, utilizando-se o Modelo 1 (normalização via valores máximos e mínimos, MLP com 2 neurônios na camada oculta). De posse das previsões, os índices MAPE, MSE, ME foram calculados. Também os valores de erros máximos (MÁX.) e mínimos (MÍN.) médios foram computados. Para realizar essas simulações, desenvolveu-se rotinas em ambiente *MATLAB* (utilizou-se também o “*toolbox*” de redes neurais). Esse procedimento será adotado em cada um dos testes realizados nessa dissertação.

As Tabelas 4.2, 4.3 e 4.4 apresentam os resultados obtidos. O MAPE horário da hora  $k$  é a média de todos os MAPE no instante  $k$ , para cada um dos períodos (Tabela 4.2). De maneira similar, o MAPE diário, é a média de todos os MAPE para cada um dos dias da semana, para cada um dos períodos (Tabela 4.3). Na Tabela 4.4, todos os índices apresentados são médios.

**Tabela 4.2** MAPE horário obtido com o Modelo 1

Hora	MAPE 1	MAPE 2	MAPE 3
0	3,23	1,97	2,47
1	4,74	2,38	3,88
2	5,30	2,55	4,03
3	4,79	2,66	3,89
4	4,54	2,61	3,63
5	4,01	2,57	3,50
6	3,33	2,47	3,12
7	3,09	1,98	2,97
8	2,58	2,00	2,51
9	2,15	1,91	2,01
10	1,93	1,80	1,91
11	2,40	1,57	1,83
12	2,38	1,55	2,22
13	2,48	1,53	2,40
14	2,88	2,18	2,64
15	3,12	2,25	2,82
16	3,45	2,41	2,91
17	3,72	2,73	3,05
18	4,27	1,95	2,27
19	3,59	3,00	2,72
20	3,72	1,26	1,71
21	2,80	1,35	1,89
22	3,56	1,87	2,65
23	5,01	2,18	3,52

**Tabela 4.3** MAPE diário obtido com o Modelo 1

MAPE DIÁRIO	Inverno (%)	Verão (%)
DOMINGO	2,61	4,47
SEGUNDA	2,38	4,31
TERÇA	2,34	3,50
QUARTA	1,94	3,24
QUINTA	2,38	3,55
SEXTA	1,91	2,61
SÁBADO	3,66	2,52

**Tabela 4.4** Índices de avaliação do Modelo 1

ÍNDICE	Período 1	Período 2	Período 3
MAPE (%)	<b>3,46</b>	2,11	2,77
MSE (MW <sup>2</sup> )	17707,20	5639,48	9443,39
ME (MW)	-2,64	-7,07	5,50
MAX (%)	8,59	5,44	6,31
MIN (%)	0,36	0,19	0,36

Para se avaliar a qualidade dos previsores, valores de MAPE abaixo de 3% são considerados bastante satisfatórios pelas concessionárias de EE [Khotanzad, 1997]. Caso esse critério fosse adotado, o Modelo 1 o violaria para o Período 1 (Tabela 4.4).

Tabela 4.5 MAPE horário obtido com o Modelo 2

hora	MAPE 1	MAPE 2	MAPE 3
0	3,11	1,75	2,56
1	4,44	2,15	3,44
2	4,82	2,08	3,49
3	3,37	1,66	2,40
4	2,68	1,64	1,95
5	2,15	1,50	2,16
6	2,04	1,96	2,12
7	2,35	1,68	2,06
8	2,32	1,56	2,35
9	2,47	1,48	2,37
10	2,90	1,30	2,37
11	3,43	1,39	2,33
12	3,68	2,42	2,59
13	3,47	2,67	2,42
14	3,91	2,17	1,92
15	3,58	2,08	3,13
16	3,79	1,36	1,50
17	3,11	1,42	2,08
18	3,67	1,70	2,80
19	5,27	1,95	3,62

### 4.13.2 Teste 2

Nesse segundo teste, avaliou-se o desempenho do Modelo 2 (normalização via escores reduzidos, MLP com 2 neurônios na camada oculta). As Tabelas 4.5, 4.6 e 4.7 apresentam os resultados.

**Tabela 4.5** MAPE horário obtido com o Modelo 2

Hora	MAPE 1	MAPE 2	MAPE 3
0	3,11	1,75	2,56
1	4,44	2,35	3,94
2	4,82	2,68	3,89
3	4,73	2,77	3,63
4	4,55	2,52	3,40
5	4,17	2,22	3,29
6	3,87	2,09	2,91
7	3,37	1,66	2,40
8	2,68	1,64	1,95
9	2,15	1,80	2,16
10	2,04	1,98	2,12
11	2,35	1,68	2,06
12	2,32	1,56	2,35
13	2,47	1,48	2,37
14	2,90	1,80	2,37
15	3,43	1,99	2,53
16	3,68	2,42	2,59
17	3,47	2,67	2,42
18	3,91	2,17	1,92
19	3,58	3,08	3,13
20	3,79	1,36	1,90
21	3,11	1,43	2,08
22	3,67	1,70	2,80
23	5,27	1,90	3,62

**Tabela 4.6** MAPE diário obtido com o Modelo 2

MAPE DIÁRIO	Inverno (%)	Verão (%)
DOMINGO	2,66	4,31
SEGUNDA	2,57	3,83
TERÇA	2,10	3,44
QUARTA	1,81	3,00
QUINTA	2,16	3,72
SEXTA	2,07	3,36
SÁBADO	3,22	2,74

**Tabela 4.7** Índices de avaliação do Modelo 2

ÍNDICE	Período 1	Período 2	Período 3
MAPE (%)	3,49	2,03	2,68
MSE (MW <sup>2</sup> )	18111,04	5153,34	8590,87
ME (MW)	-5,77	-12,19	-9,39
MAX (%)	8,62	5,25	5,91
MIN (%)	0,38	0,17	0,38

A partir da Tabela 4.7, nota-se que o MAPE para o período 1 é superior aos 3% tido como referência. No entanto, ao comparar-se os resultados dos dois modelos para os três períodos, na média, o Modelo 2 apresentou resultados ligeiramente melhores que o Modelo 1. É esse melhor desempenho que justifica o fato de o Modelo 2 servir de base para os demais.

### 4.13.3 Teste 3

Nesse terceiro teste avaliou-se a performance do Modelo 3 (normalização via escores reduzidos + Diferenciação, MLP com 2 neurônios na camada oculta). As Tabelas 4.8, 4.9 e 4.10 apresentam os resultados.

**Tabela 4.8** MAPE horário obtido com o Modelo 3

Hora	MAPE 1	MAPE 2	MAPE 3
0	1,43	1,32	1,24
1	1,98	1,53	1,94
2	2,65	2,15	2,23
3	2,91	2,30	2,48
4	2,92	2,17	2,34
5	3,14	2,23	2,40
6	3,01	2,27	2,35
7	3,28	1,98	1,98
8	3,05	2,00	1,59
9	2,64	2,21	1,83
10	2,29	2,06	1,76
11	2,25	1,62	1,62
12	2,27	1,52	1,90
13	2,58	1,50	1,94
14	2,85	1,61	1,96
15	3,14	1,98	2,48
16	3,19	2,25	2,39
17	2,85	2,58	1,87
18	2,82	1,95	1,66
19	3,11	2,52	2,15
20	2,41	1,84	1,96
21	2,66	1,80	2,12
22	3,45	1,91	2,83
23	4,19	2,01	3,62

Tabela 4.9 MAPE diário obtido com o Modelo 3

MAPE DIÁRIO	Inverno (%)	Verão (%)
DOMINGO	2,34	3,44
SEGUNDA	2,16	2,60
TERÇA	1,91	3,48
QUARTA	1,46	2,20
QUINTA	1,88	3,36
SEXTA	1,84	2,26
SÁBADO	2,76	2,08

Tabela 4.10 Índices de avaliação do Modelo 3

ÍNDICE	Período 1	Período 2	Período 3
MAPE (%)	2,79	1,97	2,11
MSE (MW <sup>2</sup> )	11903,13	4685,55	5867,75
ME (MW)	2,59	-14,58	-9,96
MAX (%)	6,98	4,72	5,04
MIN (%)	0,44	0,17	0,16

Nota-se agora que esse Modelo atendeu ao critério de MAPE < 3% para os três períodos. Além disso, constata-se ainda um ganho significativo em relação ao Modelo 2. Esse “ganho”<sup>4</sup> pode ser calculado a partir da equação abaixo:

$$ganho(\%) = \left( \frac{MAPE_{anterior} - MAPE_{atual}}{MAPE_{anterior}} \right) \cdot 100\% \quad \text{Eq.4.14}$$

Comparando-se os Modelos 2 e 3, verifica-se ganhos de 20,01%, 2,85% e 21,32% para os períodos 1, 2 e 3, respectivamente.

<sup>4</sup> A estrutura da Eq.4.14 é válida para qualquer índice

#### 4.13.4 Teste 4

Nesse quarto teste, avaliou-se o desempenho do Modelo 4 (normalização via escores reduzidos + Diferenciação + Desazonalização do ciclo diário, MLP com 2 neurônios na camada oculta). As Tabelas 4.11, 4.12 e 4.13 apresentam os resultados.

**Tabela 4.11** MAPE horário obtido com o Modelo 4

Hora	MAPE 1	MAPE 2	MAPE 3
0	3,09	1,94	2,86
1	5,00	2,76	4,34
2	5,73	3,03	4,74
3	5,28	2,84	4,34
4	4,72	2,64	3,72
5	4,12	2,45	3,45
6	3,40	2,26	3,16
7	3,26	1,84	2,81
8	2,87	2,05	2,27
9	2,73	2,28	2,41
10	2,50	2,01	2,08
11	2,41	1,54	1,76
12	2,52	1,62	2,06
13	2,29	1,67	2,17
14	2,98	1,54	2,22
15	3,60	1,83	2,58
16	3,59	2,29	2,57
17	3,36	2,47	2,38
18	3,79	2,58	2,04
19	3,97	3,27	3,10
20	4,13	1,61	1,96
21	3,64	1,22	2,32
22	3,77	1,21	2,98
23	5,12	1,50	3,59

**Tabela 4.12** MAPE diário obtido com o Modelo 4

MAPE DIÁRIO	Inverno (%)	Verão (%)
DOMINGO	2,84	5,02
SEGUNDA	3,15	2,99
TERÇA	2,16	3,77
QUARTA	1,99	3,19
QUINTA	2,16	4,03
SEXTA	1,82	3,61
SÁBADO	3,17	2,92

**Tabela 4.13** Índices de avaliação do Modelo 4

ÍNDICE	Período 1	Período 2	Período 3
MAPE (%)	3,66	2,10	2,82
MSE (MW <sup>2</sup> )	18970,84	5464,65	9675,92
ME (MW)	3,69	-7,95	-2,45
MAX (%)	9,18	5,66	6,71
MIN (%)	0,35	0,17	0,34

O critério de  $MAPE < 3\%$  para os três períodos não foi atendido. Para esse modelo em particular, esse resultado não era esperado. Com a combinação entre diferenciação e sazonalização, esperava-se facilitar o trabalho da rede, pois a sazonalidade diária estaria implícita na série sazonalizada. No entanto, com base nos resultados obtidos, pode-se concluir que a rede prefere que a informação a respeito do ciclo diário lhe seja fornecida de maneira explícita.

Realizados esses testes, conclui-se a etapa de validação do desempenho dos modelos do Grupo 1 (modelos que não utilizam recursos de filtragem digital). O modelo 3 apresentou o melhor desempenho dentre os quatro primeiros analisados. Isso justifica o fato de as previsões produzidas por ele terem sido utilizadas como “padding” para os modelos 5 e 6.

A seguir, avaliar-se-á os modelos do Grupo 2.

### 4.13.5 Teste 5

Nesse quinto teste avaliou-se o desempenho do Modelo 5 (normalização via escores reduzidos + filtragem digital, MLP com 3 neurônios na camada oculta para o período do verão e um único neurônio para o período do inverno). As Tabelas 4.14, 4.15 e 4.16 apresentam os resultados.

**Tabela 4.14** MAPE horário obtido com o Modelo 5

Hora	MAPE 1	MAPE 2	MAPE 3
0	1,80	1,78	1,16
1	1,47	1,92	1,65
2	2,18	2,08	1,91
3	2,74	1,78	1,93
4	2,81	1,78	1,67
5	2,74	1,86	1,91
6	2,66	2,06	2,02
7	2,67	1,77	2,04
8	2,37	2,00	1,62
9	2,33	2,06	1,85
10	2,24	1,94	1,69
11	2,35	1,58	1,55
12	2,44	1,43	1,79
13	2,75	1,44	1,81
14	3,01	1,61	1,75
15	3,29	1,81	2,22
16	3,31	2,08	2,30
17	2,78	2,50	1,98
18	2,83	1,97	1,85
19	3,03	2,52	2,03
20	2,35	1,60	1,52
21	2,67	1,75	1,79
22	3,35	1,65	2,24
23	3,50	1,64	2,81

**Tabela 4.15** MAPE diário obtido com o Modelo 5

MAPE DIÁRIO	Inverno (%)	Verão (%)
DOMINGO	1,99	3,52
SEGUNDA	2,21	2,69
TERÇA	1,61	3,24
QUARTA	1,40	2,17
QUINTA	1,55	3,09
SEXTA	1,61	1,96
SÁBADO	2,78	1,79

**Tabela 4.16** Índices de avaliação do Modelo 5

ÍNDICE	Período 1	Período 2	Período 3
MAPE (%)	2,65	1,85	1,87
MSE (MW <sup>2</sup> )	11148,19	4186,22	4631,86
ME (MW)	1,07	-12,62	-18,58
MAX (%)	6,74	4,64	4,28
MIN (%)	0,20	0,14	0,16

O Modelo 5 atendeu ao critério de  $MAPE < 3\%$  para os três períodos. Comparando-o com o Modelo 3, constata-se “ganhos” de 5,01 %, 5,89 % e 11,13 % para os períodos 1, 2 e 3, respectivamente.

### 4.13.6 Teste 6

Nesse sexto teste avaliou-se o desempenho do Modelo 6 (normalização via escores reduzidos + diferenciação + filtragem digital, MLP com 3 neurônios na camada oculta para o período do verão e um único neurônio para o período do inverno). As Tabelas 4.17, 4.18 e 4.19 apresentam os resultados.

**Tabela 4.17** MAPE horário obtido com o Modelo 6

Hora	MAPE 1	MAPE 2	MAPE 3
0	1,56	1,68	1,23
1	1,80	1,83	1,55
2	2,27	2,18	1,81
3	2,51	1,89	1,88
4	2,42	1,83	1,62
5	2,45	1,98	1,86
6	2,69	2,09	1,96
7	2,83	1,78	2,21
8	2,37	2,07	1,73
9	2,10	2,22	1,89
10	2,00	2,05	1,59
11	2,27	1,67	1,45
12	2,18	1,50	1,77
13	2,68	1,47	1,78
14	3,22	1,68	1,74
15	3,62	1,89	2,25
16	3,54	2,10	2,46
17	2,97	2,56	2,11
18	2,91	1,89	1,89
19	3,06	2,39	2,11
20	2,19	1,53	1,73
21	2,18	1,69	1,85
22	2,53	1,66	2,33
23	2,61	1,73	2,92

**Tabela 4.18** MAPE diário obtido com o Modelo 6

MAPE DIÁRIO	Inverno (%)	Verão (%)
DOMINGO	2,00	3,86
SEGUNDA	2,29	2,50
TERÇA	1,65	2,68
QUARTA	1,37	1,93
QUINTA	1,52	2,74
SEXTA	1,78	2,04
SÁBADO	2,76	1,96

**Tabela 4.19** Índices de avaliação do Modelo 6

ÍNDICE	Período 1	Período 2	Período 3
MAPE (%)	2,53	1,89	1,90
MSE (MW <sup>2</sup> )	9897,25	4261,25	5017,02
ME (MW)	3,60	-10,66	-17,59
MAX (%)	6,50	4,79	4,49
MIN (%)	0,18	0,12	0,15

O Modelo 6 também atendeu ao critério de  $MAPE < 3\%$  para os três períodos. Comparando-o com o Modelo 3, constata-se “ganhos” de 9,28 %, 4,3 % e 10,97 %, para os períodos 1, 2 e 3, respectivamente. Os ganhos referentes aos períodos 2 e 3 (inverno) foram comparáveis àqueles obtidos com o Modelo 5. No entanto, esse modelo o superou para o período do verão (ganho de 9,28 % contra 5,01 %).

Realizados esses testes, conclui-se a etapa de validação do desempenho dos modelos do Grupo 2 (modelos que utilizam recursos de filtragem digital).

As Tabelas 4.20, 4.21 e 4.22 apresentam um panorama geral dos resultados dos seis testes realizados para os períodos 1, 2 e 3, respectivamente.

**Tabela 4.20** Índices de avaliação dos modelos para o período 1 (verão)

ÍNDICES	MODELO 1	MODELO 2	MODELO 3	MODELO 4	MODELO 5	MODELO 6	GANHO (%) <sup>5</sup>
MAPE (%)	3,46	3,49	2,79	3,66	2,65	<b>2,539</b>	<b>9,28</b>
MSE (MW <sup>2</sup> )	17707,20	18111,04	11903,13	18970,84	11148,19	<b>9897,25</b>	<b>16,85</b>
ME (MW)	-2,64	-5,77	2,54	3,69	1,07	<b>3,60</b>	<b>-41,73</b>
MAX (%)	8,59	8,62	6,98	9,18	6,74	<b>6,50</b>	<b>6,88</b>
MIN (%)	0,36	0,38	0,44	0,35	0,20	<b>0,18</b>	<b>59,09</b>

**Tabela 4.21** Índices de avaliação dos modelos para o período 2 (inverno)

ÍNDICES	MODELO 1	MODELO 2	MODELO 3	MODELO 4	MODELO 5	MODELO 6	GANHO (%)
MAPE (%)	2,11	2,03	1,97	2,10	<b>1,85</b>	1,89	<b>5,89</b>
MSE (MW <sup>2</sup> )	5639,48	5153,34	4685,55	5464,65	<b>4186,22</b>	4261,25	<b>10,66</b>
ME (MW)	-7,07	-12,19	-14,58	-7,95	<b>-12,62</b>	-10,66	<b>13,44</b>
MAX (%)	5,44	5,25	4,72	5,66	<b>4,64</b>	4,79	<b>1,69</b>
MIN (%)	0,19	0,17	0,17	0,17	<b>0,14</b>	0,12	<b>17,65</b>

**Tabela 4.22** Índices de avaliação dos modelos para o período 3 (inverno)

ÍNDICES	MODELO 1	MODELO 2	MODELO 3	MODELO 4	MODELO 5	MODELO 6	GANHO (%)
MAPE (%)	2,77	2,68	2,11	2,82	<b>1,87</b>	1,90	<b>11,13</b>
MSE (MW <sup>2</sup> )	9443,39	8590,87	5867,75	9675,92	<b>4631,86</b>	5017,02	<b>21,06</b>
ME (MW)	5,50	-9,39	-9,96	-2,45	<b>-18,58</b>	-17,59	<b>-86,55</b>
MAX (%)	6,31	5,91	5,04	6,71	<b>4,28</b>	4,49	<b>15,08</b>
MIN (%)	0,36	0,38	0,16	0,34	<b>0,16</b>	0,15	-

<sup>5</sup> O ganho foi calculado em relação ao Modelo 3

## 4.14 Exemplo de previsões

### 4.14.1 Exemplo 1

Na Figura 4.17 é apresentado um exemplo de previsão de carga de 1 a 24 passos à frente para o dia 31/01/1995, segunda-feira. As previsões foram realizadas pelos melhores modelos de cada um dos dois grupos para o período do verão. A Figura 4.18 mostra os erros para os conjuntos de treinamento e teste durante o processo de aprendizagem de cada um dos modelos utilizados para realizar as previsões. A Tabela 4.23 apresenta os resultados obtidos.

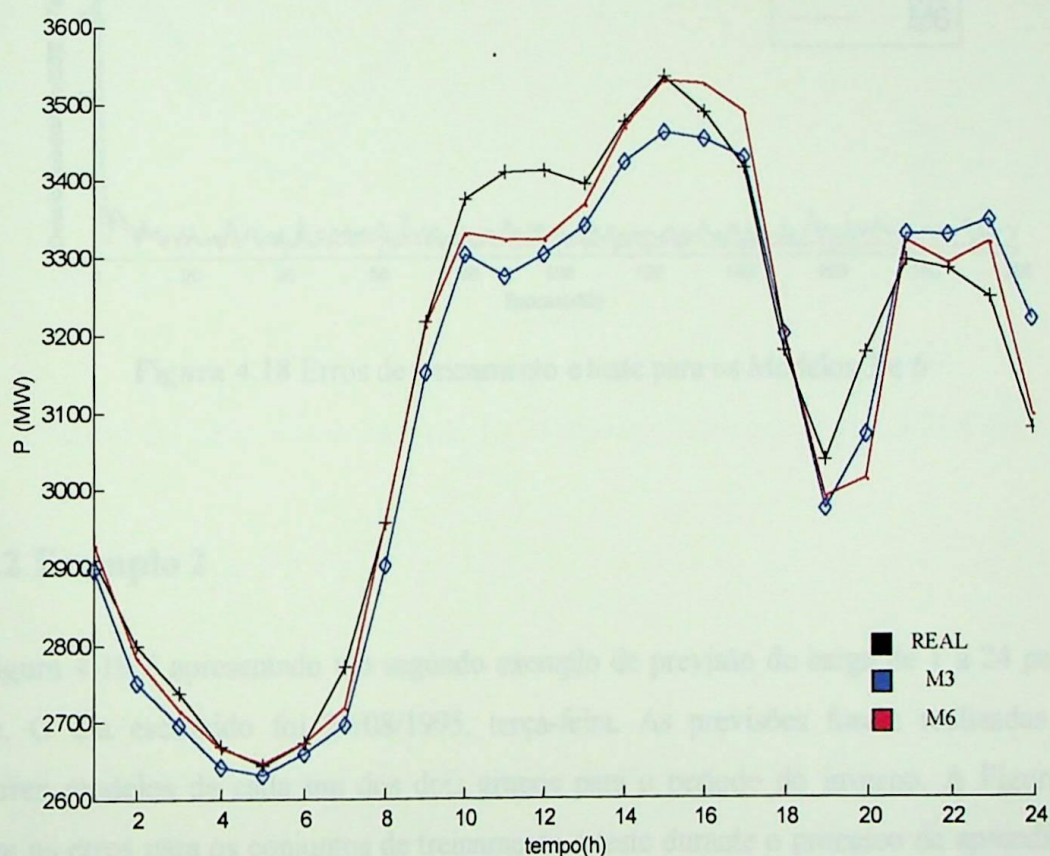


Figura 4.17 Previsão de carga de 1 a 24 passos à frente para o dia 31/01/1995

Tabela 4.23 Índices de avaliação dos Modelos 3 e 6 para o dia 31/01/1995

ÍNDICE	MODELO 3	MODELO 6	GANHO (%)
MAPE (%)	1,85	1,11	40
MSE (MW <sup>2</sup> )	4818,20	2715,06	43,65
ME (MW)	-29,01	-11,98	58,70
MAX (%)	4,62	5,19	-12,34
MIN (%)	0,37	0,03	91,89

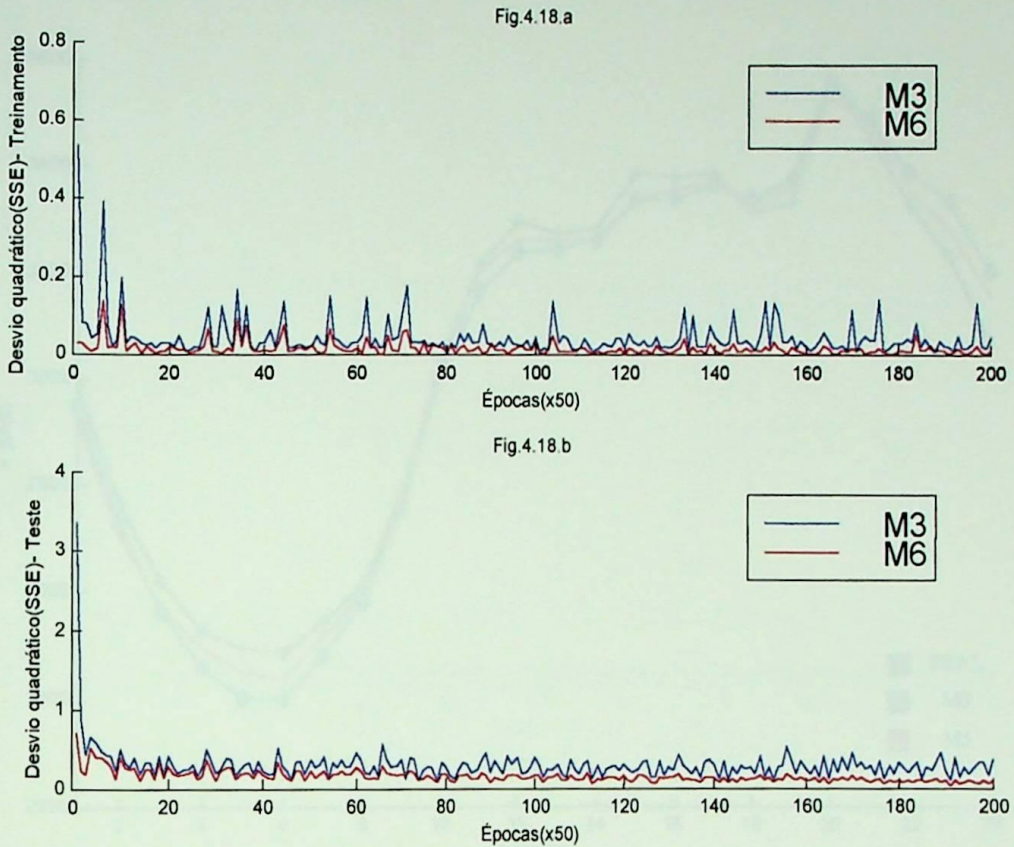


Figura 4.18 Erros de treinamento e teste para os Modelos 3 e 6

Figura 4.19 Previsão de carga de 1 a 24 passos à frente para o dia 29/08/1995

#### 4.14.2 Exemplo 2

Na Figura 4.19 é apresentado um segundo exemplo de previsão de carga de 1 a 24 passos à frente. O dia escolhido foi 29/08/1995, terça-feira. As previsões foram realizadas pelos melhores modelos de cada um dos dois grupos para o período do inverno. A Figura 4.20 mostra os erros para os conjuntos de treinamento e teste durante o processo de aprendizagem de cada um dos modelos utilizados para realizar as previsões. A Tabela 4.24 apresenta os resultados obtidos.



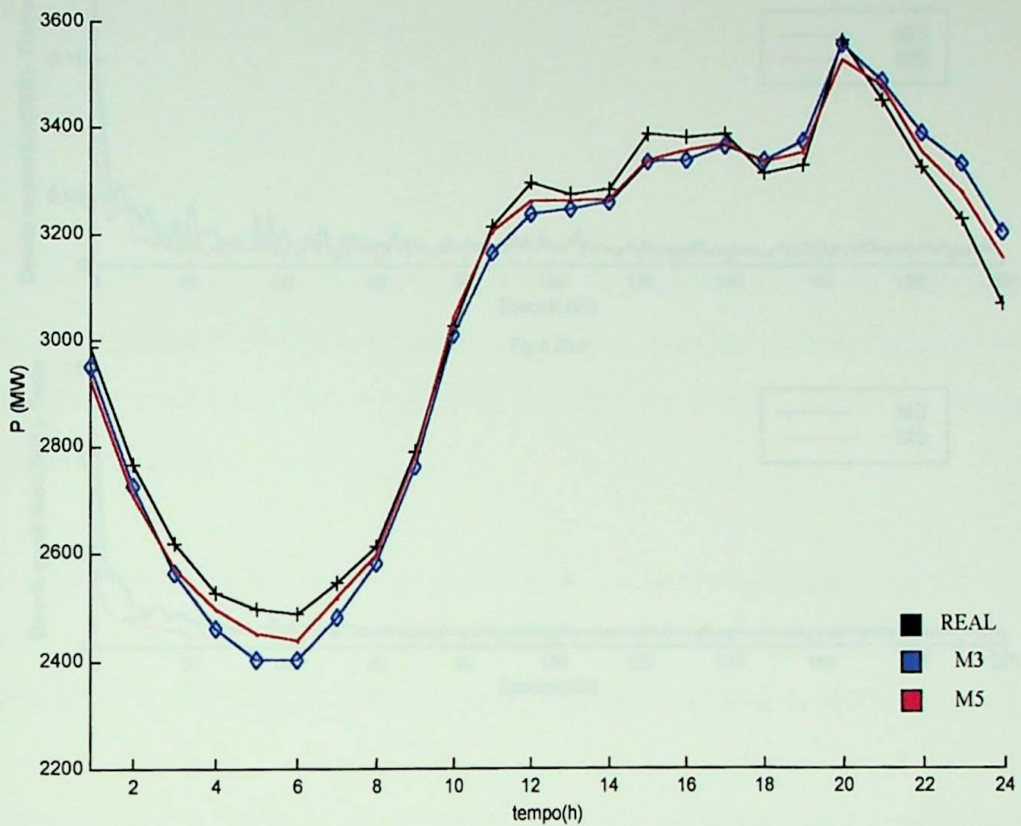


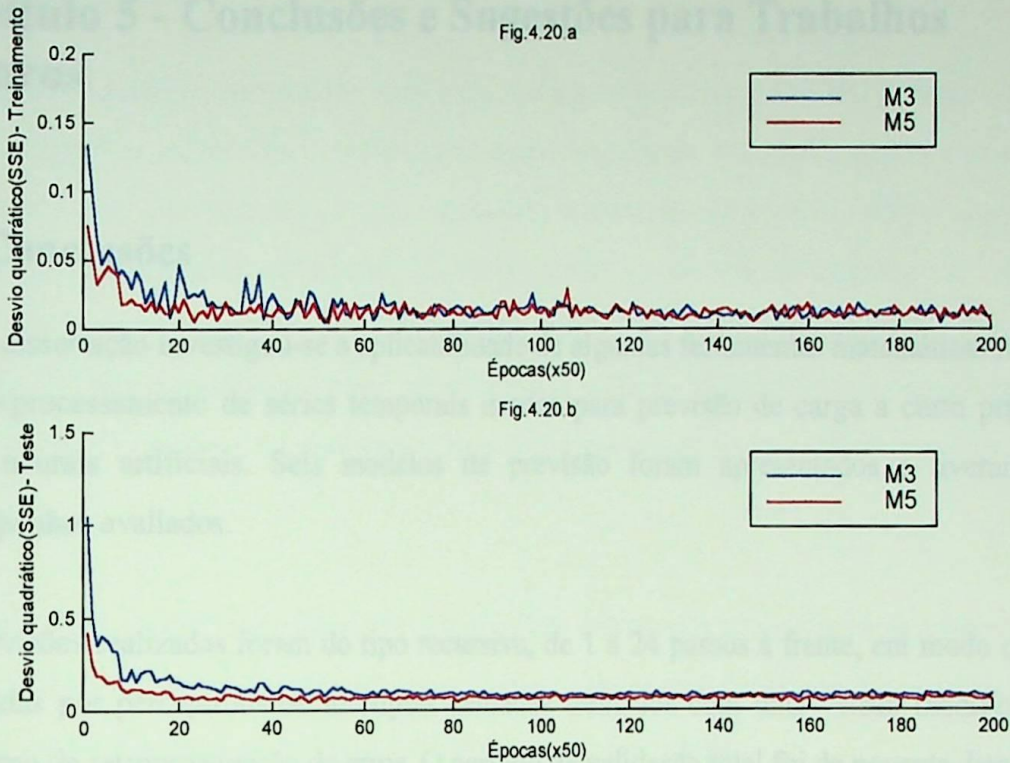
Figura 4.19 Previsão de carga de 1 a 24 passos à frente para o dia 29/08/1995

#### 4.15 Esforço computacional

O tempo médio da etapa de treinamento/teste para cada uma das redes simuladas foi de 2 minutos. A máquina utilizada contém alguns processadores de 160 Processadores MIPS, 232MBs.

Tabela 4.24 Índices de avaliação dos modelos 3 e 5 para o dia 29/08/1995

ÍNDICE	MODELO 3	MODELO 5	GANHO (%)
MAPE (%)	1,71	1,13	33,91
MSE (MW <sup>2</sup> )	3476,15	1516,27	56,38
ME (MW)	-15,90	-11,45	27,99
MAX (%)	4,47	2,95	34,00
MIN (%)	0,21	0,30	-42,86



**Figura 4.20** Erros de treinamento e teste para os Modelos 3 e 5

## 4.15 Esforço computacional

O tempo médio da etapa de treinamento/teste para cada uma das redes simuladas foi de 2 minutos. A máquina utilizada contava com um processador do tipo Pentium MMX, 233MHz, 64 Mbytes de RAM, com um HD de 1,96 Gbytes.

## Capítulo 5 - Conclusões e Sugestões para Trabalhos Futuros

### 5.1 Conclusões

Nessa dissertação investigou-se a aplicabilidade de algumas ferramentas matemáticas na etapa de pré-processamento de séries temporais usadas para previsão de carga a curto prazo via redes neurais artificiais. Seis modelos de previsão foram apresentados e tiveram seus desempenhos avaliados.

As previsões realizadas foram do tipo recursiva, de 1 a 24 passos à frente, em modo *off-line*, realizadas por perceptrons de múltiplas camadas treinados com uma versão modificada do algoritmo de retropropagação de erros. O período de validação total foi de noventa dias.

Dois tipos de normalização dos dados de entrada foram avaliados e seus resultados sugerem que a centralização dos dados é benéfica em alguns períodos do ano.

Técnicas de diferenciação de pontos adjacentes e desazonalização também foram avaliadas. A combinação entre séries de carga normalizadas via escores reduzidos e diferenciação apresentou erros percentuais absolutos médios (MAPE), na média, 20 % menores em relação a modelos alimentados com dados unicamente normalizados. Por outro lado, a combinação entre dados normalizados e desazonalizados não produziu resultados satisfatórios; o que permite concluir que a informação sazonal diária deve ser fornecida à rede de forma explícita.

Além dessas técnicas tradicionais, verificou-se a aplicabilidade de filtros digitais na previsão de séries temporais. A combinação entre série de carga normalizada, diferenciação e filtragem digital aumentou a exatidão das previsões, na média, de 5,89 % a 11,13 % em relação ao melhor modelo que não utilizou nenhum tipo de filtro. Esses resultados mostram que os filtros digitais podem contribuir para uma melhoria considerável na exatidão das previsões de carga realizadas por arquiteturas neurais.

## Referências

**5.2 Sugestões para trabalhos futuros**

A evolução natural desse trabalho seria partir para a implementação de um previsor operando em modo *on-line*. Além disso, os testes realizados deverão ser executados para um período de, no mínimo, um ano completo e para várias concessionárias de energia elétrica situadas em pontos geograficamente diferentes. Deve-se ainda considerar a inclusão de variáveis climáticas (e.g., temperatura, umidade relativa do ar, etc.). Esse tema tem sido amplamente discutido na literatura especializada, sendo que sua inclusão pode influenciar positivamente o desempenho do previsor. Seria desejável que essa nova versão também tratasse feriados e dias atípicos.

Os filtros digitais apresentados nesse trabalho são fortemente dependentes da qualidade dos dados utilizados como “padding”. Isso sugere que outros tipos de filtros possam ser empregados. Considera-se as “*Wavelets*” como fortes candidatas.

- [Bosner et al., 1997] BOSNER, J.M. et al., “Are Artificial Neural Networks Black Boxes?” *IEEE Transactions on Neural Networks*, Vol. 8, nº 3, pp. 1156-1164, Sep. 1997.
- [Box & Jenkins, 1976] BOX, G.E.P. & JENKINS, G.M., *Time Series Analysis: Forecasting and Control*. Holden Day, 1976.
- [Brockwell & Davis, 1996] BROCKWELL, P.J. & DAVIS, R.A., *Introduction to Time Series and Forecasting*. Springer-Verlag, NY, 1996.
- [Cherkassky & Muller, 1998] CHERKASSKY, V.S. & MULLER, F.M. *Learning from Data: Concepts, Theory and Methods*. John Wiley & Sons, Mar. 1998.
- [Chow & Leung, 1996] CHOW, T.W.S. & LEUNG, C.T. “Nonlinear Autoregressive Integrated Neural Network Model for Short-term Load Forecasting”. *IEE Proceedings – Generation, Transmission, Distribution*, Vol. 143, nº 5, pp. 500-506, Sep. 1996.

## Referências

---

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

[Acton, 1959] ACTON, F.S. *Analysis of Straight-Line Data*. Dover publications, NY, 1959.

[Amari et al., 1996] AMARI, S. et al., "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] BARAKAT, E.H. et al. "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.

[Benítez et al., 1997] BENÍTEZ, J.M. et al., "Are Artificial Neural Networks Black Boxes?". *IEEE Transactions on Neural Networks*, Vol.8, nº.5, pp.1156-1164, Sep.1997.

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

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

[Cherkassky & Mulier, 1998] CHERKASSKY, V.S. & MULIER, F.M. *Learning from Data: Concepts, Theory and Methods*. John Wiley & Sons, Mar.1998.

[Chow & Leung, 1996] CHOW, T.W.S & LEUNG, C.T "Nonlinear Autoregressive Integrated Neural Network Model for Short-term Load Forecasting". *IEE Proceedings – Generation, Transmission, Distribution*, Vol.143, nº.5, pp.500-506, Sep.1996.

- [Cigré, 1993] NIEBUR, D. et al. "Artificial Neural Networks for Power Systems: A Literature Survey". *International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications*, Vol.1, n°.3, pp.133-158, Dec.1993.
- [Crúz, 1996] CRÚZ, C.F.C. *Previsão da Carga Elétrica a Curto prazo via Redes Neurais*. Dissertação de Mestrado, Escola Federal de Engenharia de Itajubá, Itajubá, 1996.
- [Drezga & Rahman, 1998] DREZGA, I. & RAHMAN, S. "Input Variable Selection for ANN-Based Short-term Load Forecasting". *IEEE Transactions on power Systems*, vol.13, n°.4, pp.1238-1244, Nov.1998.
- [El-Sharkawi et al., 1999] EL-SHARKAWI, M.A. et al. "Short Term Load Forecast Using a Neuro-Fuzzy Model". *Proceedings of ISAP'99*, pp.8-13, 1999.
- [Etter, 1996] ETTER, D. M. *Introduction to Matlab for Engineers and Scientists*. Prentice Hall, inc., New Jersey, 1996.
- [Freund, 1988] FREUND, J. E. *Modern Elementary Statistics*. Prentice Hall Inc. New Jersey, 7<sup>th</sup> Ed., pp.7-8, 1988.
- [Goh et al., 1986] GOH, T.N 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.
- [Gross & Galiana, 1987] GROSS, G. & GALIANA F.D. "Short Term Load Forecasting". *Proceedings of IEEE*, vol.75, n°.4, pp. 1558-1573, Dec. 1987.
- [Haykin, 1999] HAYKIN, S. *Neural networks – A Comprehensive Foundation*. Prentice Hall, NJ, 2<sup>nd</sup> ed., 1999.
- [Heinemann et al., 1966] HEINEMANN, G.T. 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.

- [Holland, 1975] HOLLAND, J.H. *Adaptation in natural and Artificial Systems*. Ann Arbor, MI, University of Michigan Press, 1975.
- [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.
- [Kendal, 1976] KENDALL, M. & STUART, A. *The Advanced Theory of Statistics*. Hafner, Vol.III, NY, 1976.
- [Khotanzad et al., 1997] KHOTANZAD, A. et al. “ANNSTLF - A Neural-Network-Based Electric Load Forecasting System”. IEEE Transactions On Neural Networks, Vol.8, n°4, pp. 835-846, Jul. 1997.
- [Khotanzad et al., 1998] KHOTANZAD, A. et al. “ANNSTLF - Artificial Neural Network Short-Term Load Forecaster – Generation Three”. IEEE Transactions On Power Systems, Vol.13, n°4, pp.1413-1422, Nov. 1998.
- [Lamedica et al., 1996] LAMEDICA, R. et al. “A Neural Network Based Technique for Short-Term Forecasting of Anomalous Load Periods”. IEEE Transactions On Power Systems, Vol.11, n°4, pp.1749-1756, Nov. 1996.
- [Masters, 1993] MASTERS, T. *Practical Neural Networks Recipes in C++*. Academic Press, Inc., San Diego, 1993.
- [Masters, 1995] MASTERS, T. *Neural, Novel & Hybrid Algorithms for Time Series Prediction*. John Wiley & Sons, Inc., New York, 1995.
- [Morettin & Toloi, 1987] MORETTIN, P.A. & TOLOI, C.M.C. *Previsão de Séries Temporais*. Atual Editora Ltda., São Paulo, 2ª edição, 1987.
- [Mori, 1996] MORI, H. “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 Power Engineering Society, edited by M. El-Sharkawi and D. Niebur, Ch. 6, 1996, pp. 51-70.

[Moulin, 1998] MOULIN, L.S. *Intervalos de Confiança para a Previsão de Carga a Curto Prazo via Redes Neurais Artificiais*. Dissertação de mestrado, Escola Federal de Engenharia de Itajubá, Itajubá, 1998.

[Papalexopoulos, 1996] PAPALEXOPOULOS, A., "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, pp. 71-89, 1996.

[Peng et al., 1992] PENG, T.M. et al., "Advancement in the Application of Neural Networks for Short-Term Load Forecasting". IEEE Transactions on Power Systems, Vol.7, nº.1, pp.250-257, Feb.1992.

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

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

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

[Strang, 1986] STRANG, G. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Wellesley, 1986.

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

# Anexo I – Teoria sobre Redes Neurais Artificiais

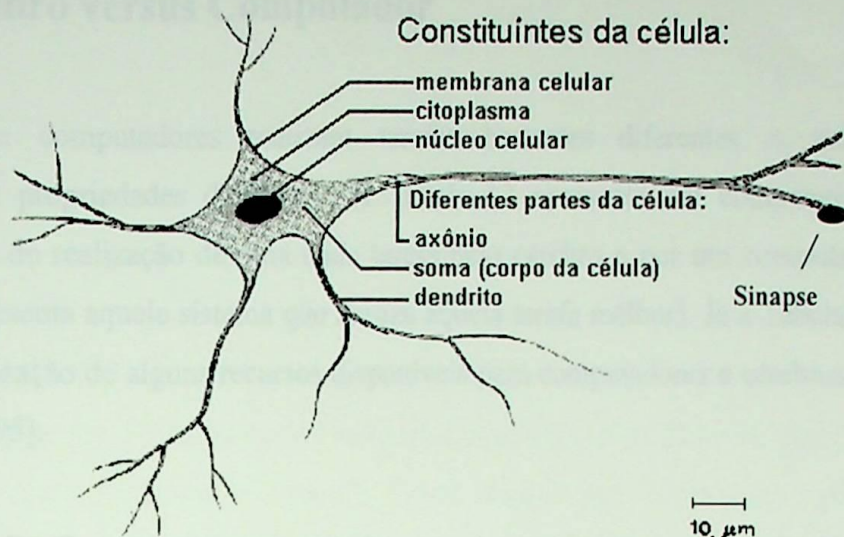
## I.1 Introdução

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.



**Figura 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 Cérebro versus Computador

Cérebros e computadores realizam tarefas bastantes diferentes e, naturalmente, apresentam propriedades diferentes. A Tabela I.1 apresenta um comparativo entre a capacidade de realização de uma dada tarefa pelo cérebro e por um computador digital (um x representa aquele sistema que realiza aquela tarefa melhor). Já a Tabela I.2 mostra uma comparação de alguns recursos disponíveis para computadores e cérebros [Russel & Norvig, 1995].

**Tabela I.1** Comparação entre o cérebro e o computador (para realização de tarefas)

<i>Tarefa</i>	<i>Cérebro</i>	<i>Computador</i>
Aritmética		X
Visão	X	
Tolerância a falhas	X	
Memória p/ fatos arbitrários		X
Memória endereçável por conteúdo para fatos do mundo real	X	

**Tabela I.2** Comparação entre o cérebro e o computador (recursos disponíveis)

	<i>Cérebro</i>	<i>Computador</i>
Unidades computacionais	$10^{11}$ neurônios	1 CPU, $10^5$ portas
Unidades de armazenamento	$10^{11}$ neurônios, $10^{14}$ sinapses	$10^9$ bits RAM, $10^{10}$ bits HD
Tempo de processamento	$10^{-3}$ s	$10^{-8}$ s
Largura de faixa	$10^{14}$ bits/s	$10^9$ bits/s
Número de atualizações/s	$10^{14}$	$10^5$

## I.4 Um Breve Histórico

McCulloch e Pitts (1943) introduziram a idéia fundamental de se analisar a atividade neural via limites e somas ponderadas. Hebb, em 1949, publicou “*The Organization of Behavior*” que foi decisivo em promover a hipótese de que a memória a longo prazo dos seres humanos e dos animais é possível devido a permanentes alterações nas sinapses. Frank Rosenblatt (1957) criou a estrutura denominada “*perceptron*”, composta por unidades treináveis. Praticamente de maneira simultânea, dispositivos similares chamados “*adalines*” (para “*Adaptive Linear*”) foram criados por Widrow and Hoff (1960), Widrow (1962).

A publicação de *Perceptrons* (Minsky e Papert, 1969) marcou o fim de uma era. Os autores criticaram severamente a experimentação desorientada e ausência de rigor matemático característicos dos primeiros trabalhos com perceptrons. Além disso, apontaram a limitação dos perceptrons de resolver somente problemas de natureza linearmente separável. Isso implicou em um período de “adormecimento” das redes neurais.

No entanto, em 1981, as publicações de Hinton e Anderson, baseadas em uma conferência realizada em San Diego (1979), marcaram o renascimento das redes neurais (ou Teoria Conexionista). Hopfield (1982) analisou redes simétricas usando mecânica estatística e analogias da física. A Máquina de Boltzmann (Hinton e Sejnowski, 1983; Hinton e Sejnowski, 1986) e a análise de redes neurais usando a “Teoria física dos *magnetic spin glasses*” estreitaram a relação entre a mecânica estatística e a teoria conexionista, dando várias idéias matemáticas úteis e também *respeitabilidade*. O algoritmo de treinamento usado nas redes neurais, “*Back-propagation*”, foi inventado por Werbos (1974). A expressividade de uma rede unidirecional de várias camadas foi investigada por Cybenko (1988; 1989), que mostrou que uma rede com duas camadas ocultas (ou intermediárias) é suficiente para representar qualquer função e uma rede de camada única é suficiente para representar qualquer função contínua.

O começo dos anos 90 foi marcado com o surgimento de sistemas híbridos. A título de exemplo, Miller et al. (1989) e Harp et al. (1990) usaram algoritmos genéticos<sup>1</sup> (GA) na determinação da estrutura ótima de uma rede neural.

Também contribuíram imensamente para o estabelecimento das redes neurais os seguintes pesquisadores: Grossberg (1968), Kohonen (1972; 1982), Rumelhart (1984), entre vários outros. Para um histórico mais abrangente ver [Russel & Norvig, 1995; Haykin, 1999].

## I.5 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.

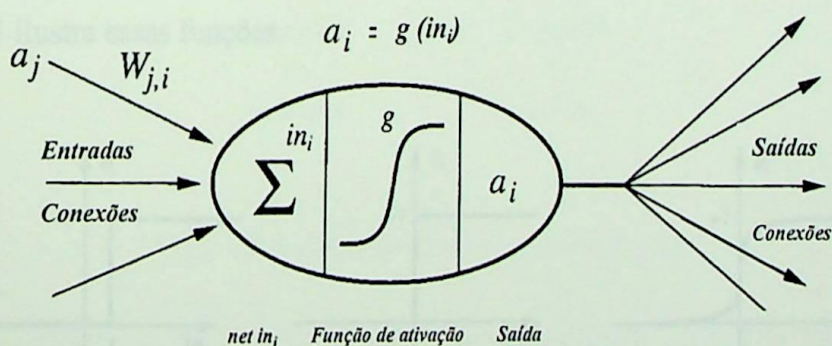


Figura I.2 Representação Matemática de uma Unidade

<sup>1</sup> Os GA são uma técnica de IA criada por J. Holland, inspirada na teoria de evolução animal [Holland, 1975].

Matematicamente falando,

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

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

### I.5.1 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:

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

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

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

A Figura I.3 ilustra essas funções.

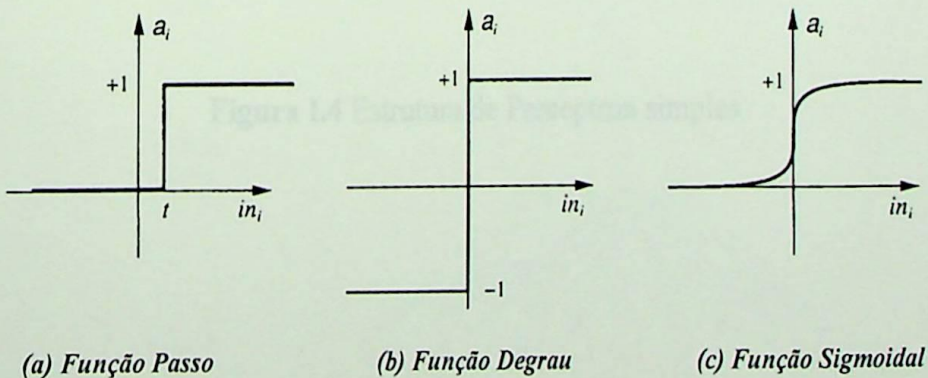
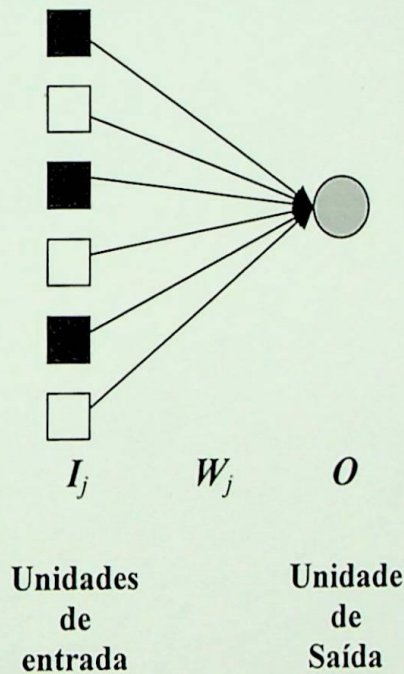


Figura I.3 Três diferentes funções de ativação

## I.6 Tipos de Redes Neurais Artificiais

Existe uma grande variedade de estruturas de redes, e.g., perceptrons de múltiplas camadas (MLP), 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.



**Figura I.4** Estrutura de Perceptron simples

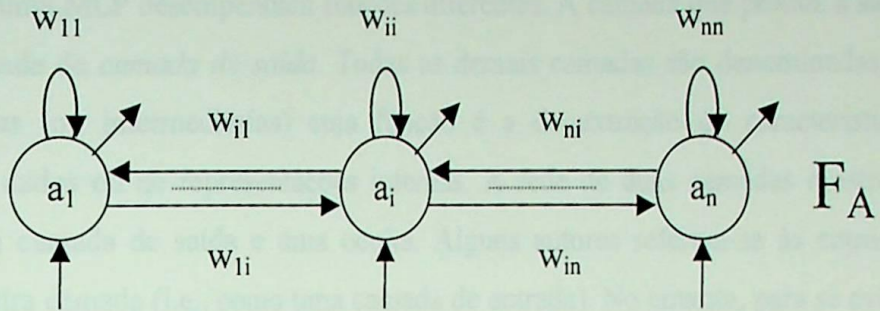


Figura I.5 Rede de Hopfield e Máquinas de Boltzman

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

## I.7 Perceptrons de Múltiplas Camadas

### I.7.1 Estrutura

Arquiteturas neurais do tipo MLP 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 MLP típica.

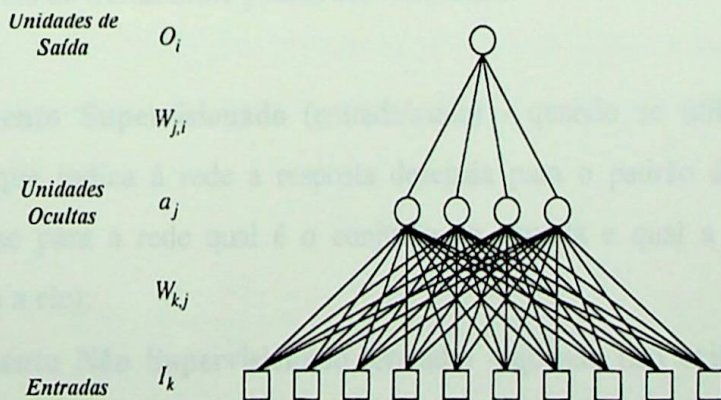


Figura I.6 Perceptron de Múltiplas Camadas

As camadas de uma MLP 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 MLP.

### I.7.2 Processo de Aprendizagem de uma MLP

Todo o “conhecimento” adquirido pela RNA está armazenado nos pesos relativos às conexões entre os seus neurônios [Rich & Knight, 1994]. Consequentemente, 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.7.3 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.7.4 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:

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

onde,

$t_{pj}$  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

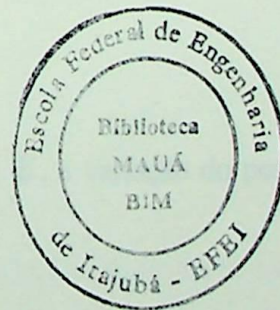
Para generalizar a regra delta padrão, primeiro define-se o erro correspondente a um padrão  $p$  de entrada-saída como sendo o erro quadrático parcial dado pela Eq.I.2. O erro quadrático global é dado por  $E = \sum E_p$  (que é calculado após a apresentação de todos os padrões utilizados no treinamento)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

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

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,

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

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

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

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

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

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:

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

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 MLP.

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:

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

onde,

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

$K$  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.

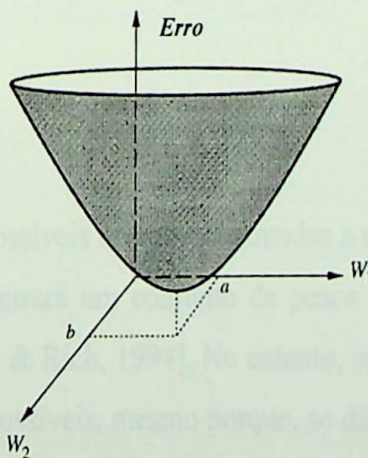
Um termo de *momentum* ( $\xi$ ) pode ser adicionado na equação I.17 com o objetivo de evitar que a rede se prenda a algum mínimo local da superfície de erro. A saber,

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

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.7.5 “Back-propagation” como busca em Gradiente Descendente

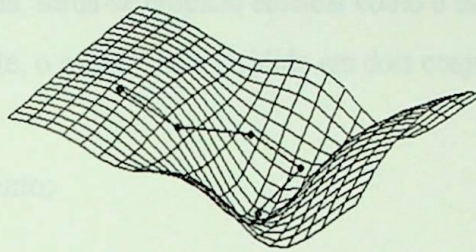
A superfície que descreve o erro em cada exemplo (ou no conjunto de treinamento como um todo) como uma função de todos os pesos na rede, é conhecida como *superfície de erro*. A Figura I.7 apresenta uma superfície de erro característica de problemas linearmente separáveis. O conjunto atual de pesos define 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  $w_2=b$ ).



**Figura I.7** Superfície de erro característica de problemas linearmente separáveis

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 (*simulated annealing*), 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.



**Figura 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.7.6 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., 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.

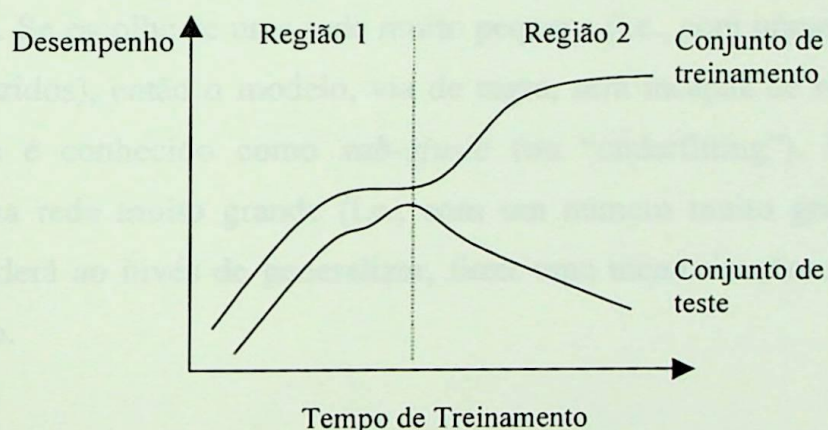
O conjunto de treinamento será formado por  $2/3$  do conjunto total<sup>2</sup> de dados e será usado efetivamente na tarefa de aprendizagem da rede conforme apresentado em seções anteriores. Por conseguinte, o conjunto de teste contará com o restante dos dados. A particularidade desse último conjunto 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

---

<sup>2</sup> esse valor é de caráter orientativo

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 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* (“Early Stopping Method”) 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.



**Figura I.9** Generalização em aprendizagem de RNA

### I.7.7 Estrutura de Rede Ótima

Até então, do ponto de vista prático, muito pouco foi dito a respeito da estrutura de RNA do tipo MLP. 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

<sup>3</sup> passagem da região 1 para a 2

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 conterá?

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 MLP. 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 MLP, 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).

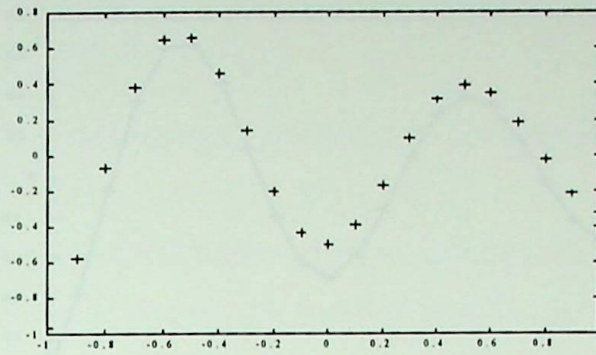


Figura I.10 Problema proposto

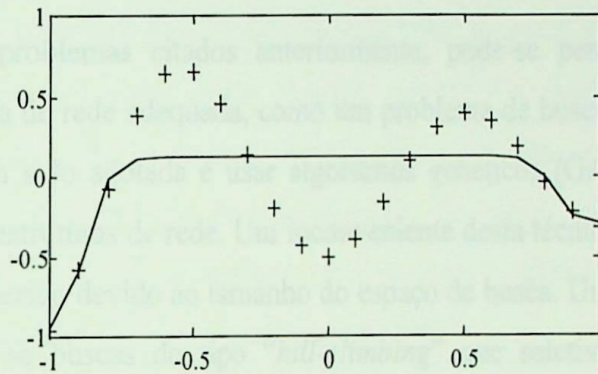


Figura I.11 “Underfitting”

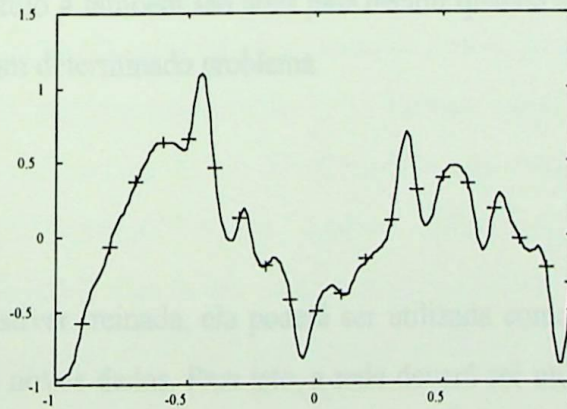
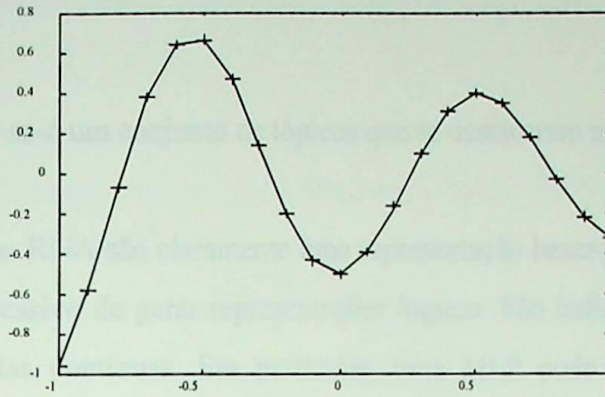


Figura I.12 “Overfitting”



**Figura I.13** Generalização adequada

Com respeito aos 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 apresentadas no capítulo 4 também são úteis para decidir quando a estrutura de rede ideal foi encontrada para um determinado problema.

### I.7.8 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.8 Discussão

A seguir, examinar-se-á um conjunto de tópicos que se destacaram nesse 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 MLP 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.9 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 [Cigré, 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.

De acordo com [Cigré, 1993], a proporção dos tipos de aplicações de RNA em SEP encontrava-se da seguinte forma:

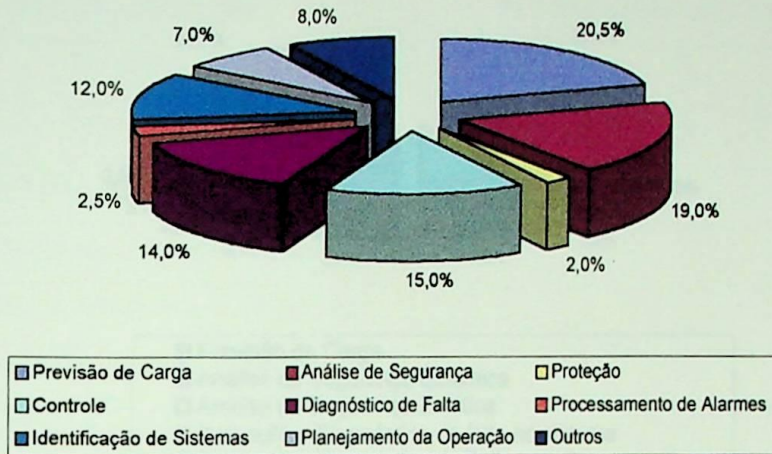


Figura I.14 Principais áreas de aplicação em abril de 1993

Já os modelos de RNA mais empregados são apresentados na Figura I.15, são eles:

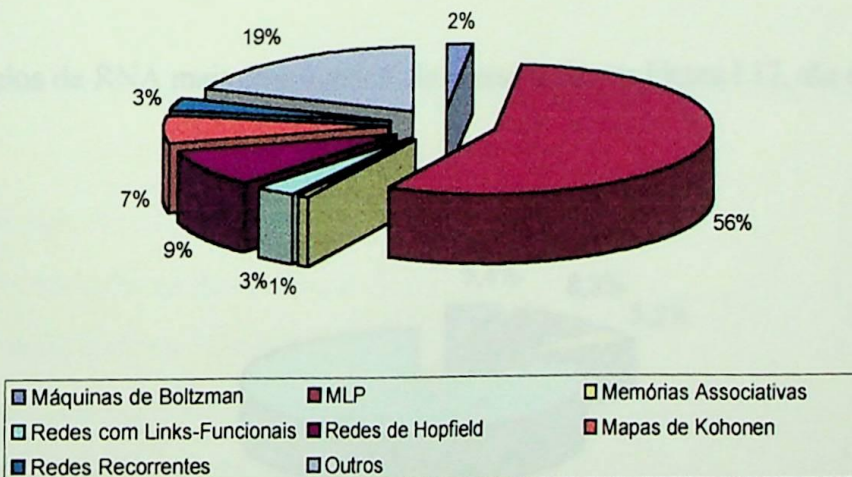


Figura I.15 Modelos de RNA usados em problemas de SEP (1993)

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

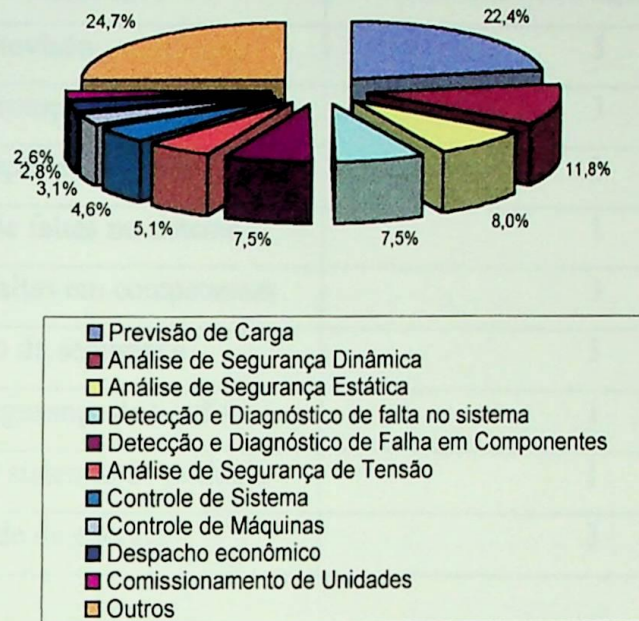


Figura I.16 Principais áreas de aplicação em 1995

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

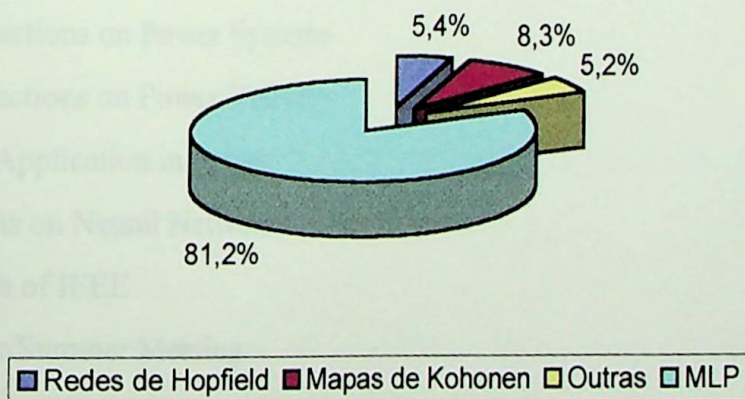


Figura I.17 Modelos de RNA usados em problemas de SEP (1995)

A Tabela I.3 mostra a área de aplicação das RNA em SEP e o respectivo número de artigos apresentados no ISAP 99, a saber:

**Tabela I.3** 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 Tabela I.3, nota-se que as RNA estão sendo aplicadas em outras áreas dos SEP (e.g., Qualidade de Energia).

## I.10 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 Application)
- PSCC(Power System Computation Conference)
- ISAP (International Conference on Intelligent System Application to Power Systems; Rio de Janeiro, Brasil, 1999)

## I.11 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 MLP podem representar qualquer função (desde que possua uma topologia adequada para tanto).
4. O algoritmo de *Back-propagation* é aplicado em redes do tipo MLP. 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*.

No entanto, esse não é o único modo de se expressar a informação contida nos dados. Na seção 3.4, mostrou-se que pode ser útil computar a *variação* de cada amostra relativa a amostra anterior. Embora as duas séries sejam equivalentes (do ponto de vista de conterem a mesma informação), elas não têm necessariamente a mesma utilidade.

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, aquelas usadas nessa dissertação serão estritamente reais.

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:

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

Essa equação 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.

---

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

**Tabela II.1** Aplicação da DFT

Ponto	Série	Parte real da DFT	Parte Imag. da DFT
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:

- $W_i$  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.

Os conceitos e equações até agora apresentados são suficientes para a implementação dos filtros digitais usados nessa dissertação. 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, grosseiramente, 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 adequado e suficiente. Captura-se todas as freqüências audíveis.

## 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 freqüência 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:

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

A Eq.II.2 é extremamente útil. Ela mostra como usar a DFT para estimar a energia contida em cada faixa de freqüência 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.

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