

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO
DEPARTAMENTO DE ENGENHARIA DE SISTEMAS

Análise e Previsão de Vazões Utilizando Modelos de Séries Temporais, Redes Neurais e Redes Neurais Nebulosas

Por: Rosangela Ballini

Orientador: Marinho Gomes de Andrade Filho

Co-orientador: Secundino Soares Filho

Comissão Julgadora:

Prof. Dr. Marinho Gomes de Andrade Filho – ICMC/USP-São Carlos

Prof. Dr. Alexandre Pinto Alves da Silva – EFEI/Itajubá

Dr. Paulo Roberto de Holanda Sales – ELETROBRÁS/Rio de Janeiro

Prof. Dr. João Bosco Ribeiro do Val – FEEC/UNICAMP

Prof. Dr. Fernando Antônio Campos Gomide – FEEC/UNICAMP

Prof. Dr. Fernando José Von Zuben – FEEC/UNICAMP

Tese apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, como parte dos requisitos exigidos para a obtenção do Título de Doutor em Engenharia Elétrica.

29 de Setembro de 2000

i

| |
|--|
| Este exemplar corresponde a redação final da tese defendida por <u>Rosangela Ballini</u> |
| aprovada pela Comissão |
| Julgada em <u>29</u> / <u>09</u> / <u>2000</u> |
|  Orientador |

71707008

N.º CHAMADA:
UNICAMP
B214a

V. _____ Ex. _____

TOMBO BC/ 44344

PROC. 16.892101

C D

PREC. R\$ 11,00

DATA 10/05/01

N.º CPD _____

CM-00155176-9

FICHA CATALOGRÁFICA ELABORADA PELA
 BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

B214a Ballini, Rosangela

 Análise e previsão de vazões utilizando modelos de séries temporais, redes neurais e redes neurais nebulosas / Rosangela Ballini, - Campinas, SP:[s.n.], 2000.

 Orientadores: Marinho Gomes de Andrade Filho, Secundino Soares Filho.

 Tese (doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

 1. Previsão Hidrológica. 2. Análise de Séries Temporais. 3. Redes Neurais (computação). 4. Conjuntos Difusos. I. Andrade Filho, Marinho Gomes de. II. Soares Filho, Secundino. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

UNICAMP
BIBLIOTECA CENTRAL
SECÃO CIRCULANTE

Aos meus pais.

Este trabalho teve suporte financeiro do **CNPq** e **FAPESP** através do projeto temático de equipe sobre o Planejamento e Operação de Sistemas de Energia Elétrica com Predominância de Geração Hidroelétrica.

Resumo

Análise e previsão de vazões são de fundamental importância no planejamento da operação de sistemas de recursos hídricos. Uma das grandes dificuldades na previsão das séries de vazões é a presença da sazonalidade devido aos períodos de cheia e seca do ano. Os modelos estocásticos foram, por um longo tempo, a alternativa mais comum aos modelos determinísticos ou hidrológicos na análise e previsão de vazões, baseados principalmente na metodologia de Box & Jenkins. Esta metodologia exige algum tipo de manuseio nos dados para tratar a não-estacionariedade ou o uso de modelos periódicos, necessitando de uma laboriosa formulação teórica para os procedimentos estatísticos.

Redes neurais artificiais, especialmente redes multi-camadas com algoritmo back-propagation vêm sendo sugeridas para análise de séries temporais devido a sua capacidade para tratar com relações não-lineares de entrada-saída, destacando sua habilidade de aprendizado e capacidade de generalização, associação e busca paralela. Estas qualidades as tornam capazes de identificar e assimilar as características mais marcantes das séries, tais como sazonalidade, periodicidade, tendência, entre outras, muitas vezes camufladas por ruídos. A capacidade de mapeamentos complexos das redes neurais cresce com o número de camadas e neurônios, acarretando maior tempo de processamento bem como considerável soma de dados. Entretanto, na prática muitas vezes os parâmetros devem ser estimados rapidamente e somente uma pequena quantidade de dados é disponível. Frequentemente, dados do mundo real apresentam ruídos, podendo conter contradições e imperfeições. Tolerância a imprecisão e incertezas é também exigida para considerar tratabilidade e robustez. Conjuntos nebulosos baseados em modelos de análise de dados vêm sendo empregados sob essas hipóteses. A aplicação de modelos de redes neurais nebulosas une os benefícios das redes neurais e da teoria de conjuntos nebulosos, combinando-os em um sistema integrado para previsão de vazões naturais médias mensais.

São realizadas análise e previsão de vazões usando modelos de séries temporais, redes neurais e redes neurais nebulosas para previsão um passo à frente e vários passos à frente para as séries das usinas hidroelétricas brasileiras localizadas em diferentes regiões. O desempenho dos modelos foi comparado e os resultados mostraram que os modelos propostos apresentaram melhor desempenho que as outras abordagens tanto para previsão um passo à frente como para previsão com vários passos à frente.

Abstract

Analysis and forecast of seasonal streamflow series are of utmost importance in the operation planning of water resources systems. One of the greatest difficulties in forecasting of those series is the seasonality nature of streamflow series due to wet and dry periods of the year. For a long time, the use of stochastic models, based on the classic Box & Jenkins methodology, were the most employed alternative to the deterministic or hydrologic models in the analysis and forecast of streamflow series. This methodology requires either some kind of data manipulation to deal with the nonstationarity or the use of periodic models. Therefore the statistical procedures, requires an arduous theoretical formulation.

Artificial Neural Networks (ANN), specially multilayer perceptrons with a back-propagation algorithm, have recently been suggested for time series analysis. They have the ability to deal with nonlinear input-output relationships. Their major assets are the learning ability and generalization, association and parallel search capability. These qualities enable them to identify and to assimilate some of the features of the series as seasonality, periodicity, tendency sometimes difficult to detect under noise. The capability of complex mapping of the ANN increases with the number of layers and neurons. The use of ANN usually requires the investment of a long period of time in the modeling process, as well as a considerable amount of data. In practice, however, the parameters usually must be quickly estimated and only a small quantity of data is available. Very often, real world data are noisy, and the collected data may contain contradictions and imperfections. Tolerance for imprecision and uncertainty is also required to achieve tractability and robustness. Fuzzy sets based data analysis models have been especially suitable for these purposes. This suggests the application of neurofuzzy network models to seasonal streamflow forecasting. These models combine the advantages of the ANN and fuzzy set based approaches in a single integrated decision-making system.

Analysis and forecast of streamflows one-step-ahead and multi-step-ahead are accomplished, using time series models, neural networks, and neurofuzzy networks. Database of average monthly inflows from Brazilian hydroelectric plants located in different river basins were used. The performance of the models was compared and the results show that the models here proposed provide a better performance than the others ones considering one-step-ahead forecasting and multi-step-ahead forecasting.

Agradecimentos

À minha família pelo carinho e apoio.

Aos Professores Marinho Gomes de Andrade e Secundino Soares pela orientação.

Aos Professores Fernando Gomide e Fernando Von Zuben pelas críticas e sugestões.

Aos amigos Pedro, Luciana, Cilene e Leonardo pelo apoio, carinho e ajuda.

Aos amigos do COSE.

Conteúdo

| | |
|--|------------|
| RESUMO | iv |
| ABSTRACT | v |
| AGRADECIMENTOS | vi |
| CONTEÚDO | vii |
| LISTA DE FIGURAS | xi |
| LISTA DE TABELAS | xv |
| 1 Introdução | 1 |
| 1.1 Descrição do Problema e Revisão Bibliográfica | 1 |
| 1.2 Motivação para a Utilização de Redes Neurais Nebulosas | 7 |
| 1.3 Objetivos do Trabalho | 8 |
| 1.4 Organização do Trabalho | 9 |
| 2 Análise de Séries Temporais | 11 |
| 2.1 Introdução | 11 |
| 2.2 Conceitos de Processos Estocásticos | 14 |
| 2.2.1 Média e Variância de um Processo Estacionário | 15 |
| 2.2.2 Coeficientes de Autocovariância e Autocorrelação | 16 |

| | | |
|----------|--|-----------|
| 2.3 | Funções de Autocovariância e Autocorrelação | 17 |
| 2.3.1 | Estimação das Funções de Autocovariância e Autocorrelação | 18 |
| 2.4 | Análise da Série de Vazões | 19 |
| 2.5 | Resumo | 24 |
| 3 | Modelos de Séries Temporais | 25 |
| 3.1 | Introdução | 25 |
| 3.2 | Modelos Estacionários Lineares | 27 |
| 3.3 | Modelos Auto-Regressivos | 29 |
| 3.3.1 | Função de Autocorrelação de Modelos Auto-Regressivos | 31 |
| 3.3.2 | Função de Autocorrelação Parcial de Modelos Auto-Regressivos | 33 |
| 3.3.3 | Erro Padrão das Estimativas de Autocorrelação Parcial | 35 |
| 3.3.4 | Ajuste de um Modelo Auto-Regressivo para a Série de Vazões | 36 |
| 3.4 | Modelos Periódicos Auto-Regressivos | 42 |
| 3.4.1 | Inferência Bayesiana | 44 |
| 3.4.2 | Densidade <i>A Priori</i> Não Informativa | 45 |
| 3.4.3 | Densidade Preditiva | 46 |
| 3.4.4 | Seleção de Modelos | 48 |
| 3.4.5 | Ajuste de Modelos Periódicos Auto-Regressivos para a Série de Vazões | 49 |
| 3.5 | Resumo | 57 |
| 4 | Redes Neurais Multi-Camadas | 59 |
| 4.1 | Introdução | 59 |
| 4.1.1 | Um Breve Histórico da Área de Redes Neurais | 60 |
| 4.2 | Rede Perceptron Multi-Camadas | 62 |
| 4.3 | Algoritmo de Retro-Propagação | 64 |
| 4.3.1 | Função de Ativação | 69 |

| | | |
|----------|---|------------|
| 4.3.2 | Taxa de Aprendizagem | 70 |
| 4.3.3 | Inicialização dos Pesos | 71 |
| 4.3.4 | Aprendizagem Entrada-a-Entrada e Aprendizagem por Ciclo | 71 |
| 4.3.5 | Critério de Parada | 73 |
| 4.3.6 | Generalização | 73 |
| 4.3.7 | Validação Cruzada | 74 |
| 4.3.8 | Aproximação de Funções | 75 |
| 4.4 | Algoritmo de Retro-Propagação Modificado | 76 |
| 4.4.1 | Regra de Aprendizagem Delta-Bar-Delta | 77 |
| 4.5 | Aplicação da Rede MLP para Previsão de Vazões | 79 |
| 4.6 | Resumo | 87 |
| 5 | Modelo de Rede Neural Nebulosa | 89 |
| 5.1 | Introdução | 89 |
| 5.2 | Rede Neural Nebulosa | 91 |
| 5.2.1 | Estrutura da Rede Neural Nebulosa | 92 |
| 5.2.2 | Métodos de Aprendizagem | 97 |
| 5.2.3 | Método de Aprendizagem <i>Off-Line</i> | 99 |
| 5.2.4 | Método de Aprendizagem <i>On-Line</i> | 104 |
| 5.2.5 | Aplicação da Rede Neural Nebulosa para Previsão de Vazões | 109 |
| 5.3 | Resumo | 115 |
| 6 | Comparação dos Modelos | 117 |
| 6.1 | Introdução | 117 |
| 6.2 | Análise e Ajuste dos Dados de Vazões para Emborcação e Sobradinho | 118 |
| 6.3 | Ajuste do Modelo PREVAZ | 127 |
| 6.4 | Análise dos Erros | 130 |

| | | |
|----------|--|------------|
| 6.4.1 | Análise dos Erros de Previsão para o Posto de Furnas | 131 |
| 6.4.2 | Análise dos Erros de Previsão para o Posto de Emborcação | 142 |
| 6.4.3 | Análise dos Erros de Previsão para o Posto de Sobradinho | 145 |
| 6.5 | Resumo | 149 |
| 7 | Conclusão | 151 |
| 7.1 | Trabalhos Futuros | 154 |
| | BIBLIOGRAFIA | 155 |
| | APÊNDICE | 165 |
| A | Fundamentos de Sistemas Nebulosos | 165 |

Lista de Figuras

| | | |
|-----|---|----|
| 1.1 | Obtenção de Séries Históricas de Vazões Afluentes. | 2 |
| 2.1 | Série de Vazões Afluentes Médias Mensais para a Usina Hidroelétrica de Furnas. | 13 |
| 2.2 | Média e Desvio Padrão Mensal para a Série de Vazões do Posto de Furnas. | 20 |
| 2.3 | Série Padronizada para os Dados de Vazões do Posto de Furnas. | 21 |
| 2.4 | Histograma da Série Estacionária para os Dados de Vazões do Posto de Furnas. | 22 |
| 2.5 | Função de Autocorrelação para a Série de Vazões do Posto de Furnas. | 23 |
| 2.6 | Diagramas de Dispersão para a Série de Vazões do Posto de Furnas. | 23 |
| 3.1 | Representação de uma Série Temporal como Saída de um Filtro. Linear | 27 |
| 3.2 | Função de Autocorrelação Parcial para a Série de Vazões do Posto de Furnas. | 38 |
| 3.3 | Gráfico Normal-Plote (Resíduo a_t). | 38 |
| 3.4 | Função de Autocorrelação do Resíduo a_t | 39 |
| 3.5 | Previsão de Vazões Usando o Modelo AR(2) para o Posto de Furnas - Período de 1952 a 1956. | 40 |
| 3.6 | Previsão de Vazões Usando o Modelo AR(2) para o Posto de Furnas - Período de 1972 a 1976. | 40 |
| 3.7 | Previsão de Vazões Usando o Modelo AR(2) para o Posto de Furnas - Período de 1981 a 1985. | 41 |

| | | |
|------|--|----|
| 3.8 | Previsão de Vazões Usando o Modelo PAR(2) para o Posto de Furnas - Período de 1952 a 1956. | 51 |
| 3.9 | Previsão de Vazões Usando o Modelo PAR(2) para o Posto de Furnas - Período de 1972 a 1976. | 52 |
| 3.10 | Previsão de Vazões Usando o Modelo PAR(2) para o Posto de Furnas - Período de 1981 a 1985. | 52 |
| 3.11 | Previsão de Vazões Usando o Modelo PAR(p_m) para o Posto de Furnas - Período de 1952 a 1956. | 55 |
| 3.12 | Previsão de Vazões Usando o Modelo PAR(p_m) para o Posto de Furnas - Período de 1972 a 1976. | 56 |
| 3.13 | Previsão de Vazões Usando o Modelo PAR(p_m) para o Posto de Furnas - Período de 1981 a 1985. | 56 |
| 4.1 | Modelo do Neurônio Biológico. | 60 |
| 4.2 | Modelo do Neurônio Artificial. | 61 |
| 4.3 | Rede Neural Perceptron Multi-Camadas. | 64 |
| 4.4 | Saída do Neurônio. | 66 |
| 4.5 | Saída do Neurônio Intermediário. | 68 |
| 4.6 | Função Logística e Derivada. | 70 |
| 4.7 | Representação do Modelo de Previsão k Passos à Frente. | 82 |
| 4.8 | Previsão de Vazões Usando uma Rede Neural MLP para o Posto de Furnas - Período de 1952 a 1956. | 83 |
| 4.9 | Previsão de Vazões Usando uma Rede Neural MLP para o Posto de Furnas - Período de 1972 a 1976. | 83 |
| 4.10 | Previsão de Vazões Usando uma Rede Neural MLP para o Posto de Furnas - Período de 1981 a 1985. | 84 |
| 4.11 | Previsão de Vazões Usando 12 Modelos de Redes Neurais MLP para o Posto de Furnas - Período de 1952 a 1956. | 85 |
| 4.12 | Previsão de Vazões Usando 12 Modelos de Redes Neurais MLP para o Posto de Furnas - Período de 1972 a 1976. | 86 |

| | | |
|------|--|-----|
| 4.13 | Previsão de Vazões Usando 12 Modelos de Redes Neurais MLP para o Posto de Furnas - Período de 1981 a 1985. | 86 |
| 5.1 | Função de Pertinência Discretizada. | 93 |
| 5.2 | Modelo Geral do Neurônio Nebuloso. | 95 |
| 5.3 | Rede Neural Nebulosa. | 96 |
| 5.4 | Função de Decodificação. | 97 |
| 5.5 | Previsão de Vazões Usando um Modelo de RNN com Algoritmo <i>On-line</i> para o Posto de Furnas - Período de 1952 a 1956. | 111 |
| 5.6 | Previsão de Vazões Usando um Modelo de RNN com Algoritmo <i>On-line</i> para o Posto de Furnas - Período de 1972 a 1976. | 112 |
| 5.7 | Previsão de Vazões Usando um Modelo de RNN com Algoritmo <i>On-line</i> para o Posto de Furnas - Período de 1981 a 1985. | 112 |
| 5.8 | Previsão de Vazões Usando um Modelo de RNN com Algoritmo <i>On-line</i> para o Posto de Furnas - Período de 1952 a 1956. | 113 |
| 5.9 | Previsão de Vazões Usando um Modelo de RNN com Algoritmo <i>On-line</i> para o Posto de Furnas - Período de 1972 a 1976. | 114 |
| 5.10 | Previsão de Vazões Usando um Modelo de RNN com Algoritmo <i>On-line</i> para o Posto de Furnas - Período de 1981 a 1985. | 114 |
| 6.1 | Série de Vazões Afluentes Médias Mensais para a Usina Hidroelétrica de Emborcação. | 119 |
| 6.2 | Série de Vazões Afluentes Médias Mensais para a Usina Hidroelétrica de Sobradinho. | 120 |
| 6.3 | Média e Desvio Padrão Mensal para as Séries de Vazões dos Postos de Emborcação (a) e Sobradinho (b). | 121 |
| 6.4 | Função de Autocorrelação para as Séries de Vazões dos Postos de Emborcação (a) e Sobradinho (b). | 121 |
| 6.5 | Diagramas de Dispersão para a Série de Vazões do Posto de Emborcação. | 122 |
| 6.6 | Diagramas de Dispersão para a Série de Vazões do Posto de Sobradinho. | 122 |

| | | |
|------|---|-----|
| 6.7 | Função de Autocorrelação Parcial para as Séries de Vazões de Emborcação (a) e Sobradinho (b). | 123 |
| 6.8 | Previsão de Vazões Usando o Modelo PREVAZ para o Posto de Furnas - Período de 1952 a 1956. | 128 |
| 6.9 | Previsão de Vazões Usando o Modelo PREVAZ para o Posto de Furnas - Período de 1972 a 1976. | 129 |
| 6.10 | Previsão de Vazões Usando o Modelo PREVAZ para o Posto de Furnas - Período de 1981 a 1985. | 129 |
| 6.11 | Erros de Previsão de Vazões para o Posto de Furnas - Período de 1952 a 1956. | 134 |
| 6.12 | Erros de Previsão de Vazões para o Posto de Furnas - Período de 1972 a 1976. | 138 |
| 6.13 | Erros de Previsão de Vazões para o Posto de Furnas - Período de 1981 a 1985. | 141 |
| 6.14 | Erros de Previsão de Vazões para o Posto de Emborcação - Período de 1972 a 1976. | 145 |
| 6.15 | Erros de Previsão de Vazões para o Posto de Sobradinho - Período de 1972 a 1976. | 148 |
| A.1 | Função de Pertinência Triangular. | 167 |
| A.2 | Função de Pertinência Gaussiana. | 168 |

Lista de Tabelas

| | | |
|-----|--|----|
| 2.1 | Média e Desvio Padrão para a Série de Vazões do Posto de Furnas. | 20 |
| 2.2 | Estimativas da Função de Autocorrelação para a Série de Vazões do Posto de Furnas. | 22 |
| 3.1 | Variância dos Resíduos para Diferentes Modelos. | 36 |
| 3.2 | Estimativas dos Coeficientes de Autocorrelação Parcial para a Série de Vazões do Posto de Furnas. | 37 |
| 3.3 | Médias Mensais e Valores Previstos com o Modelo AR(2) para a Série de Vazões do Posto de Furnas - m^3/s | 41 |
| 3.4 | Valores Estimados dos Parâmetros para a Série de Vazões do Posto de Furnas. | 50 |
| 3.5 | Estimativas para a Seleção do Modelo para a Série de Vazões do Posto de Furnas. | 51 |
| 3.6 | Médias Mensais e Valores Previstos com o Modelo PAR(2) para a Série de Vazões do Posto de Furnas - m^3/s | 53 |
| 3.7 | Ordem dos Modelos Seleccionados para a Série de Vazões do Posto de Furnas. | 54 |
| 3.8 | Valores Estimados dos Parâmetros Usando Inferência Bayesiana para a Série de Vazões do Posto de Furnas. | 55 |
| 4.1 | Escolha da Topologia da Rede Neural Multi-Camadas. | 81 |
| 4.2 | Topologia Escolhida para os Períodos Considerados Para o Posto de Furnas. | 81 |
| 4.3 | Configuração da Rede MLP para cada Mês. | 85 |

| | | |
|------|---|-----|
| 6.1 | Média e Desvio Padrão para as Série de Vazões dos Postos de Emborcação e Sobradinho - m^3/s | 120 |
| 6.2 | Estimativas dos Parâmetros para as Séries de Vazões dos Postos de Emborcação e Sobradinho. | 124 |
| 6.3 | Estimativas dos Parâmetros e Ordem dos Modelos Seleccionados para a Série de Vazões do Posto de Emborcação. | 125 |
| 6.4 | Estimativas dos Parâmetros e Ordem dos Modelos Seleccionados para a Série de Vazões do Posto de Sobradinho. | 126 |
| 6.5 | Topologia da Rede MLP para cada Mês para os Postos de Emborcação e Sobradinho - Período de 1972 - 1976. | 127 |
| 6.6 | Médias Mensais e Valores Previstos com o Modelo PREVAZ para a Série de Vazões de Furnas, Previsão 6 Passos à Frente - m^3/s | 130 |
| 6.7 | Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 1 Passo à Frente, Período de 1952 a 1956. | 131 |
| 6.8 | Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 3 Passos à Frente, Período de 1952 a 1956. | 132 |
| 6.9 | Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 6 Passos à Frente, Período de 1952 a 1956. | 133 |
| 6.10 | Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 12 Passos à Frente, Período de 1952 a 1956. | 134 |
| 6.11 | Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 1 Passo à Frente, Período de 1972 a 1976. | 135 |
| 6.12 | Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 3 Passos à Frente, Período de 1972 a 1976. | 136 |
| 6.13 | Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 6 Passos à Frente, Período de 1972 a 1976. | 137 |
| 6.14 | Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 12 Passos à Frente, Período de 1972 a 1976. | 137 |
| 6.15 | Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 1 Passo à Frente, Período de 1981 a 1985. | 139 |

| | | |
|------|--|-----|
| 6.16 | Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 3 Passos à Frente, Período de 1981 a 1985. | 140 |
| 6.17 | Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 6 Passos à Frente, Período de 1981 a 1985. | 140 |
| 6.18 | Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 12 Passos à Frente, Período de 1981 a 1985. | 141 |
| 6.19 | Estatísticas para os Modelos de Previsão para o Posto de Emborcação - Previsão 1 Passo à Frente, Período de 1972 a 1976. | 142 |
| 6.20 | Estatísticas para os Modelos de Previsão para o Posto de Emborcação - Previsão 3 Passos à Frente, Período de 1972 a 1976. | 143 |
| 6.21 | Estatísticas para os Modelos de Previsão para o Posto de Emborcação - Previsão 6 Passos à Frente, Período de 1972 a 1976. | 143 |
| 6.22 | Estatísticas para os Modelos de Previsão para o Posto de Emborcação - Previsão 12 Passos à Frente, Período de 1972 a 1976. | 144 |
| 6.23 | Estatísticas para os Modelos de Previsão para o Posto de Sobradinho - Previsão 1 Passo à Frente, Período de 1972 a 1976. | 146 |
| 6.24 | Estatísticas para os Modelos de Previsão para o Posto de Sobradinho - Previsão 3 Passos à Frente, Período de 1972 a 1976. | 147 |
| 6.25 | Estatísticas para os Modelos de Previsão para o Posto de Sobradinho - Previsão 6 Passos à Frente, Período de 1972 a 1976. | 147 |
| 6.26 | Estatísticas para os Modelos de Previsão para o Posto de Sobradinho - Previsão 12 Passos à Frente, Período de 1972 a 1976. | 148 |

Capítulo 1

Introdução

1.1 Descrição do Problema e Revisão Bibliográfica

O planejamento da operação de sistemas de produção de energia elétrica visa determinar uma política de operação para usinas hidroelétricas e termoeleétricas e para os intercâmbios, inclusive internacionais, do sistema que atenda à demanda de energia elétrica de forma econômica e confiável. Em linhas gerais, o objetivo é utilizar da forma mais eficiente possível a geração de base hidráulica para substituir ao máximo a geração térmica, cujo custo do combustível é elevado e crescente com a geração. As usinas hidroelétricas possuem características de produção não-linear e suas vazões afluentes são variáveis aleatórias, levando assim a um problema estocástico e não-linear. Devido à complexidade do problema, é comum sua decomposição em uma cadeia de planejamento que inclui as etapas de longo, médio e curto prazos (Soares, 1987).

O parque gerador de energia elétrica no Brasil é predominantemente hidráulico, sendo o setor elétrico responsável pelo gerenciamento da quantidade de água nos grandes reservatórios de acumulação existentes no país. Estes reservatórios, que até pouco tempo destinavam-se exclusivamente à geração de energia, hoje são alvos de interesses diversos e conflitantes. Neste contexto, cresce a importância na área de planejamento e controle de sistemas de recursos hídricos.

A produção energética de um sistema hidroelétrico depende da série de vazões afluentes às diversas usinas do sistema. Por exemplo, quando se adota um aproveitamento hídrico, a série histórica de vazões no local é usualmente utilizada como dado de entrada para um modelo de simulação e/ou otimização.

As séries históricas de vazões afluentes aos locais de aproveitamento hidroelétricos

são obtidas a partir de um processo laborioso baseado em medições e análises como mostra a figura 1.1 (Fortunato et al., 1990).

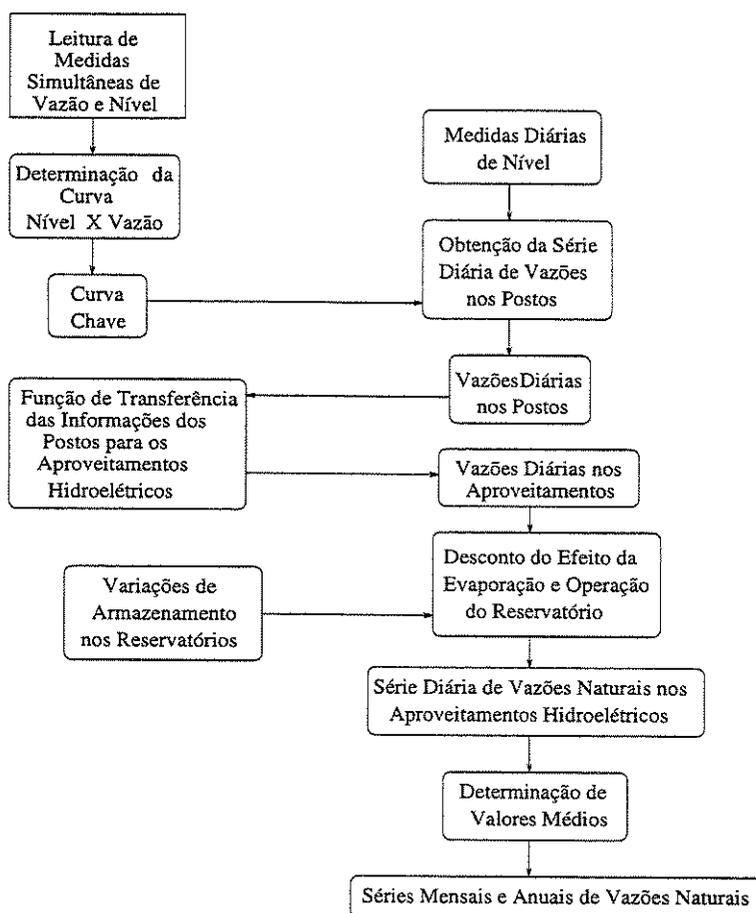


Figura 1.1: Obtenção de Séries Históricas de Vazões Afluentes.

Os modelos de simulação e/ou otimização empregados para o planejamento e operação de sistemas de produção de energia elétrica utilizam usualmente séries de vazões com intervalos de discretização mensal/semanal.

As séries de vazões naturais médias mensais têm como características o comportamento periódico, como por exemplo a média, a variância, a assimetria e a estrutura de autocorrelação. A análise destes tipos de séries pode ser feita pelo uso de formulações auto-regressivas cujos parâmetros apresentam um comportamento periódico. A esta classe de modelos costuma-se denominar modelos auto-regressivos periódicos (Salas et al., 1980). O modelo auto-regressivo periódico é proposto para ser utilizado no Modelo Estratégico de Geração Hidrotérmica a Subsistemas Equivalentes Interligados (NEWAVE) (CEPEL, 1983).

Os modelos empregados atualmente pelo setor elétrico brasileiro têm como base os modelos estocásticos, baseados na metodologia mista de Yevjevich - Box & Jenkins (Box e Jenkins, 1976), para todos os aproveitamentos considerados no programa de operação do Grupo Coordenador para a Operação Interligada (GCOI, 1983), (Sales e Vieira, 1984).

Este sistema de previsão de vazões é conhecido como Sistema PREVAZ. As previsões fornecidas por este modelo, junto com as previsões de carga, são os elementos fundamentais para a atuação do referido programa de operação (Sales e Vieira, 1984).

Em anos recentes, redes neurais artificiais tornaram-se extremamente conhecidas para previsão em várias áreas, incluindo finanças, carga elétrica e recursos hídricos. Embora o conceito de neurônios artificiais tenha sido proposto em 1943 (McCulloch e Pitts, 1943) pesquisas e aplicações de redes neurais tomaram importância após a publicação em 1986 do algoritmo *back-propagation* (Rumelhart et al., 1986b).

A utilização de redes neurais para previsão de séries temporais vem sendo proposta por diversos autores, entre eles (Lapedes e Farber, 1987), (Weigend et al., 1990), (Contrell et al., 1995), (Lachtermacher e Fuller, 1995), mostrando a viabilidade de utilizar estes modelos.

Em vários estudos, redes neurais e técnicas de séries temporais tradicionais vêm sendo comparadas (Sharda e Patil, 1990), (Tang et al., 1991). A maioria dos estudos realizados utilizam o algoritmo *back-propagation*, ou algumas de suas extensões. Este método já foi utilizado com sucesso em diversas áreas, entre as quais sistemas de potência (Srinivasan et al., 1991), (Hwang e Monn, 1991), (El-Sharkawi et al., 1991), (Peng et al., 1992), previsões econômicas (Hoptroff et al., 1991) e análise de séries temporais (Weigend et al., 1991a), (Weigend et al., 1991b).

As comparações de redes neurais com outras abordagens ainda são contraditórias. Alguns estudos concluem que redes neurais artificiais são melhores que os métodos tradicionais (Weigend et al., 1991a) (Weigend et al., 1991b), enquanto que outras pesquisas concluem o contrário (Tang et al., 1991). As explicações para tais contradições podem ser descritas por diferentes fatores, tais como: a estrutura da rede neural, o tipo de série (estacionária, não-estacionária) usada nos estudos e a relação do tamanho da rede e o número de entradas das séries temporais.

(Lapedes e Farber, 1987) aplicaram redes neurais multi-camadas (*Multi-Layer Perceptron-MLP*) para previsão de duas séries temporais caóticas geradas por um processo determinístico não-linear. Os resultados indicaram que os modelos MLP podem “aprender”, simulando sistemas não-lineares, com excelentes propriedades de previsão, comparadas a métodos tradicionais.

(Tang et al., 1991) compararam modelos MLP e Box & Jenkins, usando séries de tráfego de passageiros de vôo internacional, vendas de carros domésticos e vendas de carros importados nos EUA. Os resultados mostraram que o desempenho do modelo de Box & Jenkins foi superior ao modelo de redes neurais para previsão a curto prazo. Já para previsão de longo prazo, os modelos de redes neurais apresentaram um melhor desempenho.

(Weigend et al., 1990), (Weigend et al., 1991a) introduziram a eliminação das conexões sinápticas (pesos) no procedimento de aprendizado das redes neurais para tratar o problema de *overfitting* com aplicação às séries temporais de taxa de câmbio. O critério de parada no procedimento de validação foi discutido e os resultados foram comparados com modelos de séries temporais tradicionais. O modelo de redes neurais apresentou desempenho comparável ao modelo *Threshold Auto-Regressive* (TAR) (Tong e Lin, 1980) na previsão um passo à frente. Em previsões de múltiplos passos os modelos de redes neurais apresentaram um desempenho superior ao modelo TAR.

(Lachtermacher e Fuller, 1995) descreveram uma metodologia híbrida para a aplicação de redes neurais em análise de séries temporais. A metodologia de Box & Jenkins foi utilizada como um procedimento para explorar importantes relações nos dados da série e estas informações foram usadas para definir uma estrutura inicial para a rede neural reduzindo, assim, os parâmetros a serem estimados e os dados exigidos. Dois tipos de séries foram estudadas: séries estacionárias e não estacionárias. Observou-se que o modelo de redes neurais, para a maioria das séries testadas, obteve um desempenho igual ou superior quando comparados aos modelos de Box & Jenkins.

(Atiya et al., 1999) aplicaram redes neurais para previsões de vazões e compararam quatro técnicas para pré-processamento das entradas e saídas incluindo o problema de previsão vários passos à frente. Neste caso, foram utilizados dois métodos: direto e recursivo. No método direto a rede neural foi treinada para prever k -passos à frente, ou seja, a saída da rede será $x(t + k)$. No método recursivo a rede é treinada para prever um passo à frente e esta rede é aplicada para prever k passos à frente. Neste caso, em algum passo intermediário a rede usará alguma das previsões calculadas nos passos anteriores como entrada. Os resultados mostraram que o método direto apresentou um desempenho melhor que o método recursivo.

(Zealand et al., 1999) propuseram o desenvolvimento de modelos de redes neurais MLP's para previsão de vazões a curto prazo, determinando quais as características dos modelos que influenciam no desempenho do mesmo. Na aplicação do modelo foi ajustado o número de entradas. Um primeiro experimento foi realizado testando o mesmo número de entradas que um modelo convencional, no caso um modelo Auto-Regressivo Médias Móveis (ARMA). Um outro experimento utilizou algumas das entradas do ARMA e entradas adi-

cionais. Em ambos os casos os resultados foram satisfatórios, porém o número de entradas influenciou no tempo de treinamento, pois quanto menor o número de entradas menor o número de parâmetros que devem ser ajustado na rede. Os resultados foram comparados com os modelos de séries temporais e pode-se observar que o modelo de rede neural apresentou um bom desempenho.

(Maier e Dandy, 2000) fazem uma revisão dos trabalhos propostos para aplicação de previsão de variáveis de recursos hídricos, mostrando as várias técnicas de escolha do critério de desempenho do modelo, o pré-processamento dos dados, a determinação das entradas e a arquitetura da rede, técnicas de treinamento e validação do modelo.

(Ballini, 1996) propôs o uso de redes neurais multi-camadas com algoritmo *back-propagation* para previsão de chuva-vazão diária. Os resultados foram comparados ao Método dos Vizinhos Mais Próximos (Geva e Sitte, 1991) com algumas modificações propostas por (Andrade Filho, 1994). Ambas as abordagens apresentaram resultados comparáveis para previsão de vazões diárias.

(Francelin et al., 1996) aplicaram redes neurais MLP para previsão de vazões a longo prazo e compararam com os modelos de séries temporais Box & Jenkins. Os resultados mostraram desempenho promissor para a rede MLP, apresentando erros percentual máximo menores que os métodos de Box & Jenkins.

(Shamseldin, 1997) aplicou redes neurais multi-camadas para modelos de chuva-vazão, comparando com outras abordagens convencionais. Os resultados sugeriram que redes neurais apresentaram desempenho razoável para o problema de previsão mas estes modelos podem ter resultados variáveis dependendo da estrutura da rede e número de entradas.

(Mason et al., 1996) aplicaram redes do tipo RBF (*Radial Basis Function*) para previsão de chuva-vazão e compararam o desempenho deste modelo com rede neural multi-camadas. Os resultados mostraram precisão comparável aos modelos MLP, mas com tempo de processamento inferior.

A capacidade de mapeamentos complexos das redes neurais cresce com o número de camadas e neurônios, acarretando maior tempo de processamento, bem como considerável soma de dados. Na prática, entretanto, muitas vezes os parâmetros devem ser ajustados rapidamente e somente uma pequena quantidade de dados está disponível.

Freqüentemente, dados do mundo real apresentam ruídos, podendo conter contradições e imperfeições. Tolerância a imprecisão e incertezas é também exigido para considerar tratabilidade e robustez. Sob essas hipóteses, modelos de análise de dados baseados em conjuntos nebulosos vêm sendo empregados atualmente em diversas áreas (Kacprzyk e

Fedrizzi, 1992).

Os sistemas nebulosos são sistemas computacionais desenvolvidos a partir das teorias de conjuntos nebulosos e da lógica nebulosa, visando aproveitar de maneira pragmática o potencial destas teorias. Estes sistemas processam conhecimento e inferem conclusões a partir dos padrões apresentados (Pedrycz et al., 1995).

Os sistemas nebulosos caracterizam-se pela simplicidade no que se refere à teoria que os fundamenta e pela forma direta e intuitiva com que são empregados, ou seja, são capazes de utilizar de maneira eficaz o conhecimento codificado em expressões linguísticas, produzindo características inteligentes correspondentes à representação do conhecimento e ao raciocínio aproximado (Pedrycz, 1993). Além disso, algumas classes de sistemas nebulosos são aproximadores universais (Castro e Delgado, 1996). Em essência, estes sistemas são capazes de processar o conhecimento de forma compreensível e de manusear incertezas e imprecisões visando solucionar problemas reais complexos (Zimmermann, 1987).

Em anos recentes, surgiu um novo método unindo as vantagens de redes neurais e lógica nebulosa, resultando em um sistema integrado, onde a aprendizagem e o poder computacional dos modelos de redes neurais e a capacidade de representação e raciocínio da lógica nebulosa são combinados. Estes sistemas, chamados de redes neurais nebulosas (Gomide et al., 1995), vêm sendo aplicados sucessivamente para sistemas de identificação não-linear (Wang e Mendel, 1992), controle de processos (Lee, 1990), reconhecimento de padrões (Caminhas et al., 1999), previsão de carga (Alves da Silva et al., 1997), (Bakirtzis et al., 1995), (Liu et al., 1996).

(Bakirtzis et al., 1995) desenvolveram um modelo para previsão de carga a curto prazo. Este sistema combina os princípios da inferência nebulosa com a estrutura de redes neurais e a capacidade de aprendizado em um sistema integrado de decisão nebulosa. Os resultados mostraram que a rede neural nebulosa aplicada ao problema de previsão um passo à frente é comparável aos modelos de redes neurais, enquanto que o treinamento é muito mais rápido.

(Liu et al., 1999) desenvolveram um modelo de rede neural nebulosa com ajuste dinâmico dos parâmetros para previsão. O modelo foi aplicado ao problema de previsão de vendas e comparado com o modelo ARIMA (*AutoRegressive Integrated Moving Average*) e o modelo de rede neural MLP com algoritmo *back-propagation* para previsão a curto prazo e longo prazo. Os resultados mostraram que o modelo proposto apresentou melhor desempenho de precisão e adaptabilidade, tanto para previsão a curto prazo como para longo prazo.

(Liu et al., 1996) comparam três técnicas para previsão de carga a curto prazo. As técnicas comparadas são lógica nebulosa, redes neurais e modelos auto-regressivos,

avaliando o comportamento dos modelos através de uma simulação. Os estudos preliminares mostraram um desempenho satisfatório tanto para redes neurais como para o sistema nebuloso.

(Ballini et al., 1999c) aplicaram uma rede neural nebulosa baseada no modelo de (Yamakawa e Tomoda, 1989) para previsão de vazões médias mensais um passo à frente e comparadas aos modelos de redes neurais MLP com algoritmo *back-propagation* e modelos Periódicos Auto-Regressivos (PAR). Os resultados obtidos com o modelo RNN mostraram melhor desempenho, além de apresentar um rápido processo de treinamento.

(Ballini et al., 1999b), (Ballini et al., 2000) propuseram um modelo de rede neural nebulosa baseado em (Figueiredo e Gomide, 1998) para previsão de vazões médias mensais um passo à frente. O modelo é baseado em um método de aprendizagem competitivo, onde os parâmetros fundamentais para definir o mecanismo de inferência nebuloso são ajustados durante o treinamento. O desempenho do modelo foi comparado com os modelos MLP e PAR. Os resultados forneceram um melhor desempenho para a RNN, com erros significativamente menores.

Uma rede neural nebulosa com estrutura adaptativa foi aplicada para previsão de vazões em (Ballini et al., 1998). Este modelo, proposto por (Figueiredo e Gomide, 1998) apresenta um método de aprendizagem construtivo, onde regras nebulosas são adicionadas à estrutura da rede quando o desempenho desta não é satisfatório. Para avaliar o desempenho do modelo, uma análise comparativa foi realizada com os modelos de redes neurais MLP e PAR. Os resultados mostraram desempenho superior para previsão de vazões um passo à frente usando a rede neural nebulosa.

1.2 Motivação para a Utilização de Redes Neurais Nebulosas

A complexidade inerente aos processos hidrológicos apresenta um desafio ao engenheiro quando este se depara com o dimensionamento e operação de sistemas de recursos hídricos. Este fato se deve ao número de condicionantes físicos que estão envolvidos na definição das vazões associadas aos sistemas que devem ser dimensionados e operados. Apesar das inúmeras tentativas de representar deterministicamente, de maneira completa, as interações sofridas durante a obtenção dos dados de vazões, desde sua formação até sua transformação em vazão no curso d'água (figura 1.1), não existe nenhum método que não esteja sujeito a incertezas (Braga, 1983).

Em razão de diversos fatores, alguns locais de aproveitamentos podem apresentar previsões insatisfatórias. As causas para a deterioração de um modelo de previsão de vazões,

em um dado local de aproveitamento ou posto, passam por diferentes motivações. Um exemplo é o caso do rio Tietê, que em função da operação de saneamento considerada no Alto do Tietê, exigiu a modelagem de novas séries a partir do aproveitamento de Barra Bonita até Itaipu. A partir de revisões e extensão dos arquivos de vazões médias mensais utilizados no âmbito do setor elétrico brasileiro, é considerada oportuna a seleção de novos modelos de previsão mesmo porque as séries podem ser modificadas (Sales e Vieira, 1984).

Modelos que tratam as incertezas e apresentam uma fácil adaptabilidade a novos dados, sem a necessidade de laboriosas formulações estatísticas, são recomendados. Os modelos de Redes Neurais Nebulosas fornecem estas características, pois, diferente das abordagens clássicas, estes modelos abstraem o conhecimento físico dos sistemas a ser modelado, já que o modelo pode ser gerado a partir dos dados de entrada e saída, e ao se manter ativo o processo de geração do modelo, criam-se condições para eliminar deficiências causadas por variações dos parâmetros.

1.3 Objetivos do Trabalho

Este trabalho tem como objetivo geral a aplicação de modelos de séries temporais, redes neurais artificiais e redes neurais nebulosas para previsão de vazões médias mensais.

Os dados utilizados neste trabalho consistem em séries históricas de vazões naturais afluentes médias mensais que abrangem o período de 1931 a 1990, medidas em postos de medições de vazões onde existe aproveitamento hidroelétrico. Esses conjuntos de dados compõem um banco de vazões históricas mantido pela Eletrobrás.

Utilizando dados de postos localizados em diferentes regiões do Brasil e para diferentes períodos do histórico, é possível avaliar o desempenho dos modelos estudados. A análise estatística dos dados fornece informações sobre a natureza da série hidrológica, caracterizando seu comportamento e sugerindo informações sobre os modelos de séries temporais que podem ser ajustados aos dados.

Os modelos são analisados quando considerada a previsão de séries de vazões mensais com horizonte de curto, médio e longo prazos aplicados a três postos hidrológicos. O desempenho dos modelos é avaliado fazendo-se comparações entre as previsões obtidas. Neste caso, algumas medidas de erros para comparação do desempenho dos modelos são usadas, tais como, o erro médio quadrático, erro médio relativo percentual, erro médio absoluto e o máximo erro relativo percentual.

1.4 Organização do Trabalho

Este capítulo apresentou o contexto dentro do qual o trabalho foi desenvolvido e uma revisão dos trabalhos recentes para aplicação de modelos de redes neurais e redes neurais nebulosas para séries temporais, visando principalmente a aplicação desses modelos para previsão de vazões.

No capítulo 2, são apresentados conceitos de processos estocásticos, analisando o comportamento de uma série de vazões mensais. Esta análise tem como objetivo resumir as propriedades estatísticas e caracterizar o comportamento da série, identificando um modelo de série temporal a ser ajustado.

Dois modelos de séries temporais são apresentados no capítulo 3. O primeiro modelo a ser apresentado e ajustado à série de vazões é o modelo Auto-Regressivo (AR). A ordem do modelo é determinada através da análise das funções de auto-correlação e auto-correlação parcial, sendo os parâmetros estimados através dos métodos dos momentos. O segundo modelo estudado e ajustado à série de vazões são os modelos Periódicos Auto-Regressivos (PAR). Neste caso, é utilizada inferência Bayesiana para determinar a ordem do modelo. Os parâmetros são estimados utilizando densidade *a priori* e a ordem do modelo é dada através da densidade preditiva.

No capítulo 4, é apresentado o modelo de rede neural multi-camadas com algoritmo *back-propagation*. Também é apresentado o algoritmo *back-propagation* modificado. Neste algoritmo, a taxa de aprendizagem é ajustada utilizando o método do gradiente. Além disso, neste capítulo, este modelo é aplicado à previsão de vazões para uma série hidrológica.

Redes neurais nebulosas são apresentadas no capítulo 5, juntamente com dois algoritmos de aprendizagem. Estes dois algoritmos são baseados nos paradigmas de aprendizagem construtiva e competitiva. Estes algoritmos são aplicados a séries de vazões.

No capítulo 6, uma análise comparativa dos modelos é proposta. Também são analisados e ajustados dados de vazões de outros dois postos, comparando o desempenho dos modelos e avaliando os vários tipos de erros. Além disso, neste capítulo também é ajustado o sistema de previsão de vazões (Sistema PREVAZ) utilizado pelo setor elétrico brasileiro.

O capítulo 7 conclui o trabalho fazendo um retrospecto das suas contribuições e apresenta sugestões para pesquisas futuras.

O apêndice A apresenta noções básicas de sistemas nebulosos utilizadas no decorrer deste trabalho.

Capítulo 2

Análise de Séries Temporais

Análise de séries temporais tem como objetivos básicos a modelagem do fenômeno sob consideração, a obtenção de conclusões em termos estatísticos e a avaliação da adequação do modelo em termos de previsão.

Neste capítulo, serão primeiramente introduzidos conceitos de processos estocásticos e, a seguir, será descrito o comportamento de uma série de vazões naturais afluentes médias mensais, considerando como exemplo os dados de vazões da usina hidroelétrica de Furnas localizada no sudeste do Brasil. Em seguida, para ajudar na avaliação do modelo a ser ajustado para previsão realizaram-se a verificação da existência de variações sazonais, a obtenção das funções de autocovariância e autocorrelação e a construção de histogramas e diagramas de dispersão.

2.1 Introdução

Uma série temporal consiste de um conjunto de observações de uma variável aleatória indexada no tempo, denotada por $\{x_t, t \in \mathbb{R}_+\}$, onde \mathbb{R}_+ representa o conjunto dos números reais positivos. Em geral, as observações vizinhas são dependentes entre si e o estudo de uma série temporal consiste em analisar e modelar esta dependência.

Uma das principais razões para se modelar uma série temporal é tornar possível realizar previsões de valores futuros, assim como gerar seqüências sintéticas equiprováveis. Como em qualquer análise de dados, procura-se obter um modelo matemático que descreva o sistema de maneira parcimoniosa, principalmente no que se refere ao número de parâmetros, que deve ser mínimo, para o objetivo desejado. Além disso, o modelo deve, se possível, ser parametrizado de tal forma que cada parâmetro possa ser interpretado facilmente e

identificado com algum aspecto da realidade (Pereira et al., 1986).

O objetivo da análise de série temporal é sumarizar as propriedades estatísticas e caracterizar seu comportamento, identificando ou sugerindo um modelo adequado. Há basicamente dois enfoques usados nesta análise. No primeiro, a análise é feita no domínio do tempo, isto é, o interesse reside na magnitude de eventos que ocorrem em determinado instante e na relação entre as observações em diferentes instantes de tempo. A ferramenta utilizada é a função de autocorrelação e a análise é baseada em um modelo paramétrico.

No segundo, a análise é realizada no domínio da frequência, isto é, o interesse está na frequência com que certos eventos ocorrem em determinado período de tempo. A ferramenta utilizada é o espectro (que é a transformada de Fourier da função de autocovariância) e a análise é baseada em modelos não-paramétricos (Pereira et al., 1986).

As duas formas de análise não são alternativas, mas complementares, mostrando cada uma diferentes aspectos da natureza da série temporal. Entretanto, pode-se dizer que a análise no domínio da frequência é conveniente na procura de características determinísticas da série temporal, enquanto a no domínio do tempo é usada na análise de processos não determinísticos.

As séries temporais se desenvolvem no tempo, sujeitas às variações que podem ser descritas por leis probabilísticas. Esse tipo de sistema é objeto da teoria dos processos estocásticos. Ou seja, o mecanismo gerador de uma série temporal é considerado como um processo estocástico e a série temporal observada uma de suas realizações possíveis.

Neste trabalho, serão consideradas séries de vazões naturais afluentes médias mensais cujos dados foram observados somente para valores inteiros de t , ou seja, $t \in \mathcal{Z}_+$, onde \mathcal{Z}_+ denota o conjunto dos números inteiros positivos. Desta forma, as séries serão denotadas como $\{x_1, x_2, \dots, x_N\}$, sendo x_t a observação da série no instante $t = 1, 2, \dots, N$. A figura 2.1 mostra um exemplo de uma série de vazões médias mensais, da usina hidrelétrica de Furnas, localizada no Rio Grande, compreendendo o período de janeiro de 1931 a dezembro de 1990. Esta série será utilizada no decorrer deste capítulo para ilustrar as características de uma série temporal.

A análise clássica de séries temporais trata, basicamente, da decomposição da série em quatro componentes: tendência, sazonal, cíclica e aleatória. As séries de vazões apresentam um comportamento periódico, ou seja, são séries sazonais como ilustra a figura 2.1. Nota-se que as vazões oscilam entre um mínimo e um máximo, podendo ser observada uma variação sazonal, com período aproximado de doze meses. Esse tipo de comportamento representa um caso particular de não estacionariedade da série.

De forma geral, uma série temporal $\{x_t, t \in \mathcal{Z}_+\}$, com não estacionariedade

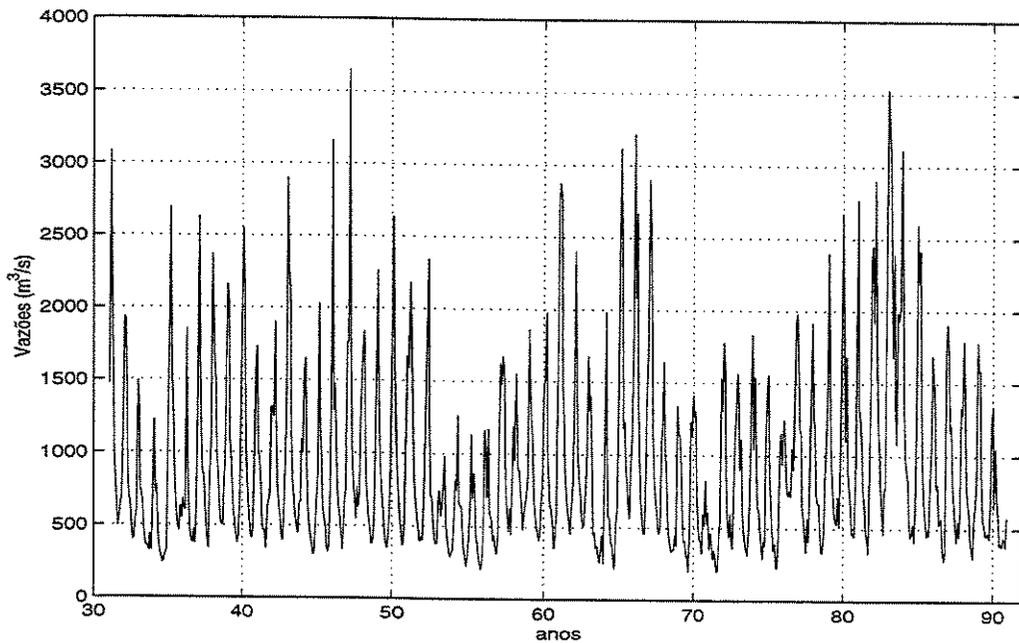


Figura 2.1: Série de Vazões Afuentes Médias Mensais para a Usina Hidroelétrica de Furnas.

sazonal pode ser representada por dois componentes, um sazonal s_t e outro estacionário \tilde{z}_t . Ou seja, a vazão x_t pode ser escrita como:

$$x_t = s_t + \tilde{z}_t + a_t \quad (2.1)$$

onde a_t é um componente aleatório, com média zero e variância constante σ_a^2 , chamado de ruído branco. O componente aleatório é denominado estacionário quando a lei probabilística que gera este componente não se altera com um deslocamento no tempo. Esta propriedade será formalizada posteriormente.

Uma forma alternativa de representação do modelo sazonal é o tipo produto:

$$y_t = q_t v_t u_t \quad (2.2)$$

onde $x_t = \ln y_t$, $s_t = \ln q_t$, $z_t = \ln v_t$ e $a_t = \ln u_t$.

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

$$z_t = \frac{x_t - \mu_m}{\sigma_m} \quad (2.3)$$

onde, μ_m e σ_m são a média e o desvio padrão de cada mês, respectivamente. A série z_t tem, aproximadamente, média zero e desvio padrão unitário.

A seguir, são apresentados alguns conceitos de processos estocásticos que serão úteis na elaboração dos modelos para uma série temporal estacionária.

2.2 Conceitos de Processos Estocásticos

DEFINIÇÃO 2.1 Seja \mathcal{Z}_+ o conjunto de inteiros positivos. Um processo estocástico é uma família $\{x_t, t \in \mathcal{Z}_+\}$ tal que, para cada $t \in \mathcal{Z}_+$, x_t é uma variável aleatória.

Em geral, as observações que caracterizam uma série temporal igualmente espaçada podem ser descritas por uma variável aleatória $\{x_t, t \in T \subset \mathcal{Z}_+\}$, $T = \{t \mid 1 \leq t \leq N\}$, com função densidade de probabilidade conjunta $p(x_1, x_2, \dots, x_N)$ (Box e Jenkins, 1976).

Um processo estocástico é estacionário se suas propriedades (função densidade de probabilidade conjunta) não são alteradas no tempo, de modo que a origem não é importante. Ou, ainda, se a função densidade de probabilidade conjunta é gaussiana o processo estacionário se resume a não existir mudanças sistemáticas na média e na matriz de covariância. Formalmente, há dois tipos de estacionariedade: forte (ou estrita) e fraca (ou ampla, ou de 2ª ordem, ou em covariância).

DEFINIÇÃO 2.2 Um processo estocástico $\{z_t, t \in T \subset \mathcal{Z}_+\}$ é estritamente estacionário se a distribuição de probabilidade conjunta associada com as N observações z_1, z_2, \dots, z_N em algum tempo $t = 1, 2, \dots, N$ é a mesma que a associada com as N observações $z_{1+k}, z_{2+k}, \dots, z_{N+k}$, no tempo $t = 1+k, 2+k, \dots, N+k$, isto é:

$$p(z_1, z_2, \dots, z_N) = p(z_{1+k}, z_{2+k}, \dots, z_{N+k})$$

para quaisquer k e $N \geq 1$.

Isso significa, em particular, que todas as distribuições unidimensionais são invariantes sob translação do tempo. Ou seja, o valor esperado $\mu_t = E[z_t]$ e a variância $\sigma_t^2 = E[z_t - \mu_t]^2$ são invariantes no tempo.

DEFINIÇÃO 2.3 Um processo estocástico $\{z_t, t \in T \subset \mathcal{Z}_+\}$ é estacionário no sentido amplo se e somente se

1. $E[z_t] = \mu(t) = \mu$, constante para qualquer t ;
2. $Var[z_t] = \sigma^2(t) = \sigma^2$, constante para qualquer t ;
3. $Cov(z_t, z_{t+k}) = \gamma(k)$, ou seja, é uma função que só depende de k .

Uma propriedade verificada em (Box et al., 1994) é que qualquer processo estacionário no sentido estrito é estacionário no sentido amplo e todo processo gaussiano estacionário no sentido amplo é também estacionário no sentido estrito.

2.2.1 Média e Variância de um Processo Estacionário

A hipótese de estacionariedade implica que a distribuição de probabilidade $p(z_t)$ é a mesma para todo tempo t , podendo ser escrita como $p(z)$. Assim, o processo estocástico tem média constante:

$$\mu = E[z_t] = \int_{-\infty}^{+\infty} z p(z) dz \quad (2.4)$$

que define o nível em torno do qual a série oscila, e uma variância constante:

$$\sigma_z^2 = E[z_t - \mu]^2 = \int_{-\infty}^{+\infty} (z - \mu)^2 p(z) dz \quad (2.5)$$

que mede a amplitude desta oscilação. Como a distribuição de probabilidade $p(z_t)$ é a mesma para todo tempo t , sua forma pode ser inferida através do histograma das observações $z_1, z_2, \dots, z_N, t = 1, \dots, N$, sobre a série temporal observada. Assim, a média μ do processo estocástico pode ser estimada por:

$$\bar{z} = \frac{1}{N} \sum_{t=1}^N z_t \quad (2.6)$$

sendo denominada média amostral, e a estimativa da variância σ_z^2 , dada por:

$$\text{Var}[z] = \frac{1}{N} \sum_{t=1}^N (z_t - \bar{z})^2 \quad (2.7)$$

é denominada variância amostral.

2.2.2 Coeficientes de Autocovariância e Autocorrelação

A hipótese de estacionariedade também implica que a função de distribuição conjunta $p(z_t, z_{t+k})$ é a mesma para todo tempo $t, t+k$, a qual é separada por um intervalo constante k . Assim, segue que a distribuição conjunta pode ser inferida plotando um diagrama de dispersão usando os valores de (z_t, z_{t+k}) , das séries temporais, separadas por um intervalo constante k .

A covariância entre z_t e o valor z_{t+k} , separada por k intervalos de tempo, é chamada de autocovariância de passo k sendo definida como:

$$\gamma_k = \text{cov}[z_t, z_{t+k}] = E[(z_t - \mu)(z_{t+k} - \mu)] \quad (2.8)$$

onde μ é a média da série dada por (2.4).

Similarmente, o coeficiente de autocorrelação ρ_k com passo k é definido como:

$$\rho_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sqrt{E[(z_t - \mu)^2(z_{t+k} - \mu)^2]}} = \frac{\gamma_k}{\sigma_z^2} \quad (2.9)$$

onde a variância $\sigma_z^2 = \gamma_0$ é a mesma tanto no tempo $t+k$ como no tempo t , o que implica que $\rho_0 = 1$.

A matriz de covariância associada a um processo estacionário, para as observações (z_1, z_2, \dots, z_N) realizadas em N vezes sucessivas, é representada por:

$$\Gamma_N = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{N-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{N-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \dots & \gamma_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{N-1} & \gamma_{N-2} & \gamma_{N-3} & \dots & \gamma_0 \end{bmatrix}$$

$$= \sigma_z^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{N-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{N-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{N-1} & \rho_{N-2} & \rho_{N-3} & \dots & 1 \end{bmatrix} = \sigma_z^2 \mathcal{P}_N \quad (2.10)$$

Desta forma, a matriz Γ_N é simétrica com elementos constantes nas diagonais sendo chamada de matriz de autocovariância e a correspondente matriz \mathcal{P}_N é chamada de matriz de correlação. Ambas as matrizes são definidas positivas para o processo estacionário (Box e Jenkins, 1976).

A condição da matriz de autocorrelação ser definida positiva implica que o determinante é maior que zero. Em particular, para $N=2$ tem-se:

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} = 1 - \rho_1^2 > 0$$

e, assim, $-1 < \rho_1 < 1$.

Similarmente, para $N=3$ deve-se ter

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} > 0; \quad \begin{vmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{vmatrix} > 0; \quad \begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix} > 0$$

as quais implicam que

$$-1 < \rho_1 < 1; \quad -1 < \rho_2 < 1; \quad -1 < \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} < 1$$

e assim por diante. Como \mathcal{P}_N deve ser definida positiva para todo N , os coeficientes de autocorrelação de um processo estacionário devem satisfazer um grande número de condições.

2.3 Funções de Autocovariância e Autocorrelação

O coeficiente de covariância γ_k de passo k , mede a covariância entre dois valores z_t e z_{t+k} . O gráfico de γ_k versus k é chamado de função de autocovariância $\{\gamma_k\}$ do

processo estocástico. Similarmente, o gráfico do coeficiente de autocorrelação ρ_k como função do intervalo k é chamado de função de autocorrelação $\{\rho_k\}$ do processo. Note que a função de autocorrelação é não-dimensionável, isto é, independe da escala de medida da série temporal.

Sendo a covariância $cov[z_t, z_{t+k}] = cov[z_{t-k}, z_t]$ então $\gamma_k = \gamma_{-k}$ e, portanto, $\rho_k = \rho_{-k}$. Ou seja, a função de autocorrelação é necessariamente simétrica.

Um processo estacionário normal z_t é completamente caracterizado por sua média μ e sua função de autocovariância $\{\gamma_k\}$ ou, equivalentemente, por sua média μ e sua função de autocorrelação $\{\rho_k\}$.

2.3.1 Estimação das Funções de Autocovariância e Autocorrelação

Seja uma série temporal estacionária z_1, z_2, \dots, z_N , de N observações. A estimativa dos coeficientes de autocorrelação ρ_k é obtida através do método dos momentos (Box e Jenkins, 1976) da seguinte forma:

$$r_k = \frac{c_k}{c_0} \quad (2.11)$$

onde,

$$c_k = \frac{1}{N-k} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z}), \quad k = 0, 1, \dots, K \quad (2.12)$$

é a estimativa do coeficiente de autocovariância γ_k e \bar{z} é a média amostral dada pela equação (2.6) da série temporal.

Nota-se que r_k é calculado para cada $k = 0, 1, 2, \dots, K$ e o número de observações usadas para estimar c_k depende do valor de k . É recomendado que o valor de K seja menor que $N/4$ quando $N \sim 100$ (Box e Jenkins, 1976). Neste trabalho, $N = 720$ e o valor de K será limitado em $K \leq 36$, ou seja, será analisada a correlação entre os dados de vazão para um intervalo de até 36 meses.

2.4 Análise da Série de Vazões

Nesta seção, serão apresentados resultados da análise estatística realizada para a série de vazões médias mensais de Furnas (figura 2.1). Como mencionado anteriormente, as séries de vazões são séries não-estacionárias. As séries de vazões sazonais serão denotadas por $\{x_{i,m}, i = 1, \dots, n; m = 1, \dots, 12\}$, onde i é o índice do ano com n denotando o número de anos do histórico, e m representa os meses sendo que, $m = 1$ equivale ao mês de janeiro e $m = 12$ é o mês de dezembro. Para transformar as séries de vazões em uma série estacionária os seguintes passos foram realizados:

Passo 1: Cálculo das estimativas da média $\hat{\mu}_m$ e da variância $\hat{\sigma}_m^2$ para cada mês m da seguinte forma:

$$\hat{\mu}_m = \frac{1}{n} \sum_{i=1}^n x_{i,m} \quad (2.13)$$

$$\hat{\sigma}_m^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{i,m} - \hat{\mu}_m)^2, \quad (2.14)$$

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

A tabela 2.1 mostra a média e o desvio padrão para cada mês da série de vazões considerada. Pode-se notar que a série apresenta um comportamento periódico, onde os valores oscilam entre um mínimo, que geralmente ocorre em agosto, e um máximo, que geralmente ocorre em janeiro, podendo ser observada uma variação sazonal, com período aproximado de doze meses. A figura 2.2 mostra as médias e desvios padrões mensais.

Passo 2: A seguir, os dados da série são padronizados, pois assim elimina-se o componente sazonal. Para isso, deve-se subtrair de cada dado de vazão a média mensal e dividir pelo desvio padrão mensal:

$$\tilde{z}_{i,m} = \frac{x_{i,m} - \hat{\mu}_m}{\hat{\sigma}_m} \quad (2.15)$$

Tabela 2.1: Média e Desvio Padrão para a Série de Vazões do Posto de Furnas.

| Meses | Média | Desvio Padrão |
|-----------|--------|---------------|
| Janeiro | 1719.3 | 687.98 |
| Fevereiro | 1669.2 | 671.34 |
| Março | 1518.0 | 645.98 |
| Abril | 1019.6 | 358.67 |
| Maiο | 743.5 | 238.96 |
| Junho | 623.0 | 278.65 |
| Julho | 515.5 | 164.77 |
| Agosto | 431.2 | 135.05 |
| Setembro | 447.5 | 252.08 |
| Outubro | 532.4 | 242.19 |
| Novembro | 747.8 | 314.34 |
| Dezembro | 1268.2 | 499.49 |

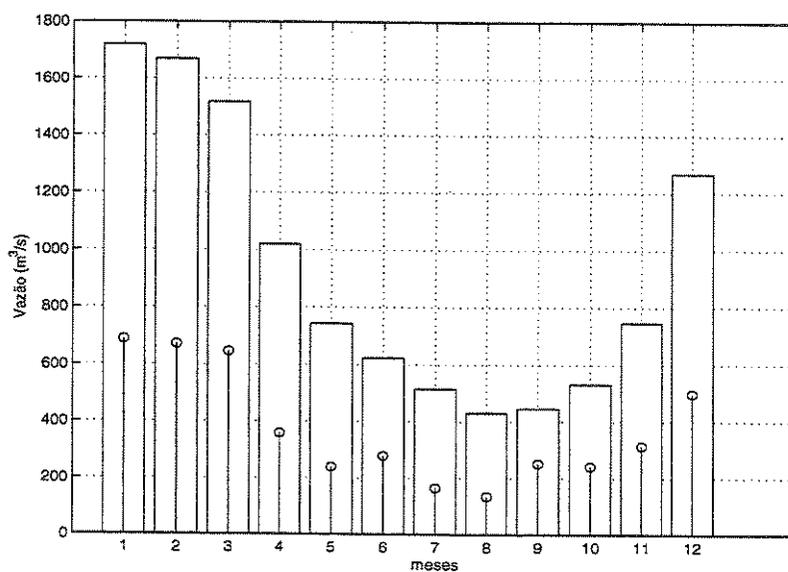


Figura 2.2: Média e Desvio Padrão Mensal para a Série de Vazões do Posto de Furnas.

A série padronizada $\tilde{z}_{i,m}$ apresenta, aproximadamente, média zero e variância unitária, ou seja:

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N \sum_{m=1}^{12} \tilde{z}_{i,m} \sim 0; \quad \text{Vaz}[\tilde{z}] = \frac{1}{N} \sum_{i=1}^N \sum_{m=1}^{12} (\tilde{z}_{i,m} - \bar{z})^2 \sim 1.0$$

onde $N = 720$. A figura 2.3 mostra a série padronizada e a figura 2.4 o histograma da série estacionária, a qual tem uma distribuição com uma certa assimetria e, portanto, a aproximação normal pode não ser a mais adequada. Isso certamente acarretará em algumas dificuldades e perda das propriedades assintóticas nas estimativas dos parâmetros para ajuste dos modelos de séries temporais, quando utiliza-se um método clássico que assume simetria na distribuição como o método de máxima verossimilhança.

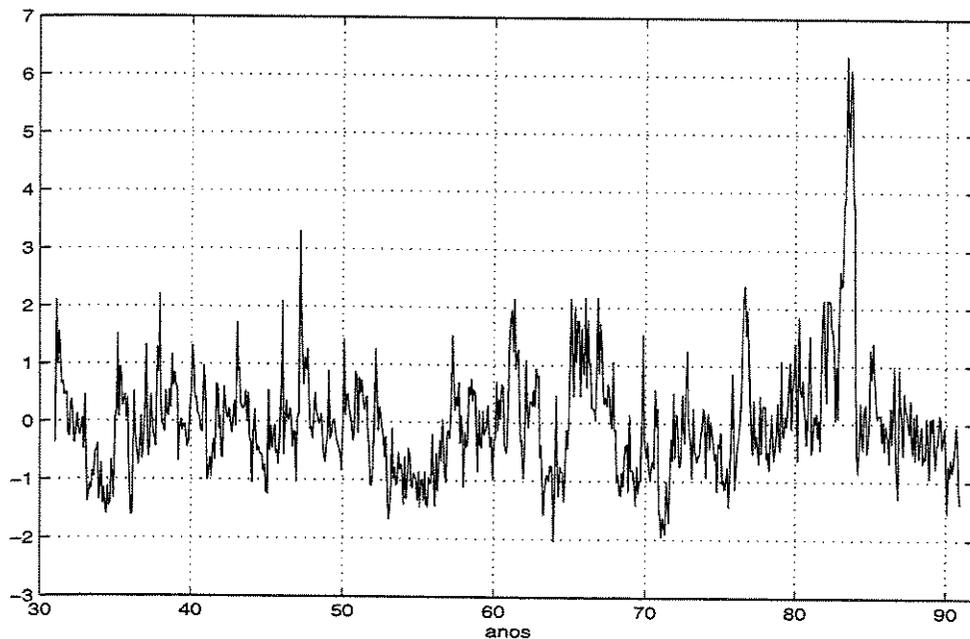


Figura 2.3: Série Padronizada para os Dados de Vazões do Posto de Furnas.

As estimativas dos coeficientes de autocovariância e autocorrelação para a série de vazões foram realizadas. Os 36 primeiros valores de r_k são dados na tabela 2.2 e plotados na figura 2.5. Pode-se notar que a estimativa da função de autocorrelação é caracterizada por um decaimento nos valores de r_k quando o intervalo k aumenta. Também observa-se uma variação no sinal da estimativa r_k fornecendo um comportamento senoidal. Esta forma de comportamento da função de autocorrelação é típica de modelos auto-regressivos. Este modelo será descrito no próximo capítulo.

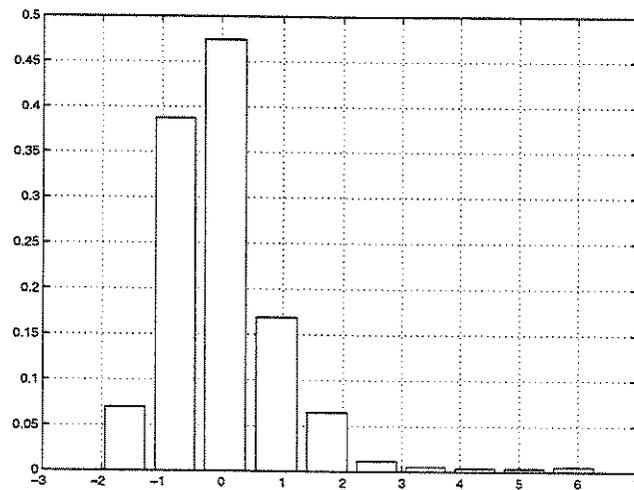


Figura 2.4: Histograma da Série Estacionária para os Dados de Vazões do Posto de Furnas.

Tabela 2.2: Estimativas da Função de Autocorrelação para a Série de Vazões do Posto de Furnas.

| k | r_k | k | r_k | k | r_k |
|-----|--------|-----|--------|-----|---------|
| 1 | 0.7511 | 13 | 0.1824 | 25 | 0.0381 |
| 2 | 0.6437 | 14 | 0.1978 | 26 | 0.0046 |
| 3 | 0.5588 | 15 | 0.1907 | 27 | -0.0078 |
| 4 | 0.4862 | 16 | 0.1925 | 28 | -0.0320 |
| 5 | 0.4296 | 17 | 0.1674 | 29 | -0.0281 |
| 6 | 0.3480 | 18 | 0.1633 | 30 | -0.0242 |
| 7 | 0.2892 | 19 | 0.1385 | 31 | -0.0078 |
| 8 | 0.2643 | 20 | 0.1058 | 32 | 0.0126 |
| 9 | 0.2424 | 21 | 0.0934 | 33 | 0.0062 |
| 10 | 0.2213 | 22 | 0.0636 | 34 | 0.0162 |
| 11 | 0.2109 | 23 | 0.0724 | 35 | 0.0314 |
| 12 | 0.1911 | 24 | 0.0479 | 36 | 0.0312 |

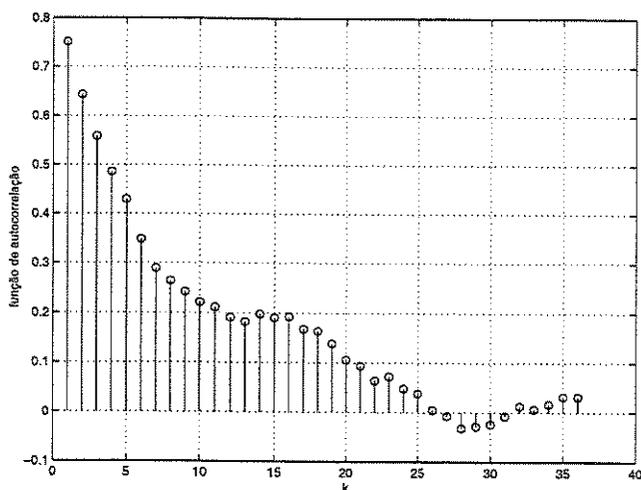


Figura 2.5: Função de Autocorrelação para a Série de Vazões do Posto de Furnas.

Através dos dados da tabela 2.2, verifica-se que a correlação entre \tilde{z}_t e \tilde{z}_{t+1} apresenta um grau de associabilidade de 75.11% e entre \tilde{z}_t e \tilde{z}_{t+2} um grau de 64.37%. Considerando esses dois valores de r_k , foram construídos os diagramas de dispersão. Ou seja, usando os valores de $(\tilde{z}_t, \tilde{z}_{t+k})$ para $k = 1$ e $k = 2$, pode-se observar uma correlação entre os dados da série com os valores vizinhos, como mostram as figuras 2.6(a) e 2.6(b), respectivamente. A análise visual desses gráficos mostra que pode-se considerar, pelo menos, uma relação linear entre as observações \tilde{z}_t e \tilde{z}_{t+k} .

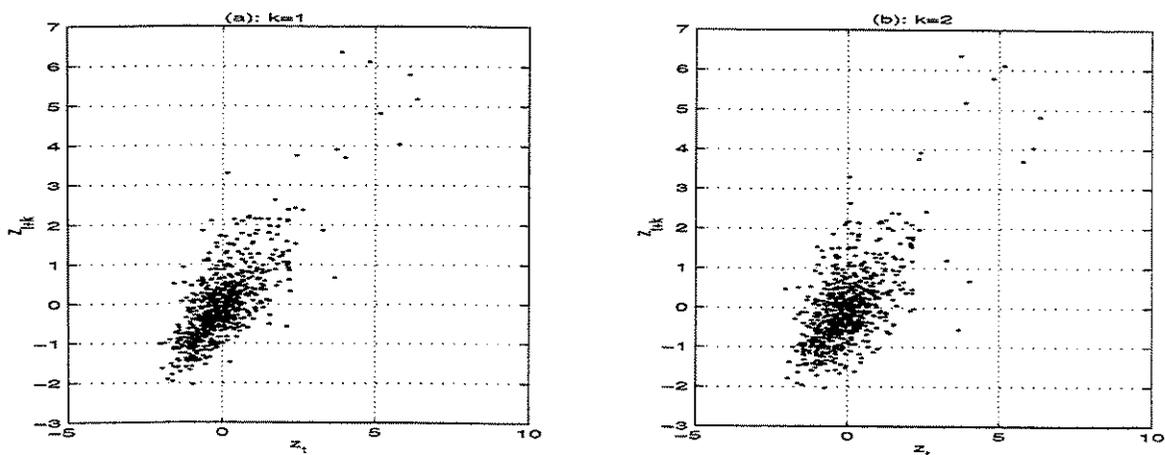


Figura 2.6: Diagramas de Dispersão para a Série de Vazões do Posto de Furnas.

2.5 Resumo

Neste capítulo foram apresentados os conceitos fundamentais relacionados à teoria de processos estocásticos e a análise de séries temporais. Foram ressaltados alguns conceitos úteis, tais como coeficientes de autocovariância e autocorrelação, na elaboração dos modelos para tratar uma série temporal estacionária.

O principal objetivo deste capítulo é alcançado com a análise de uma série de vazões naturais médias mensais do posto de Furnas, mostrando a viabilidade da análise dos coeficientes de correlação entre os dados de vazões mensais sugerindo, assim, um modelo a ser ajustado.

No próximo capítulo, os conceitos apresentados são utilizados para o ajuste dos modelos de séries temporais, aplicados aos dados de vazões do posto de Furnas.

Capítulo 3

Modelos de Séries Temporais

Na análise de métodos paramétricos, os modelos de Box & Jenkins têm recebido muita atenção nas últimas décadas. Tais métodos consistem em ajustar modelos auto-regressivos, médias móveis, modelos mistos auto-regressivos médias móveis e os modelos integrados auto-regressivos médias móveis. Neste trabalho, são utilizados modelos lineares, mais especificamente modelos auto-regressivos e modelos periódicos auto-regressivos.

O ajuste do modelo auto-regressivo é realizado através da análise das funções de autocorrelação e autocorrelação parcial e a escolha da ordem mais adequada do modelo é feita a partir dessas funções. Já para o modelo periódico auto-regressivo é empregado o método de máxima verossimilhança e inferência Bayesiana, sendo que, para este caso, os parâmetros do modelo são estimados através da distribuição de probabilidade anterior não informativa e para a seleção da ordem do modelo é usada a densidade preditiva.

3.1 Introdução

Uma importante classe de modelos estocásticos para descrever séries temporais é a chamada classe dos modelos estacionários, os quais assumem que os processos permanecem em equilíbrio sobre um nível médio constante (Box e Jenkins, 1976), como descrito no capítulo 2.

A estratégia para a construção do modelo de séries temporais é baseada em um ciclo iterativo, no qual a seleção da estrutura do modelo é baseada nos próprios dados. Os estágios do ciclo iterativo são (Morettin e Tolo, 1985):

1. Uma classe geral de modelos é considerada para a análise (especificação);

2. Identificação de um modelo, com base na análise de autocorrelação, autocorrelação parcial e outros critérios;
3. A seguir, vem a fase de estimação, na qual os parâmetros do modelo identificado são estimados;
4. Finalmente, há a verificação do modelo ajustado, através de uma análise de resíduos, para averiguar se este é adequado para os fins em vista, no caso, para a previsão.

Caso o modelo não seja adequado, o procedimento acima é repetido, voltando à fase de identificação. Um procedimento, muitas vezes utilizado, é identificar não somente um único modelo, mas alguns modelos que serão estimados e verificados. Assim, a fase crítica do procedimento acima é a identificação do modelo, sendo possível que diferentes modelos sejam identificados para a mesma série temporal (Morettin e Toloi, 1985).

Antes de descrever os modelos, é apresentada a notação de alguns operadores que serão utilizados no decorrer deste capítulo:

1. Operador translação para o passado ou, simplesmente, operador de atraso B , definido como:

$$Bz_t = z_{t-1}; \quad B^m z_t = z_{t-m} \quad (3.1)$$

2. Operador translação futura F , ou seja, operador inverso de B ($F = B^{-1}$), dado por:

$$B^{-1}z_t = Fz_t = z_{t+1}; \quad B^{-m}z_t = z_{t+m} \quad (3.2)$$

3. Operador diferença, o qual pode ser escrito em termos de B , como:

$$\nabla z_t = z_t - z_{t-1} = (1 - B)z_t \quad (3.3)$$

4. Operador soma denotado por S e definido por:

$$\begin{aligned} \nabla^{-1}z_t = Sz_t &= \sum_{k=0}^{\infty} z_{t-k} = z_t + z_{t-1} + z_{t-2} + \dots \\ &= (1 + B + B^2 + \dots)z_t \end{aligned} \quad (3.4)$$

3.2 Modelos Estacionários Lineares

Os modelos estocásticos empregados neste trabalho são baseados na idéia de que uma série temporal, na qual os valores sucessivos são altamente dependentes, pode ser considerada como sendo gerada através de uma série de independentes “impactos” a_t , $t \in \mathcal{Z}_+$. Esses “impactos” são aleatórios, com distribuição fixa assumindo, geralmente, ser uma distribuição normal, com média zero e variância σ_a^2 . Esta seqüência de variáveis aleatórias a_t, a_{t-1}, \dots é chamada de ruído branco (Box e Jenkins, 1976).

Estes modelos podem ser interpretados como sendo casos particulares de um modelo de filtro linear, supondo que a saída temporal z_t é gerada através de um filtro linear ou sistema linear, cuja entrada é um ruído branco a_t sendo $\Psi(B)$ a função de transferência (figura 3.1).

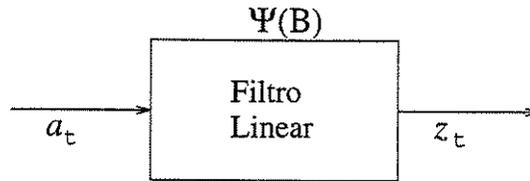


Figura 3.1: Representação de uma Série Temporal como Saída de um Filtro. Linear

Formalmente, este modelo pode ser representado da seguinte forma:

$$z_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = a_t + \sum_{k=1}^{\infty} \psi_k a_{t-k} = \Psi(B)a_t \quad (3.5)$$

onde $z_t = \tilde{z}_t - \bar{z}$ é o desvio do processo estacionário \tilde{z}_t da média estimada \bar{z} dada pela equação (2.6), a função de transferência $\Psi(B)$ pode ser escrita como:

$$\Psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots = \sum_{k=0}^{\infty} \psi_k B^k; \quad \psi_0 = 1 \quad (3.6)$$

e a_t , $t \geq 0$, é uma seqüência de variáveis aleatórias não correlacionadas com média zero e variância constante, isto é:

$$E[a_t] = 0; \quad Var[a_t] = \sigma_a^2$$

A condição de estacionariedade do modelo linear (3.5) esta associada à série $\Psi(B)$. Em (Box e Jenkins, 1976) é demonstrado que para um processo linear as condições de estacionariedade dependem da convergência da série $\Psi(B)$ a qual é verificada se $|B| \leq 1$, isto é, para valores dentro do círculo unitário.

Um processo z_t pode ser representado como uma combinação linear dos valores passados z_{t-1}, z_{t-2}, \dots mais um ruído a_t , de forma similar à equação (3.5), isto é:

$$z_t = a_t + \pi_1 z_{t-1} + \pi_2 z_{t-2} \dots = a_t + \sum_{k=1}^{\infty} \pi_k z_{t-k} \quad (3.7)$$

Segue que,

$$a_t = z_t - \sum_{k=1}^{\infty} \pi_k z_{t-k} = \Pi(B)z_t \quad (3.8)$$

Das equações (3.8) e (3.5), tem-se o seguinte resultado:

$$a_t = \Pi(B)\Psi(B)a_t \quad (3.9)$$

De modo que,

$$\Pi(B) = \Psi^{-1}(B) \quad (3.10)$$

A relação (3.10) pode ser usada para obter os parâmetros Π_k em função dos parâmetros Ψ_k e vice-versa.

A equação (3.8) está associada à propriedade de invertibilidade, ou seja, o modelo linear (3.8) é invertível se a série $\Pi(B)$ convergir ($|B| \leq 1$). Essa propriedade é independente da condição de estacionariedade, sendo também aplicável para processos lineares não-estacionários (Box e Jenkins, 1976).

Os três modelos lineares mais usados para modelar séries temporais estacionárias são os processos auto-regressivos (AR), médias móveis (MA) e os mistos auto-regressivos médias móveis (ARMA). Neste trabalho, são considerados apenas os modelos auto-regressivos, descritos na próxima seção.

3.3 Modelos Auto-Regressivos

Considerando uma série temporal estacionária z_t , $t \in \mathcal{Z}_+$, um processo auto-regressivo de ordem p , denotado por $\text{AR}(p)$, é definido como sendo um modelo onde o valor corrente do processo z_t é expresso como uma combinação linear dos p valores passados $z_{t-1}, z_{t-2}, \dots, z_{t-p}$ e de um ruído branco a_t , onde a_t é uma seqüência de variáveis aleatórias com distribuição normal de média zero e variância σ_a^2 , ou seja, $N(0, \sigma_a^2)$.

Se na equação (3.7) $\pi_k = 0$, para $k > p$, o modelo $\text{AR}(p)$ pode ser escrito como:

$$z_t = a_t + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} \quad (3.11)$$

renomeando os parâmetros π_k por ϕ_k e $z_t = \tilde{z}_t - \bar{z}$, onde \bar{z} é a média estimada (equação (2.6)) (Box e Jenkins, 1976) (Morettin e Toloí, 1985).

Usando o operador de atraso B definido na seção 3.1, a equação (3.11) pode ser rescrita como:

$$\Phi(B)z_t = a_t \quad (3.12)$$

onde $\Phi(B)$ é o operador auto-regressivo de ordem p , definido como:

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (3.13)$$

Os parâmetros $\phi_1, \phi_2, \dots, \phi_p$ devem ser ajustados, satisfazendo certas condições para o processo ser estacionário. Na seção 3.2 foi visto que para o processo ser estacionário a série $\Psi(B)$ dever convergir, isto é $|B| \leq 1$.

Considerando um caso mais simples em que a ordem do modelo é $p = 1$, isto é, um modelo $\text{AR}(1)$, a equação (3.12) pode ser escrita como:

$$(1 - \phi_1 B)z_t = a_t \quad (3.14)$$

Como $\Pi(B) = \Phi(B) = (1 - \phi_1 B)$, o processo é invertível, podendo ser rescrito como

$$z_t = (1 - \phi_1 B)^{-1} a_t = \sum_{k=0}^{\infty} \phi_1^k a_{t-k} \quad (3.15)$$

Assim, tem-se que

$$\Psi(B) = (1 - \phi_1 B)^{-1} = \sum_{k=0}^{\infty} \phi_1^k B^k \quad (3.16)$$

A equação (3.16) implica que o parâmetro ϕ_1 , de um processo AR(1), deve satisfazer a condição $|\phi_1| < 1$ para garantir estacionariedade. Como a raiz da equação $\Phi(B) = 1 - \phi_1 B = 0$ é $B = \phi_1^{-1}$, esta condição é equivalente a dizer que a raiz de $\Phi(B) = 0$ deve estar fora do círculo unitário.

Para o processo auto-regressivo de ordem p , AR(p), a condição de estacionariedade pode ser analisada escrevendo o polinômio $\Phi(B)$ como:

$$\Phi(B) = (1 - G_1 B)(1 - G_2 B) \dots (1 - G_p B) \quad (3.17)$$

onde G_i^{-1} são as raízes da equação característica $\Phi(B) = 0$. É possível expandir a equação (3.17) em frações parciais,

$$z_t = \Phi^{-1}(B) a_t = \sum_{i=1}^p \frac{A_i}{1 - G_i B} a_t \quad (3.18)$$

onde A_i são constantes, para $i = 0, \dots, p$.

Assim, se $\Psi(B) = \Phi^{-1}(B)$ é uma série convergente ($|B| \leq 1$), então deve-se ter $|G_i| < 1$, $i = 1, 2, \dots, p$. Esta condição é equivalente à condição de que a equação característica $\Phi(B) = 0$ tenha raízes fora do círculo unitário.

Como a série $\Pi(B) = \Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é finita, não há restrições sobre os parâmetros do processo AR(p) para assegurar invertibilidade.

3.3.1 Função de Autocorrelação de Modelos Auto-Regressivos

Uma importante relação de recorrência para a função de autocorrelação de um processo auto-regressivo estacionário representado pela equação (3.11), é encontrada multiplicando-se ambos os membros de (3.11) por z_{t-k} , ou seja,

$$z_{t-k}z_t = \phi_1 z_{t-k}z_{t-1} + \phi_2 z_{t-k}z_{t-2} + \dots + \phi_p z_{t-k}z_{t-p} + a_t z_{t-k} \quad (3.19)$$

Tomando os valores esperados em (3.19), e como z_{t-k} somente envolve ruídos até a_{t-k} , não correlacionados com a_t , tem-se que $E[a_t z_{t-k}] = 0$, $k > 0$, do que resulta:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} \quad (3.20)$$

onde $\gamma_k = Cov(z_{t-k}, z_t)$ é o coeficiente de covariância.

Dividindo (3.20) por γ_0 , nota-se que o coeficiente de autocorrelação satisfaz a mesma forma da equação diferença (3.11), isto é,

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}, \quad k > 0 \quad (3.21)$$

que também pode ser escrita da forma

$$\Phi(B)\rho_k = 0 \quad (3.22)$$

onde $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$. Se

$$\Phi(B) = \prod_{i=1}^p (1 - G_i B) \quad (3.23)$$

então, a solução geral de (3.21) é dada por (Box e Jenkins, 1976):

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \dots + A_p G_p^k \quad (3.24)$$

$$\gamma_0 = \phi_1\gamma_{-1} + \phi_2\gamma_{-2} + \dots + \phi_p\gamma_{-p} + \sigma_a^2 \quad (3.27)$$

Dividindo a equação (3.27) por $\gamma_0 = \sigma_z^2$ e substituindo $\gamma_k = \gamma_{-k}$, a variância σ_z^2 pode ser escrita como:

$$\sigma_z^2 = \frac{\sigma_a^2}{1 - \rho_1\phi_1 - \rho_2\phi_2 - \dots - \rho_p\phi_p} \quad (3.28)$$

Usando as estimativas de covariância c_k e substituindo ρ_k pelas estimativas r_k , obtém-se uma estimativa para a variância do resíduo a_t :

$$\hat{\sigma}_a^2 = c_0 (1 - r_1\phi_1 - r_2\phi_2 - \dots - r_p\phi_p) \quad (3.29)$$

onde c_0 é a estimativa da variância do processo σ_z^2 .

3.3.2 Função de Autocorrelação Parcial de Modelos Auto-Regressivos

Uma outra ferramenta utilizada na identificação da ordem dos modelos de série temporal é a função de autocorrelação parcial. Essa função, denotada por ϕ_{kk} para vários passos k é definida como segue:

DEFINIÇÃO 3.1 O coeficiente de autocorrelação parcial de ordem k , denotado por ϕ_{kk} , é definido como sendo o último coeficiente de um modelo AR(k), ajustado à série temporal z_t , $t = 1, 2, \dots, N$.

Supondo que para uma série z_t , $t = 1, 2, \dots, N$, um modelo AR(1), isto é, de ordem $k = 1$, foi ajustado:

$$z_t = \phi_1 z_{t-1} + a_t$$

Tem-se que o coeficiente de autocorrelação parcial deste modelo é $\phi_{11} = \phi_1$. Agora, supondo que, para a série z_t se ajuste um modelo AR(2):

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$$

Então, o coeficiente de autocorrelação parcial para este modelo é $\phi_{22} = \phi_2$. Procedendo desta maneira, para o modelo geral $AR(k)$, de ordem k :

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_k z_{t-k} + a_t$$

o coeficiente de autocorrelação parcial é dado por $\phi_{kk} = \phi_k$.

Para um processo auto-regressivo de ordem p , o coeficiente de autocorrelação parcial ϕ_{kk} será diferente de zero para k menor ou igual a p e será igual a zero para k maior que p . Essa propriedade, juntamente com o comportamento da função de autocorrelação, são duas ferramentas úteis para identificar a ordem dos modelos auto-regressivos.

Denotando por ϕ_{kj} o coeficiente de um modelo $AR(k)$, tal que, ϕ_{kk} é o último coeficiente do modelo, de (3.21) o coeficiente de autocorrelação parcial ϕ_{kj} satisfaz a equação:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k}, \quad j = 1, 2, \dots, k \quad (3.30)$$

a partir das quais obtêm-se as equações de Yule-Walker:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kp} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}$$

ou, $\mathcal{P}_k \Phi_k = \rho_k$. Resolvendo estas equações sucessivamente para $k = 1, 2, \dots$, obtêm-se:

$$\phi_{11} = \rho_1; \quad \phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}; \quad \phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

Em geral, para ϕ_{kk} , o determinante no numerador tem os mesmos elementos que no denominador, mas com a última coluna formada pelo vetor ρ_k . O coeficiente ϕ_{kk}

considerado como função do intervalo k é chamado de função de autocorrelação parcial (Box e Jenkins, 1976).

A função de autocorrelação parcial pode ser estimada através de sucessivos ajustes dos modelos auto-regressivos de ordens $1, 2, \dots$, usando o método dos momentos. Uma outra forma é usar uma aproximação da equação de Yule-Walker, substituindo as funções de autocorrelação ρ_j por suas estimativas r_j :

$$r_j = \hat{\phi}_{k1}r_{j-1} + \hat{\phi}_{k2}r_{j-2} + \dots + \hat{\phi}_{k(k-1)}r_{j-k+1} + \hat{\phi}_{kk}r_{j-k}, \quad j = 1, 2, \dots, k \quad (3.31)$$

e resolvendo as equações resultantes para $k = 1, 2, \dots$, recursivamente, como proposto por (Durbin, 1960), obtendo-se, assim, uma estimativa para os coeficientes de autocorrelação parcial.

3.3.3 Erro Padrão das Estimativas de Autocorrelação Parcial

(Quenouille, 1949) mostrou que, para as hipóteses do processo auto-regressivo de ordem p , as estimativas dos coeficientes de autocorrelação parcial de ordem $p + 1$ e superiores são, aproximadamente, independentemente distribuídas. Se N é o número de observações usadas no ajuste do modelo, tem-se que:

$$Var[\hat{\phi}_{kk}] \simeq \frac{1}{N}, \quad k \geq p + 1 \quad (3.32)$$

Assim, o desvio padrão (DP) da estimativa $\hat{\phi}_{kk}$ é

$$DP[\hat{\phi}_{kk}] = \hat{\sigma} \simeq \frac{1}{\sqrt{N}}, \quad k \geq p + 1 \quad (3.33)$$

Supondo que o número de observações N é suficientemente grande, pode-se dizer que $\hat{\phi}_{kk}$ tem distribuição aproximadamente normal, o que permite a construção de um intervalo de confiança para $\hat{\phi}_{kk}$ usando o desvio padrão $DP[\hat{\phi}_{kk}]$, considerando que a estimativa é diferente de zero quando $|\hat{\phi}_{kk}| > 2/\sqrt{N}$, ou seja, $|\hat{\phi}_{kk}| > 2DP[\hat{\phi}_{kk}]$.

3.3.4 Ajuste de um Modelo Auto-Regressivo para a Série de Vazões

Considerando a série de vazões afluentes naturais médias mensais do posto de Furnas (figura 2.1), juntamente com os resultados da análise da série apresentados na seção 2.4, ajustaram-se os coeficientes de autocorrelação parcial para 36 valores de k .

O comportamento da função de autocorrelação (figura 2.5) mostra que as estimativas dos coeficientes de autocorrelação decaem exponencialmente e oscilam na forma de uma senóide após um certo intervalo k , sendo que estas características são típicas de modelos auto-regressivos. Através da análise das funções de autocorrelação e autocorrelação parcial, um modelo AR(2) é selecionado para a série de vazões de Furnas, com ϕ_1 e ϕ_2 assumindo valores positivos. Esse modelo apresentou-se como o mais parcimonioso, ou seja, com menor número de parâmetros possíveis e com variância dos resíduos $\sigma_a^2 = 0.4150$, como pode ser visto na tabela 3.1, onde é mostrada a variância para um modelo auto-regressivo de ordem 1, 2 e 3.

Tabela 3.1: Variância dos Resíduos para Diferentes Modelos.

| Variância | AR(1) | AR(2) | AR(3) |
|--------------------|--------|--------|--------|
| $\hat{\sigma}_a^2$ | 0.4292 | 0.4150 | 0.4132 |

A tabela 3.2 mostra os 36 valores estimados para os coeficientes de autocorrelações parciais e a figura 3.2 mostra a função de autocorrelação parcial. Também é mostrado na figura 3.2 através de linhas tracejadas os limites $2DP[\hat{\phi}_{kk}]$ calculados de (3.32), na hipótese que a série se ajusta a um modelo AR(2).

Neste caso, o intervalo de confiança é:

$$|\hat{\phi}_{kk}| > 2DP[\hat{\phi}_{kk}] \simeq 2 \frac{1}{\sqrt{N}} = 2 \frac{1}{\sqrt{720}} = 0.0745$$

onde N é igual ao número de anos (60 anos) vezes os meses (12 meses), pois a série corresponde ao período de 1931 a 1990.

Assim, o modelo AR(2) estimado para a série de Furnas é da seguinte forma:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$$

onde $\phi_2 = \hat{\phi}_{22}$ e o coeficiente ϕ_1 é dado pela estimativa do coeficiente de autocorrelação parcial $\hat{\phi}_{21}$. Estes coeficientes foram estimados através das equações de Yule-Walker, ou seja:

Tabela 3.2: Estimativas dos Coeficientes de Autocorrelação Parcial para a Série de Vazões do Posto de Furnas.

| k | $\hat{\phi}_{kk}$ | k | $\hat{\phi}_{kk}$ | k | $\hat{\phi}_{kk}$ |
|-----|-------------------|-----|-------------------|-----|-------------------|
| 1 | 0.7511 | 13 | 0.0113 | 25 | -0.0000 |
| 2 | 0.1826 | 14 | 0.0687 | 26 | -0.0703 |
| 3 | 0.0627 | 15 | 0.0044 | 27 | -0.0037 |
| 4 | 0.0230 | 16 | 0.0231 | 28 | -0.0450 |
| 5 | 0.0238 | 17 | -0.0397 | 29 | 0.0349 |
| 6 | -0.0639 | 18 | 0.0196 | 30 | 0.0187 |
| 7 | -0.0143 | 19 | -0.0419 | 31 | 0.0489 |
| 8 | 0.0528 | 20 | -0.0416 | 32 | 0.0279 |
| 9 | 0.0327 | 21 | 0.0163 | 33 | -0.0328 |
| 10 | 0.0141 | 22 | -0.0323 | 34 | 0.0091 |
| 11 | 0.0333 | 23 | 0.0552 | 35 | 0.0283 |
| 12 | -0.0055 | 24 | -0.0402 | 36 | -0.0018 |

$$r_j = \hat{\phi}_{21}r_{j-1} + \hat{\phi}_{22}r_{j-2}$$

$j = 1, 2$ e $k = 2$. Resolvendo o sistema linear e considerando as estimativas dos coeficientes de autocorrelação apresentados na tabela 2.2, tem-se que:

$$\phi_1 = \hat{\phi}_{21} = \frac{r_1(1-r_2)}{1-r_1^2} = 0.6139; \quad \phi_2 = \hat{\phi}_{22} = \frac{r_2-r_1^2}{1-r_1^2} = 0.1826$$

Logo, o modelo AR(2) para a série de Furnas pode ser escrito como:

$$z_t = 0.6139 z_{t-1} + 0.1826 z_{t-2} + a_t$$

Para verificar a adequabilidade do modelo, o resíduo a_t é calculado:

$$a_t = z_t - \phi_1 z_{t-1} - \phi_2 z_{t-2} \quad (3.34)$$

onde $t = 1, \dots, N$ e verificando que $a_t \sim N(0, \sigma_a^2)$ (figura 3.3) e que a série $\{a_t\}$ é não correlacionada, como mostra a função de autocorrelação de a_t (figura 3.4).

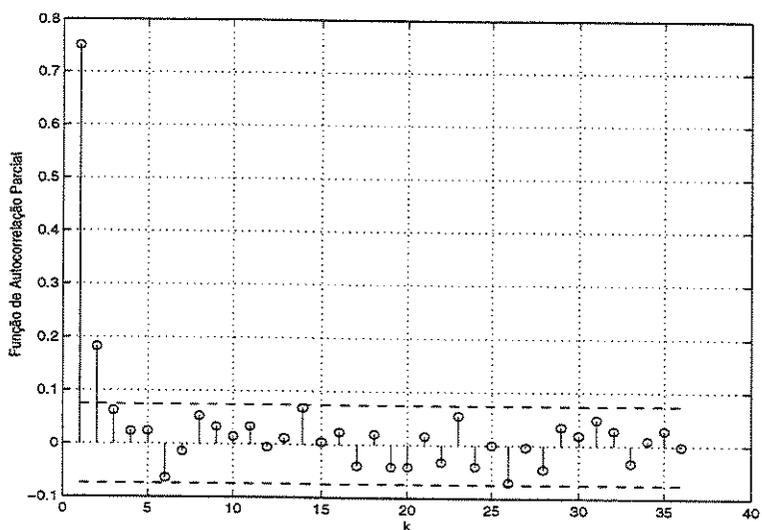


Figura 3.2: Função de Autocorrelação Parcial para a Série de Vazões do Posto de Furnas.

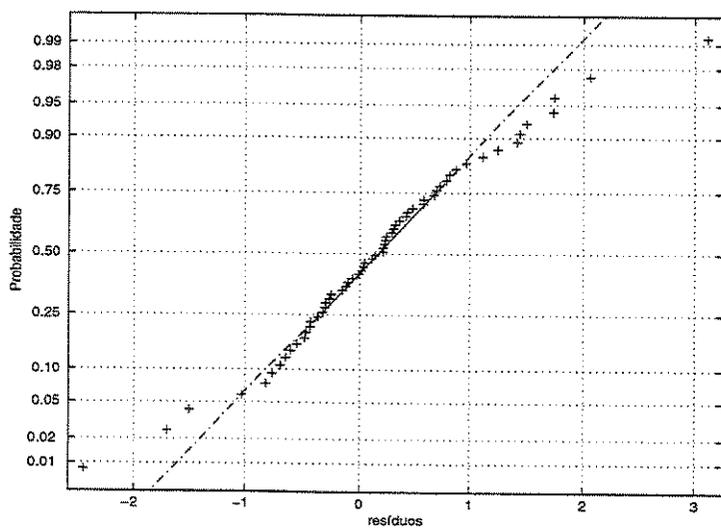


Figura 3.3: Gráfico Normal-Plote (Resíduo a_t).

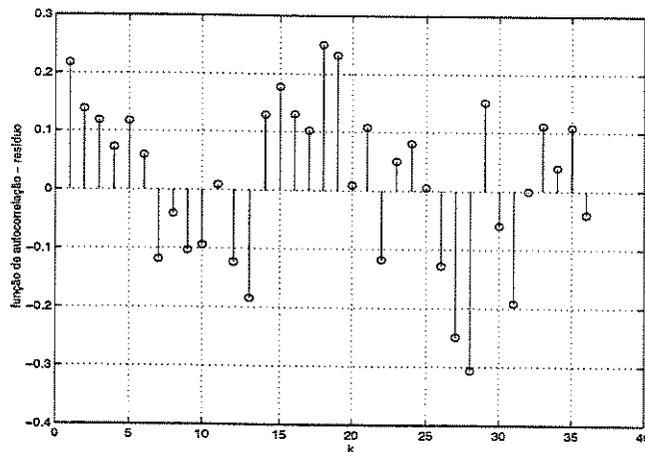


Figura 3.4: Função de Autocorrelação do Resíduo a_t .

Após o ajuste do modelo, foram realizadas previsões de vazões para cinco anos consecutivos com características distintas: um período de 1952 a 1956 considerado seco, tendo como vazão média para este período $642.73 \text{ m}^3/\text{s}$; o período de 1972 a 1976, considerado mediano com vazão média igual à $864.47 \text{ m}^3/\text{s}$; e de 1981 a 1985 representando um período úmido, com vazão média de $1374.70 \text{ m}^3/\text{s}$.

Os parâmetros dos modelos foram estimados retirando do histórico o período de teste e ajustando os parâmetros para os anos restantes. Por exemplo, para o período de 1952 a 1956 foram considerados para ajuste dos parâmetros os dados de vazões até 1951. Este processo foi repetido para os outros dois períodos de teste. Foram realizadas previsões 1, 3, 6, e 12 passos à frente. As figuras 3.5, 3.6 e 3.7 mostram as previsões para cada período e em cada figura as previsões 1, 3, 6 e 12 passos à frente são mostradas em (a), (b), (c) e (d), respectivamente.

Através das figuras 3.5(a), 3.6(a) e 3.7(a), nota-se que a previsão 1 passo à frente apresentou um melhor desempenho que as outras previsões (3, 6 e 12 passos à frente) principalmente para os períodos de seca e úmido. Também pode-se observar que, para previsões com mais de 3 passos à frente, o comportamento da série começa a se repetir, notando-se uma tendência para a Média a Longo Termo (MLT), como pode ser observado na tabela 3.3, onde são dadas as vazões médias mensais e o valor previsto para 6 passos à frente para os meses de Janeiro a Dezembro dos anos de 1952, 1972 e 1981.

A correlação da medida presente com a medida imediatamente passada e com as médias passadas no mesmo ponto de períodos anteriores das séries hidrológicas, pode tornar inadequado o uso de modelos auto-regressivos. Neste caso, os modelos periódicos auto-regressivos são propostos. Estes modelos são descritos na próxima seção.

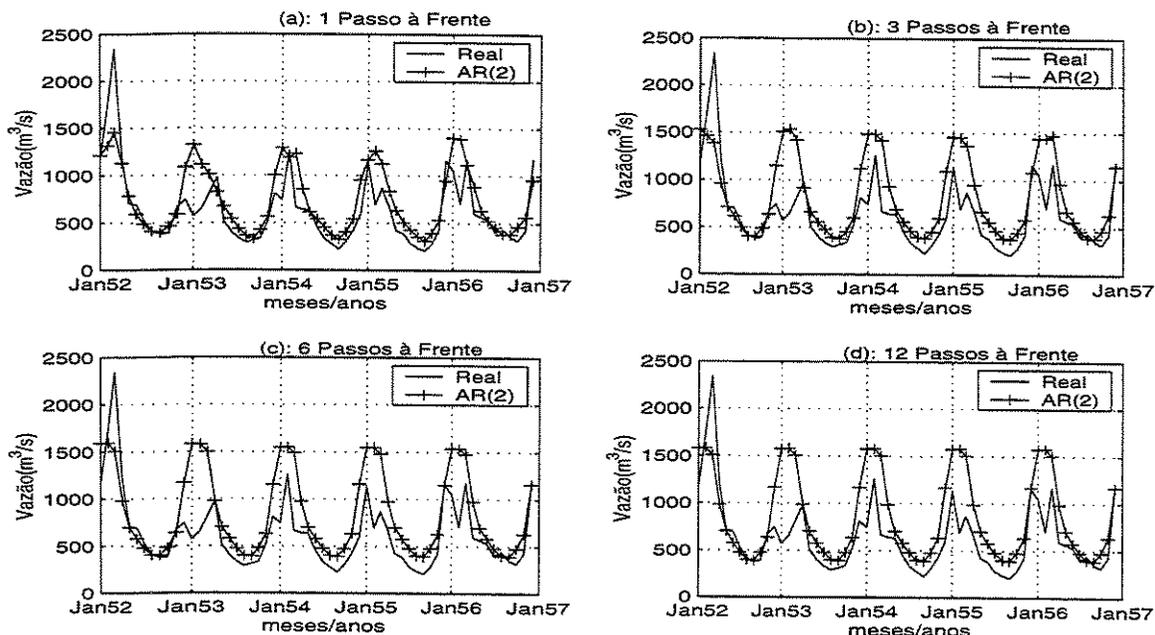


Figura 3.5: Previsão de Vazões Usando o Modelo AR(2) para o Posto de Furnas - Período de 1952 a 1956.

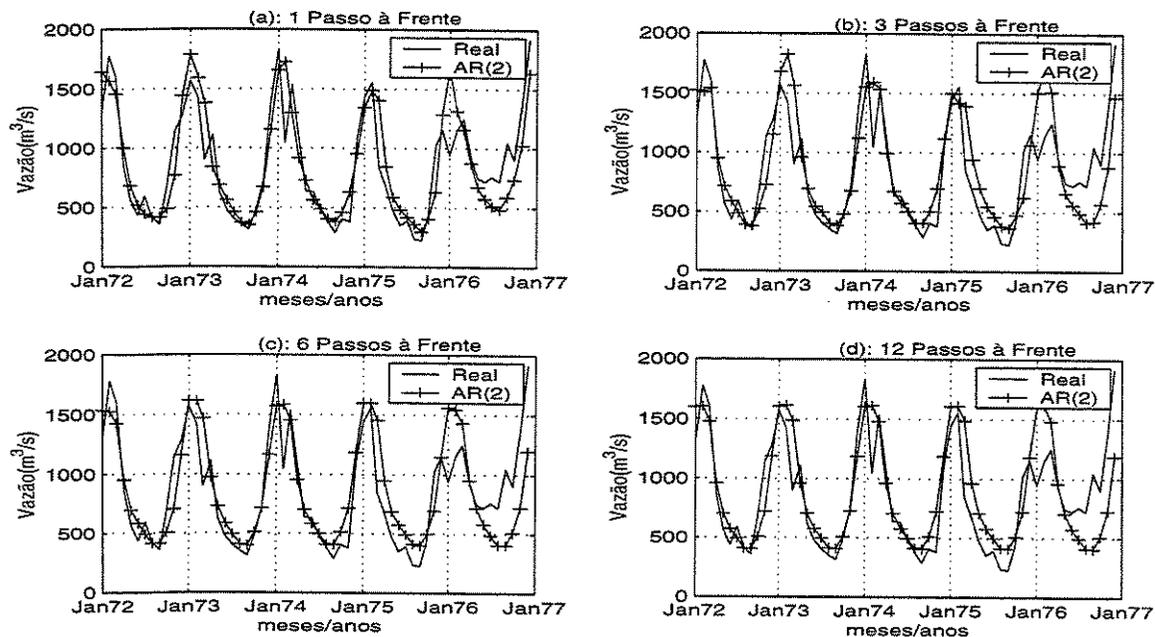


Figura 3.6: Previsão de Vazões Usando o Modelo AR(2) para o Posto de Furnas - Período de 1972 a 1976.

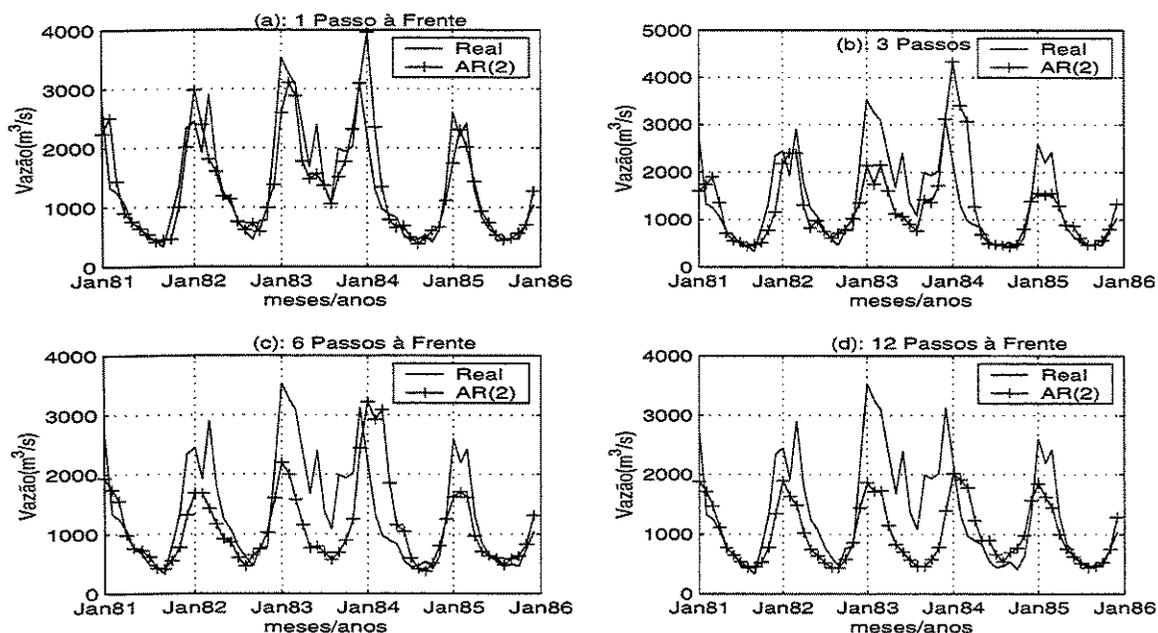


Figura 3.7: Previsão de Vazões Usando o Modelo AR(2) para o Posto de Furnas - Período de 1981 a 1985.

Tabela 3.3: Médias Mensais e Valores Previstos com o Modelo AR(2) para a Série de Vazões do Posto de Furnas - m^3/s .

| Meses | MLT | 1952 | 1972 | 1981 |
|-----------|--------|--------|--------|--------|
| Janeiro | 1719.3 | 1587.8 | 1532.6 | 1940.4 |
| Fevereiro | 1669.2 | 1593.9 | 1523.3 | 1737.3 |
| Março | 1518.0 | 1505.9 | 1424.9 | 1550.4 |
| Abril | 1019.6 | 981.1 | 951.5 | 979.3 |
| Maio | 743.5 | 695.9 | 695.6 | 760.2 |
| Junho | 623.0 | 578.4 | 585.1 | 725.9 |
| Julho | 515.5 | 478.0 | 493.2 | 613.9 |
| Agosto | 431.2 | 403.4 | 413.3 | 418.2 |
| Setembro | 447.5 | 401.6 | 412.7 | 409.6 |
| Outubro | 532.4 | 487.0 | 508.0 | 549.7 |
| Novembro | 747.8 | 638.7 | 705.9 | 774.9 |
| Dezembro | 1268.2 | 1171.7 | 1159.7 | 1312.1 |

3.4 Modelos Periódicos Auto-Regressivos

Uma outra classe de modelos para séries temporais a ser considerada são os modelos periódicos auto-regressivos (PAR). Estes modelos foram inicialmente propostos por (Jones e Brelsford, 1967) e posteriormente estudados por (Vecchia, 1985) e (Anderson e Vecchia, 1993).

As estimativas dos parâmetros dos modelos periódicos auto-regressivos podem ser realizadas através dos métodos dos momentos como proposto por (Pagano, 1978), através do método de máxima verossimilhança (Vecchia, 1985) ou, ainda, através do método de inferência Bayesiana para modelos de séries sazonais proposto em (Hutter, 1998), (Vieira et al., 1998), (Salas et al., 1982), (Sales et al., 1998).

Sejam x_1, x_2, \dots, x_{ns} observações consecutivas de uma série temporal sazonal com períodos s e número de anos n . As séries de vazões consideradas neste trabalho tratam-se de séries mensais com período de 12 meses, ou seja, $s = 12$. O índice de tempo t pode ser escrito como função do número de anos $r = 1, 2, \dots, n$ e dos meses $m = 1, 2, \dots, s$, ou seja, $t = t(r, m) = (r - 1)s + m$. Assim, $x_{t(r, m)}$ é a observação da série x_t no ano r e mês m .

Se a média $\mu_{t(r, m)} = E(x_{t(r, m)})$ e a função de autocovariância $\gamma_{t(r, m)}(j) = \text{cov}(x_{t(r, m)}, x_{t(r, m)+j})$ existem e dependem somente de j e m , então diz-se que o processo $\{x_{t(r, m)}, r = 1, \dots, n; m = 1, \dots, s\}$ é periodicamente estacionário com período s e para dois números quaisquer $j \geq 0$ e $k \geq 0$ tem-se:

$$\mu_{t(r, m)} = \mu_{t(r, m)+ks} \quad (3.35)$$

$$\gamma_{t(r, m)}(j) = \gamma_{t(r, m)+ks}(j) \quad (3.36)$$

onde $\gamma_{t(r, m)}(0) = \sigma_m^2$ é a variância do processo no mês m .

As equações (3.35) e (3.36) permitem escrever a média, a variância e a função de autocovariância somente como função do mês m . Por exemplo, para $r = 1$ tem-se que:

$$\mu_{t(r, m)} = \mu_{t(1, m)} = \mu_m$$

$$\sigma_{t(r, m)}^2 = \sigma_{t(1, m)}^2 = \sigma_m^2$$

$$\gamma_{t(r, m)}(j) = \gamma_{t(1, m)}(j) = \gamma_m(j)$$

Os modelos usuais para descrever séries temporais periodicamente estacionárias são os modelos periódicos auto-regressivos de ordem p_m , $m = 1, \dots, s$, denotados por $\text{PAR}(p_m)$, podendo ser representados da seguinte forma:

$$z_{t(r,m)} = \sum_{i=1}^{p_m} \phi_{i,m} z_{t(r,m)-i} + a_{t(r,m)}, \quad t(r,m) \geq p_m + 1 \quad (3.37)$$

onde, $\phi_{i,m}$ é uma função periódica com período s , $a_{t(r,m)}$ é um ruído branco com $N(0, \tau_m^{-1})$, $\tau_m^{-1} = \sigma_m^2$, $\tau_m > 0$ e $z_{t(r,m)}$ representa a série padronizada

$$z_{t(r,m)} = \frac{x_{t(r,m)} - \mu_m}{\sigma_m}$$

Os modelos definidos em (3.37) são ajustados usando análise Bayesiana para estimar os parâmetros $\phi_{i,m}$ e τ_m . Nesta aproximação, a função de verossimilhança condicional pode ser escrita como:

$$L_m(\Phi_m, \tau_m | \mathbf{Z}_m) \propto \tau_m^{n_m/2} \exp \left\{ -\frac{\tau_m}{2} (\mathbf{Z}_m - \mathbf{X}_m \Phi_m)' (\mathbf{Z}_m - \mathbf{X}_m \Phi_m) \right\} \quad (3.38)$$

onde,

$$n_m = \begin{cases} n-1, & m = 1, \dots, p_m \\ n, & m = p_m + s, s = 1, \dots, 12 \end{cases}$$

$$\mathbf{Z}_m = \begin{pmatrix} z_{t(1,m)} \\ z_{t(2,m)} \\ \vdots \\ z_{t(n,m)} \end{pmatrix}_{n \times 1} ; \quad \Phi_m = \begin{pmatrix} \phi_{1,m} \\ \phi_{2,m} \\ \vdots \\ \phi_{p_m,m} \end{pmatrix}_{p_m \times 1}$$

$$\mathbf{X}_m = \begin{pmatrix} z_{t(1,m)-1} & \cdots & z_{t(1,m)-p_m} \\ z_{t(2,m)-1} & \cdots & z_{t(2,m)-p_m} \\ \vdots & \ddots & \cdots \\ z_{t(n,m)-1} & \cdots & z_{t(n,m)-p_m} \end{pmatrix}_{n \times p_m}$$

Se $m \geq p_m$, o vetor \mathbf{Z}_m será de dimensão $(n - 1) \times 1$ e a matriz \mathbf{X}_m será de dimensão $(n - 1) \times p_m$, idênticos aos definidos acima, porém sem a primeira linha.

A função de verossimilhança (3.38) representa somente a densidade de probabilidade conjunta das observações \mathbf{Z}_m . Por esta razão, denota-se esta função por $L_m(\Phi_m, \tau_m | \mathbf{Z}_m)$, condicionada somente aos dados \mathbf{Z}_m .

Os parâmetros Φ_m e τ_m podem ser estimados calculando os estimadores de máxima verossimilhança tomando-se o logaritmo de (3.38), ou seja,

$$l_m(\Phi_m, \tau_m | \mathbf{Z}_m) = \frac{n_m}{2} \ln(\tau_m) - \frac{\tau_m}{2} (\mathbf{Z}_m - \mathbf{X}_m \Phi_m)' (\mathbf{Z}_m - \mathbf{X}_m \Phi_m) \quad (3.39)$$

Derivando (3.39) com relação a Φ_m e τ_m e igualando a zero, os estimadores de máxima verossimilhança são dados por:

$$\hat{\Phi}_m = (\mathbf{X}_m' \mathbf{X}_m)^{-1} (\mathbf{X}_m' \mathbf{Z}_m) \quad (3.40)$$

$$\hat{\tau}_m^{-1} = \frac{1}{n_m} (\mathbf{Z}_m - \mathbf{X}_m \hat{\Phi}_m)' (\mathbf{Z}_m - \mathbf{X}_m \hat{\Phi}_m) \quad (3.41)$$

Uma outra forma de estimar os parâmetros Φ_m e τ_m é utilizar inferência Bayesiana.

3.4.1 Inferência Bayesiana

A abordagem de inferência Bayesiana consiste em combinar a função de verossimilhança, vista como uma distribuição dos dados, condicionada ao conhecimento dos parâmetros e uma distribuição *a priori* para os parâmetros.

Neste enfoque, os parâmetros são interpretados como variáveis aleatórias, ao invés de constantes, como é feito na abordagem clássica. A combinação dessas duas distribuições através do teorema de Bayes resulta na distribuição a posteriori dos parâmetros. A inferência Bayesiana dos parâmetros é feita através das estatísticas (média, moda, mediana e desvio-padrão) da distribuição a posteriori.

Um ponto crítico dessa abordagem é a seleção da densidade *a priori* sobre os parâmetros. Este trabalho será restrito a seleção dos parâmetros ao caso da densidade *a*

priori não informativa de Jeffreys (Box e Tiao, 1973). Mas, deve-se levar em conta que a principal vantagem da abordagem Bayesiana é usar informações sobre a distribuição de probabilidade dos parâmetros.

3.4.2 Densidade *A Priori* Não Informativa

Quando se tem pouca ou nenhuma informação sobre os parâmetros, a aproximação Bayesiana sugere o uso de uma densidade de probabilidade *a priori* não informativa. Assumindo que os parâmetros Φ_m e τ_m são variáveis independentes e usando a densidade não informativa de Jeffreys (Box e Tiao, 1973), a densidade *a priori* conjugada para Φ_m e τ_m é dada por:

$$\pi_m^0(\Phi_m, \tau_m) \propto \frac{1}{\tau_m}, \quad \tau_m > 0 \quad (3.42)$$

Ajustando a função de probabilidade equação (3.38) e a densidade *a priori* equação (3.42), segue que a densidade a posteriori é dada por:

$$\pi_m(\Phi_m, \tau_m | \mathbf{Z}_m) \propto L_m(\Phi_m, \tau_m | \mathbf{Z}_m) \pi_m^0(\Phi_m, \tau_m) \quad (3.43)$$

onde $\pi_m(\Phi_m, \tau_m | \mathbf{Z}_m)$ tem distribuição Normal-Gama. Então, para $m = 1, 2, \dots, s$, a densidade a posteriori é como segue:

$$\prod_{m=1}^s \pi_m(\Phi_m, \tau_m | \mathbf{Z}_m) \propto \prod_{m=1}^s \frac{\tau_m^{\frac{n_m}{2}-1}}{\tau_m^2} \exp \left\{ -\frac{\tau_m}{2} \left[(\Phi_m - \hat{\Phi}_m)' \mathbf{X}'_m \mathbf{X}_m (\Phi_m - \hat{\Phi}_m) + (\mathbf{Z}_m - \hat{\mathbf{Z}}_m)' (\mathbf{Z}_m - \hat{\mathbf{Z}}_m) \right] \right\}$$

A independência dos parâmetros Φ_m e τ_m , $m = 1, 2, \dots, s$, permite escrever as densidades de probabilidade condicional Normal e Gama, respectivamente, como segue (Hutter, 1998):

$$\pi_m(\Phi_m | \tau_m, \mathbf{Z}_m) \sim N \left(\hat{\Phi}_m, (\tau_m V_m)^{-1} \right)$$

$$\pi_m(\tau_m | \hat{\Phi}_m, \mathbf{Z}_m) \sim \Gamma \left(\frac{n_m}{2}, B_m \right)$$

onde,

$$\widehat{\Phi}_m = (\mathbf{X}'_m \mathbf{X}_m)^{-1} \mathbf{X}'_m \mathbf{Z}_m \quad (3.44)$$

$$V_m = (n_m - 2p_m) \mathbf{X}'_m \mathbf{X}_m \left[(\mathbf{Z}_m - \widehat{\mathbf{Z}}_m)' (\mathbf{Z}_m - \widehat{\mathbf{Z}}_m) \right]^{-1} \quad (3.45)$$

$$B_m = \frac{(\mathbf{Z}_m - \widehat{\mathbf{Z}}_m)' (\mathbf{Z}_m - \widehat{\mathbf{Z}}_m)}{2} \quad (3.46)$$

Como a densidade conjunta a posteriori não é uma densidade padrão, as densidades a posteriori marginais somente podem ser avaliadas por meio de métodos de aproximação, tal como o método de Laplace (Tierney et al., 1986), ou usando métodos de simulação de Monte Carlo em Cadeia de Markov, tal como o algoritmo amostrador de Gibbs (Casella e George, 1992). Neste trabalho, é usado este algoritmo para gerar as amostras dos parâmetros Φ_m e τ_m .

O método amostrador de Gibbs é usado para gerar variáveis aleatórias de uma distribuição sem utilizar sua densidade. O algoritmo é essencialmente um esquema iterativo de uma cadeia de Markov cujo núcleo de transição é formado pelas distribuições condicionais (Oliveira, 1998), (Hutter, 1998).

3.4.3 Densidade Preditiva

Em muitas ocasiões, dada a informação amostral, pode-se ter interesse em fazer inferência sobre algum valor que ainda não foi observado, isto é, fazer uma previsão para um valor futuro. Na análise Bayesiana, a função densidade de probabilidade para um valor ainda não observado, dada a informação amostral, é conhecida como densidade preditiva para o valor futuro.

Assim, realiza-se a previsão para o mês m , $\widehat{z}_{t(n+1,m)}$, usando a função densidade preditiva Bayesiana, condicionada às observações anteriores $\mathbf{Z}_m = (z_{t(1,m)}, \dots, z_{t(n,m)})$, como segue:

$$\Pi_m(z_{t(n+1,m)} | \mathbf{Z}_m) = \int_{\phi_{1,m}} \int_{\phi_{2,m}} \cdots \int_{\phi_{p_m,m}} \int_{\tau_m} \Pi_m(z_{t(n+1,m)} | \Phi_m, \tau_m, \mathbf{Z}_m) \times$$

$$\times \Pi_m(\Phi_m, \tau_m | \mathbf{Z}_m) d\phi_{1,m} d\phi_{2,m} \dots d\phi_{p_m,m} d\tau_m \quad (3.47)$$

onde $\Pi_m(\Phi_m, \tau_m | \mathbf{Z}_m)$ é a densidade a posteriori dos parâmetros Φ_m e τ_m , e $\Pi_m(z_{t(n+1,m)} | \mathbf{Z}_m)$ é a função de densidade de $z_{t(n+1,m)}$, condicionada a estes parâmetros e aos dados \mathbf{Z}_m . Desta forma, tem-se que:

$$\Pi_m(z_{t(n+1,m)} | \mathbf{Z}_m) = E(\Pi_m(z_{t(n+1,m)} | \mathbf{Z}_m)) \quad (3.48)$$

A solução da equação (3.47) é difícil de ser obtida analiticamente. Uma alternativa é utilizar o método de simulação de Monte Carlo. Para isso, sejam as amostras $\Phi^{(j)} = (\phi_{1,m}^{(j)}, \phi_{2,m}^{(j)}, \dots, \phi_{p_m,m}^{(j)})$, $j = 1, 2, \dots, m$ geradas a partir do algoritmo amostrador de Gibbs. Um estimador de Monte Carlo para (3.47) é dado por:

$$E(z_{t(n+1,m)} | \mathbf{Z}_m) = \int_{-\infty}^{+\infty} z_{t(n+1,m)} \Pi_m(z_{t(n+1,m)} | \mathbf{Z}_m) dz_{t(n+1,m)} \quad (3.49)$$

Substituindo a equação (3.47) em (3.49) e mudando a ordem de integração a equação (3.49) pode ser rescrita da seguinte forma:

$$\begin{aligned} E(z_{t(n+1,m)} | \mathbf{Z}_m) &= \int_{\phi_{1,m}} \int_{\phi_{2,m}} \dots \int_{\phi_{p_m,m}} \int_{\tau_m} E(z_{t(n+1,m)} | \Phi_m, \tau_m, \mathbf{Z}_m) \times \\ &\times \Pi_m(\Phi_m, \tau_m | \mathbf{Z}_m) d\phi_{1,m} d\phi_{2,m} \dots d\phi_{p_m,m} d\tau_m \end{aligned} \quad (3.50)$$

Assim, um estimador de Monte Carlo para a equação (3.47), considerando as amostras $\Phi^{(j)}$ e τ_m^j é dado por:

$$\widehat{E}(\Pi_m(z_{t(n+1,m)} | \mathbf{Z}_m)) = \frac{1}{M} \sum_{j=1}^M E(z_{t(n+1,m)} | \Phi_m^{(j)}, \tau_m^{(j)}, \mathbf{Z}_m) \quad (3.51)$$

A equação (3.51) fornece a estimativa para a previsão de $z_{t(n+1,m)}$ denotada por $\widehat{z}_{t(n+1,m)} = \widehat{E}(\Pi_m(z_{t(n+1,m)} | \mathbf{Z}_m))$. Nota-se que a equação (3.51) apresenta cálculos simples de previsão devido ao uso do algoritmo de simulação de Monte Carlo em Cadeia de Markov (MCMC).

Usando o modelo PAR(p_m), escrito como segue:

$$z_{t(n+1,m)} = \sum_{i=1}^{p_m} \phi_{i,m} z_{t(n+1,m)-i} + a_{t(n-1,m)} \quad (3.52)$$

onde $a_{t(n-1,m)} \sim N(0, \tau_m^{-1})$, a função de probabilidade de $\hat{z}_{t(n+1,m)}$, condicionada aos parâmetros Φ_m e τ_m e às observações anteriores, é da seguinte forma:

$$\begin{aligned} & \Pi_m(z_{t(n+1,m)} | \Phi_m, \tau_m, \mathbf{Z}_m) \propto \\ & \tau_m^{1/2} \exp \left\{ -\frac{\tau_m}{2} \left(z_{t(n+1,m)} - \phi_{1,m} z_{t(n+1,m)-1} - \dots - \phi_{p_m,m} z_{t(n+1,m)-p_m} \right)^2 \right\} \end{aligned} \quad (3.53)$$

A densidade preditiva Bayesiana para a observação futura $z_{t(n+1,m)}$ apresenta um valor esperado como segue:

$$E(z_{t(n+1,m)} | \Phi_m, \tau_m, \mathbf{Z}_m) = \phi_{1,m} z_{t(n+1,m)-1} + \dots + \phi_{p_m,m} z_{t(n+1,m)-p_m} \quad (3.54)$$

Observa-se que $z_{t(n+1,m)} = z_{n_s+1}$ e os valores anteriores são dados por $z_{t(n+1,m)-1} = z_{n_s}$, $z_{t(n+1,m)-2} = z_{n_s-1}$, \dots , $z_{t(n+1,m)-p_m} = z_{n_s-p_m}$. Dessa forma,

$$E(z_{t(n+1,m)} | \Phi_m^{(j)}, \tau_m^{(j)}, \mathbf{Z}_m) = \phi_{1,m}^{(j)} z_{n_s} + \dots + \phi_{p_m,m}^{(j)} z_{t(n+1,m)-p_m} \quad (3.55)$$

Substituindo a equação (3.55) na equação (3.51), a previsão de $\hat{z}_{t(n+1,m)}$ pode ser computada usando o método de simulação de Monte Carlo.

3.4.4 Seleção de Modelos

Na análise Bayesiana existem vários critérios para a seleção de modelos (Chang, 1995). Neste trabalho, é considerada a técnica que utiliza densidades preditivas ordenadas (Hutter, 1998).

Usando o método de Monte Carlo para estimar a densidade preditiva $c_{t(r,m)} = \hat{\Pi}_m(z_{t(n+1,m)} | \mathbf{Z}_m)$ (equação (3.48)) o modelo que apresentar o maior valor de $c_{t(r,m)}$ é

indicado como sendo o melhor. Ou seja, o critério da preditiva Bayesiana seleciona o modelo que maximiza o produto:

$$c(p) = \prod_{r=1}^n c_{t(r,m)}(p) \quad (3.56)$$

onde p é o p -ésimo modelo auto-regressivo periódico. Por exemplo, se $p = 1$ significa que a densidade preditiva $c_{t(r,m)}$ esta sendo calculada para o modelo PAR(1) para um certo mês m .

3.4.5 Ajuste de Modelos Periódicos Auto-Regressivos para a Série de Vazões

Considere a série de vazões naturais afluentes médias mensais do reservatório de Furnas localizado no Sudeste do Brasil, analisada no capítulo 2. As funções de autocorrelação e autocorrelação parcial (seções 2.4 e 3.3.4) indicaram o ajuste de um modelo de ordem 2. Assim, para esta série de vazões será ajustado um modelo periódico auto-regressivo de ordem 2, PAR(2), escrito da seguinte forma:

$$z_{t(r,m)} = \phi_{1,m} z_{t(r,m)-1} + \phi_{2,m} z_{t(r,m)-2} + a_{t(r,m)} \quad (3.57)$$

onde

$$z_{t(r,m)} = \frac{x_{t(r,m)} - \hat{\mu}_m}{\hat{\sigma}_m}$$

sendo $\hat{\mu}_m$ a estimativa da média mensal e $\hat{\sigma}_m$ a estimativa do desvio padrão mensal.

Para inferir os parâmetros $\Phi_m = [\phi_{1,m} \quad \phi_{2,m}]'$ e τ_m , considera-se o método de máxima verossimilhança e inferência Bayesiana. As densidades *a priori* não informativas utilizaram o Método de Simulação em Cadeia de Markov para aproximar as densidades a posteriori marginais de Φ_m e τ_m . A tabela 3.4 apresenta os valores dos parâmetros estimados usando o método de máxima verossimilhança (EMV) e as estimativas usando inferência Bayesiana para cada mês.

Com os resultados obtidos na tabela 3.4, observa-se que as estimativas pontuais são próximas para os parâmetros Φ_m e τ_m , considerando o método de máxima verossimilhança que é um método clássico e usando o método Bayesiano com densidade *a priori*

Tabela 3.4: Valores Estimados dos Parâmetros para a Série de Vazões do Posto de Furnas.

| Meses | EMV | | | Inferência Bayesiana | | |
|-----------|--------------------|--------------------|----------|----------------------|--------------------|----------|
| | $\hat{\phi}_{1,m}$ | $\hat{\phi}_{2,m}$ | τ_m | $\hat{\phi}_{1,m}$ | $\hat{\phi}_{2,m}$ | τ_m |
| Janeiro | 0.1581 | 0.3702 | 1.2100 | 0.1438 | 0.3760 | 1.2017 |
| Fevereiro | 0.5030 | 0.0551 | 1.4991 | 0.4947 | 0.0357 | 1.4794 |
| Março | 0.4903 | 0.2366 | 1.5967 | 0.5162 | 0.2323 | 1.5287 |
| Abril | 0.6107 | 0.2614 | 2.5664 | 0.5836 | 0.2680 | 2.5177 |
| Maiο | 0.6448 | 0.2823 | 4.1624 | 0.6445 | 0.2779 | 4.1238 |
| Junho | 0.6867 | 0.2092 | 3.9363 | 0.6739 | 0.2242 | 3.8945 |
| Julho | 0.6011 | 0.3764 | 8.9106 | 0.6006 | 0.3932 | 8.6427 |
| Agosto | 0.9667 | -0.0254 | 8.4488 | 0.9781 | -0.0311 | 8.3586 |
| Setembro | 0.4710 | 0.3850 | 3.2394 | 0.4745 | 0.3858 | 3.1574 |
| Outubro | 0.5806 | 0.2335 | 2.7564 | 0.6025 | 0.2758 | 2.7057 |
| Novembro | 0.8329 | -0.0661 | 2.3904 | 0.8140 | -0.0544 | 2.3549 |
| Dezembro | 0.3661 | 0.3783 | 1.8395 | 0.3490 | 0.3856 | 1.8029 |

não informativa. Logo, a abordagem Bayesiana pode ser usada de forma segura além de apresentar simplicidade nos cálculos na inferência dos parâmetros. A tabela 3.5 mostra os valores de $\bar{c}_{t(r,m)}(p)$, que é a média da densidade preditiva $c_{t(r,m)}(p)$, e do produto $c_{t(r,m)}(p)$ da densidade preditiva para o modelo $p = 2$ selecionado.

Através da abordagem Bayesiana com uso da densidade *a priori* não informativa, a avaliação dos modelos ajustados é verificada para os seguintes anos: 1952 a 1956, período seco; 1972 a 1976 considerado um período mediano; e 1981 a 1985 período úmido. Foram realizados os ajustes dos parâmetros Φ_m e τ_m para cada período de teste, ou seja, para o período de 1952 a 1956, os parâmetros foram estimados para os dados de vazões até 1951; para o período de 1972 a 1976 os parâmetros foram estimados para os dados até 1971; e para o período úmido os parâmetros foram estimados para os dados até 1980. As figuras 3.8, 3.9 e 3.10 mostram as previsões obtidas para cada período e as previsões 1, 3, 6 e 12 passos à frente são mostradas na figuras (a), (b), (c) e (d), respectivamente.

Tabela 3.5: Estimativas para a Seleção do Modelo para a Série de Vazões do Posto de Furnas.

| Meses | $\bar{c}_{t(r,m)}(l)$ | $c_{t(r,m)}(l)$ |
|-----------|-----------------------|-----------------|
| Janeiro | 0.7553 | 0.0001 |
| Fevereiro | 0.8849 | 0.0146 |
| Março | 0.9437 | 0.0679 |
| Abril | 1.0535 | 0.0062 |
| Mai | 1.4082 | 219.8453 |
| Junho | 1.5522 | 0.2564 |
| Julho | 2.0634 | 182.9681 |
| Agosto | 1.6119 | 126.0080 |
| Setembro | 1.1985 | 0.4468 |
| Outubro | 0.9912 | 0.0713 |
| Novembro | 1.0286 | 0.4738 |
| Dezembro | 0.9164 | 0.0129 |

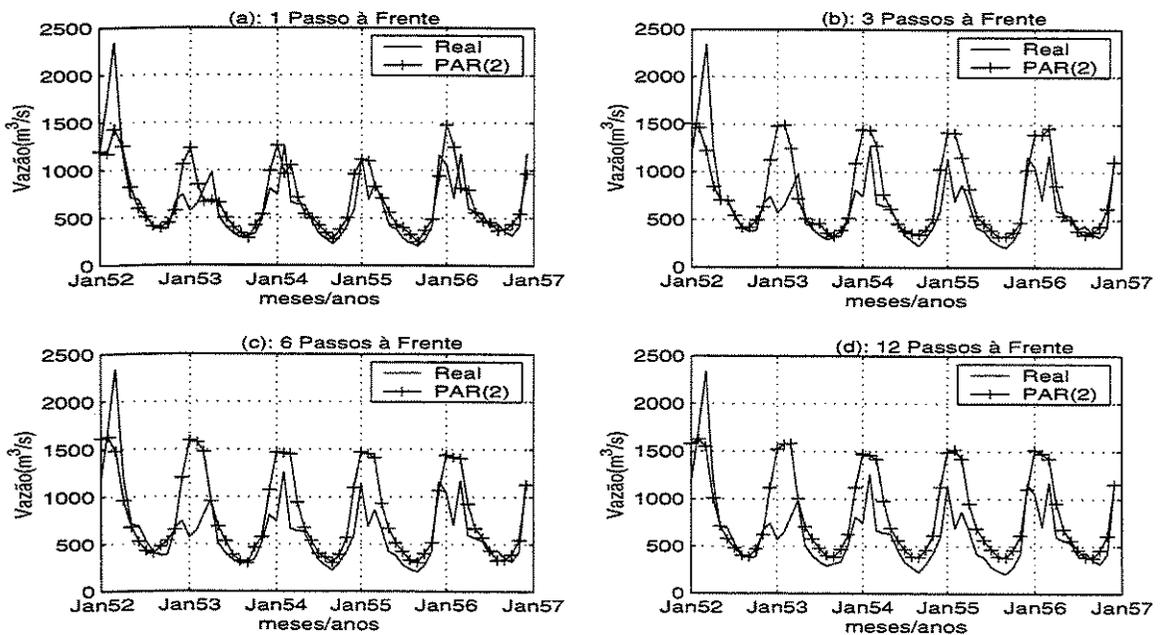


Figura 3.8: Previsão de Vazões Usando o Modelo PAR(2) para o Posto de Furnas - Período de 1952 a 1956.

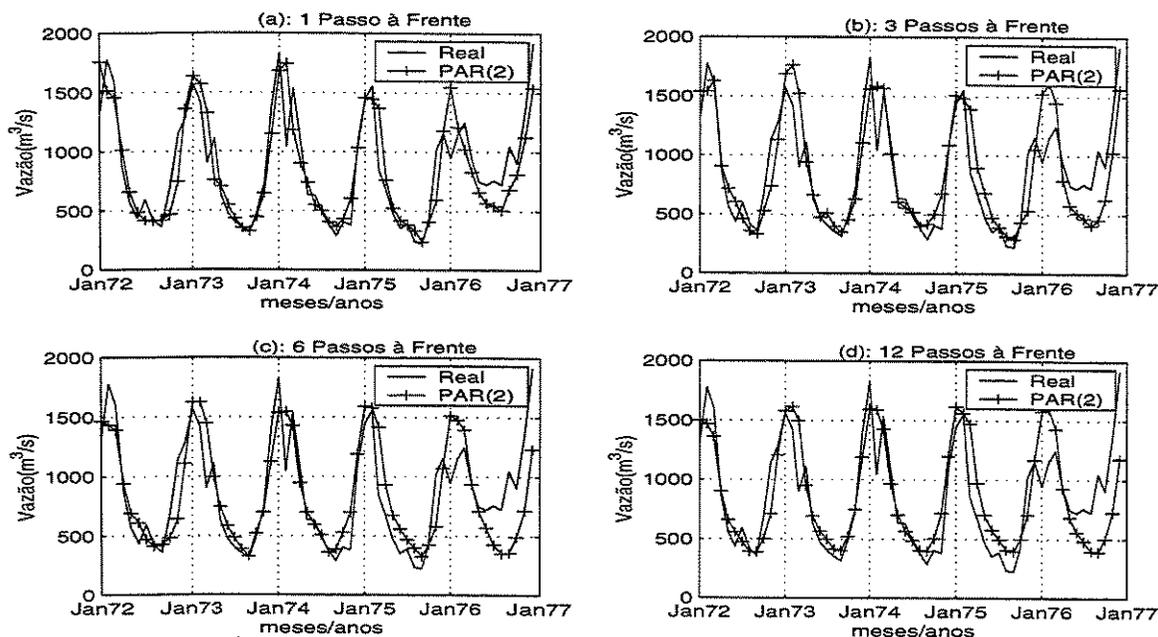


Figura 3.9: Previsão de Vazões Usando o Modelo PAR(2) para o Posto de Furnas - Período de 1972 a 1976.

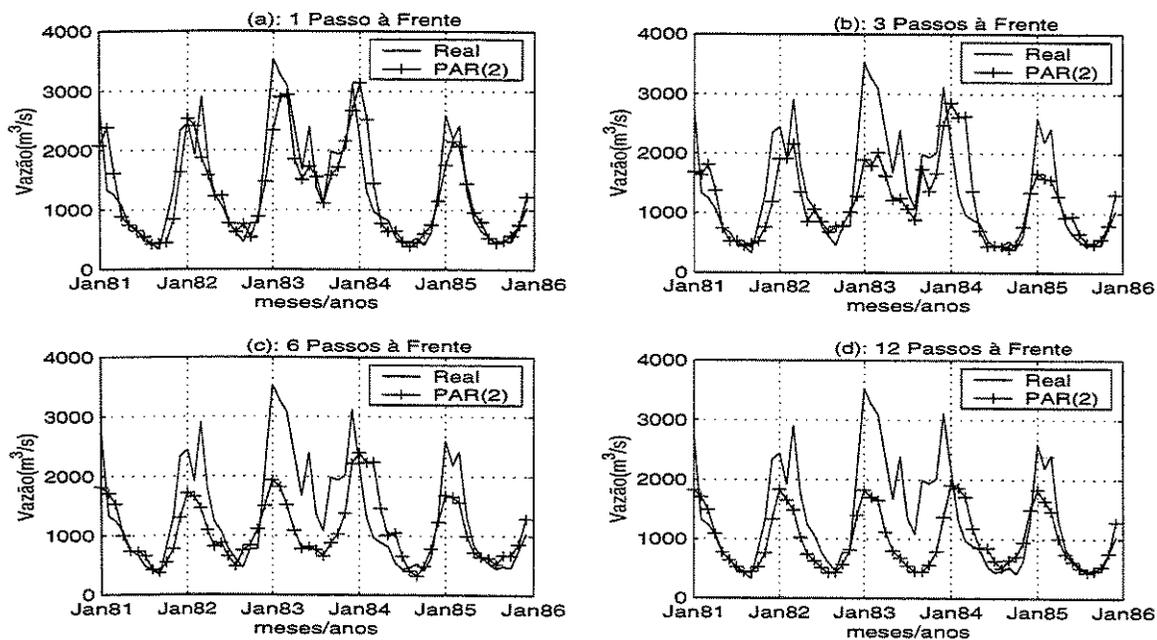


Figura 3.10: Previsão de Vazões Usando o Modelo PAR(2) para o Posto de Furnas - Período de 1981 a 1985.

Das figuras 3.8(a), 3.9(a) e 3.10(a), pode-se notar que a previsão com 1 passo à frente apresentou melhor comportamento que as previsões vários passos à frente. Também pode-se observar que, para as previsões com mais de 3 passos à frente, o comportamento começa a se repetir, notando-se uma tendência para a Média a Longo Termo como mostra a tabela 3.6. Esta tabela apresenta os dados de vazões médias mensais e as previsões 6 passos à frente para o período de Janeiro a Dezembro de 1952, 1972 e 1981. Este comportamento também foi observado quando estimado um modelo AR(2) para esta série.

Tabela 3.6: Médias Mensais e Valores Previstos com o Modelo PAR(2) para a Série de Vazões do Posto de Furnas - m^3/s .

| Meses | MLT | 1952 | 1972 | 1981 |
|-----------|--------|--------|--------|--------|
| Janeiro | 1719.3 | 1612.7 | 1463.5 | 1816.9 |
| Fevereiro | 1669.2 | 1627.9 | 1432.6 | 1709.7 |
| Março | 1518.0 | 1482.0 | 1392.0 | 1533.4 |
| Abril | 1019.6 | 963.3 | 940.4 | 1000.9 |
| Maió | 743.5 | 680.7 | 689.6 | 735.4 |
| Junho | 623.0 | 529.4 | 610.1 | 737.5 |
| Julho | 515.5 | 429.9 | 471.1 | 660.8 |
| Agosto | 431.2 | 399.7 | 414.7 | 427.3 |
| Setembro | 447.5 | 474.9 | 427.1 | 376.4 |
| Outubro | 532.4 | 536.6 | 482.9 | 551.8 |
| Novembro | 747.8 | 652.8 | 641.3 | 781.8 |
| Dezembro | 1268.2 | 1201.2 | 1108.4 | 1307.5 |

Analisando os anos de seca figura 3.8 e úmido (figura 3.10), pode-se observar que a previsão 1 passo à frente apresentou melhores resultados. Uma das razões para ocorrer isso é devido às grandes variações nos dados. Já para as outras previsões vários passos à frente, os modelos apresentaram um desempenho inferior. Para os anos medianos (figura 3.9) o modelo segue a tendência da série já que este período não apresenta grandes variações, mas pode-se observar uma tendência para a MLT.

O próximo passo é selecionar o “melhor” modelo para cada mês. Isto significa ajustar um modelo PAR(p_m) para cada mês do ano analisando a densidade preditiva e escolhendo a ordem do modelo que apresentar o maior valor do produto de $c_{t(r,m)}(p)$ e menor variância dos resíduos ($\sigma_m^2 = \tau_m^{-1}$) (Ballini et al., 1999a). A tabela 3.7 mostra os valores da densidade preditiva e a ordem dos modelos selecionados. Pode-se verificar que a ordem do modelo é maior para os meses de seca, isto devido à forte correlação entre estes

meses. A tabela 3.8 apresenta os valores das estimativas dos parâmetros $\hat{\Phi}_m$ e τ_m .

Tabela 3.7: Ordem dos Modelos Seleccionados para a Série de Vazões do Posto de Furnas.

| Meses | $\bar{c}_{t(r,m)}(p)$ | $c_{t(r,m)}(p)$ | p |
|-----------|-----------------------|-----------------|-----|
| Janeiro | 0.7985 | 0.0025 | 1 |
| Fevereiro | 0.8849 | 0.0146 | 2 |
| Março | 0.9437 | 0.0679 | 2 |
| Abril | 1.1558 | 0.2402 | 3 |
| Mai | 1.4573 | 243.6561 | 3 |
| Junho | 1.5322 | 2.1147 | 3 |
| Julho | 2.1082 | 264.8665 | 3 |
| Agosto | 1.8000 | 2704.8000 | 4 |
| Setembro | 1.3587 | 45.6093 | 4 |
| Outubro | 1.3338 | 40.6173 | 3 |
| Novembro | 1.0344 | 0.4740 | 1 |
| Dezembro | 0.9164 | 0.0129 | 2 |

Ajustando um modelo para cada mês pode-se observar que a previsão apresenta um desempenho satisfatório, tentando acompanhar a tendência da série. Este comportamento é verificado para as previsões vários passos à frente, como mostram as figuras 3.11, 3.12 e 3.13 em (a), (b), (c) e (d), respectivamente, para cada período considerado.

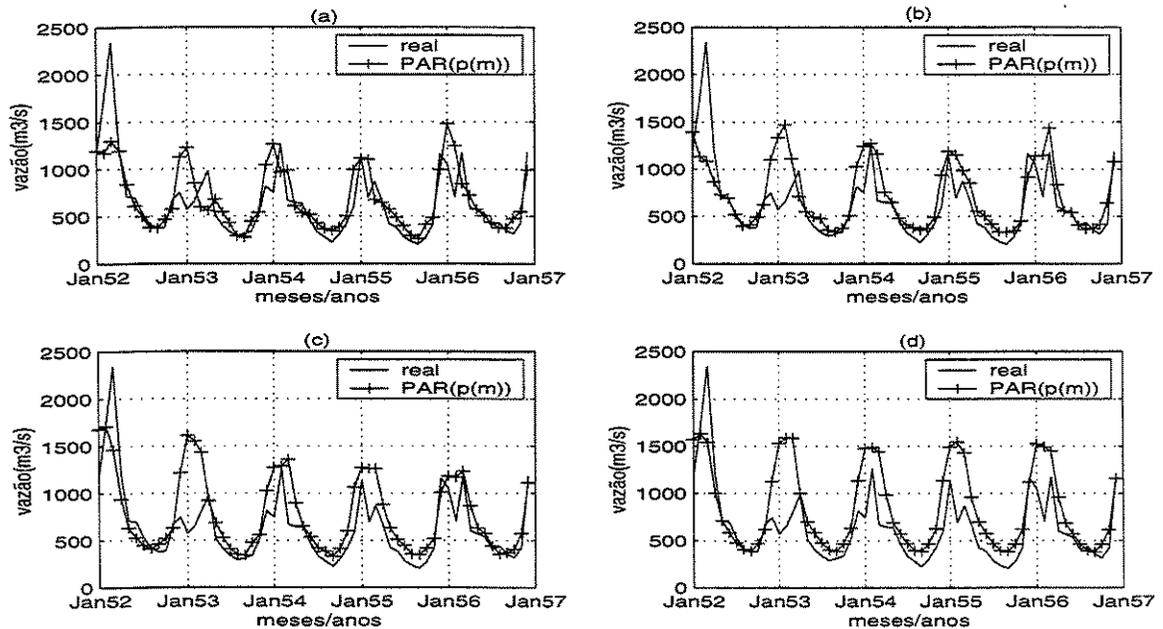
Novamente, no caso da previsão 1 passo à frente, para os três períodos com características distintas, foi verificado um melhor desempenho do modelo quando comparado com as previsões 3, 6 e 12 passos à frente.

Comparando o comportamento dos modelos $PAR(p_m)$ e $PAR(2)$ pode-se verificar que, para o período de seca e úmido, os modelos $PAR(p_m)$ apresentaram um melhor desempenho. Já para o período mediano tanto os modelos $PAR(p_m)$ como $PAR(2)$ apresentaram um desempenho semelhante. Portanto, para dados de vazões onde o período não é considerado crítico pode-se utilizar um modelo de ordem fixa para todo mês.

Também pode-se notar que, para todos os períodos, as previsões para os meses de cheia (Janeiro a Março), os modelos $PAR(p_m)$ mostraram um melhor desempenho. Uma outra característica observada é que para os meses de seca (Maio a Outubro) as previsões, para todos os modelos considerados neste capítulo, apresentaram comportamento semelhante.

Tabela 3.8: Valores Estimados dos Parâmetros Usando Inferência Bayesiana para a Série de Vazões do Posto de Furnas.

| Meses | $\hat{\phi}_{1,m}$ | $\hat{\phi}_{2,k}$ | $\hat{\phi}_{3,k}$ | $\hat{\phi}_{4,k}$ | τ_m |
|-----------|--------------------|--------------------|--------------------|--------------------|----------|
| Janeiro | 0.4146 | - | - | - | 1.0953 |
| Fevereiro | 0.4947 | 0.0357 | - | - | 1.4794 |
| Março | 0.5162 | 0.2323 | - | - | 1.5287 |
| Abril | 0.5724 | 0.1886 | 0.2429 | - | 2.7352 |
| Mai | 0.5318 | 0.2111 | 0.2631 | - | 5.1579 |
| Junho | 0.7126 | 0.2668 | -0.1010 | - | 3.9809 |
| Julho | 0.5789 | 0.3572 | 0.0255 | - | 8.8249 |
| Agosto | 0.9144 | -0.0382 | 0.2438 | 0.1993 | 9.3935 |
| Setembro | 0.6220 | 0.1722 | 0.6358 | -0.5645 | 4.7330 |
| Outubro | 0.4841 | -0.2231 | 0.5804 | - | 3.3125 |
| Novembro | 0.8829 | - | - | - | 2.4869 |
| Dezembro | 0.3490 | 0.3856 | - | - | 1.8029 |

Figura 3.11: Previsão de Vazões Usando o Modelo $PAR(p_m)$ para o Posto de Furnas - Período de 1952 a 1956.

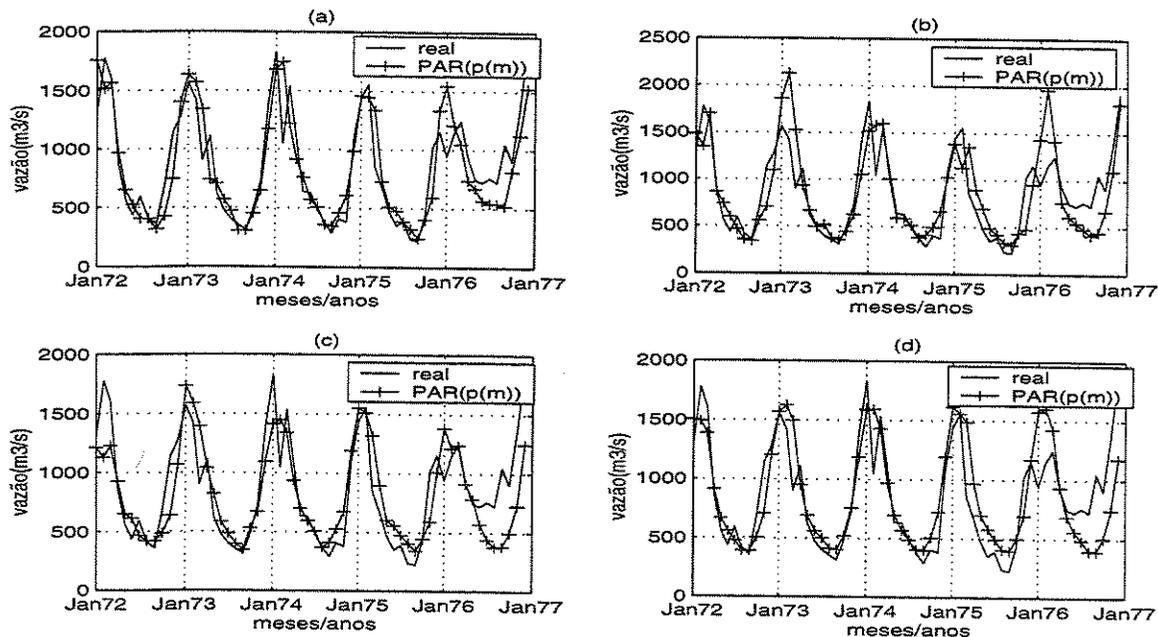


Figura 3.12: Previsão de Vazões Usando o Modelo $PAR(p_m)$ para o Posto de Furnas - Período de 1972 a 1976.

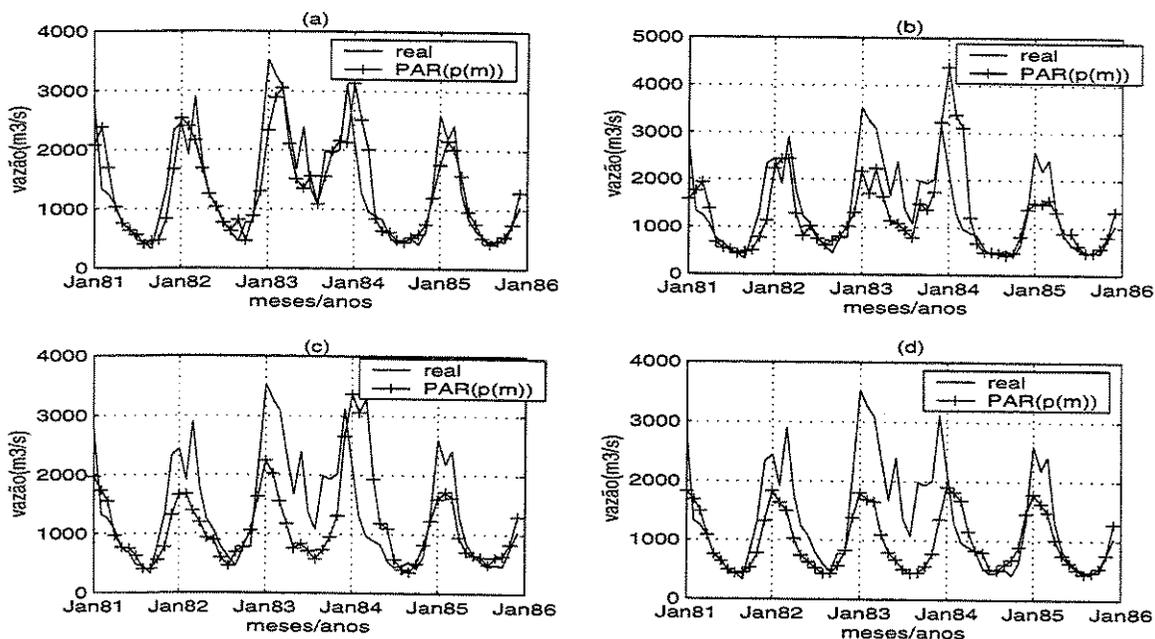


Figura 3.13: Previsão de Vazões Usando o Modelo $PAR(p_m)$ para o Posto de Furnas - Período de 1981 a 1985.

3.5 Resumo

Após a análise de uma série temporal, a identificação de um modelo de séries temporais é realizada e, assim, os parâmetros do modelo identificado são estimados.

Neste capítulo, primeiramente foi ajustado um modelo auto-regressivo, sendo a ordem do modelo determinada através da análise das funções de autocorrelação e autocorrelação parcial. Neste caso, os parâmetros foram estimados através do método dos momentos.

A seguir, modelos periódicos auto-regressivos foram propostos para a previsão de vazões afluentes naturais médias mensais, pois o uso de modelos auto-regressivos pode tornar-se inadequado quando existe correlação do valor presente com o valor passado e com as médias passadas para o mesmo mês de períodos anteriores das séries hidrológicas. Neste caso, os parâmetros foram estimados através do método de inferência Bayesiana para modelos de séries sazonais. A escolha da ordem do modelo foi feita através da densidade preditiva e a estimação dos parâmetros foi feita através da densidade *a priori* não informativa de Jeffreys.

Os modelos descritos neste capítulo foram aplicados aos dados de vazões do posto de Furnas. Nesta fase, a ênfase deste capítulo foi dada ao ajuste dos modelos e no capítulo 6 será realizado uma comparação entre os modelos, analisando os erros de previsão através de diferentes estatísticas.

No próximo capítulo é apresentado um modelo de rede neural, conhecido como rede multi-camadas com algoritmo de retro-propagação, que vem sendo aplicado a previsão de séries temporais.

Capítulo 4

Redes Neurais Multi-Camadas

Este capítulo apresenta uma breve introdução sobre os conceitos da teoria de redes neurais e o modelo de rede neural perceptron multi-camadas. A rede perceptron multi-camadas vem sendo frequentemente aplicada para resolver uma variedade de problemas de processamento de informação, incluindo classificação de padrões, aproximação de funções e previsão de séries temporais.

O algoritmo fundamental para a aprendizagem da rede multi-camadas é o algoritmo de retro-propagação, o qual é baseado no método do gradiente descendente. A derivação do algoritmo de retro-propagação e considerações sobre as virtudes e limitações das redes perceptron multi-camadas aplicadas ao problema de previsão de vazões também serão comentadas neste capítulo.

4.1 Introdução

Redes Neurais Artificiais (RNA) podem ser definidas, basicamente, como sendo sistemas de processamento paralelo e distribuído baseadas no sistema nervoso biológico. Os modelos de redes neurais são compostos de elementos computacionais, chamados neurônios, arranjados em padrões semelhantes a redes neurais biológicas.

A figura 4.1 ilustra um modelo de neurônio biológico. Um neurônio recebe os sinais de entrada através dos dendritos, processa-os no corpo celular, gerando sinais de saída que são transmitidos para outros neurônios por meio dos axônios e suas ramificações.

Os neurônios comunicam-se através de sinapses. Sinapse é o nome dado à região onde o axônio de um neurônio encontra o dendrito de outro neurônio, sendo de extrema importância para a transmissão de sinais entre os neurônios.

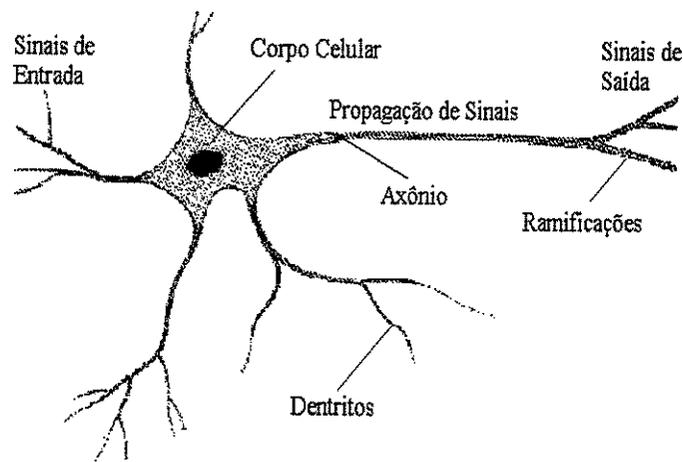


Figura 4.1: Modelo do Neurônio Biológico.

4.1.1 Um Breve Histórico da Área de Redes Neurais

A teoria de redes neurais artificiais teve início em 1943 com o trabalho de McCulloch e Pitts (McCulloch e Pitts, 1943) quando foi desenvolvido o primeiro modelo matemático para um neurônio biológico. O modelo foi proposto como uma unidade de processamento binária descrevendo um cálculo lógico que procura simular a forma, o comportamento e as funções de um neurônio biológico.

Assim sendo, os dentritos foram substituídos por entradas, cujas ligações com o corpo celular artificial são realizadas através dos pesos das conexões entre os neurônios (simulando as sinapses). Os estímulos captados pelas entradas são processados pela função soma e o limiar (*threshold*) de disparo do neurônio biológico é representado por uma função de ativação ou de transferência (figura 4.2).

Este modelo, apesar de muito simples, apresentou grande poder computacional mas faltavam-lhe características adaptativas e de aprendizagem, imprescindíveis em modelos de sistemas inteligentes.

Em 1949, o primeiro método de aprendizagem em redes neurais foi proposto pelo

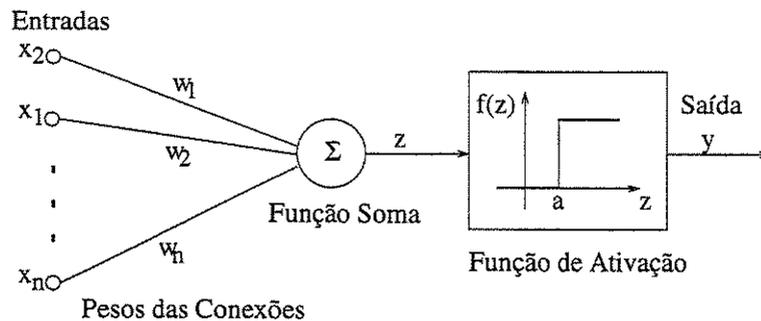


Figura 4.2: Modelo do Neurônio Artificial.

neurofisiologista Donald Hebb (Hebb, 1949). Este modelo sugere que as conexões ativas entre os neurônios tendem a se fortalecer, enquanto que as outras tendem a se enfraquecer. O trabalho de Hebb foi influente entre os psicólogos, mas infelizmente teve pouco ou nenhum impacto na comunidade de engenharia (Haykin, 1994).

No final dos anos 50, F. Roseblatt (Roseblatt, 1958) introduziu uma nova abordagem para o problema de reconhecimento de padrões com o desenvolvimento do perceptron. Roseblatt defendia a idéia de que o cérebro humano trabalha como um associador de padrões adaptável, e não como um circuito lógico determinístico como foi proposto por Minsky e von Neumann (Minsky, 1954), (von Neumann, 1956). O “Teorema da Convergência para Perceptron” foi demonstrado, sendo proposto um algoritmo para ajuste dos pesos, provando a convergência quando os padrões são linearmente separáveis (Roseblatt, 1958).

Widrow e Hoff, em 1960, introduziram o algoritmo *Least-Mean-Square* (LMS) e formularam o *Adaline* (*Adaptive Linear Element*) baseado neste algoritmo de aprendizagem (Widrow e Hoff, 1960). Em 1962, Widrow e seus colaboradores desenvolveram a estrutura de múltiplos elementos adaptativos chamado *Madaline* (Widrow, 1962).

Estes modelos obtiveram expressivo sucesso até a publicação do trabalho de Minsky e Papert (Minsky e Papert, 1969), onde foram provados matematicamente as limitações do perceptron linear, mostrando que este modelo de neurônio artificial não era capaz de solucionar problemas que não fossem separáveis, estabelecendo assim o fim de um período de grande desenvolvimento e investigação no campo de redes neurais.

A falta de um método de aprendizagem para aplicação de redes mais complexas e de maior potencial computacional desestimulou os pesquisadores, fazendo com que as redes neurais permanecessem “esquecidas” durante os anos 70. Entretanto, devem ser citados

alguns pesquisadores que realizaram importantes trabalhos em redes neurais durante esta década:

1. (Grossberg, 1967) e (Grossberg, 1968) no final dos anos 60 propôs um modelo aditivo de um neurônio, envolvendo equações a diferenças e diferenciais, e exploraram o uso do modelo como uma base para memória a curto-prazo;
2. (Werbos, 1974) descreveu pela primeira vez o algoritmo de retro-propagação do erro o qual foi aplicado para modelos gerais de redes, mas não foi aplicado a redes neurais, que representaria um caso particular;
3. (Little e Shaw, 1975) descreveram um modelo probabilístico de um neurônio, e usaram este modelo para desenvolver uma teoria de memória a curto-prazo ;
4. (Willshaw e von der Malsburg, 1976) publicaram o primeiro artigo na formação de mapas auto-organizáveis;
5. (Anderson et al., 1977) publicaram o conceito de *Brain-State-in-a-Box* (BSB), consistindo de uma rede associativa simples acoplada a um sistema dinâmico não-linear.

O interesse nas pesquisas de redes neurais ressurgiu principalmente após a publicação dos trabalhos do físico John Hopfield em 1982. Hopfield apresentou um modelo de rede neural recorrente composto de elementos interconectados, que buscam atingir um estado de energia mínima, garantindo a estabilidade da rede. O modelo representa a operação do neurônio a partir de um determinado nível de excitação, mostrando que a memória da rede neural está armazenada nas interconexões entre os neurônios (Hopfield, 1982).

Em 1986, a redescoberta do algoritmo de retro-propagação por (Rumelhart et al., 1986a) e a publicação do livro *Parallel Distributed Processing* (Rumelhart et al., 1986b) forneceram o grande impulso para o interesse na área de redes neurais. A partir de então, o algoritmo de retro-propagação passou a ser o método de aprendizagem mais popular para o treinamento de redes neurais perceptron multi-camadas, fazendo com que pesquisadores tivessem interesses em aplicações de redes neurais artificiais nas mais diferentes áreas, entre as quais Previsão de Séries Temporais (Weigend et al., 1991a), (Tang et al., 1991). Na próxima seção, é descrita a rede perceptron multi-camadas.

4.2 Rede Perceptron Multi-Camadas

A rede perceptron multi-camadas (*MultiLayer Perceptron* - MLP) é uma das mais conhecidas e aplicadas arquiteturas de redes neurais, a qual representa uma generalização do perceptron proposto por Rosemblatt.

A arquitetura (ou topologia) deste modelo consiste de uma camada de entrada, uma ou mais camadas intermediárias e uma camada de saída. A camada de entrada, cujos neurônios pertencentes a esta camada são denominados unidades de entrada, difunde as entradas para as camadas seguintes sem nenhuma modificação. As camadas intermediárias transmitem informações através das conexões entre as camadas de entrada e saída. Os neurônios pertencentes a esta camada são denominados unidades intermediárias. Finalmente, a camada de saída, cujos neurônios são chamados unidades de saída, transmite a resposta da rede neural à entrada aplicada na camada de entrada. A figura 4.3 mostra a arquitetura de uma rede neural com duas camadas intermediárias.

Os neurônios são conectados por arcos e a cada arco está associado um peso. A comunicação entre os neurônios é realizada através destas interconexões ponderadas. Entretanto, não há comunicação entre as unidades de uma mesma camada.

Cada neurônio da rede tem uma função não-decrescente e diferenciável, conhecida como função de ativação, a qual pode ser uma função linear, logística, ou tangente hiperbólica.

Assim, a arquitetura da rede neural MLP apresenta três características distintas:

1. Variações quanto ao número de camadas e neurônios intermediários da rede;
2. Tipo de conexão determinada pelas sinapses;
3. Função de ativação.

O modelo MLP apresenta um treinamento supervisionado, ou seja, é necessário um “professor” para indicar a resposta desejada para o padrão de entrada apresentado à rede durante a fase de aprendizagem. Um sinal de erro é definido como sendo a diferença entre a resposta desejada e a resposta observada. Os parâmetros da rede (pesos e limiares) são ajustados de acordo com este sinal. Assim, a rede se adapta através de um processo iterativo de ajustes sucessivos dos pesos, o qual pode ser expresso na forma de um método de aprendizagem.

Para o treinamento da rede MLP, o método de aprendizagem mais utilizado é o algoritmo de retro-propagação do erro. Este algoritmo consiste basicamente de duas fases. A primeira fase, conhecida como fase *forward*, as entradas são apresentadas e propagadas através da rede, camada por camada, calculando a saída de cada neurônio. Durante esta fase, os pesos são fixos e a saída calculada é comparada com a saída desejada, resultando em um erro para cada unidade. Na segunda fase, o erro calculado é propagado da camada de saída para a camada de entrada, fase *backward*, e os pesos são ajustados de acordo com a regra de correção do erro, originando o termo “retro-propagação do erro”.

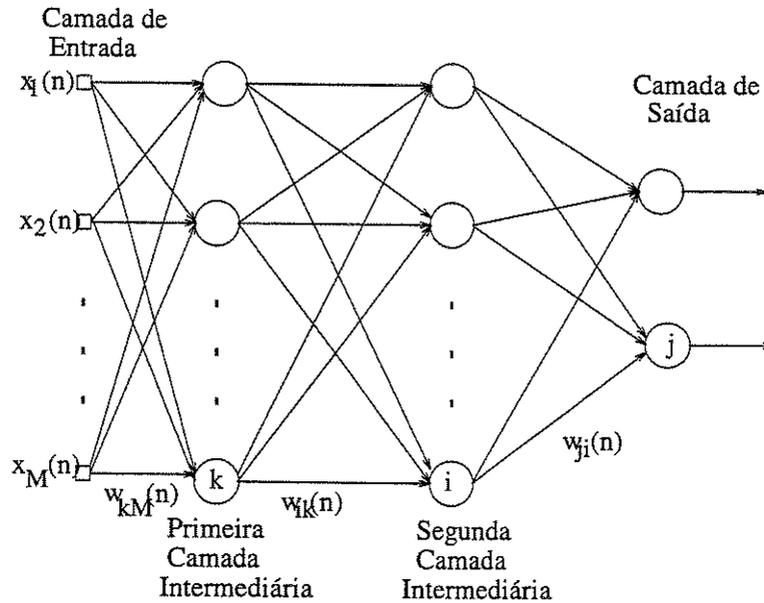


Figura 4.3: Rede Neural Perceptron Multi-Camadas.

4.3 Algoritmo de Retro-Propagação

O algoritmo de aprendizagem de retro-propagação (*back-propagation*) consiste em alterar os pesos da rede a fim de minimizar a diferença entre as saídas desejadas e as saídas fornecidas pelo modelo. Se o erro na camada de saída não é menor que uma dada tolerância, então o erro é retro-propagado através da rede, tendo como base para a atualização dos pesos a Regra Delta, a qual implementa o Método do Gradiente Descendente (Haykin, 1994).

Seja $y_{d_j}(n)$ a saída desejada do neurônio j e $y_j(n)$ a saída calculada pela rede para uma entrada n . O erro $e_j(n)$ na camada de saída é:

$$e_j(n) = y_{d_j}(n) - y_j(n) \quad (4.1)$$

O valor do erro quadrático para o neurônio j é definido como sendo $\frac{1}{2}(e_j(n))^2$. A soma dos erros quadráticos é obtida para todos os neurônios da camada de saída, isto é:

$$\mathcal{E}(n) = \frac{1}{2} \sum_j (e_j(n))^2 \quad (4.2)$$

Seja N o número total de pares entrada/saída-desejada contidos no conjunto de dados de treinamento. O erro quadrático médio é obtido pela soma do erro $\mathcal{E}(n)$ sobre todo n , normalizado com relação ao número de padrões N :

$$E = \frac{1}{N} \sum_{n=1}^N \mathcal{E}(n) \quad (4.3)$$

O objetivo do processo de treinamento é ajustar os parâmetros livres (pesos) da rede, minimizando uma função objetivo representada pelo erro quadrático médio, ou seja,

$$\min E = \min \frac{1}{N} \sum_{n=1}^N \mathcal{E}(n) \quad (4.4)$$

Para a derivação do algoritmo de retro-propagação, é considerado, inicialmente, um método de treinamento em que os pesos são ajustados entrada-a-entrada, ou seja, o ajuste dos pesos é realizado de acordo com o erro calculado para cada entrada apresentada à rede. A posteriori, será apresentada uma outra forma de ajustar os pesos conhecida como aprendizagem por ciclo ou em lote (*batch*).

O nível de ativação interno do neurônio j , $\nu_j(n)$, é uma função linear das saídas y_i dos neurônios que estão conectados ao neurônio j através dos pesos w_{ji} dado por:

$$\nu_j(n) = \sum_{i=0}^p w_{ji}(n)y_i(n) \quad (4.5)$$

onde, p é o número de neurônios da camada imediatamente anterior ao neurônio j , excluindo o limiar (figura 4.4). O peso sináptico w_{j0} é igual ao limiar θ_0 e corresponde à entrada $y_0 = -1$,

A saída $y_j(n)$ do neurônio j é um valor real, dada por uma função de ativação não-linear:

$$y_j(n) = f_j(\nu_j(n)) \quad (4.6)$$

Para minimizar (4.4) usando o método do gradiente descendente, é necessário calcular a derivada parcial de \mathcal{E} em relação a cada peso da rede. De acordo com a regra da cadeia, a derivada parcial pode ser escrita como:

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = \frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial \nu_j(n)} \frac{\partial \nu_j(n)}{\partial w_{ji}(n)} \quad (4.7)$$

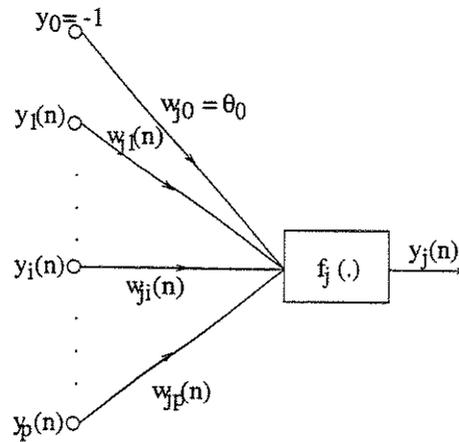


Figura 4.4: Saída do Neurônio.

Diferenciando as equações (4.1) em relação a $y_j(n)$, (4.2) em relação a $e_j(n)$, (4.5) em relação a $w_{ji}(n)$, (4.6) em relação a $\nu_j(n)$ e substituindo na equação (4.7), tem-se:

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = -e_j(n) f'_j(\nu_j(n)) y_i(n) \quad (4.8)$$

O algoritmo de retro-propagação aplica uma correção $\Delta w_{ji}(n)$ aos pesos sinápticos w_{ji} proporcional à direção oposta ao gradiente $\partial \mathcal{E}(n) / \partial w_{ji}(n)$, sendo definida pela Regra Delta como segue:

$$\Delta w_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} \quad (4.9)$$

onde, η é uma constante que determina a taxa de aprendizagem ou velocidade de aprendizagem, sendo chamada de parâmetro taxa de aprendizagem. Substituindo a equação (4.8) na equação (4.9), a atualização dos pesos pode ser reescrita como:

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \quad (4.10)$$

onde, $\delta_j(n)$ é o gradiente local definido por:

$$\delta_j(n) = - \frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial \nu_j(n)} = e_j(n) f'_j(\nu_j(n)) \quad (4.11)$$

Das equações (4.10) e (4.11), pode-se observar que a atualização dos pesos depende do sinal de erro $e_j(n)$ para o cálculo do gradiente local. Neste caso, deve ser considerado onde está localizado o neurônio j , ou seja, se o neurônio j está na camada de saída ou se está na camada intermediária.

Considerando, primeiramente, o caso em que o neurônio j está na camada de saída e conhecendo a saída desejada, pode-se calcular o erro (equação (4.1)). Assim, o gradiente local, $\delta_j(n)$, é igual ao produto da derivada da função de ativação $f'_j(\nu_j(n))$ e o erro $e_j(n)$, ou seja,

$$\delta_j(n) = e_j(n) f'_j(\nu_j(n)) \quad (4.12)$$

Caso o neurônio j esteja localizado na camada intermediária, não existe nenhuma saída desejada específica atribuída a este neurônio. Então, o sinal de erro $e_j(n)$ é calculado em termos dos erros de todos os neurônios aos quais o neurônio j está conectado.

Redefinindo o gradiente local $\delta_j(n)$ para o neurônio j , a equação (4.11) pode ser rescrita como:

$$\delta_j(n) = - \frac{\partial \mathcal{E}(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial \nu_j(n)} = - \frac{\partial \mathcal{E}(n)}{\partial y_j(n)} f'_j(\nu_j(n)) \quad (4.13)$$

Para calcular a derivada parcial $\partial \mathcal{E}(n) / \partial y_j(n)$ procede-se da seguinte forma: da figura 4.5 o erro quadrático (equação (4.2)) do neurônio k é dado por:

$$\mathcal{E}(n) = \frac{1}{2} \sum_k (e_k(n))^2 \quad (4.14)$$

Diferenciando a equação (4.14) em relação a $y_j(n)$

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n)}{\partial y_j(n)} \quad (4.15)$$

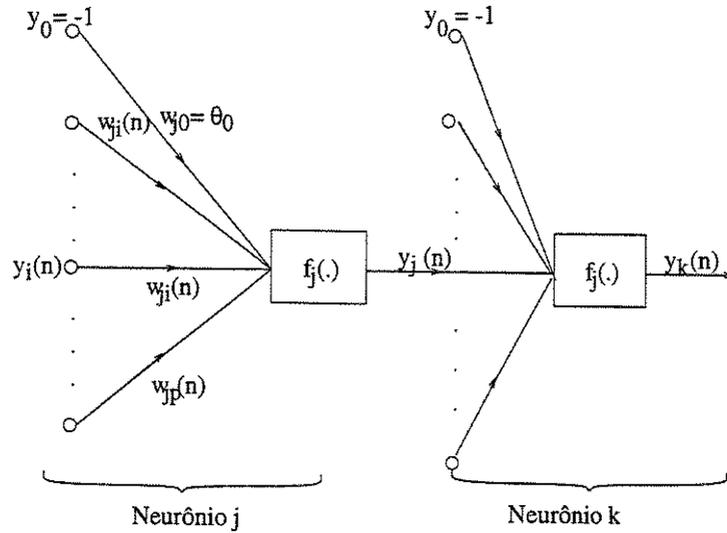


Figura 4.5: Saída do Neurônio Intermediário.

e usando a regra da cadeia para o cálculo da derivada parcial $\partial e_k(n)/\partial y_j(n)$, a equação (4.15) pode ser rescrita como:

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n)}{\partial \nu_k(n)} \frac{\partial \nu_k(n)}{\partial y_j(n)} \quad (4.16)$$

Da figura 4.5, tem-se que o sinal de erro é dado por:

$$e_k(n) = y_{d_k}(n) - y_k(n) = y_{d_k}(n) - f_k(\nu_k(n)) \quad (4.17)$$

e o nível de ativação do neurônio k é

$$\nu_k(n) = \sum_{j=0}^q w_{kj}(n) y_j(n) \quad (4.18)$$

onde q é o número de entradas aplicadas ao neurônio k .

Diferenciando as equações (4.17) em relação a $\nu_k(n)$ e (4.18) em relação a $y_j(n)$ e substituindo na equação (4.15), a derivada parcial pode ser escrita como:

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = - \sum_k e_k(n) f'_k(\nu_k(n)) w_{kj}(n) = - \sum_k \delta_k(n) w_{kj}(n) \quad (4.19)$$

onde $\delta_k(n)$ é o gradiente local do neurônio k .

Finalmente, substituindo a equação (4.19) na equação (4.13), o gradiente local $\delta_j(n)$ para o neurônio intermediário j é calculado da seguinte forma:

$$\delta_j(n) = f'_j(\nu_j(n)) \sum_k \delta_k(n) w_{kj}(n) \quad (4.20)$$

Assim, se o neurônio j é uma unidade intermediária, o gradiente local $\delta_j(n)$ é igual ao produto da derivada da função de ativação, associada ao neurônio j , e a soma ponderada dos gradientes locais calculados para os neurônios na próxima camada ou camada de saída que estão conectados ao neurônio j .

Portanto, na aplicação do algoritmo de retro-propagação, na primeira fase (fase *forward*) os pesos sinápticos permanecem inalterados através da rede e a saída é dada pelas equações (4.5) e (4.6). Os erros são calculados para cada unidade de saída (equações (4.1) e (4.2)). Assim, a fase *forward* inicia-se na primeira camada com a apresentação das entradas, e termina na camada de saída pelo cálculo do erro para cada neurônio nesta camada.

A fase *backward* inicia-se na camada de saída para a camada de entrada, através da propagação dos gradientes locais para cada neurônio da rede. Para os neurônios localizados na camada de saída, o gradiente é dado pela equação (4.12) e os pesos são atualizados usando a equação (4.9). A seguir, são calculados os gradientes de todos os neurônios da próxima camada (equação (4.19)) e são realizadas as mudanças nos pesos de todas as conexões desta camada (equação (4.9)). O cálculo recursivo é realizado, camada por camada, propagando as mudanças de todos os pesos sinápticos até a camada de entrada.

4.3.1 Função de Ativação

Para o cálculo dos gradientes locais de cada neurônio da rede MLP é preciso conhecer a derivada da função de ativação. A diferenciabilidade da função de ativação é, portanto, uma condição necessária. Uma das funções de ativação mais utilizadas em aplicações é a função logística, definida como segue:

$$y_j(n) = f_j(\nu_j(n)) = \frac{1}{1 + \exp^{-\nu_j(n)}} \quad (4.21)$$

onde $\nu_j(n)$ é o nível de ativação do neurônio j no instante n . Esta função é limitada, assumindo valores no intervalo $(0, 1)$, como mostra a figura 4.6(a).

Diferenciando a equação (4.21) em relação a $\nu_j(n)$, tem-se:

$$f'_j(\nu_j(n)) = \frac{\exp(-\nu_j(n))}{[1 + \exp(-\nu_j(n))]^2} = y_j(n)[1 - y_j(n)] \quad (4.22)$$

permitindo concluir que o cálculo da derivada requer poucos recursos computacionais.

A figura 4.6(b) mostra a derivada da função. Outras funções de ativação com as correspondentes derivadas podem ser vistas em (Silva, 1998).

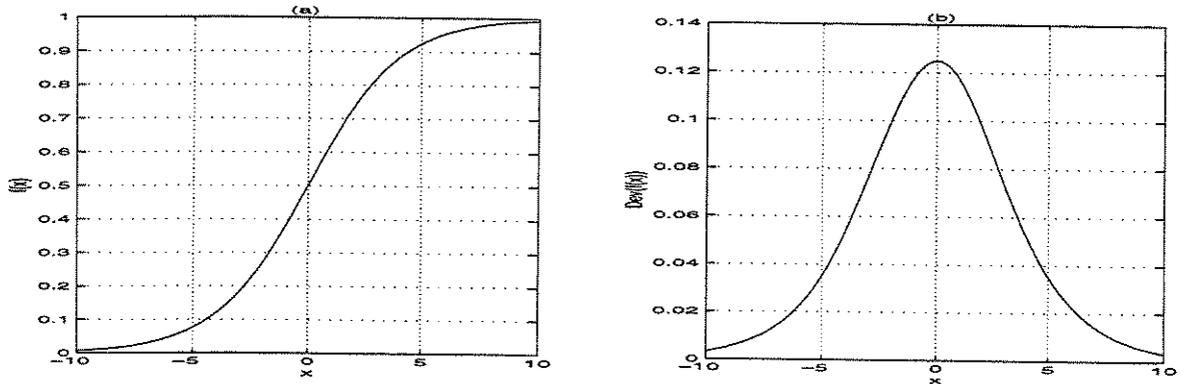


Figura 4.6: Função Logística e Derivada.

4.3.2 Taxa de Aprendizagem

A taxa de aprendizagem η é uma constante de proporcionalidade, que assume valores no intervalo $(0, 1)$, e que determina a natureza de adaptação dos pesos.

Um alto valor de η produzirá uma rápida aprendizagem, resultando em grandes alterações nos pesos. Entretanto, aumentará o risco da rede tornar-se instável, podendo oscilar ao redor da solução e provocar a saturação dos neurônios.

Por outro lado, se a taxa de aprendizagem assumir valores pequenos, as mudanças nos pesos sinápticos serão menores de uma iteração para outra e as trajetórias no espaço definido pelos pesos serão mais suaves. Este benefício é alcançado com um alto custo

no processo de treinamento, pois este torna-se mais lento por requerer muito mais iterações até a convergência (Ng, 1997).

Assim, um método simples para aumentar a velocidade de aprendizagem é modificar a Regra Delta, dada pela equação (4.10), adicionando um termo *momentum*, como proposto por (Rumelhart et al., 1986a):

$$\Delta w_{ji}(n) = \alpha \Delta w_{ji}(n-1) + \eta \delta_j(n) y_i(n) \quad (4.23)$$

onde α é um número positivo, restrito ao intervalo $[0,1)$, chamado constante *momentum*.

A equação (4.23) é conhecida como Regra Delta Generalizada, pois inclui a Regra Delta (equação (4.10)) como sendo um caso especial quando $\alpha = 0$.

O termo *momentum* relaciona as alterações dos pesos sinápticos na iteração n com as alterações realizadas na iteração $(n-1)$. A utilização desta constante dificulta a mudança de tendência na atualização dos pesos, podendo ainda ser usada para reduzir a possibilidade de ocorrer mínimos locais.

4.3.3 Inicialização dos Pesos

O primeiro passo no algoritmo de retro-propagação é a inicialização dos pesos da rede. Uma boa escolha dos parâmetros iniciais é fundamental para um bom desempenho do algoritmo. Caso a inicialização dos pesos seja feita inadequadamente, pode ocorrer um fenômeno chamado “saturação prematura”. Este fenômeno corresponde a um ponto de sela na superfície de erro pois a soma dos erros quadráticos permanece constante por algum período, continuando a decrescer após este período durante o processo de aprendizagem (Lee et al., 1991).

Quando existe alguma informação a priori sobre os dados, esta informação pode ser utilizada para inicializar os pesos. Quando esta informação não é utilizada, recomenda-se que a inicialização dos parâmetros seja feita aleatoriamente, com distribuição uniforme e magnitude pequena (Haykin, 1994).

4.3.4 Aprendizagem Entrada-a-Entrada e Aprendizagem por Ciclo

As aplicações práticas do algoritmo de retro-propagação consistem na apresentação dos padrões (pares entrada/saída) pertencentes ao conjunto de treinamento repetidas

vezes. Denomina-se de época a apresentação de todos os N pares do conjunto de treinamento durante o processo de aprendizagem. Este processo é repetido até que os pesos e limiares estejam estáveis e o erro quadrático médio convirja para um valor mínimo aceitável.

Segundo (Ridmiller, 1994), existem basicamente dois possíveis métodos para se calcular e executar a atualização dos pesos durante o processo de aprendizagem:

1. **Aprendizagem Entrada-a-Entrada:** este método atualiza os pesos da rede depois de cada apresentação de um par entrada/saída do conjunto de treinamento e do cálculo do respectivo gradiente, isto é, a atualização é feita imediatamente após a obtenção do gradiente referente a um único par apresentado à rede. Para um conjunto de treinamento contendo N pares entrada-saída são realizadas N correções de pesos em uma época, sendo que cada correção é baseada somente no erro do par apresentado naquela iteração. Este tipo de método, também conhecido como aprendizagem *on-line* ou aprendizagem local, produz melhores resultados para os conjuntos de amostras apresentados aleatoriamente a cada época, pois a ordem de apresentação das amostras não deve representar nenhuma informação significativa para condicionar o ajuste dos pesos.
2. **Aprendizagem por Ciclo:** neste método de aprendizagem, apenas uma correção dos pesos é feita por época. Primeiramente, todas as amostras do conjunto de treinamento são propagadas pela rede e então o erro é calculado. Em seguida, calcula-se o erro quadrático médio, definido na equação (4.3), para os N pares como segue:

$$E = \frac{1}{2N} \sum_{n=1}^N \sum_j (e_j(n))^2 \quad (4.24)$$

Para uma taxa de aprendizagem η , o ajuste do peso sináptico w_{ji} é definido pela regra delta da seguinte forma:

$$\Delta w_{ji}(n) = -\eta \frac{\partial E}{\partial w_{ji}} = -\frac{\eta}{N} \sum_{n=1}^N \frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} \quad (4.25)$$

A derivada parcial $\partial \mathcal{E}(n)/\partial w_{ji}$ é calculada da mesma forma como descrito na derivação do algoritmo retro-propagação. Este método também é conhecido como aprendizagem *off-line* ou aprendizagem em lote (*batch*).

O modo de aprendizagem entrada-a-entrada apresenta melhores resultados para grandes conjuntos de treinamento que contenham quantias significativas de informações

redundantes. Como as amostras são apresentadas aleatoriamente, o uso deste método torna a busca no espaço de pesos estocástica, reduzindo a possibilidade do algoritmo de retro-propagação ficar preso em um mínimo local. Já a utilização do método de aprendizagem por ciclo fornece informações mais seguras com relação ao vetor gradiente (Møller, 1993). A eficiência dos dois métodos depende do problema a ser tratado (Hertz et al., 1991). Neste trabalho, foi utilizado o modo de aprendizagem entrada-a-entrada.

4.3.5 Critério de Parada

O critério de parada do algoritmo de retro-propagação não é bem definido e a convergência não é garantida. Um critério de parada é interromper o treinamento após um número fixo de iterações, mas esta forma não é recomendável por não levar em conta as informações sobre o estado do processo de aprendizagem. Então, deve-se formular um critério de convergência que considere a existência de mínimos locais (Kramer e Sangiovanni-Vincentelli, 1989).

Um dos critérios discutidos em (Haykin, 1994) e que considera tal propriedade é o fato de que a função objetivo (erro quadrático médio) é estacionária no ponto $w = w^*$, onde w^* é um vetor peso que denota o mínimo, podendo ser local ou global. Este critério de parada pode ser definido da seguinte forma: O algoritmo de retro-propagação converge se a variação do erro quadrático médio de uma época para outra assumir um valor suficientemente pequeno.

Um outro critério de parada é testar, após cada iteração, o desempenho de generalização do algoritmo. O treinamento é encerrado quando o desempenho de generalização apresentar tendência de piora. Este critério será discutido na subseção 4.3.7 quando será apresentado a técnica estatística conhecida como validação cruzada.

4.3.6 Generalização

Generalização é uma das mais importantes características de redes neurais, podendo ser influenciada por três fatores:

1. O tamanho e a eficiência do conjunto de treinamento, que deve ser muito bem escolhido, pois deve expressar todas as características relevantes acerca do problema no universo considerado;
2. A arquitetura da rede neural: números de camadas e neurônios intermediários;
3. A complexidade do problema a ser tratado.

A rede neural é dita ter uma boa generalização se, fornecidos os conjuntos de dados de treinamento e teste, a relação entre os dados de entrada/saída calculada pela rede já treinada está correta ou tão próxima quanto desejável para os padrões pertencentes ao conjunto de teste, podendo ser obtida através da validação cruzada.

4.3.7 Validação Cruzada

O processo de aprendizagem consiste de uma seleção adequada dos parâmetros livres da rede para um determinado conjunto de dados de treinamento. A determinação desses parâmetros pode ser baseada em um conjunto de arquiteturas de redes neurais candidatas a serem os “melhores” modelos para realizar o ajuste dos parâmetros.

Uma técnica estatística utilizada para determinar os parâmetros da rede é a chamada validação cruzada (Stone, 1974), descrita como segue.

Primeiramente, um conjunto de dados é aleatoriamente particionado em um conjunto de treinamento e um conjunto de teste. O conjunto de treinamento é então particionado em dois subconjuntos distintos:

1. Um subconjunto usado durante a aprendizagem da rede para a estimação do modelo, denominado de subconjunto de treinamento;
2. Um subconjunto usado para avaliar a evolução do desempenho do modelo, ou seja para a validação da rede, denominado de subconjunto de validação. Este subconjunto corresponde a uma parte pequena do conjunto de treinamento, variando entre 10% a 20% de seu tamanho.

Esta técnica consiste em validar o modelo em um conjunto diferente do usado para estimar os parâmetros da rede, podendo, assim, verificar o desempenho de vários modelos e escolher o que apresentar melhor desempenho. Após encontrada a melhor arquitetura, a rede é novamente treinada com todo o conjunto de treinamento. Só, então, o conjunto de teste é utilizado para verificar a capacidade de generalização da rede. Estes resultados auxiliam na escolha da melhor topologia, em termos de generalização.

Uma outra forma na qual a validação cruzada pode ser usada é para decidir quando o processo de aprendizagem pode ser encerrado. Esta decisão é tomada com base no desempenho do erro quadrático médio correspondente ao ponto de erro mínimo em relação ao número de épocas. A validação cruzada foi utilizada para a escolha da arquitetura mais adequada para cada série de vazão e para determinar o fim do processo de treinamento.

4.3.8 Aproximação de Funções

Uma rede MLP com algoritmo retro-propagação realiza um mapeamento de entrada/saída não-linear o qual é continuamente diferenciável. Para este mapeamento, é considerada a seguinte questão:

Qual o número mínimo de camadas intermediárias em um perceptron multi-camadas com um mapeamento entrada/saída que fornece uma realização aproximada de algum mapeamento contínuo?

(Cybenko, 1989) demonstrou pela primeira vez e de forma rigorosa que uma única camada intermediária é suficiente para aproximar uniformemente qualquer função contínua em um hipercubo unitário. Em 1989, (Funahashi, 1989) publicou que os modelos perceptron multi-camadas são aproximadores universais. O *teorema da aproximação universal* é descrito a seguir:

TEOREMA 4.1 Seja $f(\cdot)$ uma função contínua, monotônica-crescente, limitada e não-constante. Seja I_M um hipercubo unitário de dimensão M e seja $C(I_M)$ o espaço de funções contínuas em I_M . Então, dada alguma função $f \in C(I_M)$ e $\epsilon > 0$, existe um inteiro p e conjuntos de constantes reais α_j , θ_j , e w_{ji} , onde $i = 0, \dots, M$ e $j = 0, \dots, p$ tal que

$$F(x_1, \dots, x_M) = \sum_{j=0}^p \alpha_j f \left(\sum_{i=0}^M w_{ji} x_i - \theta_j \right) \quad (4.26)$$

é uma aproximação da função $f(\cdot)$, isto é,

$$|F(x_1, \dots, x_M) - f(x_1, \dots, x_M)| < \epsilon$$

para todo $\{x_1, \dots, x_M\} \in I_M$.

Demonstração: (Cybenko, 1989).

Este teorema é diretamente aplicável a redes MLP. A função logística é uma função não-linear, não-constante, limitada e monotônica-crescente satisfazendo as condições impostas à função $f(\cdot)$. Além disso, a equação (4.26) representa a saída da rede como segue:

1. A rede possui M entradas e uma única camada intermediária consistindo de p neurônios;

2. O neurônio intermediário j tem pesos sinápticos w_{j1}, \dots, w_{jM} e limiares θ_j ;
3. A saída da rede é uma combinação linear das saídas dos neurônios intermediários, com α_j definindo os coeficientes desta combinação.

O *teorema da aproximação universal* é um teorema de existência, pois é uma justificativa matemática para a aproximação de uma função contínua arbitrária. O resultado acima afirma que uma única camada intermediária é suficiente para um perceptron multi-camadas realizar uma aproximação uniforme ϵ , para um dado conjunto de treinamento, representado por um conjunto de entradas x_1, x_2, \dots, x_M e uma saída desejada $f(x_1, x_2, \dots, x_M)$. Entretanto, o teorema não afirma que uma única camada intermediária é ótima no sentido de tempo de aprendizagem e custo computacional, além de não indicar um valor adequado para p , que representa o número de neurônios na camada intermediária (Haykin, 1994).

4.4 Algoritmo de Retro-Propagação Modificado

Na seção 4.3.2, foi apresentada uma breve discussão sobre a velocidade de convergência do algoritmo retro-propagação. (Jacobs, 1988) sugere quatro heurísticas para acelerar a convergência do algoritmo através da taxa de aprendizagem, descritas como segue:

1. Todos os pesos da rede devem ter sua própria taxa de aprendizagem, já que uma boa taxa de aprendizagem para ajustar um determinado peso não é necessariamente adequada para o ajuste de outro peso da rede;
2. A taxa de aprendizagem deve poder variar de uma iteração para outra;
3. Quando a derivada da função custo com relação a um determinado peso sináptico tem o mesmo sinal algébrico para várias iterações consecutivas do algoritmo, o parâmetro de taxa de aprendizagem para este peso deve aumentar;
4. Quando o sinal algébrico da derivada da função objetivo com relação a um peso particular alterna em várias iterações consecutivas, o parâmetro de taxa de aprendizagem para este determinado peso deve diminuir.

O uso de um parâmetro diferente e variável no tempo para cada peso sináptico, de acordo com as heurísticas descritas acima, modifica o algoritmo de retro-propagação de uma maneira fundamental, já que os ajustes aplicados aos pesos são baseados nas derivadas parciais do erro com relação aos pesos e nas estimativas das curvaturas da superfície de erro em relação aos pesos em cada iteração (Haykin, 1994).

4.4.1 Regra de Aprendizagem Delta-Bar-Delta

Esta técnica usa o gradiente descendente para a pesquisa da direção para atualizar os pesos e, então, para aplicar o tamanho dos passos individuais para cada peso da rede, aumentando a velocidade de convergência do algoritmo.

Para derivar a modificação para o algoritmo de retro-propagação, um procedimento similar àquele descrito na seção 4.3 é realizado. Inicialmente, uma função objetivo é definida como sendo a soma dos erros quadráticos da seguinte forma:

$$\mathcal{E}(n) = \frac{1}{2} \sum_j (e_j(n))^2 = \frac{1}{2} \sum_j [y_{d_j}(n) - y_j(n)]^2 \quad (4.27)$$

Embora a equação (4.27) seja matematicamente similar à equação (4.2), deve-se considerar que a nova função objetivo consiste de diferentes taxas de aprendizagem. Seja $\eta_{ji}(n)$ o parâmetro de taxa de aprendizagem relacionado ao peso $w_{ji}(n)$ na iteração n . Aplicando a regra da cadeia para $\mathcal{E}(n)$, tem-se:

$$\frac{\partial \mathcal{E}(n)}{\partial \eta_{ji}(n)} = \frac{\partial \mathcal{E}(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial \nu_j(n)} \frac{\partial \nu_j(n)}{\partial \eta_{ji}(n)} \quad (4.28)$$

onde, o nível interno de ativação do neurônio j é dado por:

$$\nu_j(n) = \sum_i w_{ji}(n) y_i(n) \quad (4.29)$$

sendo a saída do neurônio dada por:

$$y_j(n) = f(\nu_j(n)) \quad (4.30)$$

onde f é uma função de ativação utilizada pelo neurônio j .

A atualização dos pesos é dada pela regra delta:

$$w_{ji}(n) = w_{ji}(n-1) - \eta_{ji}(n) \frac{\partial \mathcal{E}(n-1)}{\partial w_{ji}(n-1)} \quad (4.31)$$

Substituindo a equação (4.31) na equação (4.29) e diferenciando em relação a $\eta_{ji}(n)$ tem-se:

$$\frac{\partial \nu_j(n)}{\partial \eta_{ji}(n)} = -y_i(n) \frac{\partial \mathcal{E}(n-1)}{\partial w_{ji}(n-1)} \quad (4.32)$$

Diferenciando as equações (4.27) com relação a $y_j(n)$, e (4.30) com relação a $\nu_j(n)$, substituindo e rearranjando os termos na equação (4.28), tem-se:

$$\frac{\partial \mathcal{E}(n)}{\partial \eta_{ji}(n)} = f'_j(\nu_j(n)) e_j(n) y_i(n) \frac{\partial \mathcal{E}(n-1)}{\partial w_{ji}(n-1)} \quad (4.33)$$

Usando a equação (4.8) pode-se rescrever a equação (4.33) com:

$$\frac{\partial E(n)}{\partial \eta_{ji}(n)} = -\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} \frac{\partial \mathcal{E}(n-1)}{\partial w_{ji}(n-1)} \quad (4.34)$$

A equação (4.34) define a derivada da superfície de erro em relação ao parâmetro de taxa de aprendizagem $\eta_{ji}(n)$ aplicado a cada neurônio da rede. Assim, o ajuste da taxa de aprendizagem aplicado a $\eta_{ji}(n)$ pode ser definido como:

$$\Delta \eta_{ji}(n+1) = -\gamma \frac{\partial E(n)}{\partial \eta_{ji}(n)} = \gamma \frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} \frac{\partial \mathcal{E}(n-1)}{\partial w_{ji}(n-1)} \quad (4.35)$$

onde, γ é uma constante positiva. Duas observações podem ser consideradas em relação à equação (4.35):

1. Quando a derivada da superfície de erro com relação ao peso $w_{ji}(n)$ tem o mesmo sinal algébrico em duas iterações consecutivas, o ajuste de $\Delta \eta_{ji}(n)$ tem um valor positivo, aumentando o parâmetro de taxa de aprendizagem e, conseqüentemente, o algoritmo nesta direção é mais rápido.
2. Quando a derivada de superfície de erro com relação ao peso $w_{ji}(n)$ alterna de sinal em duas iterações consecutivas, o ajuste $\Delta \eta_{ji}(n)$ assume um valor negativo, decrescendo o parâmetro $\eta_{ji}(n)$ com relação ao peso $w_{ji}(n)$, tornando o algoritmo mais lento nesta direção.

Embora estas duas observações satisfaçam as heurísticas 3 e 4, este procedimento de ajuste do parâmetro $\eta_{ji}(n)$ apresenta alguns problemas, tais como: se a derivada do erro em relação a um determinado peso tem o mesmo sinal mas, magnitudes pequenas em duas iterações consecutivas, o ajuste aplicado ao parâmetro $\eta_{ji}(n)$ é muito pequeno para este peso; e, se a derivada apresenta sinal oposto com magnitudes grandes, o ajuste aplicado à taxa de aprendizagem pode ser muito grande (Haykin, 1994). Considerando estas duas circunstâncias, torna-se difícil escolher um valor para o parâmetro γ . Para suprir esta dificuldade, a seguinte modificação foi proposta por (Jacobs, 1988):

$$\Delta\eta_{ji}(n+1) = \begin{cases} \kappa, & \psi_{ji}(n-1)\phi_{ji}(n) > 0; \\ -\beta\eta_{ji}(n), & \psi_{ji}(n-1)\phi_{ji}(n) < 0; \\ 0, & \text{caso contrário.} \end{cases} \quad (4.36)$$

onde, $\psi_{ji}(n-1)$ e $\phi_{ji}(n)$ são definidos como:

$$\psi_{ji}(n) = \frac{\partial\mathcal{E}(n)}{\partial w_{ji}(n)} \quad (4.37)$$

$$\phi_{ji}(n) = (1 - \xi)\psi_{ji}(n-1) + \xi\phi_{ji}(n-1) \quad (4.38)$$

onde ξ é uma constante positiva e κ e β são parâmetros definidos previamente. O termo $\psi_{ji}(n-1)$ é o valor atual da derivada da superfície de erro com relação ao peso e o termo $\phi_{ji}(n)$ é o decaimento exponencial dos valores dos gradientes. Pode-se observar que, se um dos parâmetros $\psi_{ji}(n-1)$ ou $\phi_{ji}(n)$ for igual a zero a taxa de aprendizagem assume um valor constante, como no algoritmo de retro-propagação descrito na seção 4.3.

4.5 Aplicação da Rede MLP para Previsão de Vazões

Nesta seção, a rede MLP é ajustada para a série de vazões do posto de Furnas analisada no capítulo 2. São considerados para teste os seguintes anos: 1952 a 1956, 1972 a 1976 e 1981 a 1985, correspondendo a períodos com características distintas: seco, mediano e úmido, respectivamente.

Uma rede com uma camada intermediária será considerada devido ao número de padrões N e à precisão exigida ϵ , pois o número de pesos na estrutura da rede MLP

deve ser menor que ϵN (Haykin, 1994). Assim, os parâmetros que devem ser ajustados são: número de neurônios intermediários, taxa de aprendizado e termo *momentum*.

Para cada conjunto de teste foram realizados os ajustes dos parâmetros livres da rede. Por exemplo, o período de teste de 1952 a 1956 é retirado do conjunto de treinamento e os parâmetros são ajustados para o restante dos dados. Após selecionar a estrutura da rede, o conjunto de teste é utilizado para verificar o desempenho da rede ajustada. Este processo foi repetido para os outros dois períodos de teste considerado.

Em (Ballini et al., 1997a) foram utilizadas as informações obtidas pelas funções de autocorrelação e autocorrelação parcial para determinar o número de entradas na rede neural. Ou seja, o número de entradas foi determinado a partir dos dados padronizados como mostrado nos capítulos 2 e 3, sendo este número igual a dois. Observou-se que a rede neural apresentou o mesmo comportamento que o modelo auto-regressivo ajustado (AR(2)), com erros da mesma ordem de grandeza.

Em (Kadowaki et al., 1997), a seleção do número de entradas foi determinada a partir dos dados históricos não padronizados. O estudo das funções de autocorrelação e autocorrelação parcial indicaram que para o posto de Furnas deveriam ser escolhidas seis entradas. Este número de entrada forneceu, para a rede MLP, um desempenho satisfatório quando comparado com o desempenho dos modelos de Box & Jenkins.

Em (Ballini et al., 1997b) foram comparados o desempenho dos modelos de redes neurais com 2 e 6 entradas e verificou-se que o modelo proposto em (Kadowaki et al., 1997) apresentou um melhor desempenho.

Neste trabalho, os modelos de redes neurais são ajustados utilizando seis entradas, sendo estas dadas pelas vazões dos meses $t - 1$, $t - 2$, $t - 3$, $t - 11$, $t - 12$, $t - 13$. A seleção dessas entradas foi devido ao fato de fornecer ao modelo a tendência da série dada pelos três meses imediatamente anteriores ($t - 1$, $t - 2$, $t - 3$) e também informações sobre a sazonalidade representada pelos três meses $t - 11$, $t - 12$ e $t - 13$.

Para a seleção adequada dos parâmetros livres da rede para um determinado conjunto de dados, foi utilizado um conjunto de arquiteturas candidatas e através da técnica de validação cruzada os parâmetros foram estimados, verificando o desempenho dos modelos sob o conjunto de validação, sendo escolhido o modelo que apresentou melhor desempenho em relação ao erro quadrático médio (EQM).

As taxas de aprendizado foram ajustadas usando a regra delta-bar-delta (seção 4.4), fornecendo para cada peso da rede um tamanho de passo específico. Os parâmetros usados para ajuste da taxa de aprendizado, que devem ser definidos previamente, foram fixados em $\xi = 0.01$, $\kappa = 0.7$ e $\beta = 0.2$. Assim, para uma rede com uma única camada interme-

diária foram realizadas combinações do parâmetro termo *momentum* $\alpha = \{0.0, 0.25, 0.5, 0.9\}$ e número de neurônios intermediários $M = 2, 3, 4, 5, 6$.

O processo de aprendizagem foi encerrado avaliando o desempenho do EQM obtido do conjunto de treinamento, correspondente ao ponto de mínimo em relação ao número de épocas, ou, o EQM menor que um certo limiar ϵ . O número de épocas máximo usado foi 1000 e o valor do limiar $\epsilon < 0.001$. A tabela 4.1 mostra os valores dos EQM considerando o número de neurônios intermediários e o termo *momentum* sob o conjunto de validação. Estes resultados foram obtidos para o período de 1952 a 1956.

Tabela 4.1: Escolha da Topologia da Rede Neural Multi-Camadas.

| M | $\alpha = 0.0$ | $\alpha = 0.25$ | $\alpha = 0.5$ | $\alpha = 0.9$ |
|-----|----------------|-----------------|----------------|----------------|
| 2 | 0.00423 | 0.00373 | 0.00398 | 0.00339 |
| 3 | 0.00332 | 0.00287 | 0.00320 | 0.00319 |
| 4 | 0.00326 | 0.00309 | 0.00315 | 0.00224 |
| 5 | 0.00318 | 0.00306 | 0.00213 | 0.00239 |
| 6 | 0.00338 | 0.00316 | 0.00319 | 0.00320 |

Analisando os dados da tabela 4.1, pode-se notar que não ocorrem grandes variações nos erros. Também pode-se observar que para $\alpha = 0.9$ o EQM apresentou resultados satisfatórios porém o aumento na velocidade de convergência pode provocar instabilidade na rede. Assim, a topologia escolhida foi $M = 5$ e $\alpha = 0.5$, como mostra a tabela 4.2. Nesta mesma tabela é mostrada a topologia para os períodos de 1972 a 1976 e 1981 a 1985. Nota-se que, para cada conjunto de dados de treinamento, os parâmetros M e α são reajustados, sendo esta uma importante característica de redes neurais.

Tabela 4.2: Topologia Escolhida para os Períodos Considerados Para o Posto de Furnas.

| Período | 1952 a 1956 | 1972 a 1976 | 1981 a 1985 |
|----------|-------------|-------------|-------------|
| M | 5 | 3 | 3 |
| α | 0.5 | 0.25 | 0.5 |

A estrutura do modelo de rede neural usada neste trabalho é uma estrutura de rede não-recorrente, ou seja, é um modelo estático. Este tipo de estrutura não apresenta memória dinâmica e sua utilização para representar sistemas dinâmicos somente é possível

com a inclusão, no vetor de entradas, de valores previstos em instantes anteriores de tempo ($t - 1, t - 2, \dots$), conforme mostra a figura 4.7. Este vetor de entradas é conhecido como linha de atraso e transforma um problema de aprendizagem do comportamento dinâmico de um sistema no domínio do tempo em um mapeamento estático das entradas e saídas, juntamente com seus respectivos retardos no tempo, na próxima saída (Von Zuben, 1996).

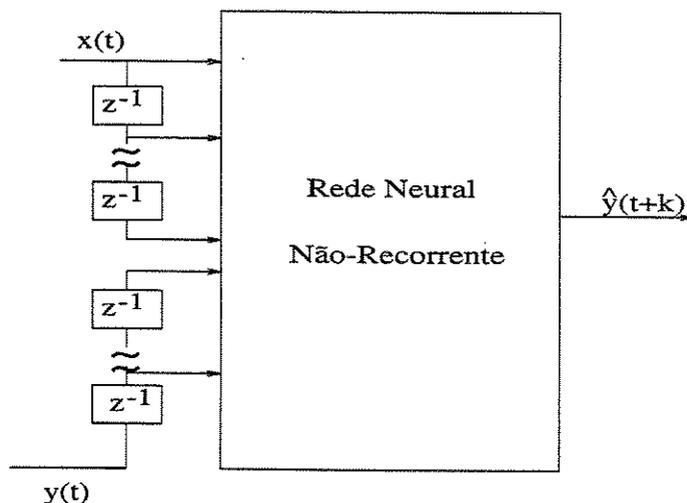


Figura 4.7: Representação do Modelo de Previsão k Passos à Frente.

As figuras 4.8, 4.9 e 4.10 mostram as previsões 1, 3, 6 e 12 passos à frente representadas em (a), (b), (c) e (d), obtidas para os períodos de teste de 1952 a 1956, 1972 a 1976 e 1981 a 1985, respectivamente. Para todos os casos o número de entradas na rede foi igual a 6 pois, desta forma, a dimensão do espaço de aproximação não é alterada.

Como pode ser observado pelas figuras 4.8, 4.9 e 4.10, a previsão 1 passo à frente apresentou um melhor comportamento que a previsão vários passos à frente. Uma das justificativas para este fato é que o vetor de entrada apresenta dados de atraso que foram previstos e assim ocorre uma degradação no desempenho da rede. Esta perda de desempenho é causada pelos ruídos decorrentes dos erros de previsão.

Uma tentativa de amenizar este comportamento foi construir um modelo de rede neural para cada mês e, assim, tornar o modelo mais específico para prever o dado de vazão de cada mês. Neste caso, são ajustados 12 modelos de redes neurais. A tabela 4.3 mostra o número de neurônios intermediários para cada mês para os respectivos períodos. Neste caso, os parâmetros taxa de aprendizado foram obtidos pela regra delta-bar-delta e o termo *momentum* foi fixado em $\alpha = 0.5$.

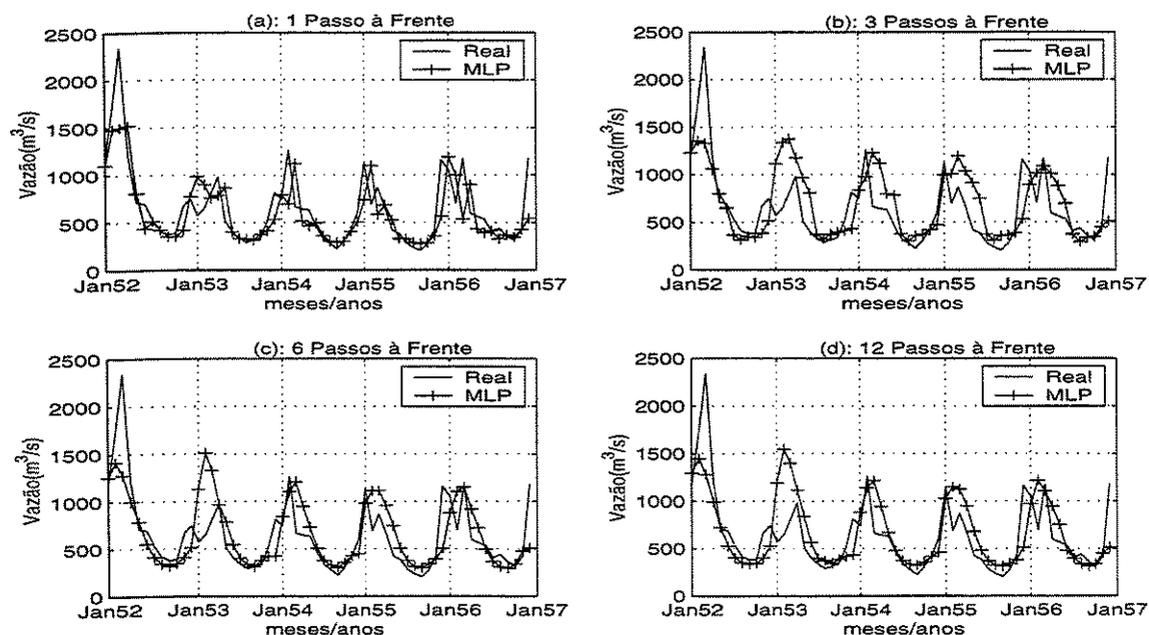


Figura 4.8: Previsão de Vazões Usando uma Rede Neural MLP para o Posto de Furnas - Período de 1952 a 1956.

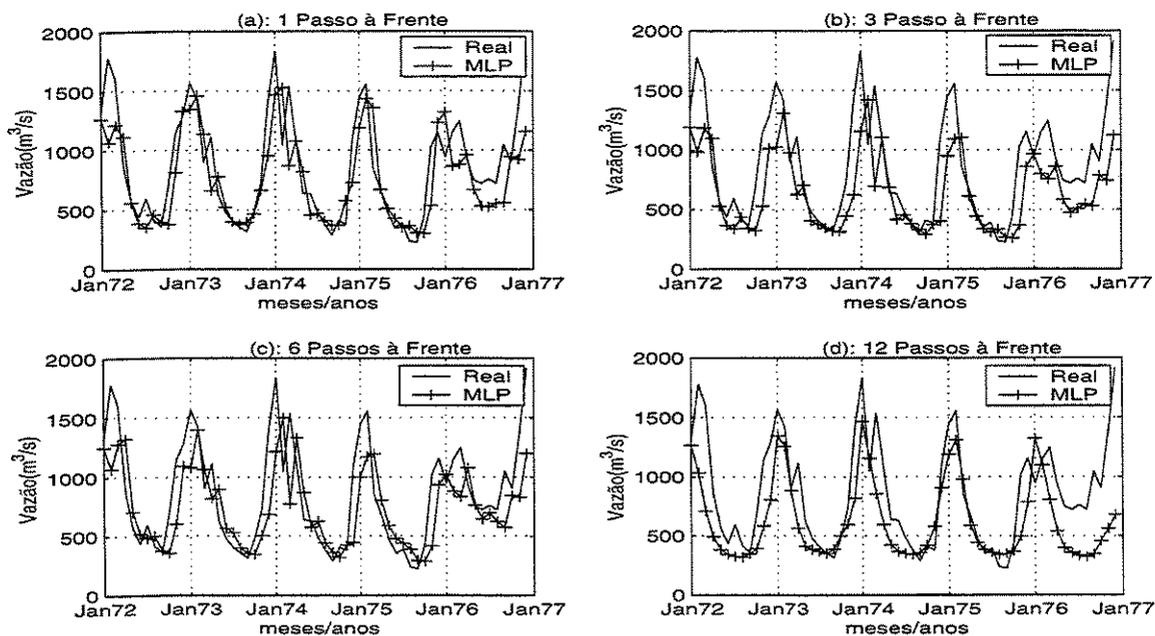


Figura 4.9: Previsão de Vazões Usando uma Rede Neural MLP para o Posto de Furnas - Período de 1972 a 1976.

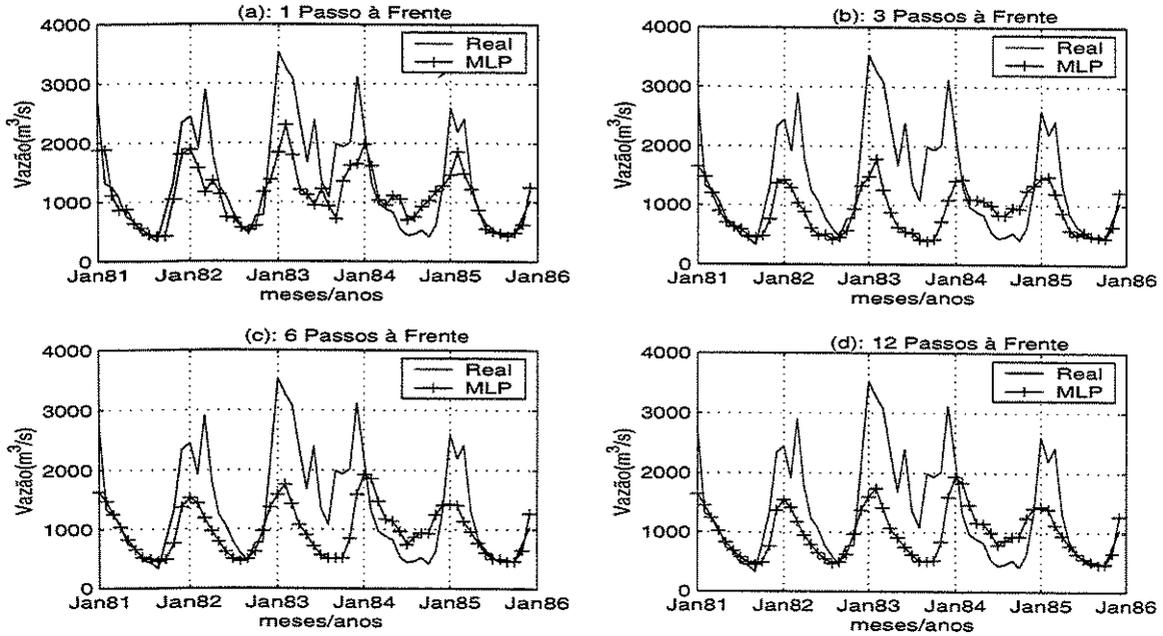


Figura 4.10: Previsão de Vazões Usando uma Rede Neural MLP para o Posto de Furnas - Período de 1981 a 1985.

As figuras 4.11, 4.12 e 4.13 mostram as previsões para 1, 3, 6 e 12 passos à frente através dos gráficos (a), (b), (c) e (d), respectivamente. Pode-se notar que o comportamento do modelo foi satisfatório, apresentando uma maior tendência em acompanhar o comportamento da série, com exceção para o período de 1972 a 1976, onde a previsão vários passos à frente apresentou um desempenho inferior, principalmente para a previsão no início do período de teste (janeiro/1972 a abril/1972). Isto ocorre porque os dados de vazões utilizados como entrada para prever a vazão de janeiro/1972 a abril/1972 são vazões baixas em relação ao mês que esta sendo previsto.

Para os períodos seco e úmido pode-se notar que as curvas de previsão apresentaram um desempenho satisfatório para os meses de cheia, verificando uma tendência em prever os picos das séries.

Comparando as figuras 4.8, 4.9 e 4.10, quando foi ajustado um único modelo, e as figuras 4.11, 4.12 e 4.13, ajuste de 12 modelos, pode-se observar que as previsões usando o ajuste de 12 modelos foram mais satisfatórias tanto para os meses de cheia como para os meses de seca para os três períodos considerados, com exceção para o período de janeiro/1972 a abril/1972.

Tabela 4.3: Configuração da Rede MLP para cada Mês.

| Meses | 1952-1956 | 1972-1976 | 1981-1985 |
|-----------|-----------|-----------|-----------|
| Janeiro | 4 | 5 | 6 |
| Fevereiro | 3 | 4 | 3 |
| Março | 4 | 4 | 3 |
| Abril | 3 | 3 | 3 |
| Maio | 4 | 4 | 4 |
| Junho | 3 | 3 | 5 |
| Julho | 4 | 3 | 4 |
| Agosto | 3 | 3 | 3 |
| Setembro | 4 | 4 | 4 |
| Outubro | 3 | 4 | 3 |
| Novembro | 5 | 3 | 4 |
| Dezembro | 3 | 3 | 4 |

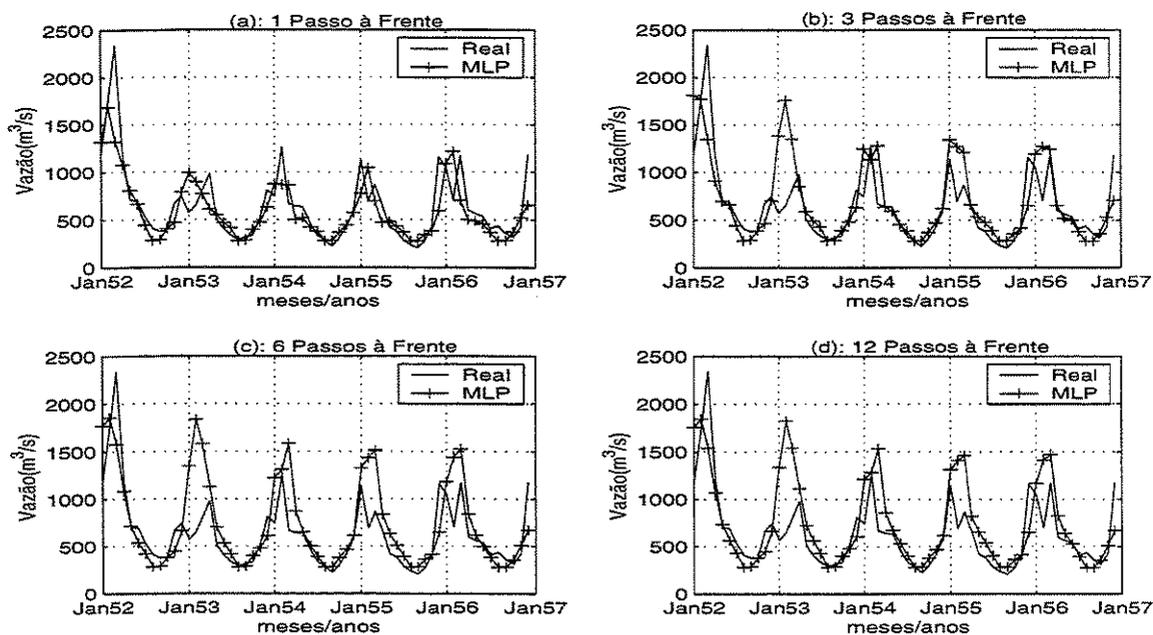


Figura 4.11: Previsão de Vazões Usando 12 Modelos de Redes Neurais MLP para o Posto de Furnas - Período de 1952 a 1956.

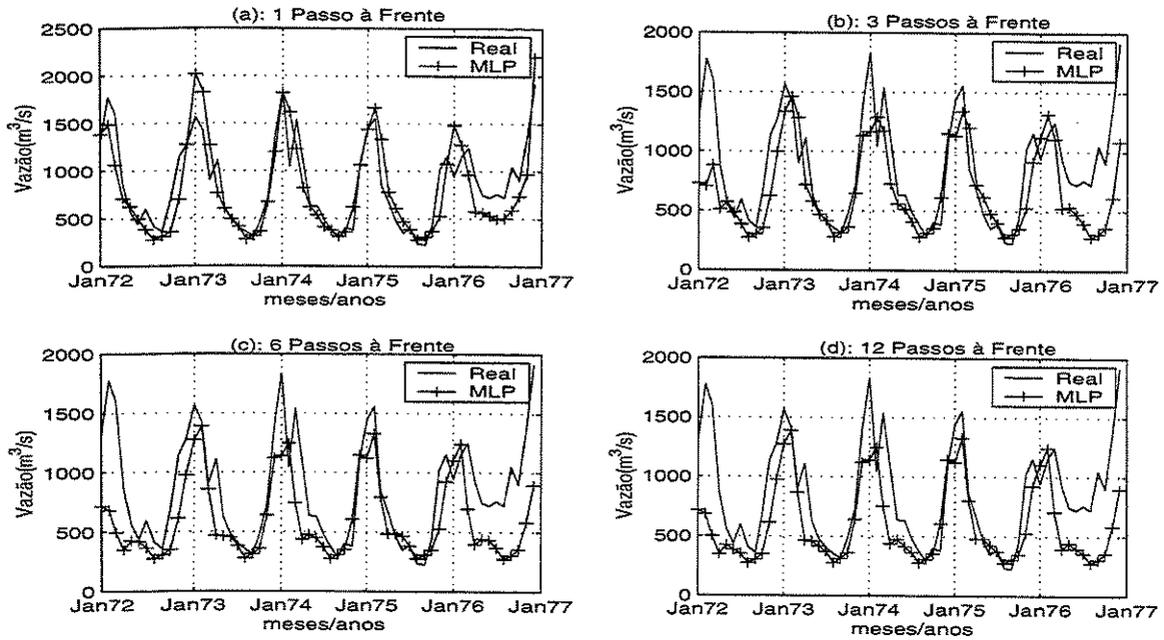


Figura 4.12: Previsão de Vazões Usando 12 Modelos de Redes Neurais MLP para o Posto de Furnas - Período de 1972 a 1976.

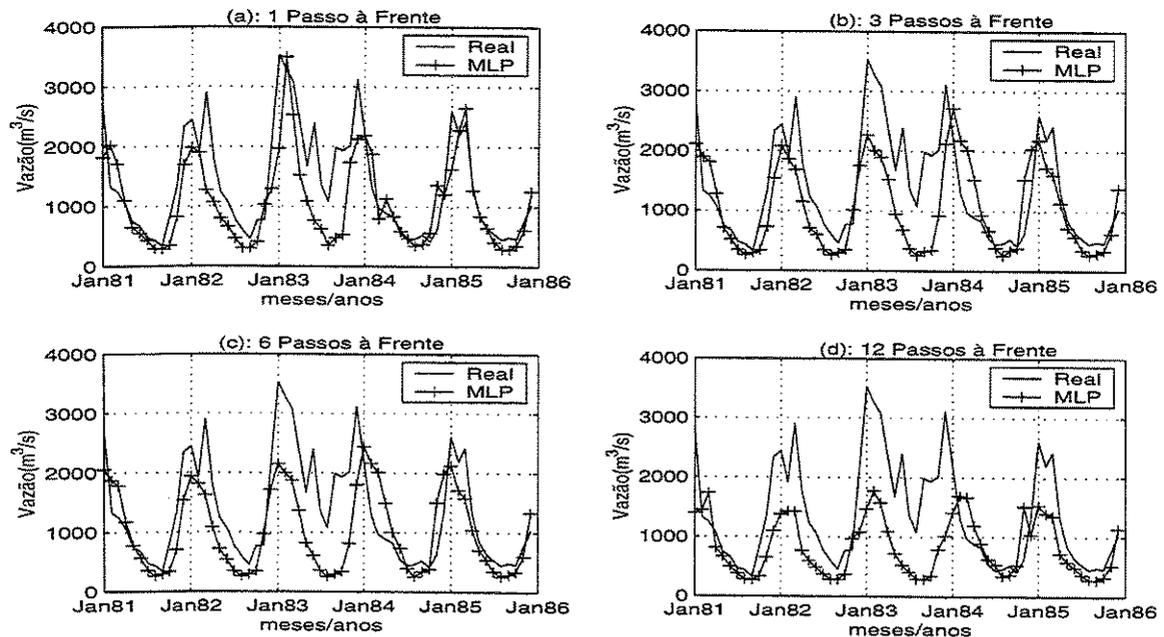


Figura 4.13: Previsão de Vazões Usando 12 Modelos de Redes Neurais MLP para o Posto de Furnas - Período de 1981 a 1985.

4.6 Resumo

Os modelos de redes neurais artificiais vêm sendo extensivamente aplicados aos problemas de previsão de séries temporais devido à sua capacidade de generalização e habilidade de aprendizado, tornando-se assim capazes de identificar e assimilar as características das séries.

Neste capítulo, apresentou-se o modelo de rede neural multi-camadas com algoritmo de retro-propagação. Este modelo é aplicado nas mais diferentes áreas incluindo previsão de séries temporais devido as propriedades citadas acima.

O principal objetivo deste capítulo foi a aplicação da rede multi-camadas ao problema de previsão de vazões afluentes médias mensais da série de dados da usina hidroelétrica de Furnas. Foram realizadas previsões 1, 3, 6 e 12 passos à frente para três períodos de teste com características distintas: seco, mediano e úmido.

Como a estrutura do modelo de rede neural usada neste trabalho é não-recorrente, ou seja, é um modelo estático, para realizar as previsões vários passos à frente, o vetor de entrada é composto por dados de atraso que foram previstos, provocando, assim, uma degradação no desempenho da rede devido à presença de ruídos decorrentes dos erros de previsão.

Uma tentativa de amenizar este problema foi ajustar um modelo de rede neural para cada mês. Porém observou-se que para os períodos medianos, ou seja, para os períodos onde a variação nos dados é menor, o ajuste de um único modelo forneceu um desempenho semelhante ao ajuste de 12 modelos. No capítulo 6, serão analisados os erros de previsão comparando com os erros obtidos com os modelos de séries temporais.

No próximo capítulo, será apresentado um modelo de rede neural nebulosa, o qual combina as características de redes neurais e a teoria de sistemas nebulosos em um sistema integrado. Este modelo será aplicado ao problema de previsão de vazões afluentes médias mensais.

Capítulo 5

Modelo de Rede Neural Nebulosa

Redes neurais nebulosas representam ser uma abordagem eficiente para projetar facilmente e automaticamente sistemas nebulosos. Estes modelos computacionais apresentam habilidade não somente de aprender os parâmetros que definem os sistemas nebulosos mas também de codificar regras em sua estrutura.

Recentemente, diversas estruturas e estratégias de aprendizagem vêm sendo propostas visando construir modelos de redes neurais nebulosas mais eficientes para projetar sistemas nebulosos. Neste capítulo, será apresentada uma estrutura de rede neural nebulosa, juntamente com dois algoritmos de aprendizagem, aplicada ao problema de previsão de vazões médias mensais.

5.1 Introdução

Os sistemas nebulosos são sistemas computacionais desenvolvidos a partir dos conjuntos nebulosos e da lógica nebulosa, visando aproveitar de maneira pragmática o potencial destas teorias. Estes sistemas processam conhecimento e inferem conclusões a partir dos padrões apresentados (Pedrycz et al., 1995).

Os sistemas nebulosos caracterizam-se pela sua simplicidade no que se refere à teoria que os fundamenta e pela forma direta e intuitiva com que são empregados, ou seja, são capazes de utilizar de maneira pragmática o conhecimento codificado em expressões lingüísticas, reproduzindo características inteligentes correspondentes à representação do conhecimento e ao raciocínio aproximado (Pedrycz, 1993). Além disso, algumas classes de sistemas nebulosos são aproximadores universais (Castro, 1995). Em essência, estes sistemas são capazes de processar o conhecimento de forma compreensível e de manusear incertezas

e imprecisões visando solucionar problemas reais complexos (Zimmermann, 1987).

À medida que as teorias relacionadas às redes neurais e aos sistemas nebulosos se desenvolviam, observou-se que estas abordagens poderiam ser unificadas, resultando em sistemas de características mais completas no que se refere às suas capacidades de processamento inteligente. Além destas abordagens exibirem características comuns, isto é, tratamento de incertezas e raciocínio aproximado, cada uma apresenta propriedades exclusivas que contribuem para compensar deficiências individuais da outra. De um lado, as redes neurais falham em representar o conhecimento de forma explícita, mas exibem sua inclinação para o aprendizado. Por outro lado, os sistemas nebulosos são fracos sob o ponto de vista da aquisição automática de conhecimento, porém valem-se amplamente de suas características com relação à manipulação de termos lingüísticos. O resultado desta simbiose se concretizou nas redes neurais nebulosas (Pedrycz e Gomide, 1998).

Estes modelos herdam as características das redes neurais, suas arquiteturas e seus métodos de aprendizagem e adquirem dos sistemas nebulosos a capacidade de representação e processamento compreensível do conhecimento, bem como os demais recursos oferecidos pelas teorias de conjuntos nebulosos e lógica nebulosa.

Esta abordagem vem sendo sucessivamente aplicada para modelagem de sistemas não-lineares e estimação de parâmetros (Figueiredo e Gomide, 1998), (Wang e Mendel, 1992), reconhecimento de padrões (Caminhas et al., 1996), controle de processo (Lee e Lin, 1991), previsão de séries econômicas (Kuo e Xue, 1999), previsão de carga (Bakirtzis et al., 1995), (Faria Neto et al., 1998) e previsão de séries de vazões (Ballini et al., 1999c) (Ballini et al., 2000).

Para diferentes propósitos e usos, muitas arquiteturas e algoritmos de aprendizagem de redes neurais nebulosas vêm sendo propostos na literatura (Buckley e Hayashi, 1994). (Lee e Lee, 1974) e (Lee e Lee, 1975) desenvolveram um dos primeiros trabalhos introduzindo redes neurais e conjuntos nebulosos. O modelo do neurônio nebuloso proposto foi uma generalização do modelo de McCulloch e Pitts (McCulloch e Pitts, 1943), sendo que o resultado da agregação das entradas no neurônio é mapeado no intervalo $[0, 1]$. (Keller e Hunt, 1985) propuseram adicionar funções de pertinências nebulosas ao perceptron e (Yamakawa e Tomoda, 1989) desenvolveram um neurônio nebuloso que apresenta pesos representados por conjuntos nebulosos ao invés de pesos pertencentes aos números reais. Novas propostas de neurônios nebulosos mais adequadas a esta classe de redes neurais foram pesquisadas por (Gupta e Qi, 1992).

Um outro modelo desenvolvido por (Yamakawa et al., 1992), chamado de *Neo-Fuzzy-Neuron*, apresenta uma coleção de conjuntos nebulosos fixos e pesos pertencentes aos números reais, ao invés de um único peso em cada arco associado ao neurônio. Nesta

estrutura, a agregação dos sinais sinápticos é determinada simplesmente por uma soma algébrica. Do ponto de vista computacional, este modelo apresenta tempo de processamento inferior à estrutura da rede neural multi-camadas, pois o número de operações envolvidas no processamento é consideravelmente menor e o cálculo das funções é mais simples.

Esta estrutura de rede neural nebulosa foi aplicada ao problema de previsão de vazões médias mensais um passo à frente em (Ballini et al., 1999c). Os resultados obtidos foram comparados aos modelos de redes neurais multi-camadas e auto-regressivos periódicos. Em geral, estes resultados mostraram uma melhor desempenho para a rede neural nebulosa, porém pode-se observar que o número de regras nebulosas obtidas através da análise Bayesiana sob os dados de vazões não cobriam completamente o espaço de entradas, provocando uma degradação no desempenho do modelo. Este fato pode ser provocado devido a complexidade do problema real (no caso, previsão de vazões), pois quanto maior a complexidade associada ao sistema real, maiores são as dificuldades para se determinar os parâmetros que definem a base de regras para o sistema nebuloso.

Em (Ballini et al., 1998) e (Ballini et al., 1999b) foi utilizada uma estrutura de rede neural nebulosa para previsão de vazões, em que os parâmetros que definem as funções de pertinência do sistema nebuloso são obtidos através de um método de aprendizagem construtivo. Este modelo foi proposto por (Figueiredo et al., 1995) e será descrito a seguir juntamente com seus algoritmos de aprendizagem.

5.2 Rede Neural Nebulosa

(Figueiredo et al., 1995) desenvolveram um modelo onde os sinais de entrada e os pesos sinápticos do neurônio nebuloso parametrizam funções de pertinência de conjuntos nebulosos. As atividades sinápticas, a agregação de sinais no corpo celular e a geração de sinais no axônio são modelados, respectivamente, por operadores monotônicos, operador de agregação e uma função de codificação não-linear. Este modelo define uma classe de neurônios bastante geral, pois dependendo da escolha conveniente dos operadores é possível representar diferentes modelos de neurônios nebulosos, inclusive o modelo básico de neurônio. Esta seção apresenta a estrutura da rede neural nebulosa e os algoritmos de aprendizagem propostos em (Figueiredo, 1997).

As principais características da rede neural nebulosa (RNN) são (Figueiredo e Gomide, 1997):

1. Fácil inserção e extração de conhecimento na forma de regras lingüísticas nebulosas;
2. O processamento neural é semelhante ao mecanismo de raciocínio nebuloso;

3. Regras são geradas automaticamente, sem qualquer procedimento adicional;
4. A partição do espaço de entradas garante que as regras cobrem todo o domínio de operação do modelo;
5. A RNN é um aproximador universal.

A aprendizagem da RNN não se limita a um único período de treinamento, pois sua estrutura adquire conhecimento enquanto e sempre que o desempenho não satisfizer um índice de desempenho específico. Este fato implica na alteração da estrutura da rede, correspondendo à inserção de regras nebulosas e ajuste das funções de pertinência.

Esta estrutura apresenta um método de aprendizagem construtivo, o qual capacita a rede compensar degradações de desempenho provocadas por variações dos parâmetros do modelo ou outras perturbações devido a mudanças provocadas no sistema real.

5.2.1 Estrutura da Rede Neural Nebulosa

A estrutura da RNN foi desenvolvida com base em duas características essenciais: o mapeamento das regras nebulosas na estrutura é direto e o processamento neural deve equivaler totalmente a um mecanismo de inferência nebulosa. Essas propriedades são condições necessárias para garantir que o sistema proposto tenha uma natureza dual, ou seja, o sistema pode ser visto ora como um sistema baseado em regras nebulosas, ora como uma rede neural nebulosa (Pedrycz e Gomide, 1998).

A rede é baseada no mecanismo de inferência nebulosa, codificando uma base de regras na forma de “Se um conjunto de condições é satisfeito Então um conjunto de consequentes é inferido”, o qual pode ser definido da seguinte forma:

| | |
|-----------|---|
| Entradas: | $x_1 \text{ é } A_1 \text{ e } \dots \text{ e } x_M \text{ é } A_M$ |
| | Se $x_1 \text{ é } A_1^1 \text{ e } \dots \text{ e } x_M \text{ é } A_M^1$ Então $y \text{ é } w_1$ |
| | |
| Regras: | Se $x_1 \text{ é } A_1^i \text{ e } \dots \text{ e } x_M \text{ é } A_M^i$ Então $y \text{ é } w_i$ |
| | |
| | Se $x_1 \text{ é } A_1^N \text{ e } \dots \text{ e } x_M \text{ é } A_M^N$ Então $y \text{ é } w_N$ |
| Saída: | $y = y(\mathbf{x})$ |

onde x_j , $j = 1, \dots, M$, é uma variável nebulosa e M o número de entradas; A_j e A_j^i , $i = 1, \dots, N$, são conjuntos nebulosos representados por funções de pertinência sendo N o número de regras nebulosas; y é um número real definido no espaço de saída. Os valores de $w_i(\mathbf{x}) = w_i$, são números reais e representam os pesos sinápticos correspondentes aos consequentes das regras nebulosas, sendo que, para cada vetor de entrada $\mathbf{x} = x_1, \dots, x_j, \dots, x_M$, está associado um único peso sináptico w_i .

Para efeito de cálculo e representação de conhecimento na RNN, todas as funções de pertinência são discretizadas, ou seja, para cada conjunto nebuloso Z define-se uma função de pertinência $Z(\cdot) : U \rightarrow [0, 1]$, onde U é o universo de discurso, tal que $Z(x_j) = z_k$, se $x_j \in I_{jk} = (x_{jk_I}, x_{jk_F}]$, onde I_{jk} é o k -ésimo intervalo de discretização em relação à entrada x_j , $k = 1, \dots, Q$ (figura 5.1).

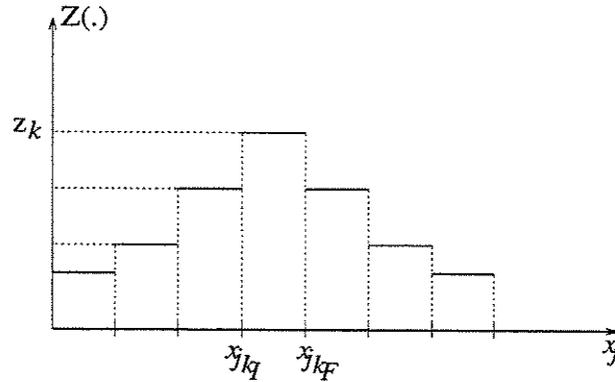


Figura 5.1: Função de Pertinência Discretizada.

Desta forma, define-se a_{jk} como sendo o valor da função de pertinência do conjunto nebuloso A_j , para todo $x_j \in I_{jk}$, ou seja, $A_j(x_j) = a_{jk}$. O valor de a_{jk}^i é também definido como sendo o valor da função de pertinência do conjunto nebuloso A_j^i , isto é, $A_j^i(x_j) = a_{jk}^i$, se $x_j \in I_{jk}$.

A implementação do mecanismo de inferência nebulosa pode ser obtida de várias formas (Pedrycz e Gomide, 1998). Em (Figueiredo, 1997) foi utilizado um dos métodos mais conhecidos, proposto por (Yager e Filev, 1994), onde a saída y é determinada a partir de três estágios, como segue:

1. *Comparação*: Para cada regra i , $i = 1, \dots, N$, e para cada antecedente j , $j = 1, \dots, M$, calcula-se a medida de possibilidade P_j^i entre os conjuntos nebulosos A_j e A_j^i , da seguinte

forma:

$$P_j^i(\mathbf{x}) = S_k \{ T(A_j(\mathbf{x}), A_j^i(\mathbf{x})) \} \quad (5.1)$$

onde S e T são operadores, correspondentes a uma s -norma e uma t -norma, respectivamente, e $\mathbf{x} = (x_1, \dots, x_j, \dots, x_M)'$ é o vetor de entrada. A notação $S_k(\cdot, \cdot)$ é a aplicação da norma $S(\cdot, \cdot)$ para todo intervalo de discretização $k = 1, \dots, Q$ relativos à variável do vetor de entrada x_j .

2. *Agregação dos Antecedentes:* Para cada regra i , $i = 1, \dots, N$, agregam-se os valores de cada um dos antecedentes através de uma t -norma. O resultado desta agregação é chamado de nível de ativação H^i , definido por:

$$H^i(\mathbf{x}) = T_j \{ P_j^i(\mathbf{x}) \} \quad (5.2)$$

onde a notação $T_j(\cdot, \cdot)$ é a aplicação da norma $T(\cdot, \cdot)$ para todo $j = 1, \dots, M$ relativo ao vetor de entrada \mathbf{x} .

3. *Agregação das Regras:* A saída y do sistema de inferência, ou seja, a etapa de defuzzificação dos dados, é realizada através de uma média ponderada entre os consequentes e os níveis de ativação calculados em (5.2) para cada regra, dada por:

$$y(\mathbf{x}) = \frac{\sum_{i=1}^N H^i(\mathbf{x}) w_i}{\sum_{i=1}^N H^i(\mathbf{x})} \quad (5.3)$$

A RNN é constituída por unidades de processamento, chamadas de neurônios nebulosos. Um modelo geral do neurônio nebuloso é mostrado na figura 5.2. Matematicamente, este neurônio pode ser representado da seguinte forma:

$$y(\mathbf{x}) = \varphi(\phi(\Psi(\mathbf{x}))) = \varphi(\phi(\psi_1(x_1), \dots, \psi_M(x_M)))$$

onde $\Psi : \mathfrak{R}^M \rightarrow \mathfrak{R}^M$ é o operador sináptico, $\phi : \mathfrak{R}^M \rightarrow \mathfrak{R}$ é o operador de agregação e $\varphi : \mathfrak{R} \rightarrow \mathfrak{R}$ é a função de ativação.

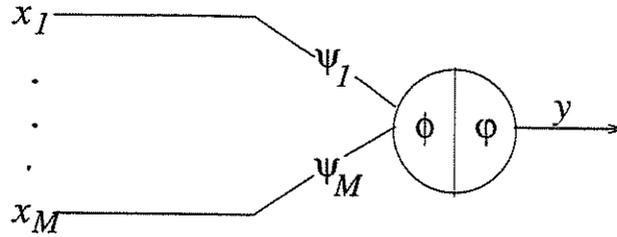


Figura 5.2: Modelo Geral do Neurônio Nebuloso.

O modelo RNN apresenta uma arquitetura não recorrente com 5 camadas, como mostra a figura 5.3. Pode-se verificar que o mecanismo de inferência descrito através dos procedimentos de comparação, agregação dos antecedentes e agregação das regras é isomórfico à estrutura da rede.

A primeira camada é dividida em M grupos de neurônios, cada um associado a uma variável nebulosa de entrada. Cada neurônio nesta camada representa um intervalo de discretização I_{j_k} , correspondente ao espaço de entrada. Ou seja, há tantos neurônios nesta camada quantos forem os intervalos de discretização. Cada grupo de neurônios é responsável por transformar as entradas não nebulosas para que as camadas seguintes possam tratar a informação adequadamente. Estes grupos geram os valores a_{j_k} referentes à função de pertinência do conjunto nebuloso A_j . Assim, um neurônio de entrada recebe um sinal simples, decodifica e o transmite para a segunda camada. O sinal a_{j_k} é transmitido pelo k -ésimo neurônio localizado no j -ésimo grupo.

Definindo-se ψ e ϕ como funções identidade, a saída $a_{j_k} \in \{0,1\}$ é dada por $a_{j_k} = \phi(x_j)$. Para um intervalo $I_{j_k} = (x_{j_{k_I}}, x_{j_{k_F}}]$, a função decodificada $\phi(\cdot)$ para este neurônio, através do k -ésimo intervalo, mostrada na figura 5.4, é dada por:

$$a_{j_k} = \phi(x_j) = \begin{cases} 1, & \text{se } x_j \in (x_{j_{k_I}}, x_{j_{k_F}}] \\ 0, & \text{caso contrário} \end{cases}$$

A segunda camada contém N grupos (número de regras pré-definido), cada qual contendo M neurônios (número de antecedentes de cada regra). Esta camada representa o primeiro estágio de inferência chamado comparação. O j -ésimo neurônio do i -ésimo grupo representa, através dos pesos sinápticos $a_{j_k}^i$, a função de pertinência do conjunto nebuloso A_j^i . Portanto, o i -ésimo grupo é composto por neurônios que representam as funções de pertinência de todos os conjuntos nebulosos correspondentes aos antecedentes

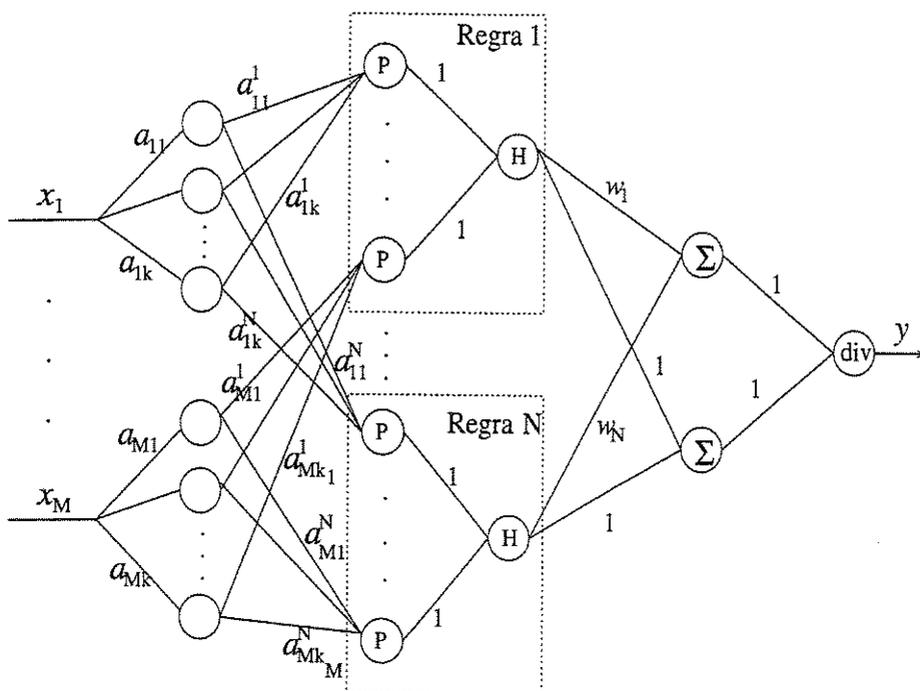


Figura 5.3: Rede Neural Nebulosa.

da i -ésima regra. Além disso, cada neurônio da i -ésima regra calcula a comparação P_j^i entre os conjuntos nebulosos A_j e A_j^i . Assim, o k -ésimo neurônio do j -ésimo grupo da primeira camada, cuja saída é dada por a_{jk} , se conecta com o i -ésimo grupo da segunda camada através de pesos sinápticos a_{jk}^i . Considerando-se que $\psi = t$ -norma, $\phi = s$ -norma e φ é a função identidade, a saída do j -ésimo neurônio é a medida de possibilidade P_j^i dada pela equação (5.1).

A terceira camada contém N neurônios, cada um com M entradas. Para cada grupo i da segunda camada, um neurônio na terceira camada realiza agregação dos antecedentes (segundo estágio de inferência). Todas as conexões possuem peso sináptico unitário. A saída desta camada é dada pela equação (5.2), assumindo que $\psi =$ função identidade, $\phi = t$ -norma e $\varphi =$ função identidade.

A quarta camada possui dois neurônios, ambos com N entradas. Ou seja, cada neurônio i da terceira camada está conectado com os dois neurônios da quarta camada. O operador de agregação associado com estes neurônios é a soma algébrica. Um dos neurônios conecta-se com o i -ésimo neurônio da camada anterior através das sinapses, com pesos w_i .

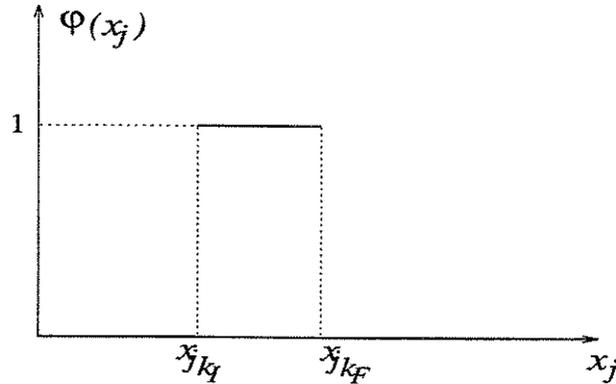


Figura 5.4: Função de Decodificação.

Os operadores ψ e ϕ são, respectivamente, o operador algébrico e a soma algébrica, e φ é a função identidade. As saídas são constituídas pelo numerador da equação (5.3).

O outro neurônio também é conectado com todos os neurônios da camada anterior com pesos sinápticos unitários. Considerando ψ =função identidade, ϕ =soma algébrica e φ = função identidade, a saída equivale ao denominador da inferência nebulosa definida em (5.3).

Finalmente, a última camada consiste de um único neurônio para calcular o quociente dos sinais de entrada, ou seja, para calcular $y(\mathbf{x})$ (equação (5.3)). Portanto, as conexões possuem pesos unitários e os operadores ψ , ϕ e φ são função identidade, operador de divisão e a função identidade, respectivamente.

Cada uma das camadas da RNN possui uma associação com os diversos estágios da inferência nebulosa, descritos de (5.1) a (5.3). Além disso, os pesos sinápticos representam tanto o conhecimento da rede neural como também representam as funções de pertinência dos antecedentes e conseqüentes das regras nebulosas. Assim, a RNN codifica um conjunto de N regras em sua estrutura, processa as informações através de neurocomputação, executando raciocínio nebuloso.

5.2.2 Métodos de Aprendizagem

A estratégia de aprendizagem da RNN se divide em duas fases. A primeira fase corresponde à aprendizagem não supervisionada, ou seja, a aprendizagem se processa

sem que as saídas sejam fornecidas, envolvendo somente uma parte da arquitetura, mais especificamente a segunda camada. Nesta fase, a rede aprende as funções de pertinência dos antecedentes das regras nebulosas enquanto ajusta os pesos sinápticos associados a esta camada. O método utilizado durante esta etapa é do tipo competitivo, baseado nas redes do tipo Kohonen. A segunda fase adota um método supervisionado, baseado no método do gradiente para aprendizagem dos conseqüentes das regras, e envolve apenas o ajuste dos pesos da quarta camada. Assim, a RNN deve aprender os pesos sinápticos a_{jk}^i relacionados aos neurônios da segunda camada, bem como os pesos sinápticos w_i da quarta camada.

O algoritmo de aprendizagem consiste na apresentação dos pares de entrada/saída à rede, sendo o t -ésimo par dado por $(\mathbf{x}(t), y_d(t))$, onde $\mathbf{x}(t) = (x_1(t), \dots, x_M(t))'$ é o vetor de entrada com M coordenadas e $y_d(t)$ é a saída desejada.

Considera-se que os conjuntos nebulosos, $A_j^i(\cdot)$, são representados por funções de pertinência do mesmo tipo, simétricas e completamente definidas pelos valores modais c_j^i e de dispersão ρ^i . Por exemplo, funções do tipo Gaussianas ou funções triangulares. Determinando-se o valor modal e de dispersão de cada função de pertinência $F_j^i(\cdot)$ e definidos os intervalos de discretização I_{jk} , é possível determinar os pesos sinápticos a_{jk}^i na forma:

$$a_{jk}^i = F_j^i(x_j), \quad \text{tal que } x_j \in I_{jk} = (x_{jk_I}, x_{jk_F}]$$

Os valores modal e de dispersão de cada função $F_j^i(\cdot)$ devem ser aprendidos durante o treinamento, segundo o método de aprendizado competitivo, para determinar os pesos sinápticos a_{jk}^i da segunda camada.

A função erro que determina o índice de desempenho do modelo, necessária à aplicação do método baseado no gradiente descendente, é definida por:

$$E(\mathbf{x}(t)) = \frac{1}{2} (y_d(t) - y(t))^2 \quad (5.4)$$

onde $y(t)$ é a saída da rede em relação à entrada $\mathbf{x}(t)$, dada pela equação (5.3).

O objetivo do processo de treinamento é ajustar os parâmetros (valor modal, dispersão e pesos) da rede, minimizando $E(\mathbf{x}(t))$. Desta forma, o ajuste dos pesos sinápticos w_i é proporcional à direção oposta ao gradiente $\partial E(\mathbf{x}(t))/\partial w_i(t)$ e é dado da seguinte forma:

$$w_i(t+1) = w_i(t) - \eta \frac{\partial E(\mathbf{x}(t))}{\partial w_i(t)} \quad (5.5)$$

onde η é a taxa de aprendizagem. Derivando a equação (5.4) em relação a w_i , tem-se:

$$\frac{\partial E(\mathbf{x}(t))}{\partial w_i} = \frac{(y_d(t) - y(t)) H^i(\mathbf{x}(t))}{\sum_{i=1}^n H^i(\mathbf{x}(t))} \quad (5.6)$$

Substituindo a equação (5.6) na equação (5.5), o ajuste dos pesos sinápticos $w_i(t)$ é descrito da seguinte forma:

$$w_i(t+1) = w_i(t) + \eta \left(\frac{(y_d(t) - y(t)) H^i(\mathbf{x}(t))}{\sum_{h=1}^n H^h(\mathbf{x}(t))} \right) \quad (5.7)$$

A equação (5.7) é utilizada nos dois métodos de aprendizagem chamados de métodos de aprendizagem *off-line* e *on-line* propostos por (Figueiredo, 1997). Estes dois algoritmos são baseados no paradigma de aprendizagem competitiva e serão descritos nas próximas seções.

5.2.3 Método de Aprendizagem *Off-Line*

No método de aprendizagem *off-line*, o número de regras nebulosas N , às quais correspondem os grupos de neurônios da segunda camada, é pré-estabelecido no início do treinamento. Assim, existem N grupos e, para cada grupo, M funções de pertinência $F_j^i(\cdot)$, como descrito na subseção 5.2.1. Os grupos de neurônios competem entre si sempre que a RNN recebe uma nova entrada. O grupo vencedor estabelece quais e como os parâmetros da rede são ajustados. O resultado da competição depende da existência ou não de regras tais que o nível de ativação seja diferente de zero.

Este método de aprendizagem apresenta duas fases. Na primeira fase a RNN aprende com os pares de entrada/saída pertencentes ao conjunto de treinamento, apresentados aleatoriamente à rede. Assim, os valores modais c_j^i são posicionados adequadamente em seus respectivos universos de discurso e a dispersão de cada função $F_j^i(\cdot)$ é ampliada convenientemente, garantindo que todo o espaço de entrada seja coberto pelas regras estabelecidas. Ou seja para qualquer entrada apresentada à RNN sempre há um grupo neural i pertencente à segunda camada tal que $H^i(\mathbf{x}(t)) \neq 0$. A primeira fase se encerra após a rede

cobrir todo o espaço de entradas e o fator de aprendizagem $\beta^i(t)$, descrito a seguir, atingir valores menores que um certo mínimo para todo i . A partir desta fase, os pesos $w_i(t)$ da quarta camada são ajustados a cada par entrada/saída apresentada à rede, correspondendo à segunda fase do algoritmo.

O fator de aprendizagem $\beta^i(t)$ é definido como sendo uma função monotônica decrescente positiva, tal que $0 < \beta^i(t) < 1$. Esta função é escolhida no início do treinamento, sendo utilizada para ajustar os valores modais c_j^i . Neste trabalho, a seguinte função é considerada:

$$\beta^i(t) = \frac{\epsilon}{\alpha^i(t) + 1} \quad (5.8)$$

onde $\alpha^i(t)$ é o número de vezes que o valor modal c_j^i é ajustado até a t -ésima entrada, $i = 1, \dots, N$, e $\epsilon \in (0, 1]$.

A taxa de aprendizagem η é utilizada para ajustar os pesos $w_i(t)$, ou seja, os conseqüentes das regras, devendo assumir valores no intervalo $(0, 1)$. Para ajuste da dispersão ρ^i é empregada uma taxa de ampliação $\lambda > 1$. Estas taxas também são definidas no início do treinamento.

A primeira fase do algoritmo de aprendizagem *off-line* pode ser descrita como segue:

1. Inicialização:

Defina o número de regras N .

Defina o perfil das funções de pertinência $F_j^i(\cdot)$. Os valores modais c_j^i e de dispersão ρ^i , $i = 1, \dots, N$ e $j = 1, \dots, M$ são inicializados aleatoriamente. A dispersão $\rho^i > 0$ deve assumir valores pequenos e o valor modal c_j^i deve pertencer ao universo de discurso da respectiva variável x_j ;

Defina os valores para ϵ e η , ambos pertencentes ao intervalo $(0, 1)$;

Defina a taxa de ampliação λ próxima à unidade.

2. Estímulo:

Apresente à rede um vetor de entrada $\mathbf{x}(t) = (x_1(t), \dots, x_n(t))'$ selecionado aleatoriamente do conjunto de treinamento.

3. Competição:

Calcule o nível de ativação $H^i(\mathbf{x}(t))$ para cada regra i e determine a regra vencedora. Esta regra guia o procedimento de adaptação dos parâmetros da rede da seguinte forma:

(a) A entrada $\mathbf{x}(t)$ é caracterizada por pelo menos uma regra, ou seja, $\exists i | H^i(\mathbf{x}(t)) \neq 0$.

O grupo $\xi(t)$ vence a competição se $H^{\xi(t)}(\mathbf{x}(t)) \geq H^i(\mathbf{x}(t)), \forall i \in \{1, \dots, N\}$.

Prossiga o processo de aprendizagem em 4(a).

(b) A entrada $\mathbf{x}(t)$ não pertence a nenhuma regra, ou seja, $H^i(\mathbf{x}(t)) = 0, \forall i \in \{1, \dots, N\}$.

Neste caso, determinam-se as distâncias $D(\mathbf{c}^i(t), \mathbf{x}(t))$ entre os vetores de entrada $\mathbf{x}(t)$ e modal $\mathbf{c}^i(t) = (c_1^i(t), \dots, c_M^i(t))$, para $i \in \{1, \dots, N\}$. O grupo vencedor $\xi(t)$ é tal que $D(\mathbf{c}^{\xi(t)}, \mathbf{x}(t)) \leq D(\mathbf{c}^i(t), \mathbf{x}(t)), \forall i \in \{1, \dots, N\}$.

Prossiga o processo de aprendizagem em 4(b).

4. Ajuste e Atualização da Rede:

(a) Se o vencedor for determinado segundo o caso 3(a), então as equações para ajuste do valor modal $c_j^i(t)$, dispersão $\rho^i(t)$, conseqüente $w_i(t)$ e para atualização $\alpha^i(t)$ dos parâmetros são dadas por:

$$c_j^{\xi(t)}(t+1) = c_j^{\xi(t)}(t) + \beta^{\xi(t)}(t) \left(x_j(t) - c_j^{\xi(t)}(t) \right), \quad j = 1, \dots, M \quad (5.9)$$

$$c_j^i(t+1) = c_j^i(t), \quad i = 1 \dots, N, \quad i \neq \xi(t) \quad (5.10)$$

$$\rho^i(t+1) = \rho^i(t), \quad i = 1 \dots, N \quad (5.11)$$

$$w_i(t+1) = w_i(t) + \eta \left(\frac{H^i(\mathbf{x}(t)) (y_d(t) - y(t))}{\sum_{h=1}^N H^h(\mathbf{x}(t))} \right), \quad i = 1 \dots, N \quad (5.12)$$

$$\alpha^{\xi(t)}(t+1) = \alpha^{\xi(t)}(t) + 1 \quad (5.13)$$

$$\alpha^i(t+1) = \alpha^i(t), \quad i = 1 \dots, N, \quad i \neq \xi(t) \quad (5.14)$$

Prossiga o processo de aprendizagem em 5.

- (b) Se o vetor de entrada $\mathbf{x}(t)$ não foi classificado por nenhuma regra (caso 3(b)), os parâmetros da rede são ajustados como segue:

$$c_j^{\xi(t)}(t+1) = c_j^{\xi(t)}(t) + \beta^{\xi(t)}(t) (x_j(t) - c_j^{\xi(t)}(t)), \quad j = 1, \dots, M \quad (5.15)$$

$$c_j^i(t+1) = c_j^i(t), \quad i = 1 \dots, N, \quad i \neq \xi(t) \quad (5.16)$$

Recalcular o valor do nível de ativação $H^{\xi(t)}(\mathbf{x}(t))$ da regra vencedora $\xi(t)$, para os valores ajustados de $c^{\xi(t)}(t+1)$. Se para o vetor de entrada $\mathbf{x}(t)$, o nível de ativação $H^{\xi(t)}(\mathbf{x}(t))$ continuar a ser zero, então amplifique a dispersão $\rho^{\xi(t)}(t)$ de cada função de pertinência $F_j^i(\cdot)$ até que $H^{\xi(t)}(\mathbf{x}(t)) \neq 0$, ou seja:

Enquanto $H^{\xi(t)}(\mathbf{x}(t)) = 0$ faça:

$$\rho^{\xi(t)}(t) = \lambda \rho^{\xi(t)}(t) \quad (5.17)$$

Ajuste e atualize os valores de dispersão ρ^i e os conseqüentes w_i das regras, de acordo com as equações abaixo:

$$\rho^i(t+1) = \rho^i(t), \quad i = 1, \dots, N, \quad i \neq \xi(t) \quad (5.18)$$

$$w_i(t+1) = w_i(t) + \eta \left(\frac{H^i(\mathbf{x}(t)) (y_d(t) - y(t))}{\sum_{h=1}^N H^h(\mathbf{x}(t))} \right), \quad \text{para } i = \xi(t) \quad (5.19)$$

$$w_i(t+1) = w_i(t), \quad i = 1, \dots, N, \quad i \neq \xi(t) \quad (5.20)$$

$$\alpha^{\xi(t)}(t+1) = \alpha^{\xi(t)}(t) + 1 \quad (5.21)$$

$$\alpha^i(t+1) = \alpha^i(t), \quad i = 1, \dots, N, \quad i \neq \xi(t) \quad (5.22)$$

Prossiga o processo de aprendizagem em 5.

5. Avaliação de Convergência:

Se para qualquer vetor de entrada $\mathbf{x}(t)$ pertencente ao conjunto de treinamento, $\exists i \mid H^i(\mathbf{x}(t)) \neq 0$ e $\beta^i(t)$ menor que um certo valor mínimo, vá para 6.

Caso contrário, retorne para 2.

6. Fim:

Aprendizagem encerrada.

Após a primeira fase de aprendizagem, as regras nebulosas codificadas pela rede cobrem todo o espaço de entradas. Para cada entrada apresentada à rede, pelo menos uma regra é disparada. Considera-se então que as funções de pertinência, relativas aos antecedentes das regras, estejam ajustadas convenientemente. Por este motivo, na segunda fase de aprendizagem somente os consequentes são ajustados. A partir do início da segunda fase, a rede permanece em aprendizagem contínua. O algoritmo de aprendizagem para a segunda fase é como segue:

1. Estímulo:

Apresente à rede um vetor de entrada $\mathbf{x}(t)$.

2. Ajuste:

Ajuste os pesos sinápticos de acordo com a equação abaixo:

$$w_i(t+1) = w_i(t) + \eta \left(\frac{H^i(\mathbf{x}(t)) (y_d(t) - y(t))}{\sum_{h=1}^N H^h(\mathbf{x}(t))} \right), \quad i = 1, \dots, N \quad (5.23)$$

3. Retorne para 1.

No algoritmo descrito acima, utilizou-se uma função particular para o fator de aprendizagem $\beta^i(t)$ (equação (5.8)). Entretanto, outras funções podem ser escolhidas, desde que sejam do tipo monotônica decrescente positiva, tal que $0 < \beta^i < 1$.

5.2.4 Método de Aprendizagem *On-Line*

O método de aprendizagem *on-line* capacita a RNN a adquirir novos conhecimentos codificando novas regras nebulosas em sua estrutura, alterando a arquitetura da rede. Portanto, o número de regras nebulosas codificadas pela rede, correspondente ao número de grupos de neurônios na segunda camada, não é constante, ou seja, $N = N(t)$.

A estrutura da rede é alterada basicamente devido a dois motivos: o desempenho desejado não é satisfeito pela rede ou todos os grupos neurais da segunda camada são tais que o nível de ativação é igual a zero. Nestes casos, a rede acrescenta novos elementos à sua estrutura de maneira a codificar uma nova regra nebulosa. Portanto, os grupos de neurônios da segunda camada competem entre si sempre que a rede recebe uma nova entrada $\mathbf{x}(t)$. O grupo do neurônio vencedor guia os procedimentos de ajuste dos pesos sinápticos.

O algoritmo *on-line* utiliza fatores de aprendizagem análogos aos do algoritmo *off-line*. Para ajuste dos valores modais $c_j^i(t)$ das funções de pertinências $F_j^i(\cdot)$ e conseqüentes das regras nebulosas $w_i(t)$, utilizam-se os fatores de aprendizagem $\beta^i(t)$ e $\eta \in (0, 1)$, respectivamente. O fator $\beta^i(t)$ é definido da mesma maneira que no método *off-line*, ou seja, $\beta^i(t) = \epsilon / (\alpha^i(t) + 1)$, onde $\alpha^i(t)$ é o número de vezes que o valor modal é ajustado até a t -ésima entrada e $\epsilon \in (0, 1]$. Ao contrário do método *off-line*, que possui um fator de ampliação, o método *on-line* utiliza um fator de redução $\gamma \in (0, 1)$ para ajuste da dispersão dos conjuntos nebulosos.

O algoritmo de aprendizagem *on-line* pode ser definido através dos seguintes passos:

1. Inicialização:

Inicialize o número de regras. Neste caso, o número de regras inicial é $N(0) = 0$ para $t = 0$.

Defina os perfis das funções de pertinência $F_j^i(\cdot)$ que formam as regras nebulosas.

Defina a função $\beta^i(t)$.

Estabeleça os valores para ϵ , η , e γ

Defina o valor inicial da dispersão das funções de pertinência, $\Theta \in \mathfrak{R}$.

Apresente o par entrada/saída inicial ($\mathbf{x}(t)$, $y_d(t)$) à RNN. A rede codifica a primeira regra nebulosa após receber o primeiro par de entrada/saída. A seguir, o valor de N é atualizado e os valores modais são definidos como sendo cada componente do vetor de entrada e o conseqüente como a saída desejada. Inicialize o fator de aprendizagem

$\alpha^i(t)$, ou seja:

$$N(t+1) = N(t) + 1 \quad (5.24)$$

Para $i = 1$ faça:

$$c_j^i(t+1) = x_j(t), \quad j = 1, \dots, M \quad (5.25)$$

$$\rho^i(t+1) = \Theta \quad (5.26)$$

$$w_i(t+1) = y_d(t) \quad (5.27)$$

$$\alpha^i(t+1) = 0 \quad (5.28)$$

$$t = t + 1 \quad (5.29)$$

2. Estímulo:

Apresente à rede um novo par $(\mathbf{x}(t), y_d(t))$ proveniente do sistema a ser modelado.

3. Competição:

A entrada $\mathbf{x}(t)$ é propagada através da rede até que seja calculada a saída da terceira camada, determinando-se, assim, o resultado da agregação dos antecedentes de cada uma das $N(t)$ regras. Dois casos são possíveis:

- (a) A entrada $\mathbf{x}(t)$ é caracterizada por pelo menos uma regra, ou seja, $\exists i | H^i(\mathbf{x}(t)) \neq 0$. A regra $\xi(t)$ vence a competição se $H^{\xi(t)}(\mathbf{x}(t)) \geq H^i(\mathbf{x}(t)), \forall i \in \{1, \dots, N(t)\}$.

Prossiga o processo de aprendizagem em 4(a).

- (b) A entrada $\mathbf{x}(t)$ não é caracterizada por nenhuma regra, isto é, $H^i(\mathbf{x}(t)) = 0, \forall i \in \{1, \dots, N(t)\}$. Neste caso, a distância $D(\mathbf{c}^i(t), \mathbf{x}(t))$ é determinada entre o vetor de entrada e o vetor de valores modais $\mathbf{c}^i(t) = (c_1^i(t) \dots c_M^i(t))'$ e o grupo vencedor $\xi(t)$ é tal que $D(\mathbf{c}^{\xi(t)}(t), \mathbf{x}(t)) \leq D(\mathbf{c}^i(t), \mathbf{x}(t)), \forall i \in \{1, \dots, N(t)\}$.

Prossiga o processo de aprendizagem em 4(b).

4. Ajuste e Atualização da Rede:

- (a) Se a entrada $\mathbf{x}(t)$ foi classificada por alguma regra (caso 3(a)), a saída da terceira camada, determinada para realizar a competição, é propagada até a última camada, calculando-se, assim, a saída da rede. Então os seguintes procedimentos são realizados:

- i. Se a saída da rede satisfizer o desempenho desejado, ou seja, $\|y_d(t) - y(t)\| \leq \delta$, então os parâmetros são ajustados de acordo com as seguintes equações:

$$c_j^{\xi(t)}(t+1) = c_j^{\xi(t)}(t) + \beta^i(t)(x_j(t) - c_j^{\xi(t)}(t)), \quad j = 1, \dots, M \quad (5.30)$$

$$c_j^i(t+1) = c_j^i(t), \quad i = 1, \dots, N(t), \quad i \neq \xi(t), \quad j = 1, \dots, M \quad (5.31)$$

$$\rho^i(t+1) = \rho^i(t), \quad i = 1, \dots, N(t) \quad (5.32)$$

$$w_i(t+1) = w_i(t) + \eta \left(\frac{H^i(\mathbf{x}(t))(y_d(t) - y(t))}{\sum_{h=1}^{N(t)} H^h(\mathbf{x}(t))} \right), \quad i = 1, \dots, N(t) \quad (5.33)$$

$$\alpha^{\xi(t)}(t+1) = \alpha^{\xi(t)}(t) + 1 \quad (5.34)$$

$$\alpha^i(t+1) = \alpha^i(t), \quad i = 1, \dots, N(t), \quad i \neq \xi(t) \quad (5.35)$$

$$N(t+1) = N(t) \quad (5.36)$$

Prossiga o processo de aprendizagem em 5.

- ii. Pode ser que a saída da rede não satisfizer o desempenho esperado, $\|y_d(t) - y(t)\| > \delta$ e exista um vetor de valores modais de uma determinada regra q igual ao padrão de entrada $\mathbf{x}(t) = \mathbf{c}^q(t) = (c_1^q(t) \dots c_M^q(t))'$, ou seja, a rede possui uma regra com antecedentes iguais ao padrão de entrada porém o conseqüente não está sendo calculado corretamente devido à interferência indesejável de outras regras. Para solucionar este problema, os valores modais atuais são mantidos e as dispersões são reduzidas para que as regras nebulosas diferentes de q não influenciem no resultado da entrada $\mathbf{x}(t)$. Assim, os seguintes ajustes são realizados:

$$c_j^i(t+1) = c_j^i(t), \quad i = 1, \dots, N(t), \quad j = 1, \dots, M \quad (5.37)$$

Para $i = 1$ até $N(t)$ e $i \neq q$ faça:

Enquanto $H^i(\mathbf{x}(t)) \neq 0$ faça:

$$\rho^i(t) = \gamma \rho^i \quad (5.38)$$

$$\rho^i(t+1) = \rho^i(t) \quad (5.39)$$

Seja $k \neq q$ tal que $D(\mathbf{c}^k(t), \mathbf{x}(t)) \leq D(\mathbf{c}^i(t), \mathbf{x}(t))$, $\forall i \in \{1, \dots, N(t)\}$, então:

$$\rho^q(t+1) = D(\mathbf{c}^k(t), \mathbf{x}(t)) \quad (5.40)$$

Os demais ajustes são definidos pelas seguintes equações:

$$w_q(t+1) = y_d(t) \quad (5.41)$$

$$w_i(t+1) = w_i(t), \quad i = 1, \dots, N(t), \quad i \neq q \quad (5.42)$$

$$\alpha^q(t+1) = \alpha^q(t) + 1 \quad (5.43)$$

$$\alpha^i(t+1) = \alpha^i(t), \quad i = 1, \dots, N(t), \quad i \neq q \quad (5.44)$$

$$N(t+1) = N(t) \quad (5.45)$$

Prossiga o processo de aprendizagem em 5.

- iii. Se a saída da rede não satisfizer o desempenho desejado, $\|y_d(t) - y(t)\| > \delta$ e não existir um vetor de valores modais igual ao vetor de entrada, ou seja, $\mathbf{x}(t) \neq \mathbf{c}^i(t) = (c_1^i(t) \dots c_M^i(t))'$, $\forall i \in \{1, \dots, N(t)\}$, isto significa que uma nova regra deve ser adicionada à estrutura da rede. As equações de ajuste dos parâmetros são:

$$c_j^i(t+1) = c_j^i(t), \quad i = 1, \dots, N(t), \quad j = 1, \dots, M \quad (5.46)$$

Reduza a dispersão de cada função $F_j^i(\cdot)$ para que as regras existentes deixem de influenciar no resultado da entrada $\mathbf{x}(t)$.

Para $i = 1$ até $N(t)$ faça:

Se $H^i(\mathbf{x}(t)) \neq 0$ então:

$$\rho^i(t) = \gamma \rho^i \quad (5.47)$$

$$\rho^i(t+1) = \rho^i(t) \quad (5.48)$$

$$w_i(t+1) = w_i(t), \quad i = 1, \dots, N(t) \quad (5.49)$$

$$\alpha^{\xi(t)}(t+1) = \alpha^{\xi(t)}(t) + 1 \quad (5.50)$$

$$\alpha^i(t+1) = \alpha^i(t), \quad i = 1, \dots, N(t), \quad i \neq \xi(t) \quad (5.51)$$

Realiza-se a inserção de uma nova regra à estrutura da rede. Os novos valores modais são iguais a cada componente do vetor de entrada, a dispersão é igual à distância $D(\mathbf{c}^{\xi(t)}(t), \mathbf{x}(t))$ e o conseqüente é igual à saída desejada $y_d(t)$:

$$N(t+1) = N(t) + 1 \quad (5.52)$$

Para $i = N(t+1)$ faça:

$$c_j^i(t+1) = x_j(t), \quad j = 1, \dots, M \quad (5.53)$$

$$\rho^i(t+1) = D(\mathbf{c}^{\xi(t)}(t), \mathbf{x}(t)) \quad (5.54)$$

$$w_i(t+1) = y_d(t) \quad (5.55)$$

$$\alpha^i(t+1) = 0 \quad (5.56)$$

Prossiga o processo de aprendizagem em 5.

- (b) Se o vetor de entrada $\mathbf{x}(t)$ não foi classificado por nenhuma regra, ou seja, o grupo vencedor $\xi(t)$ foi escolhido de acordo com o caso 3(b), então é necessário alterar a estrutura da rede, acrescentando uma nova regra capaz de codificar o padrão $(\mathbf{x}(t), y_d(t))$. Os ajustes são realizados da seguinte forma:

$$c_j^i(t+1) = c_j^i(t), \quad i = 1, \dots, N(t), \quad j = 1, \dots, M \quad (5.57)$$

$$\rho^i(t+1) = \rho^i(t), \quad i = 1, \dots, N(t) \quad (5.58)$$

$$w_i(t+1) = w_i(t), \quad i = 1, \dots, N(t) \quad (5.59)$$

$$\alpha^i(t+1) = \alpha^i(t), \quad i = 1, \dots, N(t) \quad (5.60)$$

$$N(t+1) = N(t) + 1 \quad (5.61)$$

Os valores modais para a nova regra são iguais a cada componente do vetor de entrada, a dispersão é igual a $D(\mathbf{c}^{\xi(t)}(t), \mathbf{x}(t))$ e o conseqüente é igual a saída desejada:

Para $i = N(t+1)$ faça:

$$c_j^i(t+1) = x_j^i(t), \quad j = 1, \dots, M \quad (5.62)$$

$$\rho^i(t+1) = D(\mathbf{c}^{\xi(t)}(t), \mathbf{x}(t)) \quad (5.63)$$

$$w_i(t+1) = y_d(t) \quad (5.64)$$

$$\alpha^i(t+1) = 0 \quad (5.65)$$

Prossiga o processo de aprendizagem em 5.

5. Teste de Parada:

Se ainda há padrões a serem treinados, volte ao passo 2. Caso contrário, fim do treinamento.

Os métodos de aprendizagem *off-line* e *on-line*, descritos neste capítulo, capacitam a rede a adquirir novos conhecimentos quando necessário, fornecendo uma forma automática de aprender os parâmetros para um modelo nebuloso sem a interferência ou participação de um especialista. A próxima seção apresenta a aplicação dessas estratégias de aprendizagem no problema de previsão de vazões médias mensais.

5.2.5 Aplicação da Rede Neural Nebulosa para Previsão de Vazões

Neste seção, o modelo de rede neural nebulosa (RNN) com os métodos de aprendizagem *off-line* e *on-line*, é ajustado para a série de vazões do posto de Furnas. Dados de vazões de 1952 a 1956, 1972 a 1976 e 1981 a 1985 foram considerados para teste, correspondendo a três períodos com características distintas: seco, mediano e úmido, respectivamente.

Os dados de vazões foram padronizados para o intervalo $[0, 1]$. Esta padronização é realizada para simplificar a entrada na rede pois, desta forma, pode-se considerar que a primeira camada da RNN contém M grupos e cada grupo contém 1 neurônio.

O vetor de entrada $\mathbf{x}(t)$ contém 6 componentes representadas pelos dados de vazões $t - 1$, $t - 2$, $t - 3$, $t - 11$, $t - 12$, $t - 13$. A escolha dessas entradas foi devido ao fato de fornecerem ao modelo a tendência e informações sobre a sazonalidade da série.

Os parâmetros da rede foram ajustados para cada conjunto de teste. Ou seja, cada período de teste foi retirado do conjunto de treinamento e os parâmetros foram ajustados para este conjunto. A seguir, o conjunto de teste é utilizado para verificar o desempenho do modelo. Por exemplo, o período de teste de 1952 a 1956 foi retirado do conjunto de dados para ajustar os parâmetros da rede. Após o treinamento da rede, é verificado o desempenho da rede usando o conjunto de teste. Este processo é repetido para cada conjunto de teste, resultando para cada conjunto de treinamento um determinado número de regras codificadas em sua estrutura da rede.

As funções de pertinência das regras nebulosas codificadas na RNN foram selecionadas as funções Gaussianas, definidas como segue:

$$F_j^i(x_j) = \begin{cases} 0, & \text{se } |x_j - c_j^i| > 2\rho^i \\ \exp\left(-|x_j - c_j^i|/\rho^i\right), & \text{caso contrário} \end{cases} \quad (5.66)$$

A taxa de aprendizagem para atualização dos pesos w_i foi fixada em $\eta = 0.5$ e o parâmetro $\epsilon = 1.0$ foi utilizado para atualização do fator de aprendizagem $\beta^i(t)$, para os dois métodos de aprendizagem. Os neurônios da segunda camada possuem como operador sináptico a t -norma mínima e como operador de agregação a s -norma máxima. Os neurônios na terceira camada possuem como operador de agregação a t -norma produto.

Um fator de ampliação $\lambda = 1.1$ para a dispersão foi adotado para o algoritmo *off-line*. Os valores modais c_j^i e de dispersão ρ^i são inicializados aleatoriamente.

Para a RNN com o algoritmo *on-line* o fator de redução $\gamma = 0.9$ foi considerado e a inicialização da dispersão das funções de pertinência foi adotada $\Theta = 2.0$. O número de regras nebulosas iniciais foi $N(0)=1$. Os grupos de neurônios foram adicionados à estrutura da rede para um desempenho $\delta = 0.09$, ou seja, novas regras foram adicionadas quando

$$\|y_d(t) - y(t)\| > 0.09$$

A estrutura da RNN é não-recorrente, ou seja, o modelo é estático. Assim, para realizar previsões vários passos à frente, houve a inclusão no vetor de entrada de valores previstos em instantes anteriores como descrito em 4.5.

Em (Ballini et al., 2000) e (Ballini et al., 1999b) foram ajustados um modelo para cada mês, ou seja, foram construídos 12 modelos de redes neurais nebulosas. Entretanto, o desempenho do modelo é semelhante quando ajustado um único modelo, pois a rede apresenta características de abstrair informações das séries codificando regras nebulosas em sua estrutura quando necessário. Além disso, como a arquitetura é adaptável, mesmo para previsão vários passos à frente o desempenho dos modelos é satisfatório. Neste trabalho, são apresentadas previsões obtidas ajustando-se um único modelo à série.

As figuras 5.5, 5.6 e 5.7 mostram as previsões para 1, 3, 6 e 12 passos à frente em (a), (b), (c) e (d), respectivamente, para o ajuste da RNN com algoritmo *off-line*. Neste caso, o número de regras N codificadas na estrutura da rede, o qual deve ser definido pelo usuário no início do treinamento, foi escolhido $N = 18$ para todos os períodos de teste. Pode-se observar que a previsão usando RNN com algoritmo *off-line* apresentou desempenho satisfatório, para os três períodos de teste.

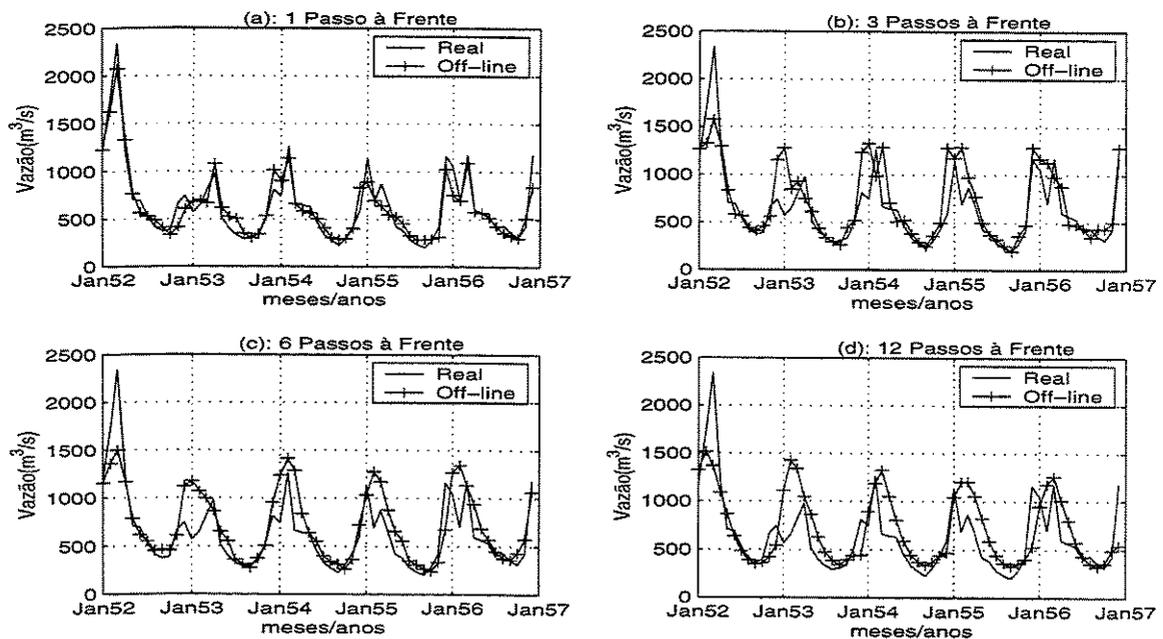


Figura 5.5: Previsão de Vazões Usando um Modelo de RNN com Algoritmo *On-line* para o Posto de Furnas - Período de 1952 a 1956.

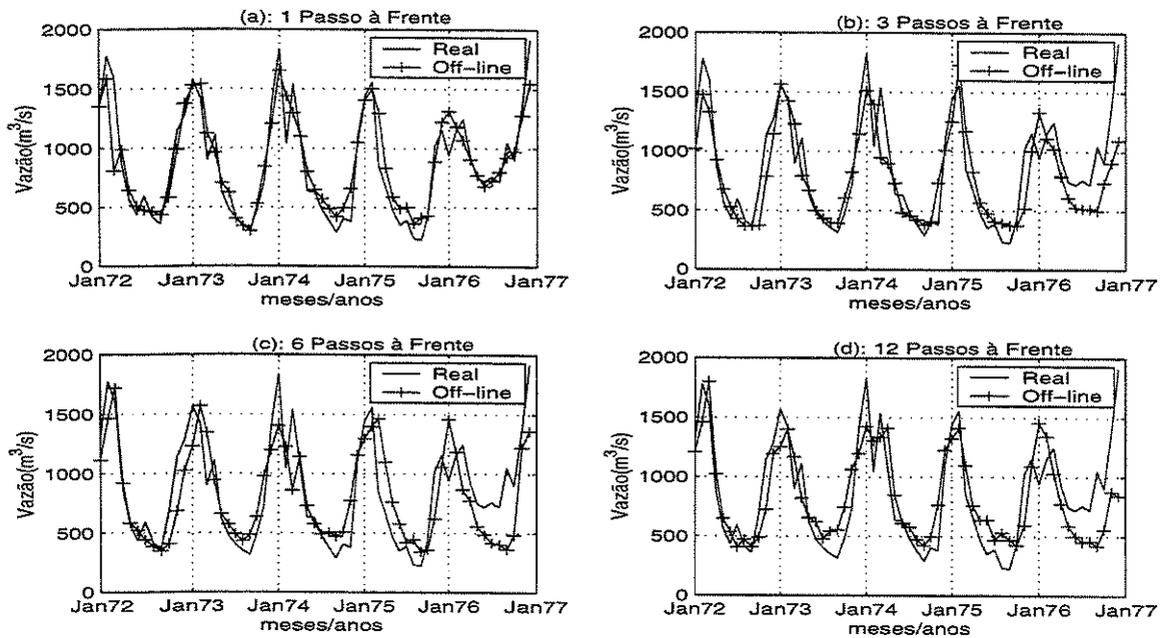


Figura 5.6: Previsão de Vazões Usando um Modelo de RNN com Algoritmo *On-line* para o Posto de Furnas - Período de 1972 a 1976.

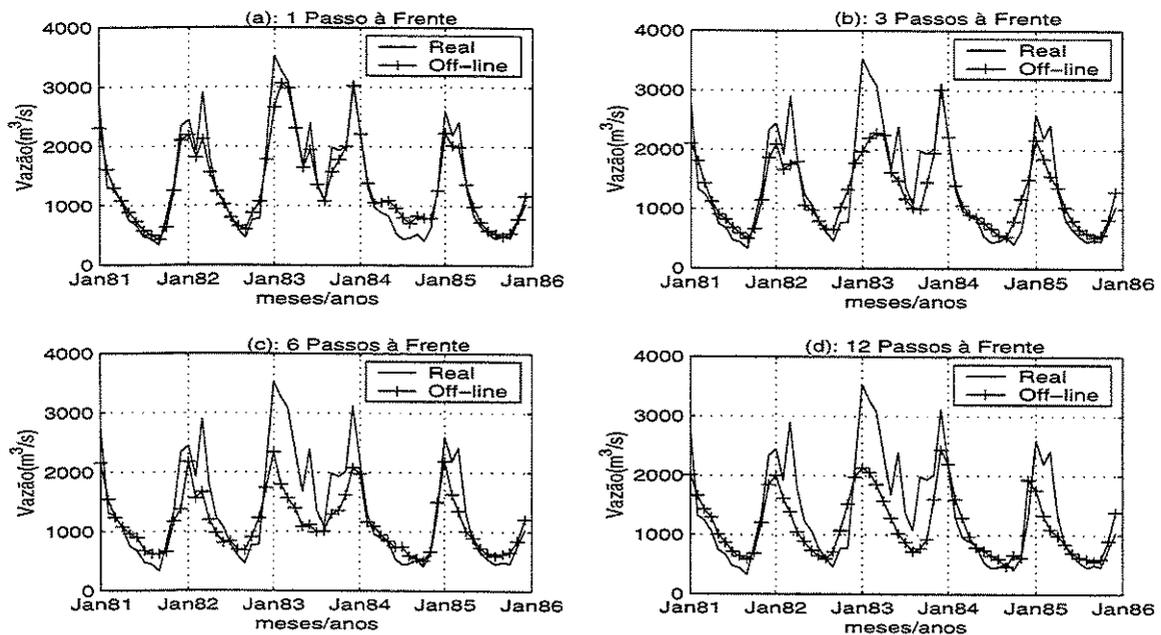


Figura 5.7: Previsão de Vazões Usando um Modelo de RNN com Algoritmo *On-line* para o Posto de Furnas - Período de 1981 a 1985.

Analisando as figuras 5.5(a), 5.6(a) e 5.7(a), pode-se observar que a previsão 1 passo à frente segue a tendência da série apresentando um bom desempenho tanto para os meses de seca (Maio a Outubro) como para os meses de cheia (Novembro a Abril), para os três períodos de teste.

O próximo passo foi ajustar o modelo RNN com algoritmo *on-line*. O número de regras ajustadas para os períodos de 1952 a 1956, 1972 a 1976 e 1981 a 1985, foram 38, 39 e 27, respectivamente. As figuras 5.8, 5.9 e 5.10 mostram as previsões para 1, 3, 6 e 12 passos à frente em (a), (b), (c) e (d), respectivamente. Pode-se observar que o modelo de RNN com algoritmo *on-line* apresentou um desempenho satisfatório.

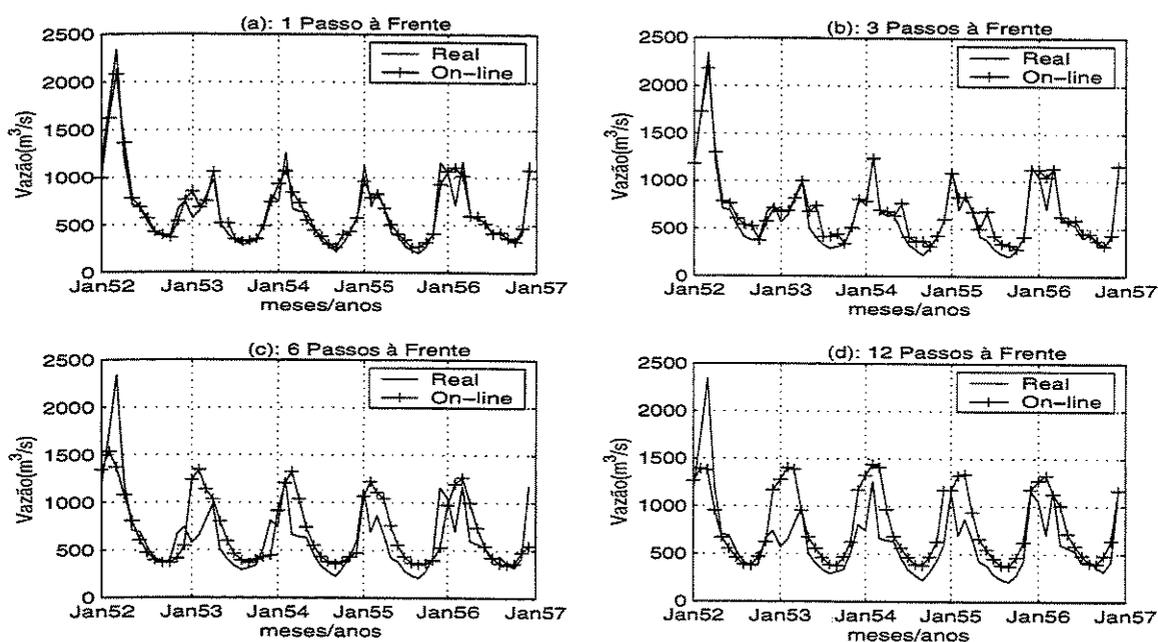


Figura 5.8: Previsão de Vazões Usando um Modelo de RNN com Algoritmo *On-line* para o Posto de Furnas - Período de 1952 a 1956.

Novamente, as previsões 1 passo à frente forneceram resultados satisfatórios. Para as previsões vários passos à frente pode-se observar que a RNN com algoritmo *on-line* apresentou um bom desempenho, indicando que as regras codificadas na estrutura da rede conseguem se adaptar às alterações nos dados de entrada, provocadas pelos ruídos nos dados decorrentes dos erros de previsão.

Os resultados obtidos com os modelos de redes neurais nebulosas são promissores, principalmente para prever os picos (meses de cheia) pois, nestes meses, a variação nos dados é maior o que dificulta realizar boas previsões com outras abordagens.

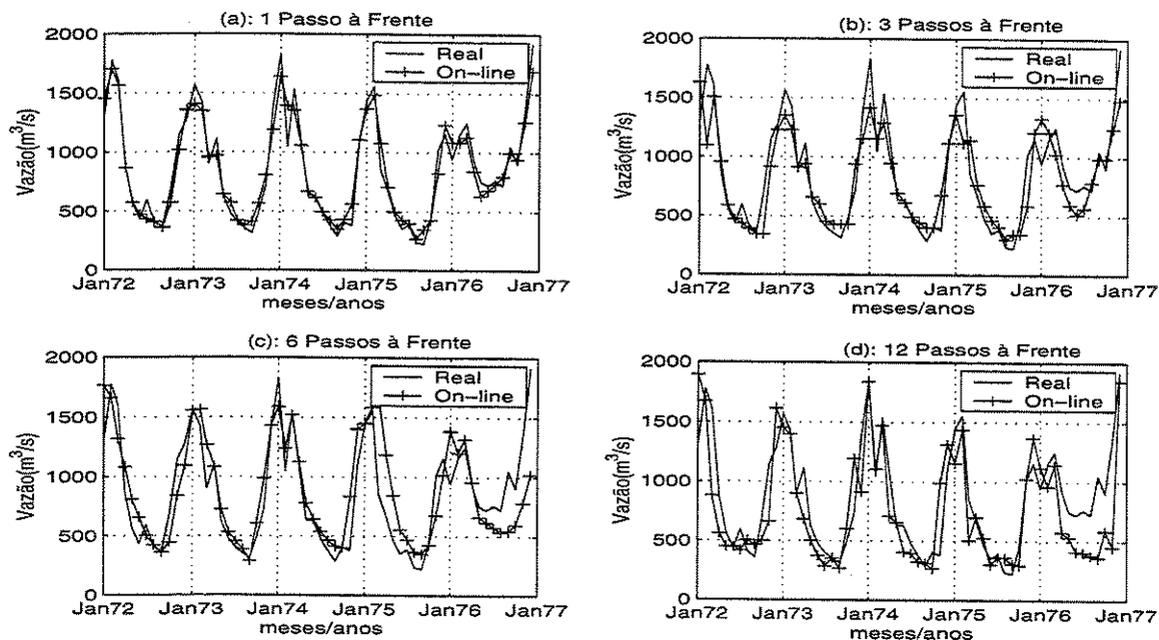


Figura 5.9: Previsão de Vazões Usando um Modelo de RNN com Algoritmo *On-line* para o Posto de Furnas - Período de 1972 a 1976.

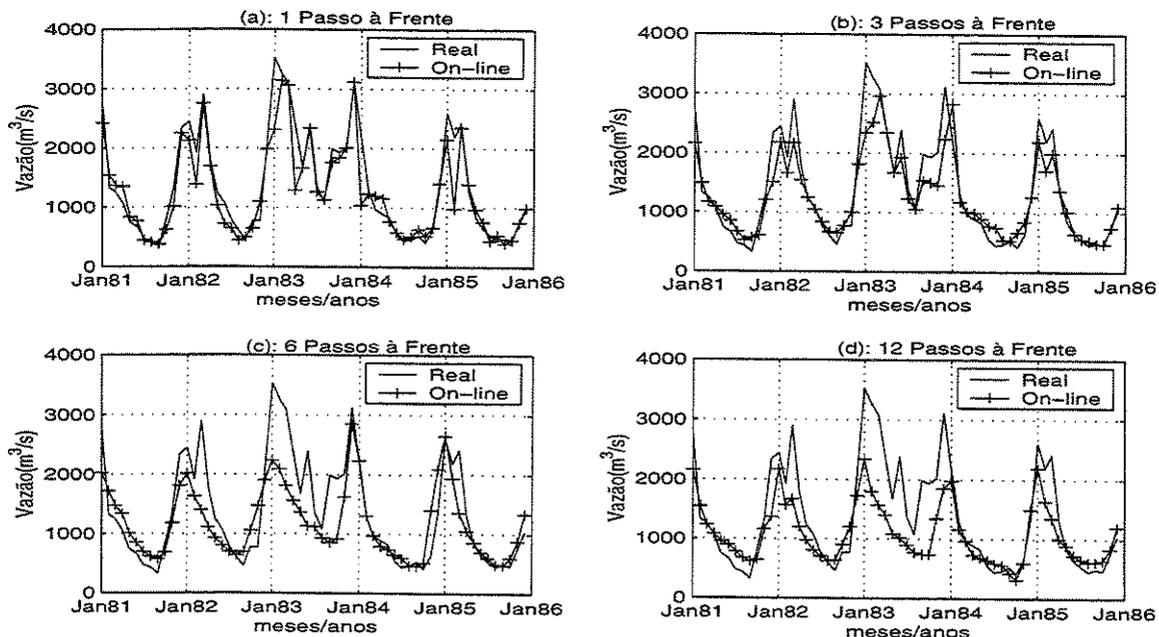


Figura 5.10: Previsão de Vazões Usando um Modelo de RNN com Algoritmo *On-line* para o Posto de Furnas - Período de 1981 a 1985.

5.3 Resumo

Redes neurais nebulosas apresentam como característica principal a capacidade de extrair conhecimento a partir de um conjunto de dados e representar este conhecimento na forma de regras nebulosas.

Neste capítulo, uma arquitetura de rede neural nebulosa adaptativa e dois métodos de aprendizagem construtivos e competitivos são propostos para realizar previsão de vazões afluentes médias mensais.

Este modelo de rede neural nebulosa aprende os parâmetros que definem as funções de pertinência para cada variável de entrada dos dados de treinamento, processa os dados seguindo os princípios de raciocínio nebuloso, codificando um conjunto de regras em sua estrutura, fornecendo uma relação dual entre a rede neural e o sistema nebuloso.

O primeiro método de aprendizagem apresentado neste capítulo é chamado de *off-line*. Neste algoritmo, o número de regras nebulosas é definido no início do treinamento. Estas regras nebulosas competem entre si, sendo que o grupo vencedor guia o processo de adaptação dos parâmetros da rede.

O segundo algoritmo de aprendizagem é chamado de *on-line*. Este método capacita a rede a adquirir novos conhecimentos codificando novas regras nebulosas em sua estrutura e, assim, alterando a arquitetura da rede. Portanto, o número de regras nebulosas é variável. Através de um método de aprendizagem competitivo, os parâmetros da rede são ajustados e novas regras são adicionadas a estrutura da rede.

A rede neural nebulosa é aplicada ao problema de previsão de vazões afluentes do posto de Furnas e previsões 1, 3, 6 e 12 passos à frente são realizadas para três períodos de teste. O desempenho do modelo com os dois métodos de aprendizagem foi satisfatório para os três períodos de teste. O método de aprendizagem *on-line* apresentou melhor desempenho que o algoritmo *off-line*, porém o número de regras nebulosas codificadas na estrutura da rede foi maior.

No próximo capítulo, os erros de previsão serão analisados, sendo também realizada uma análise comparativa entre todos os métodos estudados neste trabalho.

Capítulo 6

Comparação dos Modelos

Modelos de séries temporais, redes neurais artificiais e redes neurais nebulosas foram estudados neste trabalho e aplicados ao problema de previsão de vazões mensais. Nos capítulos anteriores, estes modelos foram ajustados aos dados de vazões afluentes naturais médias mensais do posto de Furnas.

Neste capítulo, será feita uma análise dos erros de previsão dos modelos ajustados para os postos de Furnas, Emborcação e Sobradinho e uma análise comparativa será realizada.

6.1 Introdução

Para efeito de estudo de casos, utilizaram-se os dados de duas usinas hidroelétricas Furnas e Emborcação, com reservatórios do sistema Sudeste brasileiro, localizadas na bacia do rio Paraná, e a usina hidroelétrica de Sobradinho, com reservatório do sistema Nordeste, localizada na bacia do rio São Francisco.

A usina hidroelétrica de Furnas está localizada no rio Grande, pertence à empresa FURNAS - Centrais Elétricas S.A.. Esta usina possui capacidade instalada igual à 1.312 MW e volume útil de 12.217 hm³. A usina hidroelétrica de Emborcação, localizada no rio Paranaíba, da empresa Centrais Elétricas de Minas Gerais - CEMIG, possui capacidade instalada de 1.192 MW e volume útil igual à 12.521 hm³. A usina hidroelétrica de Sobradinho, localizada no rio São Francisco, pertence à empresa Companhia HidroElétrica do São Francisco - CHESF, com capacidade instalada de 1.050 MW e volume útil igual à 28.669 hm³.

Na próxima seção, serão analisados os dados de vazões das usinas de Emborcação

e Sobradinho. Estes dados serão ajustados aos modelos auto-regressivos, auto-regressivos periódicos, rede neural multi-camadas e rede neural nebulosa. Para as três usinas e para todos os modelos, foram utilizados dados do histórico de 1931 a 1990.

Os dados de vazões da usina de Furnas foram analisados e ajustados ao longo deste trabalho. Para verificar o desempenho dos modelos foram escolhidos os períodos de 1952 a 1956, 1972 a 1976 e 1981 a 1985. Estes três períodos apresentam características distintas, sendo um período de seca, um período mediano e um período de cheia, respectivamente.

Para analisar o desempenho dos modelos para os postos de Emborcação e Sobradinho foi escolhido um período de teste de 60 meses, correspondendo aos anos de 1972 a 1976.

Também serão ajustados aos dados das usinas de Furnas e Emborcação os modelos de séries temporais empregados pelo setor elétrico brasileiro, para todos os aproveitamentos no programa da operação, conhecido como sistema PREVAZ. Este sistema é baseado na metodologia de Yevjevich - Box & Jenkins (Sales e Vieira, 1984).

6.2 Análise e Ajuste dos Dados de Vazões para Emborcação e Sobradinho

As figuras 6.1 e 6.2 mostram as séries de vazões afluentes médias mensais para os postos de Emborcação e Sobradinho, compreendendo o período de janeiro de 1931 a dezembro de 1990.

A tabela 6.1 mostra a média e o desvio padrão para os dois postos. Pode-se verificar que para a usina de Emborcação ocorre maior variação nos dados para o mês de fevereiro e menor variação para o mês de setembro. Para a usina de Sobradinho, o maior desvio padrão ocorre no mês de março, mas a média das vazões é maior no mês de fevereiro, já o mês de setembro apresenta menores desvio padrão e média. Isto ocorre devido às regiões onde estão localizadas estas usinas. Estes dados estão mostrados na figura 6.3. Pode-se notar um comportamento periódico para as duas séries, onde os valores oscilam entre um valor máximo e um valor mínimo.

Os dados de vazões $x_{i,m}$ denotam as séries de vazões sazonais, onde $i = 1, \dots, N$ representam os anos e $m = 1, 2, \dots, 12$ os meses. Estes dados foram padronizados como descrito no capítulo 2. As estimativas das funções de autocovariância e autocorrelação para as séries foram calculadas. Os valores dos coeficientes de autocorrelação r_k estão na figura 6.4. Observa-se que as estimativas das funções de autocorrelação são caracterizadas por

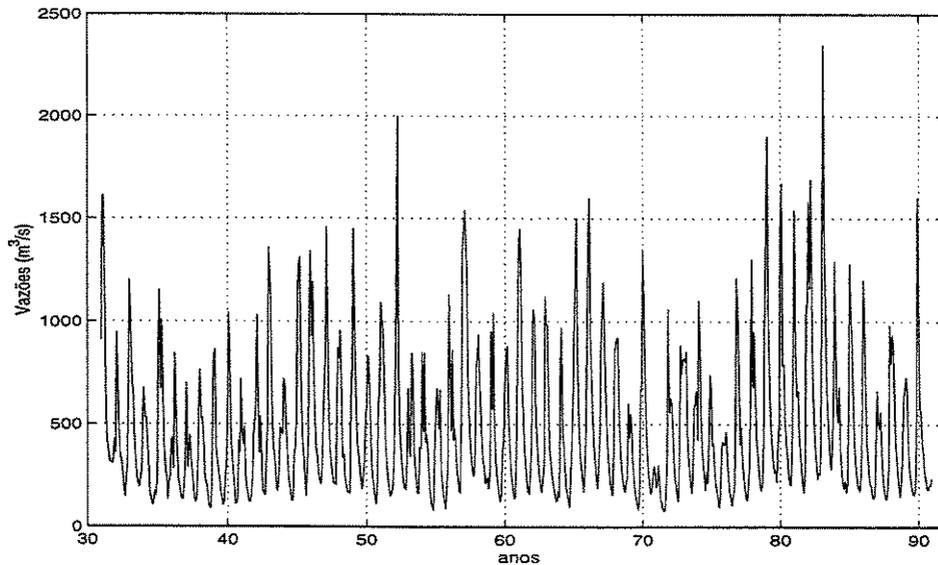


Figura 6.1: Série de Vazões Afuentes Médias Mensais para a Usina Hidroelétrica de Emborcação.

um decaimento nos valores de r_k quando o passo k aumenta. Também pode-se observar a variação no sinal da estimativa r_k para a série de vazões de Emborcação. Para os dados de vazões de Sobradinho, os coeficientes de autocorrelação assumem somente valores positivos.

Considerando os dois primeiros valores para as séries transformadas z_t e z_{t+1} foram construídos os diagramas de dispersão (figuras 6.5 e 6.6). Ou seja, tomando os valores de (z_t, z_{t+k}) para $k = 1$ e $k = 2$ observa-se a correlação entre os valores vizinhos, podendo, assim, ser considerada uma relação linear entre as observações. Pode-se notar que, para as duas séries, a correlação é mais acentuada quando $k = 1$ especialmente nos meses secos ($z(t) \leq 0$).

Os coeficientes de autocorrelação parcial foram ajustados para as séries de vazões de Emborcação e Sobradinho, como mostram as figuras 6.7(a) e 6.7(b), respectivamente. Através da análise dos diagramas de dispersão, das funções de autocorrelação e autocorrelação parcial um modelo auto-regressivo de ordem 1 (AR(1)) foi ajustado para as duas séries, podendo ser representado da seguinte forma:

$$z_t = \phi_1 z_{t-1} + a_t$$

onde $\phi_1 = \hat{\phi}_{11}$ é calculado como mostrado na subseção 3.3.2 e a_t é um ruído branco com distribuição $N(0, \sigma_a^2)$. As estimativas dos parâmetros para os dados de vazões da usina de Emborcação foi $\phi_1 = 0.6835$ e para os dados da usina de Sobradinho foi $\phi_1 = 0.7432$.

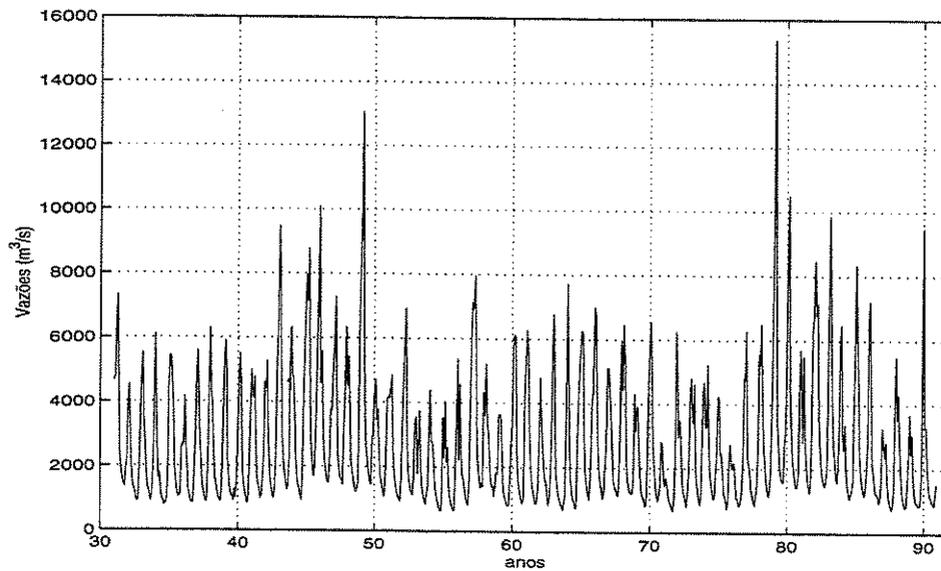


Figura 6.2: Série de Vazões Afluentes Médias Mensais para a Usina Hidroelétrica de Sobradinho.

Tabela 6.1: Média e Desvio Padrão para as Série de Vazões dos Postos de Emborcação e Sobradinho - m^3/s .

| Meses | Emborcação | | Sobradinho | |
|-----------|------------|---------------|------------|---------------|
| | Média | Desvio Padrão | Média | Desvio Padrão |
| Janeiro | 885.9 | 348.9 | 4858.9 | 1430.8 |
| Fevereiro | 907.0 | 413.1 | 5174.5 | 2062.7 |
| Março | 845.6 | 363.5 | 4937.5 | 2538.4 |
| Abril | 625.8 | 236.7 | 3973.1 | 1787.8 |
| Maio | 403.3 | 125.3 | 2485.2 | 1228.7 |
| Junho | 308.8 | 92.6 | 1697.5 | 559.4 |
| Julho | 244.2 | 70.9 | 1401.6 | 384.2 |
| Agosto | 189.9 | 57.2 | 1201.5 | 299.3 |
| Setembro | 161.8 | 49.9 | 1063.2 | 263.9 |
| Outubro | 206.9 | 67.9 | 1179.0 | 298.6 |
| Novembro | 356.7 | 165.9 | 1909.2 | 667.8 |
| Dezembro | 655.0 | 310.7 | 3477.9 | 1233.2 |

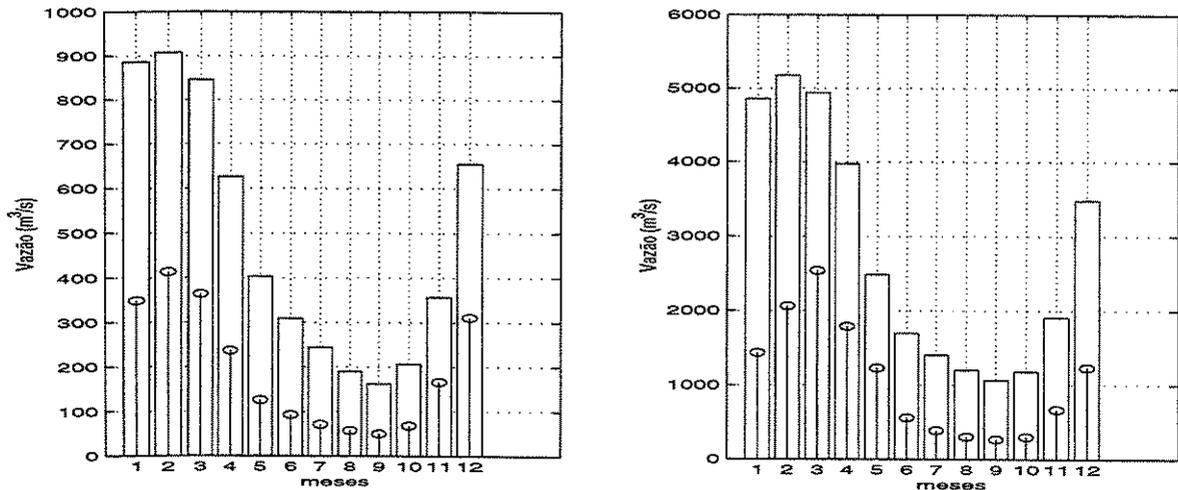


Figura 6.3: Média e Desvio Padrão Mensal para as Séries de Vazões dos Postos de Emborcação (a) e Sobradinho (b).

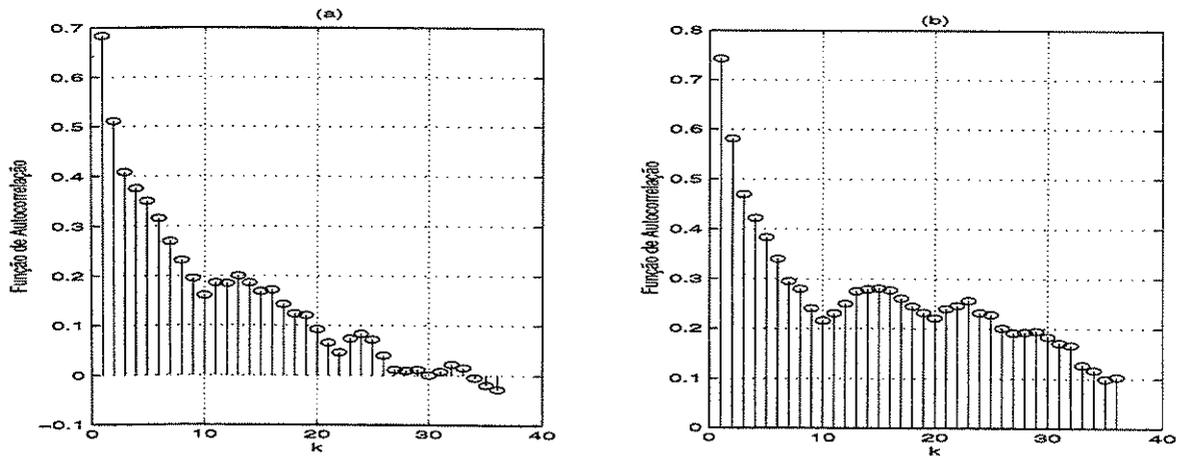


Figura 6.4: Função de Autocorrelação para as Séries de Vazões dos Postos de Emborcação (a) e Sobradinho (b).

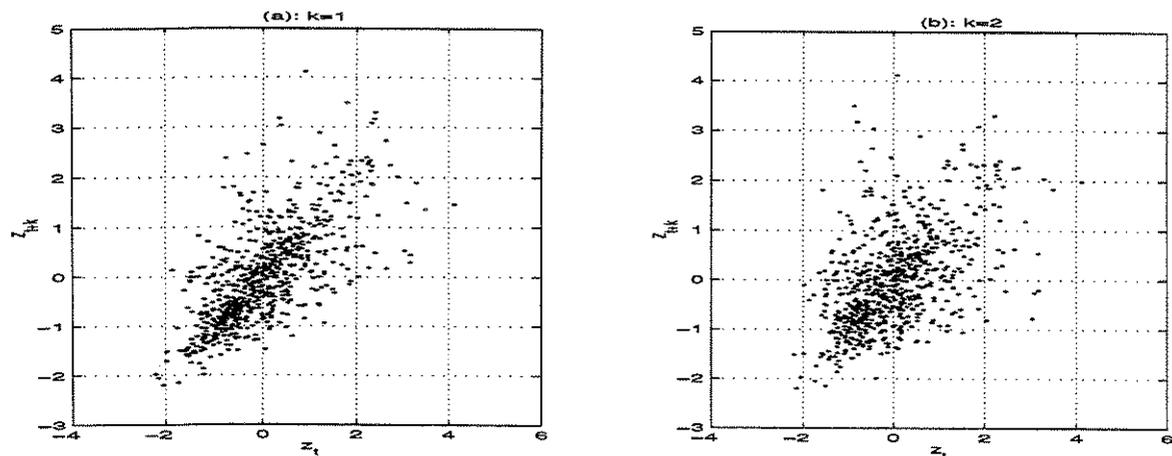


Figura 6.5: Diagramas de Dispersão para a Série de Vazões do Posto de Emborcação.

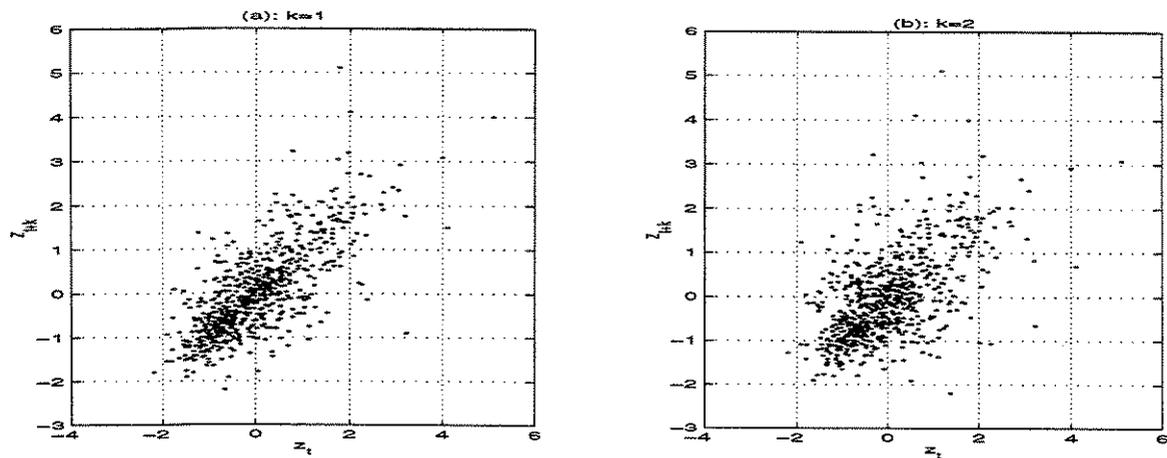


Figura 6.6: Diagramas de Dispersão para a Série de Vazões do Posto de Sobradinho.

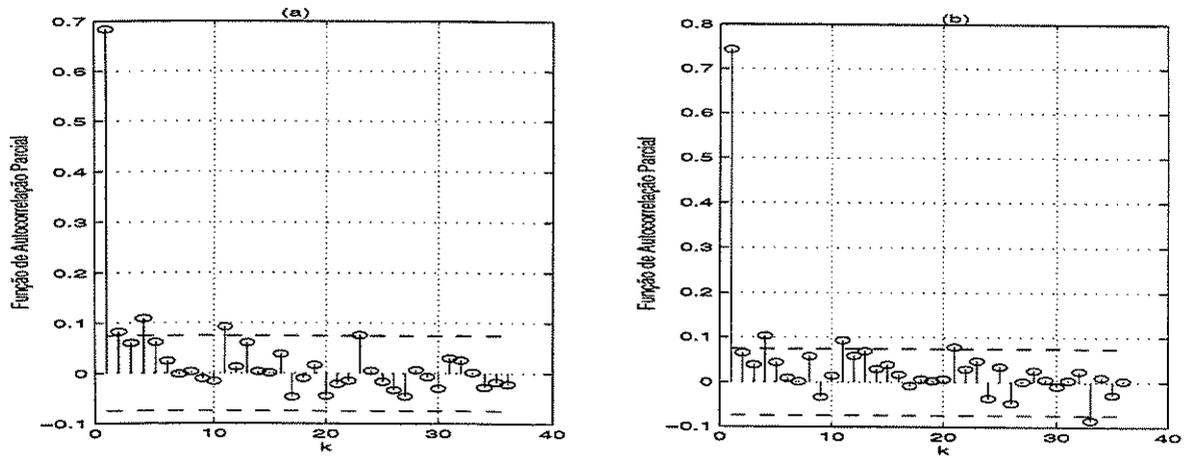


Figura 6.7: Função de Autocorrelação Parcial para as Séries de Vazões de Emborcação (a) e Sobradinho (b).

A partir das informações das funções de autocorrelação e autocorrelação parcial, um modelo auto-regressivo periódico de ordem 1 (PAR(1)) foi ajustado para as séries de vazões afluentes das usinas de Emborcação e Sobradinho. Este modelo pode ser escrito da seguinte forma:

$$z_{t(r, m)} = \phi_{1, m} z_{t(r, m)-1} + a_{t(r, m)}$$

onde $a_{t(r, m)}$ é um ruído branco com distribuição $N(0, \tau_m^{-1})$. Para inferir os parâmetros $\Phi_m = \phi_{1, m}$ e τ_m foi utilizada inferência Bayesiana, como descrito na seção 3.4. A tabela 6.2 apresenta os valores estimados dos parâmetros para as usinas de Emborcação e Sobradinho.

A seguir foram ajustados um modelo PAR(p_m) para cada mês, ou seja, a ordem de um modelo PAR(p_m) foi determinada para cada mês m do ano, analisando a densidade preditiva e escolhendo o modelo que maximiza o produto da densidade preditiva e apresente menor variância dos resíduos ($\sigma_m^2 = \tau_m^{-1}$), como apresentada na seção 3.4. As tabelas 6.3 e 6.4 mostraram a ordem dos modelos p_m e o inverso da variância τ_m para os postos de Emborcação e Sobradinho, respectivamente.

O próximo modelo ajustado foi a rede neural multi-camadas com algoritmo de retro-propagação (MLP). Este modelo foi ajustado utilizando seis entradas, dadas pelas vazões dos meses $t-1$, $t-2$, $t-3$, $t-11$, $t-12$, $t-13$ para as séries de vazões dos Postos de Emborcação e Sobradinho.

A escolha dos parâmetros livres da rede (pesos) foi feita a partir da técnica de validação cruzada, como descrito na subseção 4.3.7, sendo escolhido o modelo que apresentou

Tabela 6.2: Estimativas dos Parâmetros para as Séries de Vazões dos Postos de Emborcação e Sobradinho.

| Meses | Emborcação | | Sobradinho | |
|-----------|--------------|----------|--------------|----------|
| | $\phi_{1,m}$ | τ_m | $\phi_{1,m}$ | τ_m |
| Janeiro | 0.4975 | 1.1493 | 0.5395 | 1.8851 |
| Fevereiro | 0.6570 | 1.6343 | 0.4727 | 1.1450 |
| Março | 0.4996 | 1.2878 | 0.7006 | 1.8636 |
| Abril | 0.7083 | 1.9057 | 0.6791 | 1.6704 |
| Maiο | 0.9129 | 4.9930 | 0.8092 | 2.6307 |
| Junho | 0.8782 | 3.9923 | 0.9249 | 6.6250 |
| Julho | 0.9444 | 10.9005 | 0.9643 | 16.6390 |
| Agosto | 0.9601 | 9.9716 | 0.9807 | 20.9730 |
| Setembro | 0.8666 | 3.7966 | 0.9477 | 8.3830 |
| Outubro | 0.5057 | 1.1929 | 0.7934 | 2.5743 |
| Novembro | 0.4589 | 1.1442 | 0.6120 | 1.4846 |
| Dezembro | 0.5091 | 1.2770 | 0.6066 | 1.4822 |

melhor desempenho em relação ao erro quadrático médio. O critério de parada do algoritmo foi determinado através da técnica de validação cruzada.

Primeiramente, foi ajustado um único modelo de rede neural para realizar a previsão. Os parâmetros taxas de aprendizagem foram ajustados usando a regra delta-bar-delta (seção 4.4). Os parâmetros utilizados para ajuste da taxa de aprendizagem foram definidos com os mesmos valores para o modelo de ajuste da série de vazões da usina de Furnas, ou seja, $\xi = 0.01$, $\kappa = 0.7$ e $\beta = 0.2$. Uma rede com uma única camada foi ajustada e combinações do parâmetro termo *momentum* α e número de neurônios intermediários M foram feitas. O número de neurônios intermediários $M = 4$ e o valor do parâmetro $\alpha = 0.5$ foram ajustados para a série de vazões do posto de Emborcação, e para os dados de vazões do posto de Sobradinho foram ajustados $M = 5$ e $\alpha = 0.1$.

Também foram ajustados para a rede neural multi-camadas um modelo para cada mês, ou seja, foram ajustados 12 modelos MLP. A tabela 6.5 mostra o número de neurônios intermediários para cada mês e para cada um dos postos. Os parâmetros taxa de aprendizagem η foram obtidos pela regra delta-bar-delta. O termo *momentum* foi fixado para os dados de vazões do posto de Emborcação $\alpha = 0.5$ e para a série de vazões do posto de Sobradinho $\alpha = 0.1$.

Tabela 6.3: Estimativas dos Parâmetros e Ordem dos Modelos Selecionados para a Série de Vazões do Posto de Emborcação.

| Meses | $\phi_{1,m}$ | $\phi_{2,m}$ | $\phi_{3,m}$ | $\phi_{4,m}$ | τ_m | p_m |
|-----------|--------------|--------------|--------------|--------------|----------|-------|
| Janeiro | 0.4975 | - | - | - | 1.1493 | 1 |
| Fevereiro | 0.6570 | - | - | - | 1.6343 | 1 |
| Março | 0.5337 | -0.0136 | - | - | 1.2970 | 2 |
| Abril | 0.5780 | 0.2162 | - | - | 2.0440 | 2 |
| Maió | 0.6735 | 0.1405 | 0.1598 | 0.1187 | 8.2406 | 4 |
| Junho | 0.8244 | -0.0828 | 0.1720 | - | 4.3154 | 3 |
| Julho | 0.6738 | 0.3349 | - | - | 14.8907 | 2 |
| Agosto | 0.9601 | - | - | - | 9.9716 | 1 |
| Setembro | 0.8666 | - | - | - | 3.7966 | 1 |
| Outubro | 0.5057 | - | - | - | 1.1929 | 1 |
| Novembro | 0.4589 | - | - | - | 1.1442 | 1 |
| Dezembro | 0.5091 | - | - | - | 1.2770 | 1 |

O último modelo a ser ajustado para as séries de vazões afluentes médias mensais das usinas de Emborcação e Sobradinho foi a rede neural nebulosa (RNN) com os algoritmos *off-line* e *on-line*.

O número de entradas usado para ajuste do modelo RNN para as duas séries foi igual ao número de entradas usado na rede MLP, ou seja, foram utilizadas seis entradas dadas pelas vazões dos meses $t - 1$, $t - 2$, $t - 3$, $t - 11$, $t - 12$, $t - 13$.

As funções de pertinência das regras nebulosas codificadas na estrutura da RNN foram definidas como sendo a função Gaussiana (equação (5.66)).

O parâmetro taxa de aprendizagem utilizado para ajuste dos pesos w_i foi fixado em $\eta = 0.5$ e o parâmetro usado para atualização do fator de aprendizagem $\beta^i(t)$ foi $\epsilon = 1.0$. As normas usadas pelos operadores sinápticos da segunda camada foram escolhidas como a *t-norma* mínima e a *s-norma* máxima. Os neurônios da terceira camada possuem como operador de agregação a *t-norma* produto. Estes parâmetros e normas foram definidos iguais para o ajuste do modelo as duas séries de vazões.

O número de regras codificadas na estrutura da rede, quando utilizado o algoritmo *off-line* para a usina de Emborcação, foi escolhido $N = 12$. Para o posto de Sobradinho, o número de regras ajustadas foi $N = 18$. O fator de ampliação da dispersão das regras para os dados de vazões dos dois postos foi fixado em $\lambda = 1.1$.

Tabela 6.4: Estimativas dos Parâmetros e Ordem dos Modelos Seleccionados para a Série de Vazões do Posto de Sobradinho.

| Meses | $\phi_{1,m}$ | $\phi_{2,m}$ | $\phi_{3,m}$ | τ_m | p_m |
|-----------|--------------|--------------|--------------|----------|-------|
| Janeiro | 0.5395 | - | - | 1.8851 | 1 |
| Fevereiro | 0.4727 | - | - | 1.1450 | 1 |
| Março | 0.4996 | - | - | 1.8636 | 1 |
| Abril | 0.6711 | 0.0205 | - | 1.6734 | 2 |
| Maio | 0.8041 | 0.0065 | - | 2.5681 | 2 |
| Junho | 0.6573 | 0.1847 | 0.2407 | 13.0880 | 3 |
| Julho | 0.9643 | - | - | 16.6390 | 1 |
| Agosto | 0.9601 | - | - | 20.9730 | 1 |
| Setembro | 0.8666 | - | - | 8.3830 | 1 |
| Outubro | 0.7934 | - | - | 2.5743 | 1 |
| Novembro | 0.7844 | -0.2055 | - | 1.5218 | 2 |
| Dezembro | 0.6066 | - | - | 1.4822 | 1 |

O próximo passo foi ajustar o modelo RNN com o método de aprendizagem *on-line*. Neste método, a rede adiciona neurónios sempre que o desempenho não é alcançado ou para uma específica entrada não existe nenhuma regra nebulosa ativa. A inicialização da dispersão das funções de pertinência foi definida como $\Theta = 2.0$, o fator de redução usado para ajustar a dispersão das regras foi $\gamma = 0.9$ e o desempenho $\delta = 0.09$. O número de regras inicial foi $N(0) = 1$. Estes parâmetros foram inicializados iguais para o ajuste das séries de vazões dos dois postos.

No final do treinamento, o número de regras ajustadas foi 32 para os dados de vazões afluentes do posto de Emborcação e para a série de vazões do posto de Sobradinho foram ajustadas 37 regras.

Previsões 1, 3, 6 e 12 passos à frente foram realizadas para o período de teste de 1972 a 1976, para todos os modelos. O desempenho destes modelos para as séries de vazões dos postos de Emborcação e Sobradinho são avaliados nas seções 6.4.2 e 6.4.3 através da análise dos erros de previsão.

Tabela 6.5: Topologia da Rede MLP para cada Mês para os Postos de Emborcação e Sobradinho - Período de 1972 - 1976.

| Meses | Emborcação | Sobradinho |
|-----------|------------|------------|
| Janeiro | 5 | 3 |
| Fevereiro | 3 | 3 |
| Março | 3 | 3 |
| Abril | 3 | 3 |
| Maio | 4 | 4 |
| Junho | 3 | 5 |
| Julho | 3 | 3 |
| Agosto | 3 | 5 |
| Setembro | 5 | 5 |
| Outubro | 4 | 3 |
| Novembro | 5 | 3 |
| Dezembro | 3 | 5 |

6.3 Ajuste do Modelo PREVAZ

O sistema PREVAZ consiste em modelos de previsão de vazões médias mensais utilizados rotineiramente para a elaboração do programa mensal de operação do sistema interligado brasileiro para todos os aproveitamentos considerados pelo setor elétrico brasileiro (GCOI, 1983). O processo de seleção dos modelos de previsão é baseado na metodologia de Yevjevich - Box & Jenkins (Sales e Vieira, 1984).

O sistema PREVAZ é composto por cinco programas computacionais que visam fornecer modelos de previsão univariada para as séries de vazões médias mensais. Devido à característica univariada do modelo, são selecionados para modelagem apenas as séries dos postos considerados principais e mais representativos nas bacias a que pertencem. Estes postos são identificados como postos base e para os quais é ajustado um modelo Auto-Regressivo Média Móvel (ARMA). As previsões para os postos não-básicos são obtidas através das correlações com os postos básicos relacionados.

Estes modelos foram ajustados às séries de vazões afluentes das usinas de Furnas e Emborcação. Não foi possível realizar o ajuste destes modelos a séries de vazões do posto de Sobradinho devido à falta de informações sobre a estrutura da série e os componentes

estocásticos dependentes. Estes arquivos são incorporados aos modelos para realizar as previsões.

Para a série de Furnas, foram realizadas previsões 1, 3, 6 e 12 passos à frente para períodos de teste: 1952 a 1956 (período seco); 1972 a 1976 (período mediano); e 1981 a 1985 (período úmido), como mostram as figuras 6.8, 6.9 e 6.10 em (a), (b), (c) e (d), respectivamente.

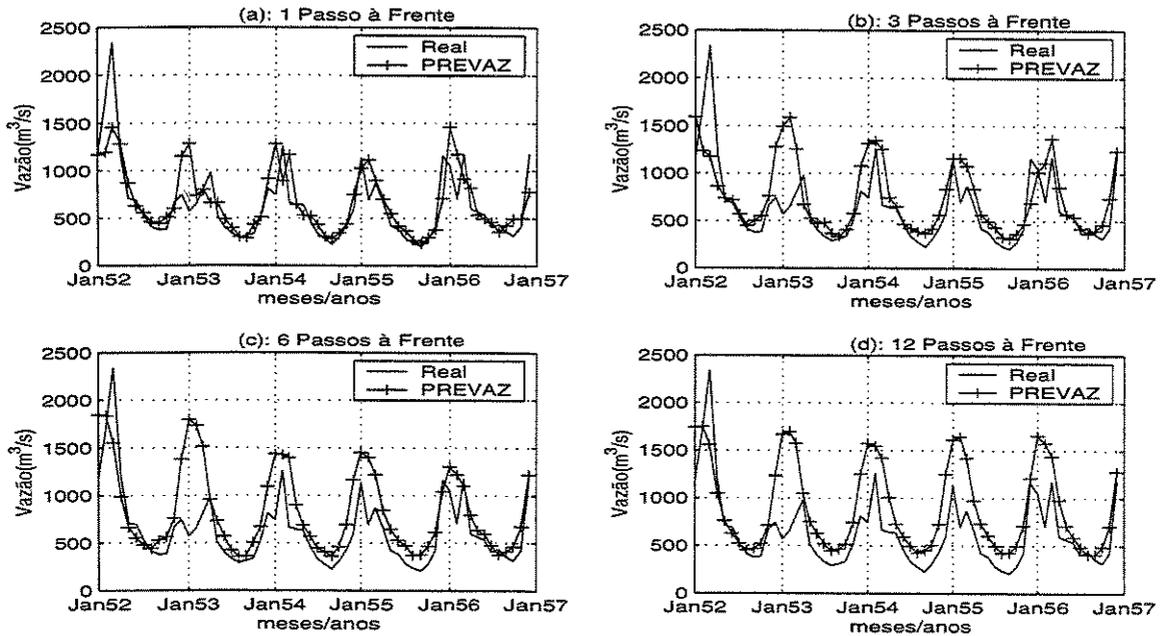


Figura 6.8: Previsão de Vazões Usando o Modelo PREVAZ para o Posto de Furnas - Período de 1952 a 1956.

O período de teste para a série de vazões afluentes médias mensais para o posto de Emborcação foi de 1972 a 1976 e também foram realizadas previsões 1, 3, 6 e 12 passos à frente. A análise dos erros de previsão com este modelo será realizada na próxima seção.

Analisando as figuras 6.8(a), 6.9(a) e 6.10(a) nota-se que a previsão 1 passo à frente apresentou melhor desempenho para os três períodos de teste. Também pode-se observar que para as previsões vários passos à frente o comportamento da série começa a se repetir, notando-se uma tendência à média das vazões mensais (Média a Longo Termo - MLT), como pode ser observado na tabela 6.6, onde são dadas as vazões médias mensais e o valor previsto para 6 passos à frente para os meses de Janeiro a Dezembro dos anos de 1952, 1972 e 1981.

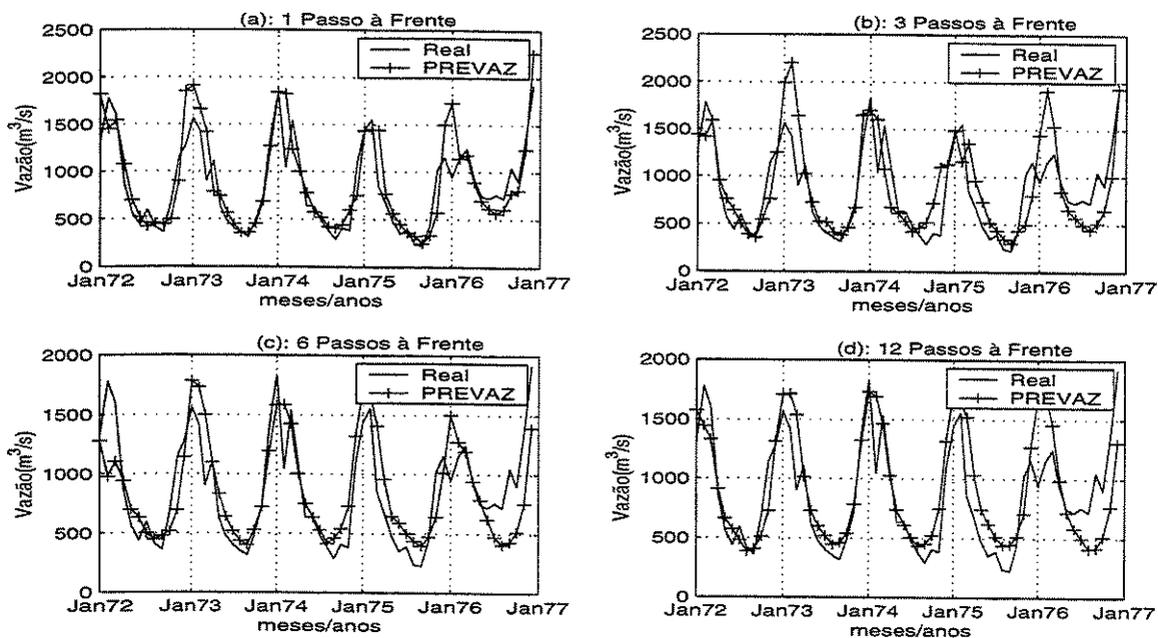


Figura 6.9: Previsão de Vazões Usando o Modelo PREVAZ para o Posto de Furnas - Período de 1972 a 1976.

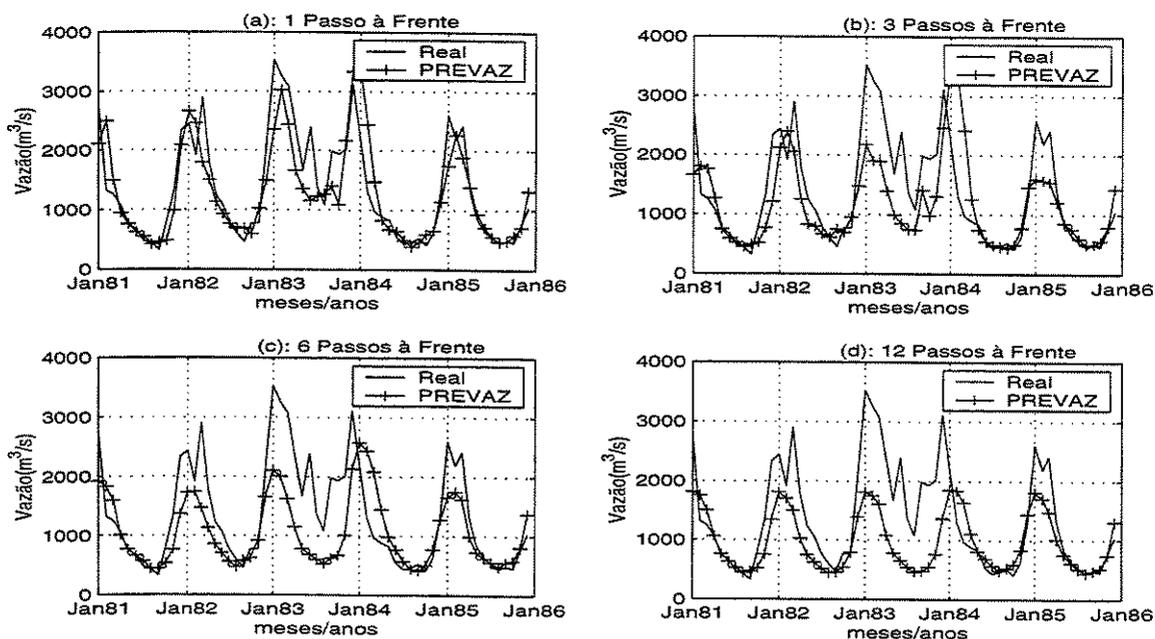


Figura 6.10: Previsão de Vazões Usando o Modelo PREVAZ para o Posto de Furnas - Período de 1981 a 1985.

Tabela 6.6: Médias Mensais e Valores Previstos com o Modelo PREVAZ para a Série de Vazões de Furnas, Previsão 6 Passos à Frente - m^3/s .

| Meses | MLT | 1952 | 1972 | 1981 |
|-----------|--------|--------|--------|--------|
| Janeiro | 1719.3 | 1843.0 | 1280.0 | 1925.0 |
| Fevereiro | 1669.2 | 1835.0 | 976.0 | 1827.0 |
| Março | 1518.0 | 1553.0 | 1106.0 | 1599.0 |
| Abril | 1019.6 | 988.0 | 940.0 | 1008.0 |
| Mai | 743.5 | 662.0 | 699.0 | 770.0 |
| Junho | 623.0 | 553.0 | 633.0 | 662.0 |
| Julho | 515.5 | 479.0 | 501.0 | 571.0 |
| Agosto | 431.2 | 439.0 | 448.0 | 442.0 |
| Setembro | 447.5 | 531.0 | 478.0 | 443.0 |
| Outubro | 532.4 | 564.0 | 516.0 | 534.0 |
| Novembro | 747.8 | 763.0 | 696.0 | 763.0 |
| Dezembro | 1268.2 | 1381.0 | 1142.0 | 1367.0 |

6.4 Análise dos Erros

Para avaliar o desempenho dos modelos estudados neste trabalho são analisados os erros quadrático médio (EQM), absoluto médio (EAM), relativo percentual médio (EPM) e relativo percentual máximo (EP_{max}), definidos como segue:

$$EQM = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2 \quad (6.1)$$

$$EAM = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (6.2)$$

$$EPM = \frac{100}{n} \sum_{j=1}^n \frac{|y_j - \hat{y}_j|}{y_j} \quad (6.3)$$

$$EP_{max} = \max 100 \left| \frac{y_j - \hat{y}_j}{y_j} \right| \quad (6.4)$$

onde $n = 1, \dots, 60$, y_j é o valor real e \hat{y}_j é o valor previsto.

6.4.1 Análise dos Erros de Previsão para o Posto de Furnas

Nesta seção, são analisados os erros de previsão dos modelos auto-regressivo (AR), periódico auto-regressivo (PAR(p) e PAR(p_m)), modelo PREVAZ, rede multi-camadas (MLP e MLP(m)) e rede neural nebulosa (RNN) com os algoritmos *off-line* e *on-line*.

Estes modelos foram ajustados para a série de vazões do posto de Furnas no decorrer deste trabalho. Os erros para o período de 1952 a 1956 para todos os modelos são mostrados nas tabelas 6.7, 6.8, 6.9 e 6.10, para previsão 1, 3, 6, e 12 passos à frente, respectivamente.

Tabela 6.7: Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 1 Passo à Frente, Período de 1952 a 1956.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(2) | 7.16×10^4 | 186.42 | 30.92 | 131.78 |
| PAR(2) | 6.02×10^4 | 164.01 | 24.76 | 115.10 |
| PAR(p_m) | 6.77×10^4 | 174.23 | 26.23 | 115.10 |
| PREVAZ | 6.40×10^4 | 163.45 | 23.59 | 125.39 |
| MLP | 6.57×10^4 | 168.36 | 22.28 | 72.00 |
| MLP(m) | 5.54×10^4 | 151.54 | 21.49 | 74.04 |
| RNN - <i>Off-line</i> | 1.47×10^4 | 91.45 | 14.98 | 50.56 |
| RNN - <i>On-line</i> | 1.19×10^4 | 73.67 | 11.48 | 57.54 |

Analisando os erros de previsão 1 passo à frente para o período de 1952 a 1956 (tabela 6.7), pode-se notar que, para os modelos de séries temporais, o EQM apresentou maior variação para o modelo AR(2) sendo que, para modelos PAR(2), PAR(p_m) e PREVAZ, a variação deste erro foi menor, mostrando, assim, que para este período os modelos periódicos auto-regressivos (PAR(2) e PAR(p_m)) e o modelo PREVAZ acompanham melhor a tendência da série. Para os EAM e EPM o modelo de séries temporais que apresentou melhor desempenho foi o modelo PREVAZ. Ou seja, este modelo “acerta” melhor os pontos da série. Já os EP_{max} foram menores para os modelos PAR(2) e PAR(p_m).

Para a rede neural o melhor desempenho para este período e para previsão 1 passo à frente foi obtido através do ajuste de 12 modelos (MLP(m)) ao invés de um único (MLP) (tabela 6.7). Neste caso, apenas o EP_{max} foi ligeiramente menor para o ajuste de um único modelo MLP.

Tabela 6.8: Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 3 Passos à Frente, Período de 1952 a 1956.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(2) | 1.24×10^5 | 251.67 | 41.89 | 164.48 |
| PAR(2) | 1.05×10^5 | 208.94 | 31.65 | 159.52 |
| PAR(p_m) | 8.63×10^4 | 183.35 | 28.34 | 133.09 |
| PREVAZ | 1.03×10^5 | 210.25 | 33.27 | 162.00 |
| MLP | 9.86×10^4 | 230.41 | 36.71 | 116.59 |
| MLP(m) | 9.78×10^4 | 186.59 | 27.17 | 169.57 |
| RNN - <i>Off-line</i> | 2.28×10^4 | 124.91 | 19.53 | 59.51 |
| RNN - <i>On-line</i> | 1.20×10^4 | 101.28 | 16.55 | 83.61 |

A RNN com algoritmo *on-line* apresentou melhor desempenho que a RNN com o algoritmo *off-line*, com exceção para o EP_{max} . Estes modelos apresentaram erros significativamente menores que os demais modelos.

Considerando as previsões 3 passos à frente (tabela 6.8), o modelo PAR(p_m) apresentou melhor desempenho em relação aos modelos de séries temporais, para todos os erros analisados. O ajuste de 12 modelos MLP(m) apresentou melhores resultados que o ajuste de um único modelo MLP. Apenas para o EP_{max} o ajuste de um único modelo MLP forneceu um erro menor.

A RNN com algoritmo *off-line* apresentou EP_{max} menor que o algoritmo *on-line* e para as outras estatísticas o modelo RNN com algoritmo *on-line* apresentou melhor desempenho, repetindo o que havia ocorrido para previsão 1 passo à frente.

Para a previsão 6 passos à frente (tabela 6.9) o modelo PAR(p_m) apresentou erros menores com exceção ao EP_{max} , onde o modelo AR(2) obteve menor erro. O ajuste de 12 modelos da rede neural MLP(m) apresentou melhor resultado que o ajuste de um único modelo para todas as estatísticas. A RNN com algoritmo *on-line* apresentou um desempenho melhor que a RNN com algoritmo *off-line* para todos os erros analisados.

O ajuste de um modelo PAR(2) para previsão 12 passos à frente para o período de 1952 a 1956 (tabela 6.10) apresentou erros menores que os outros modelos de séries temporais, porém a variação dos erros é pequena, com exceção para o modelo PREVAZ.

A rede neural apresentou melhor desempenho quando ajustados 12 modelos

Tabela 6.9: Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 6 Passos à Frente, Período de 1952 a 1956.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(2) | 1.44×10^5 | 273.37 | 46.29 | 177.09 |
| PAR(2) | 1.24×10^5 | 239.93 | 37.98 | 179.24 |
| PAR(p_m) | 9.91×10^4 | 211.99 | 35.46 | 181.95 |
| PREVAZ | 1.34×10^5 | 256.65 | 44.09 | 214.71 |
| MLP | 1.23×10^5 | 223.01 | 33.09 | 180.99 |
| MLP(m) | 8.69×10^4 | 198.66 | 29.71 | 131.13 |
| RNN - <i>Off-line</i> | 6.40×10^4 | 154.70 | 23.08 | 87.01 |
| RNN - <i>On-line</i> | 3.36×10^4 | 132.69 | 21.16 | 86.75 |

MLP(m) para todas as estatísticas analisadas.

A RNN com algoritmo *on-line* apresentou EPM e EP_{max} menores e o ajuste do modelo com algoritmo *off-line* apresentou EQM e EAM menores. Estes modelos apresentaram menores erros tanto para previsão 1 passo à frente como para previsão vários passos à frente.

Comparando todos os modelos ajustados aos dados para prever o período de 1952 a 1956, pode-se observar que o modelo PAR(p_m) apresentou, em geral, erros menores que os outros modelos de séries temporais. O ajuste de 12 modelos de redes neurais (MLP(m)) apresentou melhor desempenho tanto para previsão 1 passo à frente como para previsão vários passos à frente. A rede neural nebulosa, por sua vez, apresentou erros menores, independente do algoritmo utilizado em relação aos demais modelos ajustados.

A figura 6.11 mostra os erros obtidos para as previsões 1, 3, 6 e 12 passos à frente usando as estatísticas EQM , EAM , EPM , e EP_{max} , para os modelos PAR(p_m), MLP(m), RNN com os algoritmos *off-line* e *on-line*. Pode-se notar que conforme o horizonte de previsão aumenta os erros também aumentam. Nota-se que a RNN tanto com o algoritmo *off-line* como com o algoritmo *on-line* apresentou erros significativamente menores, independente da estatística usada e do número de passos à frente. Pode-se observar que o erro EPM para previsão 1 passo à frente, o modelo RNN apresentou um erro 50% menor em relação aos demais modelos ajustados.

Tabela 6.10: Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 12 Passos à Frente, Período de 1952 a 1956.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(2) | 1.51×10^5 | 278.73 | 46.91 | 176.03 |
| PAR(2) | 1.29×10^5 | 257.38 | 43.60 | 166.33 |
| PAR(p_m) | 1.34×10^5 | 261.06 | 44.21 | 167.67 |
| PREVAZ | 1.70×10^5 | 306.37 | 53.52 | 192.47 |
| MLP | 1.18×10^5 | 219.57 | 32.92 | 179.14 |
| MLP(m) | 9.21×10^4 | 203.45 | 30.63 | 136.83 |
| RNN - <i>Off-line</i> | 6.32×10^4 | 164.14 | 25.06 | 95.08 |
| RNN - <i>On-line</i> | 7.15×10^4 | 165.17 | 23.58 | 93.67 |

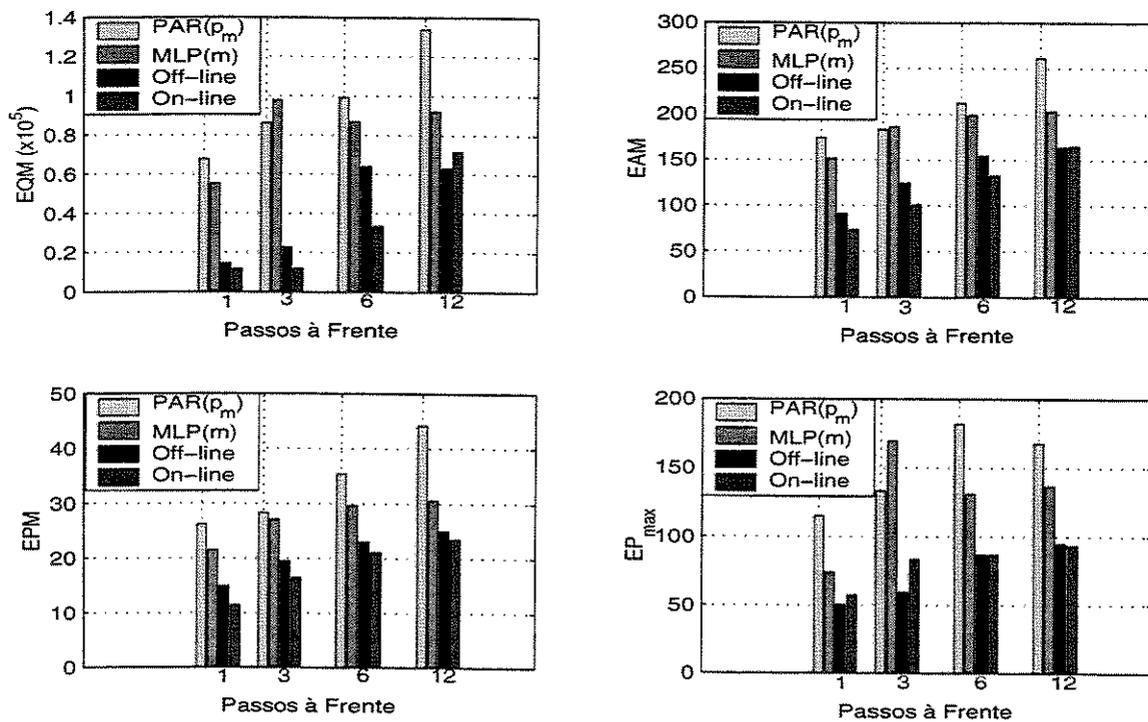


Figura 6.11: Erros de Previsão de Vazões para o Posto de Furnas - Período de 1952 a 1956.

As estatísticas para o período de 1972 a 1976 para todos os modelos são mostradas nas tabelas 6.11, 6.12, 6.13 e 6.14, para previsão 1, 3, 6, e 12 passos à frente, respectivamente.

Tabela 6.11: Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 1 Passo à Frente, Período de 1972 a 1976.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(2) | 5.49×10^4 | 176.00 | 21.39 | 76.64 |
| PAR(2) | 5.18×10^4 | 163.47 | 18.78 | 66.70 |
| PAR(p_m) | 5.47×10^4 | 170.18 | 19.86 | 66.70 |
| PREVAZ | 6.69×10^4 | 181.70 | 21.47 | 83.05 |
| MLP | 7.83×10^4 | 203.74 | 22.54 | 69.39 |
| MLP(m) | 6.26×10^4 | 187.62 | 21.79 | 63.64 |
| RNN - <i>Off-line</i> | 4.34×10^4 | 103.98 | 14.11 | 58.43 |
| RNN - <i>On-line</i> | 1.19×10^4 | 83.56 | 10.96 | 55.39 |

Analisando os modelos de séries temporais, pode-se notar que para a previsão 1 passo à frente para o período de 1972 a 1976 (tabela 6.11), o modelo PAR(2) apresentou erros menores. Para a rede neural os melhores resultados foram obtidos com o ajuste de 12 modelos (MLP(m)). A RNN com algoritmo *on-line* apresentou erros menores em relação ao algoritmo *off-line*.

Para as previsões 3 passos à frente (tabelas 6.12) pode-se observar que o modelo PAR(2) apresentou erros menores, com exceção do EP_{max} , onde o modelo PAR(p_m) obteve erro menor. A rede neural apresentou pequena variação no erro EQM tanto para o ajuste de um único modelo (MLP) como para 12 modelos ajustados (MLP(m)). Os erros EAM e EPM apresentaram melhor desempenho quando ajustado um único modelo. A RNN com algoritmo *on-line* apresentou erros menores em relação ao algoritmo *off-line*.

Para as previsões 6 e 12 passos à frente (tabelas 6.13 e 6.14), a variação nos erros para os modelos de séries temporais foi relativamente pequena, com exceção do modelo PREVAZ. O ajuste de um único modelo (MLP) para a rede neural apresentou melhor desempenho que o ajuste de 12 modelos (MLP(m)). Analisando os resultados obtidos pela rede RNN com os dois algoritmos de treinamento, pode-se observar que o método de aprendizagem *on-line* apresentou melhor desempenho. Comparando o desempenho de todos os modelos, pode-se observar que a rede neural nebulosa apresenta erros menores em relação aos outros modelos.

Tabela 6.12: Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 3 Passos à Frente, Período de 1972 a 1976.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| Modelos | EQM | EAM | EPM | EP_{max} |
| AR(2) | 7.01×10^4 | 197.87 | 25.16 | 83.67 |
| PAR(2) | 6.24×10^4 | 186.43 | 23.02 | 77.71 |
| PAR(p_m) | 7.99×10^4 | 207.89 | 24.66 | 73.83 |
| PREVAZ | 9.73×10^4 | 231.77 | 30.03 | 190.26 |
| MLP | 1.26×10^5 | 256.76 | 25.47 | 93.76 |
| MLP(m) | 1.23×10^5 | 258.27 | 27.44 | 70.54 |
| RNN - <i>Off-line</i> | 6.28×10^4 | 110.42 | 18.12 | 64.27 |
| RNN - <i>On-line</i> | 4.16×10^4 | 91.66 | 14.63 | 58.81 |

Como o período de 1972 a 1976 é um período mediano, a variação nos dados de previsão é menor e, portanto, o ajuste um único modelo torna-se mais adequado.

Isto pode ser notado nas tabelas 6.11, 6.12, 6.13 e 6.14, onde observa-se que a variação nos erros de previsão tanto para o ajuste de um modelo MLP como para o ajuste de 12 modelos (MLP(m)) é pequena. Neste caso, o ajuste de um único modelo é mais adequado já que o esforço computacional é menor. A variação nos erros para os modelos de séries temporais não é significativa, com exceção para o modelo PREVAZ, onde pode-se observar que para previsão vários passos à frente os erros são maiores. Assim, o ajuste de um modelo PAR(2) é mais adequado para prever este período.

A figura 6.12 mostra os erros de previsão para os modelos PAR(2), MLP e RNN com os algoritmos *off-line* e *on-line* para as quatro estatísticas analisadas. Pode-se observar que o modelo RNN apresentou melhores resultados com erros menores independente do horizonte de previsão.

Tabela 6.13: Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 6 Passos à Frente, Período de 1972 a 1976.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(2) | 7.43×10^4 | 201.90 | 26.28 | 88.41 |
| PAR(2) | 7.59×10^4 | 204.07 | 25.41 | 84.14 |
| PAR(p_m) | 7.44×10^4 | 202.15 | 24.87 | 77.59 |
| PREVAZ | 8.32×10^4 | 218.95 | 28.59 | 92.37 |
| MLP | 1.06×10^5 | 247.01 | 26.98 | 73.87 |
| MLP(m) | 1.68×10^5 | 296.10 | 30.58 | 93.84 |
| RNN - <i>Off-line</i> | 6.77×10^4 | 171.07 | 20.17 | 70.84 |
| RNN - <i>On-line</i> | 5.30×10^4 | 121.09 | 16.49 | 64.45 |

Tabela 6.14: Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 12 Passos à Frente, Período de 1972 a 1976.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(2) | 7.72×10^4 | 203.82 | 26.68 | 90.50 |
| PAR(2) | 7.76×10^4 | 203.50 | 26.27 | 89.38 |
| PAR(p_m) | 7.74×10^4 | 203.20 | 26.34 | 90.06 |
| PREVAZ | 8.71×10^4 | 223.98 | 30.19 | 100.45 |
| MLP | 1.44×10^5 | 281.00 | 30.07 | 83.21 |
| MLP(m) | 1.68×10^5 | 298.27 | 30.98 | 90.84 |
| RNN - <i>Off-line</i> | 6.91×10^4 | 179.08 | 21.43 | 71.02 |
| RNN - <i>On-line</i> | 5.79×10^4 | 124.09 | 18.49 | 66.91 |

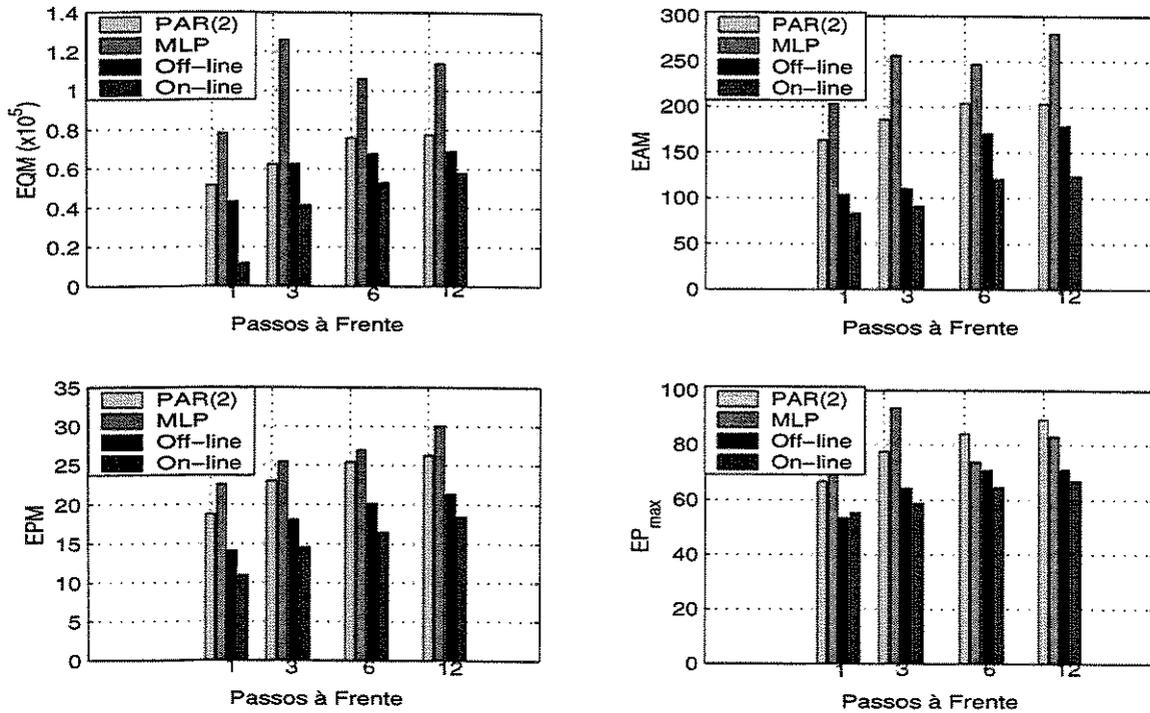


Figura 6.12: Erros de Previsão de Vazões para o Posto de Furnas - Período de 1972 a 1976.

Os erros obtidos para o período de 1981 a 1985 para todos os modelos ajustados, para previsão 1, 3, 6, e 12 passos à frente, são mostrados nas tabelas 6.15, 6.16, 6.17 e 6.19, respectivamente.

Analisando a tabela 6.15 pode-se observar que o modelo PAR(2) apresenta EQM menor e os erros EAM , EPM e EP_{max} apresentam pequena diferença em relação ao modelo AR(2). Analisando os modelos de redes neurais, nota-se que a variação nos erros é pequena, sendo que o ajuste de 12 modelos apresentou erros ligeiramente menores em relação ao ajuste de um único modelo. Para o modelo RNN o algoritmo *on-line* apresenta um melhor desempenho que o algoritmo *off-line*, porém os erros EQM e EAM apresentam pequena variação em relação ao algoritmo *off-line*.

Para as previsões 3 passos à frente (tabela 6.16), o modelo PAR(2) apresentou erros menores quando comparado com os modelos de séries temporais. A rede neural tem um melhor desempenho quando ajustados 12 modelos (MLP(m)) em relação ao ajuste de um único modelo, com exceção do EP_{max} . O modelo RNN com algoritmo *on-line* apresentou melhor desempenho que o algoritmo *off-line*. O EQM , neste caso, apresenta pouca variação para os dois métodos de aprendizagem.

Tabela 6.15: Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 1 Passo à Frente, Período de 1981 a 1985.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(2) | 2.07×10^5 | 289.81 | 20.21 | 89.10 |
| PAR(2) | 1.84×10^5 | 292.13 | 20.59 | 93.52 |
| PAR(p_m) | 2.15×10^5 | 302.11 | 21.28 | 107.16 |
| PREVAZ | 2.31×10^5 | 326.30 | 22.01 | 88.02 |
| MLP | 3.91×10^5 | 434.74 | 30.28 | 151.19 |
| MLP(m) | 3.73×10^5 | 429.51 | 30.56 | 115.53 |
| RNN - <i>Off-line</i> | 6.25×10^4 | 177.38 | 17.04 | 93.14 |
| RNN - <i>On-line</i> | 6.30×10^4 | 172.91 | 14.63 | 55.36 |

As previsões 6 e 12 passos à frente (tabelas 6.17 e 6.19), apresentaram um comportamento semelhante. Nestes casos, o modelo PREVAZ apresenta um melhor desempenho em relação aos modelos de séries temporais. O ajuste de 12 modelos para a rede neural (MLP(m)) apresentou erros EQM , EAM e EPM menores com exceção do EP_{max} , onde o ajuste de uma única rede neural forneceu melhor resultado. A RNN com algoritmo *off-line* apresenta EP_{max} menor e, em relação as demais estatísticas, a variação nos erros é pequena.

Comparando todos os modelos, pode-se observar que, para as previsões 1 e 3 passos à frente, o modelo PAR(2) apresentou melhor resultado em relação aos modelos de séries temporais. O EPM para previsão 3 passos à frente foi o mesmo que o modelo PREVAZ. Para as previsões 6 e 12 passos à frente o modelo PREVAZ apresentou melhor desempenho. Para este período o ajuste de 12 modelos (MLP(m)) de redes neurais apresentou melhor resultado que o ajuste de um único modelo (MLP). A RNN apresentou melhores resultados que os outros modelos, sendo que a variação nos erros em relação aos algoritmos *off-line* e *on-line* é pequena.

A figura 6.13 mostra as quatro estatísticas para este período em relação aos modelos PREVAZ, MLP(m) e RNN com os algoritmos *off-line* e *on-line*. Pode-se notar que os menores erros são obtidos quando ajustada a rede neural nebulosa, tanto com o algoritmo *off-line* como com o algoritmo *on-line*.

Tabela 6.16: Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 3 Passos à Frente, Período de 1981 a 1985.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(2) | 5.09×10^5 | 467.28 | 30.63 | 215.42 |
| PAR(2) | 3.67×10^5 | 421.74 | 27.97 | 170.12 |
| PAR(p_m) | 5.11×10^5 | 466.92 | 30.83 | 219.79 |
| PREVAZ | 4.59×10^5 | 480.12 | 30.32 | 153.34 |
| MLP | 7.31×10^5 | 619.77 | 40.55 | 128.78 |
| MLP(m) | 4.79×10^5 | 539.67 | 39.62 | 143.29 |
| RNN - <i>Off-line</i> | 1.89×10^5 | 293.01 | 20.07 | 93.32 |
| RNN - <i>On-line</i> | 1.21×10^5 | 236.57 | 15.83 | 72.44 |

Tabela 6.17: Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 6 Passos à Frente, Período de 1981 a 1985.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(2) | 5.49×10^5 | 537.78 | 36.35 | 217.71 |
| PAR(2) | 4.62×10^5 | 502.43 | 33.78 | 130.33 |
| PAR(p_m) | 5.65×10^5 | 540.75 | 31.56 | 114.81 |
| PREVAZ | 4.73×10^5 | 500.00 | 31.56 | 114.81 |
| MLP | 6.18×10^5 | 574.80 | 40.41 | 127.28 |
| MLP(m) | 5.21×10^5 | 539.67 | 38.16 | 138.57 |
| RNN - <i>Off-line</i> | 2.88×10^5 | 374.73 | 23.06 | 86.63 |
| RNN - <i>On-line</i> | 2.09×10^5 | 373.87 | 21.99 | 92.77 |

Tabela 6.18: Estatísticas para os Modelos de Previsão para o Posto de Furnas - Previsão 12 Passos à Frente, Período de 1981 a 1985.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(2) | 5.32×10^5 | 524.11 | 33.17 | 86.74 |
| PAR(2) | 5.40×10^5 | 522.36 | 32.14 | 77.49 |
| PAR(p_m) | 5.44×10^5 | 516.95 | 30.81 | 77.36 |
| PREVAZ | 5.39×10^5 | 510.48 | 29.85 | 76.02 |
| MLP | 6.16×10^5 | 573.47 | 41.41 | 127.86 |
| MLP(m) | 5.21×10^5 | 559.47 | 38.17 | 138.57 |
| RNN - <i>Off-line</i> | 3.45×10^5 | 429.84 | 28.07 | 87.22 |
| RNN - <i>On-line</i> | 3.16×10^5 | 409.78 | 27.75 | 96.20 |

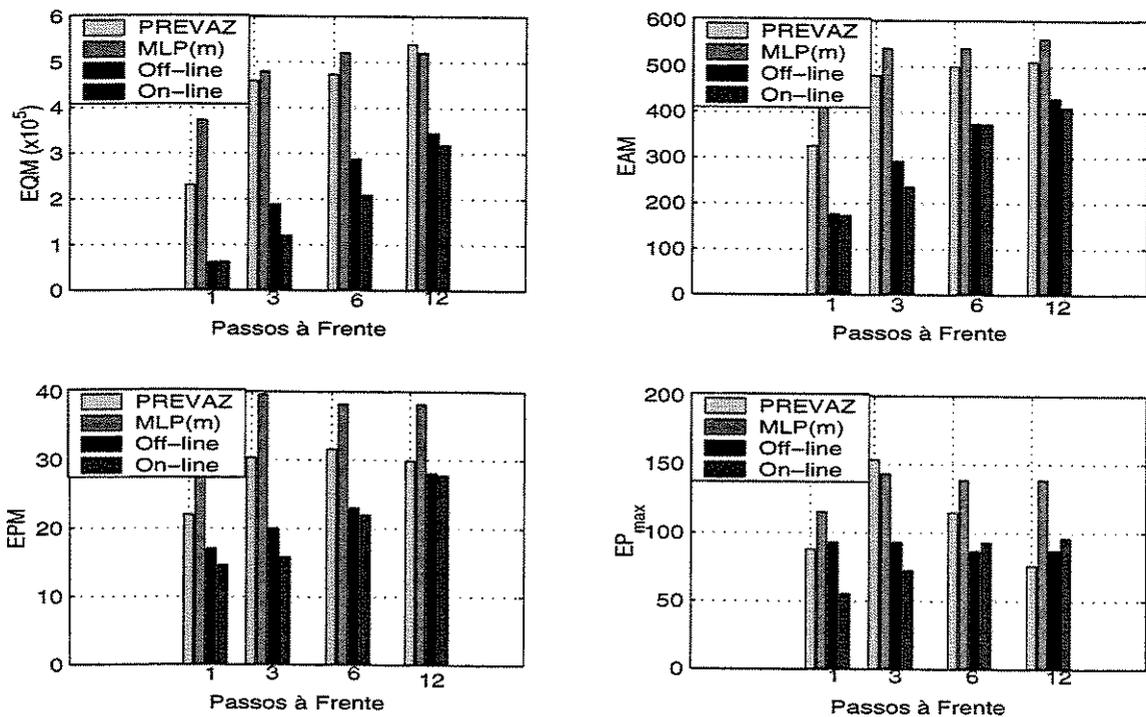


Figura 6.13: Erros de Previsão de Vazões para o Posto de Furnas - Período de 1981 a 1985.

6.4.2 Análise dos Erros de Previsão para o Posto de Emborcação

Nesta seção, são analisados os erros de previsão dos modelos auto-regressivo (AR), periódico auto-regressivo (PAR e $PAR(p_m)$), modelo PREVAZ, rede multi-camadas (MLP e $MLP(m)$) e rede neural nebulosa (RNN) com os algoritmos *off-line* e *on-line* para a série de vazões afluentes médias mensais da usina hidroelétrica de Emborcação. Estes modelos foram ajustados aos dados nas seções 6.2 e 6.3 e testados no período de 1972 a 1976.

Os erros obtidos para o período de teste para todos os modelos ajustados são mostrados nas tabelas 6.19, 6.20, 6.21 e 6.22, para as previsões 1, 3, 6, e 12 passos à frente, respectivamente.

Tabela 6.19: Estatísticas para os Modelos de Previsão para o Posto de Emborcação - Previsão 1 Passo à Frente, Período de 1972 a 1976.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(1) | 5.64×10^4 | 134.02 | 26.29 | 132.16 |
| PAR(1) | 3.11×10^4 | 111.25 | 22.86 | 93.11 |
| $PAR(p_m)$ | 3.11×10^4 | 110.87 | 22.46 | 93.11 |
| PREVAZ | 3.51×10^4 | 118.77 | 25.90 | 105.90 |
| MLP | 4.51×10^4 | 140.79 | 30.79 | 122.13 |
| $MLP(m)$ | 3.49×10^4 | 114.95 | 28.47 | 121.94 |
| RNN - <i>Off-line</i> | 5.88×10^3 | 89.56 | 11.62 | 74.34 |
| RNN - <i>On-line</i> | 2.06×10^3 | 93.56 | 12.62 | 71.34 |

Para as previsões 1 passo à frente (tabela 6.19), os modelos PAR(1) e $PAR(p_m)$ apresentaram resultados similares, já que na maioria dos meses, foram ajustados modelos de ordem 1. Estes modelos forneceram menores erros em relação aos modelos de séries temporais. Os modelos PAR(1), $PAR(p_m)$ e PREVAZ apresentaram pequena variação no erro EQM , o que mostra que estes modelos acompanham melhor a tendência da série. Para a rede neural os melhores resultados foram obtidos com o ajuste de 12 modelos ($MLP(m)$), com exceção do EP_{max} , onde o ajuste de um único modelo MLP apresentou um erro menor. A RNN com algoritmo *on-line* apresentou erros EQM e EP_{max} menores em relação à RNN com algoritmo *off-line*. Já os erros EAM e EPM foram menores quando ajustada à RNN com algoritmo *off-line*. Também pode-se observar que a RNN apresentou erros menores que os outros modelos ajustados, independente do algoritmo utilizado para treinamento.

Tabela 6.20: Estatísticas para os Modelos de Previsão para o Posto de Emborcação - Previsão 3 Passos à Frente, Período de 1972 a 1976.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(1) | 4.71×10^4 | 157.24 | 35.84 | 136.88 |
| PAR(1) | 3.89×10^4 | 128.21 | 26.65 | 102.57 |
| PAR(p_m) | 3.92×10^4 | 127.88 | 25.79 | 102.52 |
| PREVAZ | 5.73×10^4 | 169.98 | 38.42 | 174.88 |
| MLP | 4.55×10^4 | 149.12 | 32.03 | 116.64 |
| MLP(m) | 4.89×10^4 | 166.64 | 33.59 | 121.58 |
| RNN - <i>Off-line</i> | 1.31×10^4 | 97.42 | 17.16 | 89.76 |
| RNN - <i>On-line</i> | 1.19×10^4 | 87.47 | 14.68 | 78.81 |

Analisando os erros de previsão 3 passos à frente (tabelas 6.20) pode-se observar, novamente, que a variação nos erros em relação aos modelos PAR(1) e PAR(p_m) foi pequena. Comparando o ajuste de um único modelo (MLP) e 12 modelos (MLP(m)) de redes neurais, pode-se observar que o modelo MLP apresentou melhor desempenho. A RNN com algoritmo *on-line* apresentou erros menores em relação ao algoritmo *off-line*, mas com pequena variação para o EQM . Esse modelo forneceu menores erros em relação a todos os modelos ajustados.

Tabela 6.21: Estatísticas para os Modelos de Previsão para o Posto de Emborcação - Previsão 6 Passos à Frente, Período de 1972 a 1976.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(1) | 4.01×10^4 | 144.35 | 36.32 | 114.71 |
| PAR(1) | 4.38×10^4 | 146.65 | 32.28 | 102.35 |
| PAR(p_m) | 4.24×10^4 | 143.16 | 31.15 | 102.50 |
| PREVAZ | 4.47×10^4 | 150.83 | 38.27 | 126.07 |
| MLP | 4.79×10^4 | 149.66 | 33.01 | 116.64 |
| MLP(m) | 5.82×10^4 | 162.73 | 35.32 | 121.58 |
| RNN - <i>Off-line</i> | 1.33×10^4 | 96.74 | 25.51 | 85.23 |
| RNN - <i>On-line</i> | 1.44×10^4 | 89.96 | 24.76 | 83.27 |

As previsões 6 e 12 passos à frente (tabelas 6.21 e 6.22) apresentaram pequenas variações nos erros para os modelos de séries temporais. O ajuste de um único modelo para a rede MLP apresentou melhor desempenho que o ajuste de 12 modelos (MLP(m)). Comparando os dois algoritmos utilizados pela rede RNN, pode-se observar que os métodos de aprendizagem *off-line* e *on-line* apresentaram melhores resultados em relação aos demais modelos.

Tabela 6.22: Estatísticas para os Modelos de Previsão para o Posto de Emborcação - Previsão 12 Passos à Frente, Período de 1972 a 1976.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|------------|------------|-------------------------|
| | <i>EQM</i> | <i>EAM</i> | <i>EPM</i> | <i>EP_{max}</i> |
| AR(1) | 4.11×10^4 | 147.55 | 37.11 | 110.35 |
| PAR(1) | 4.31×10^4 | 153.04 | 37.67 | 112.35 |
| PAR(p_m) | 4.15×10^4 | 148.59 | 35.69 | 102.47 |
| PREVAZ | 5.71×10^4 | 143.98 | 30.19 | 100.45 |
| MLP | 4.69×10^4 | 149.50 | 32.95 | 116.64 |
| MLP(m) | 5.83×10^4 | 163.70 | 35.64 | 122.75 |
| RNN - <i>Off-line</i> | 3.04×10^4 | 112.43 | 26.27 | 89.51 |
| RNN - <i>On-line</i> | 3.65×10^4 | 104.09 | 25.59 | 87.26 |

O período de teste considerado é caracterizado como um período mediano. Neste caso, o ajuste dos modelos de séries temporais apresenta pequena variação nos erros tanto para previsão 1 passo à frente como para previsão vários passos à frente. Desta forma, o ajuste de um modelo PAR(1) para previsão 1 e vários passos à frente torna-se mais adequado em relação aos modelos de séries temporais. O ajuste do modelo MLP apresentou melhores resultados quando ajustado um único modelo, com exceção da previsão 1 passo à frente. A RNN, tanto com algoritmo *off-line* como com o algoritmo *on-line*, apresenta desempenho superior em relação a todos os modelos ajustados.

A figura 6.14 mostra os erros obtidos para as previsões 1, 3, 6 e 12 passos à frente para os modelos PAR(1), MLP e RNN com os algoritmo *off-line* e *on-line*. Pode-se notar que o modelo RNN apresentou um erro *EPM* 50% menor para a previsão 1 passo à frente. Observa-se também que para todas as estatísticas, a rede neural nebulosa apresentou melhor desempenho que os modelos PAR(1) e AR(1).

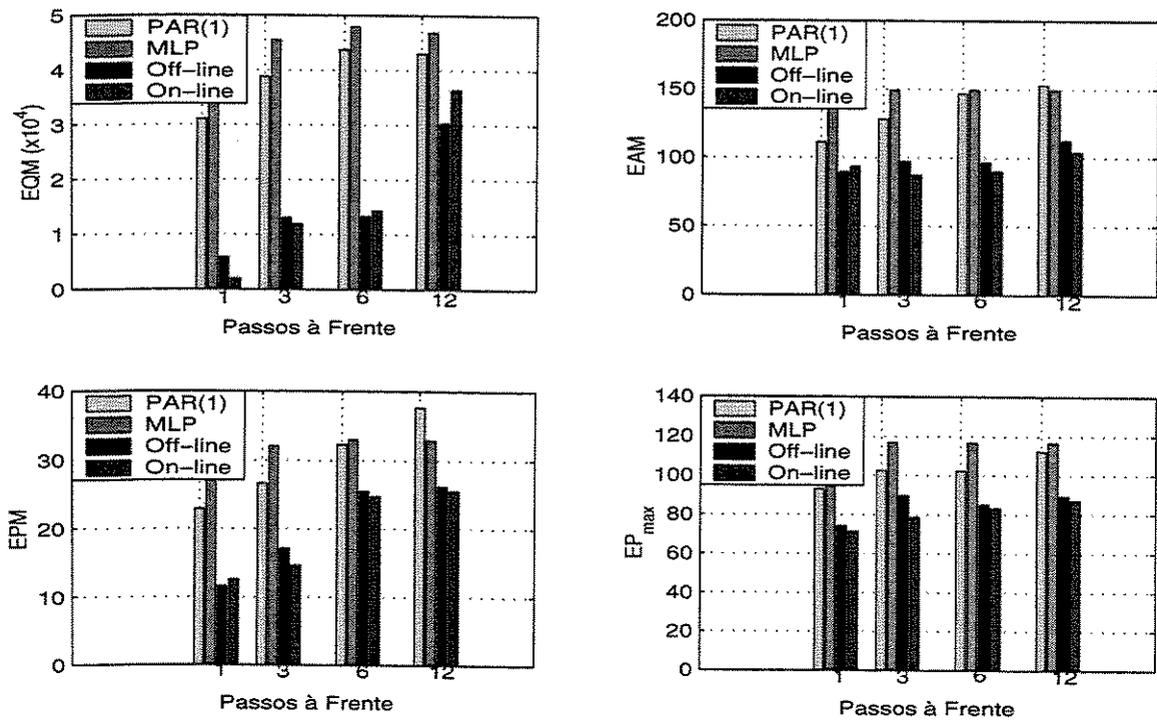


Figura 6.14: Erros de Previsão de Vazões para o Posto de Emborcação - Período de 1972 a 1976.

6.4.3 Análise dos Erros de Previsão para o Posto de Sobradinho

Nesta seção, são analisados os erros de previsão dos modelos auto-regressivo (AR), periódico auto-regressivo (PAR e $PAR(p_m)$), rede multi-camadas (MLP e $MLP(m)$) e rede neural nebulosa (RNN) com os algoritmos *off-line* e *on-line* para o posto de Sobradinho.

Os modelos foram ajustados na seção 6.3 e o desempenho é analisado para o período de 1972 a 1976. As tabelas 6.23, 6.24, 6.25 e 6.26 mostram os resultados para as previsões 1, 3, 6, e 12 passos à frente, respectivamente.

Analisando os erros de previsão 1 passo à frente (tabela 6.23), nota-se que os modelos de séries temporais apresentaram pequena variação nos erros, sendo que o modelo PAR(1) apresentou melhor desempenho. Para a rede neural o melhor desempenho foi obtido através do ajuste de 12 modelos ($MLP(m)$). A RNN treinada com o algoritmo *on-line* apresentou melhor desempenho que o treinamento com o algoritmo *off-line*, com exceção para o *EQM* que forneceu um erro ligeiramente menor.

Tabela 6.23: Estatísticas para os Modelos de Previsão para o Posto de Sobradinho - Previsão 1 Passo à Frente, Período de 1972 a 1976.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(1) | 6.03×10^5 | 487.79 | 18.49 | 89.56 |
| PAR(1) | 5.72×10^5 | 461.68 | 17.36 | 99.79 |
| PAR(p_m) | 5.87×10^5 | 470.15 | 18.10 | 99.79 |
| MLP | 8.09×10^5 | 600.54 | 23.85 | 110.02 |
| MLP(m) | 6.92×10^5 | 526.15 | 19.38 | 68.71 |
| RNN - <i>Off-line</i> | 8.19×10^4 | 244.30 | 11.98 | 68.56 |
| RNN - <i>On-line</i> | 9.39×10^4 | 212.61 | 9.15 | 65.83 |

Considerando as previsões 3 passos à frente (tabela 6.24), o modelo PAR(1) apresentou melhor desempenho em relação aos modelos de séries temporais, para todos os erros analisados. Novamente, a rede neural multi-camadas apresentou erros menores quando foi ajustado um modelo para cada mês (MLP(m)). A RNN com o método de aprendizagem *on-line* forneceu melhor desempenho que o método de aprendizagem *off-line* para todas as estatísticas analisadas.

Para a previsão 6 e 12 passos à frente (tabelas 6.25 e 6.26) os erros obtidos pelos modelos PAR(1) e PAR(p_m) forneceram uma pequena diferença, com exceção do EP_{max} , onde o modelo PAR(1) apresentou um erro menor. A rede neural apresentou melhores resultados quando foram ajustados 12 modelos (MLP(m)) para todas as estatísticas analisadas. A RNN com o algoritmo *on-line* forneceu erros menores que o treinamento com o algoritmo *off-line*, mas a diferença é relativamente pequena.

Comparando todos os modelos, pode-se verificar que tanto para previsão 1 passo à frente como para previsão vários passos à frente o modelo PAR(1) apresentou um bom desempenho quando comparado com os modelos AR(1) e PAR(1). A rede neural apresentou melhor desempenho quando foram ajustados 12 modelos (MLP(m)). A RNN forneceu melhores resultados quando ajustado o algoritmo *on-line*. Em relação a todos os modelos, a RNN com os métodos de aprendizagem *off-line* e *on-line* apresentou erros significativamente menores tanto para previsão 1 passo à frente como para previsão vários passos à frente.

A figura 6.15 mostra os erros EQM , EAM , EPM e EP_{max} para as previsões 1, 3, 6 e 12 passos à frente, para os modelos PAR(1), MLP(m) e RNN com os algoritmos *off-line* e *on-line*. Através da figura pode-se notar que os erros obtidos com o modelo RNN,

independente do algoritmo utilizado, são significativamente menores que os erros obtidos com os modelos PAR(1) e MLP(m). Nota-se que para previsão 1 passo à frente o EPM é 50% menor e o EQM apresentou um erro com mais de 50% de ganho em relação aos dois modelos comparados na figura 6.15.

Tabela 6.24: Estatísticas para os Modelos de Previsão para o Posto de Sobradinho - Previsão 3 Passos à Frente, Período de 1972 a 1976.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(1) | 1.34×10^6 | 729.17 | 27.70 | 147.17 |
| PAR(1) | 1.07×10^6 | 662.01 | 24.84 | 129.46 |
| PAR(p_m) | 1.07×10^6 | 666.62 | 25.35 | 134.14 |
| MLP | 9.46×10^5 | 600.54 | 26.03 | 136.14 |
| MLP(m) | 8.09×10^5 | 557.23 | 23.86 | 74.21 |
| RNN - <i>Off-line</i> | 3.42×10^5 | 306.03 | 15.11 | 68.46 |
| RNN - <i>On-line</i> | 2.65×10^5 | 298.95 | 12.38 | 70.13 |

Tabela 6.25: Estatísticas para os Modelos de Previsão para o Posto de Sobradinho - Previsão 6 Passos à Frente, Período de 1972 a 1976.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(1) | 1.42×10^6 | 786.44 | 32.29 | 147.85 |
| PAR(1) | 1.11×10^6 | 715.84 | 29.91 | 128.82 |
| PAR(p_m) | 1.05×10^6 | 710.20 | 29.83 | 137.92 |
| MLP | 9.56×10^5 | 665.38 | 26.77 | 136.14 |
| MLP(m) | 9.18×10^5 | 601.19 | 24.05 | 97.50 |
| RNN - <i>Off-line</i> | 4.29×10^5 | 339.42 | 17.52 | 82.34 |
| RNN - <i>On-line</i> | 3.93×10^5 | 318.24 | 15.64 | 78.32 |

Tabela 6.26: Estatísticas para os Modelos de Previsão para o Posto de Sobradinho - Previsão 12 Passos à Frente, Período de 1972 a 1976.

| Modelos | Estatísticas | | | |
|-----------------------|--------------------|--------|-------|------------|
| | EQM | EAM | EPM | EP_{max} |
| AR(1) | 1.41×10^6 | 797.71 | 34.31 | 149.51 |
| PAR(1) | 1.09×10^6 | 727.63 | 32.36 | 137.67 |
| PAR(p_m) | 1.08×10^6 | 725.24 | 32.32 | 141.52 |
| MLP | 9.72×10^5 | 615.48 | 28.74 | 108.73 |
| MLP(m) | 9.46×10^5 | 621.59 | 23.85 | 107.50 |
| RNN - <i>Off-line</i> | 6.40×10^5 | 395.04 | 20.47 | 95.46 |
| RNN - <i>On-line</i> | 4.87×10^5 | 352.57 | 19.27 | 93.18 |

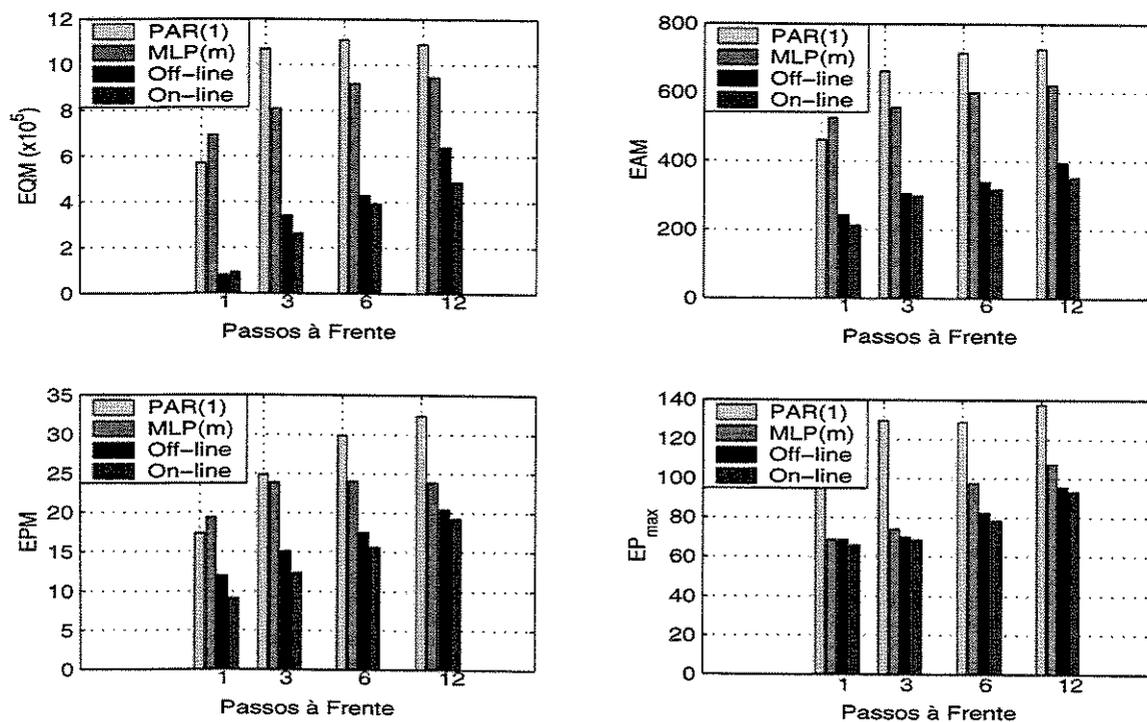


Figura 6.15: Erros de Previsão de Vazões para o Posto de Sobradinho - Período de 1972 a 1976.

6.5 Resumo

Este capítulo apresentou uma comparação dos modelos apresentados neste trabalho, realizando uma análise dos erros quadrático médio (EQM), absoluto médio (EAM), relativo percentual médio (EPM) e relativo percentual máximo (EP_{max}).

O desempenho dos modelos foi analisado após o ajuste dos dados de vazões afluentes médias mensais das usinas hidroelétricas de Furnas, Emborcação e Sobradinho.

Além dos modelos apresentados nos capítulos anteriores, neste capítulo o modelo de Previsão de Vazões (PREVAZ) utilizado pelo setor elétrico brasileiro também foi ajustado aos dados de vazões afluentes médias mensais do posto de Furnas e Emborcação.

Para a série de Furnas foram considerados três períodos com características distintas do histórico de vazões: seco (1952 a 1956), mediano (1972 a 1976) e úmido (1981 a 1985). Para estes três períodos observou-se que todos os modelos ajustados apresentaram um bom desempenho para previsão 1 passo à frente. Neste caso, a rede neural nebulosa com os métodos de aprendizagem *off-line* e *on-line* apresentaram erros 50% menores que os demais modelos ajustados.

Para as previsões vários passos à frente as estatísticas analisadas apresentaram variações quanto ao desempenho dos modelos ajustados para os períodos seco, mediano e úmido, sendo que a rede neural nebulosa apresentou erros menores. Estes resultados são devido ao fato deste modelo apresentar, como uma das características da estrutura da rede, uma fácil adaptabilidade às mudanças ocorridas nas séries.

Para os períodos seco e úmido, observou-se que o ajuste de 12 modelos para previsão, ou seja, o ajuste de um modelo para cada mês, para os modelos de séries temporais e para os modelos de redes neurais, apresentou um melhor resultado tanto para previsão 1 passo à frente como para previsão vários passos à frente.

O desempenho das previsões para os postos de Emborcação e Sobradinho foi analisado para o período mediano (1972 a 1976). Os modelos de séries temporais para este período apresentaram erros com pouca variação. Este resultado foi verificado para os três conjuntos de dados de vazões, indicando que o ajuste de um modelo periódico de ordem p foi o mais adequado entre os modelos de séries temporais. Para este período o modelo de rede neural multi-camadas com algoritmo de retro-propagação apresentou melhor desempenho quando ajustado um único modelo para as séries de vazões de Furnas e Emborcação. Já para a série de vazões do posto de Sobradinho, o ajuste de 12 modelos de redes neurais apresentou um melhor resultado que o ajuste de um único modelo. O modelo de rede neural nebulosa apresentou melhor desempenho para as três séries de vazões, sendo que o ajuste do modelo

com o método de aprendizagem *on-line* forneceu erros menores que o ajuste do modelo com o algoritmo de treinamento *off-line*. Também observou-se que, para este período, nos erros de previsões para a série de vazões de Furnas a diferença das estatísticas analisadas entre os modelos ajustados é menos destacada que nos períodos secos e úmidos.

A rede neural nebulosa treinada com o algoritmo *on-line* apresentou erros significativamente menores. Para a previsão 1 passo à frente o erro quadrático médio foi em média cinco vezes menor. Os erros absoluto médio, relativo percentual médio e erro percentual máximo apresentaram em média valores duas vezes menores em relação ao modelo que apresentou maior erro. Para as previsões 3 passos à frente este modelo apresentou, para a série de vazões de Furnas e no período seco um erro quadrático médio dez vezes menor que o modelo AR(2).

Para o período mediano, a diferença do erro quadrático médio foi três vezes menor em relação aos modelos com maior erro de previsão. Neste período o desempenho dos modelos foi melhor pois a variação nos dados de vazões é relativamente pequena. Para o período úmido o erro quadrático médio obtido pelo modelo de rede neural nebulosa com algoritmo *on-line* foi cinco vezes menor em relação à rede neural multi-camadas. As demais estatísticas apresentaram erros duas vezes menores independente do período previsto. Para as previsões 6 e 12 passos à frente as previsões apresentaram em média erros duas vezes menores usando o algoritmo *on-line* comparando com os modelos que apresentaram maiores erros.

Capítulo 7

Conclusão

Durante as duas últimas décadas, extensivas pesquisas têm sido voltadas para o desenvolvimento de métodos para análise das características estocásticas das séries temporais hidrológicas. Uma das dificuldades na previsão de vazões é a presença da sazonalidade devido aos períodos seco e úmido do ano. Modelos estocásticos foram por um longo tempo a alternativa mais comum para gerar modelos hidrológicos na análise de previsão de vazões.

Os modelos empregados pelo setor elétrico brasileiro têm como base os modelos de séries temporais baseados na metodologia de Box & Jenkins. Estes modelos necessitam de uma laboriosa formulação estatística para tratar a não-estacionariedade, identificar e estimar os dados de vazões aos modelos de séries temporais, sendo que os resultados, principalmente para previsão vários passos à frente, nem sempre são satisfatórios.

Pesquisas por abordagens alternativas, capazes de superar tais dificuldades têm sido intensas. Uma das atuais tendências é a aplicação de modelos de redes neurais para previsão de vazões, visando mostrar a eficiência desta metodologia quando comparada com os resultados dos modelos de séries temporais.

Modelos de análise de dados que consideram incertezas e apresentam tolerância a imprecisões são necessários para tratabilidade e robustez. Sistemas nebulosos são baseados nestas hipóteses.

A integração de redes neurais com outras abordagens da área de inteligência artificial vêm sendo investigada atualmente. Um dos resultados desta integração são as chamadas redes neurais nebulosas. Estes modelos unem os benefícios das redes neurais, tais como, capacidade de aprendizagem e habilidade de adaptação, com a capacidade de representação e raciocínio da lógica nebulosa. A possibilidade de usufruir destas características unificadamente é um fator que desperta grande motivação para o estudo e aplicação destes

modelos em diferentes áreas.

O principal objetivo deste trabalho concentrou-se na análise e previsão de vazões médias mensais utilizando modelos auto-regressivos, periódicos auto-regressivos, redes neurais multi-camadas com algoritmo de retro-propagação e redes neurais nebulosas com dois algoritmos de aprendizagem, chamados de *off-line* e *on-line*. Comparações dos resultados obtidos através da análise dos erros quadrático médio, absoluto médio, relativo percentual médio e relativo percentual máximo, foram realizadas.

O desempenho dos modelos foi avaliado considerando as séries de vazões naturais médias mensais de três usinas hidroelétricas brasileiras localizadas em diferentes regiões: Furnas, localizada no rio Grande; Emborcação, localizada no rio Paranaíba; e Sobradinho, localizada no rio São Francisco.

A análise das séries de vazões afluentes médias mensais para as três usinas hidroelétricas brasileiras mostra o comportamento distinto das séries de vazões para cada uma das usinas.

Modelos auto-regressivos foram ajustados a cada uma das séries através da análise das funções de auto-correlação e auto-correlação parcial. Usando o método dos momentos, foram estimados os parâmetros do modelo identificado. A análise das previsões vários passos à frente mostrou que o comportamento da série prevista tende à média a longo termo.

Modelos periódicos auto-regressivos foram estudados. Neste caso, a identificação dos parâmetros foi realizada através da inferência Bayesiana. Esta abordagem consiste em combinar a função de verossimilhança condicionada ao conhecimento dos parâmetros e de uma densidade *a priori* para os parâmetros.

Também foram ajustados modelos periódicos auto-regressivos para cada mês. Ou seja, foram selecionados através da análise da densidade preditiva a ordem do modelo a ser ajustado a cada mês, e os parâmetros para cada mês foram estimados utilizando a densidade *a priori* não-informativa. Os resultados mostraram que nas previsões um passo à frente, as três séries de vazões obtiveram bom desempenho, porém para previsões com vários passos à frente as previsões tendem à média das vazões mensais.

Modelos de redes neurais multi-camadas com algoritmo de retro-propagação foram estudados e aplicados para previsão de vazões. A estrutura deste modelo é não-recorrente e portanto não apresenta memória dinâmica. A utilização desta estrutura para representar sistemas dinâmicos somente é possível com a inclusão, no vetor de entradas, de valores em instantes anteriores de tempo. Para as previsões 1 passo à frente o desempenho do modelo foi satisfatório. Entretanto, para previsão vários passos à frente o modelo

ajustado começa a se deteriorar. Uma das justificativas para este fato é que o vetor de entrada apresenta dados de atraso que foram previstos e assim ocorre uma degradação no desempenho da rede, decorrente dos erros de previsão.

Uma tentativa de amenizar este comportamento foi ajustar um modelo de rede neural para cada mês e, assim, tornar o modelo mais específico para prever o dado de vazão de cada mês. Para este caso, o desempenho dos modelos apresentou resultados mais satisfatórios, principalmente para períodos em que a série de vazões apresenta variações mais acentuadas nos dados, ou seja, para os períodos críticos da série de vazões. Novamente, observou-se que o desempenho da série começa a se degradar para previsões vários passos à frente, devido à presença de erros de previsão.

Neste trabalho, foi estudado e aplicado um modelo de rede neural nebulosa a séries de vazões afluentes médias mensais. Este modelo apresenta um método de aprendizado construtivo, no qual grupos de neurônios são adicionados à estrutura da rede quando novo conhecimento é necessário. A rede aprende os parâmetros da função de pertinência para cada variável de entrada dos dados de treinamento, codificando um conjunto de regras nebulosas em sua estrutura, propriedade que fornece uma relação dual entre a rede e um sistema nebuloso associado. Dois algoritmos de treinamento baseados no paradigma de aprendizado competitivo foram utilizados. No primeiro, chamado de algoritmo *off-line*, o número de regras é definido no início do treinamento; no segundo, chamado de algoritmo *on-line*, grupos de neurônios são adicionados à estrutura da rede quando novos conhecimentos são necessários.

A aplicação deste modelo à previsão de vazões médias mensais apresentou resultados satisfatórios, tanto para previsão um passo à frente como para previsão vários passos à frente, pois estes modelos apresentam métodos de aprendizagem do tipo construtivo que capacitam a rede a compensar degradações de desempenho provocadas por variações nos parâmetros do sistema.

O método de aprendizagem *off-line* pode levar a uma restrição inconveniente. Dependendo da limitação do espaço de entrada a ser coberto pelas regras nebulosas codificadas na estrutura da rede, a dispersão associada às funções de pertinência de uma particular regra pode tornar-se excessivamente ampla. Esta regra pode influenciar no desempenho da rede, pois um mesmo conseqüente pode contribuir para a resposta da rede. O método de aprendizagem *on-line* supre esta deficiência, introduzindo novas regras nebulosas sempre que o desempenho não for alcançado, permitindo uma melhor distribuição das regras pelo espaço de entrada. A rede neural nebulosa com algoritmo *on-line* apresentou erros significativamente menores em relação aos demais modelos, como pode ser verificado no capítulo 6.

7.1 Trabalhos Futuros

Os resultados apresentados mostraram que os modelos de redes neurais e redes neurais nebulosas são ferramentas que fornecem bom desempenho. A degradação do desempenho dos modelos para previsão vários passos à frente conduz ao estudo de modelos de redes neurais recorrentes para aplicação na previsão vários passos à frente.

Critérios para avaliar se o modelo selecionado, tanto para os modelos de redes neurais como para o modelos de redes neurais nebulosas, é o mais adequado para representar uma série temporal, a chamada medida de parcimônia, merecem ser investigados.

Ao fazer previsões de vazões para um determinado posto deve-se levar em conta a influência de postos na mesma região hidrológica, que podem afetar as previsões. Neste sentido, usar somente o histórico do posto isoladamente para ajustar modelos de séries temporais pode ser pouco representativo. Quando consideram-se postos vizinhos as correlações espaciais são estabelecidas, provocando um aumento no grau de complexidade dos modelos de Box & Jenkins. Nestes casos, deve-se considerar as vazões como processos estocásticos multivariados, cujos componentes são as vazões de cada posto. O ajuste dos parâmetros desses modelos vetoriais exige grande esforço computacional. Este fato deve ser considerado futuramente, utilizando modelos de redes neurais e redes neurais nebulosas.

O estudo de séries de vazões semanais, cujo comportamento se diferencia das séries de vazões mensais, deve ser investigado. Estas séries são de fundamental importância para o planejamento a curto prazo.

Uma avaliação da influência dos modelos de previsão sob o ponto de vista do planejamento energético é de fundamental importância, pois auxilia na tomada de decisão baseada nos modelos determinísticos.

A geração de séries sintéticas utilizando redes neurais e redes neurais nebulosas deve ser estudada. Estas séries são utilizadas tanto para otimização como para simulação no planejamento energético.

BIBLIOGRAFIA

- Alves da Silva, A. P., Ferreira, C., Torres, G. L. e Souza, A. C. Z. (1997). A new constructive ANN and its application to electric load representation, *IEEE Transactions on Power Systems* **12**(4): 1569–1575.
- Anderson, J., Silverstein, J., Ritz, S. e Jones, R. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural networks, *Psychological Review* **84**: 413–451.
- Anderson, P. L. e Vecchia, A. V. (1993). Asymptotic results for periodic autoregressive moving average processes, *Journal Time Series Analyses* pp. 1–18.
- Andrade Filho, A. G. (1994). *Uma Investigação sobre o Método de Vizinhos Mais Próximos na Previsão Chuva/Vazão*, Tese de Doutorado, Escola de Engenharia de São Carlos – USP, São Carlos.
- Atiya, A. F., El-Shoura, S. M., Shaheen, S. I. e El-Sherif, M. S. (1999). A comparison between neural-network forecasting techniques - case study: River flow forecasting, *IEEE Transactions on Neural Networks* **10**(2): 402–409.
- Bakirtzis, A. G., Theocharis, J. B., Kiartzis, S. J. e Satsios, K. J. (1995). Short term load forecasting using fuzzy neural networks, *IEEE Transactions on Power Systems* **10**(3): 1518–1524.
- Ballini, R. (1996). *Redes Neurais para Previsão Chuva-Vazão*, Tese de Mestrado, Universidade de São Paulo, São Carlos, SP.
- Ballini, R., França, E. H., Kadowaki, M., Soares, S. e Andrade, M. G. (1997b). Modelos de redes neurais e Box & Jenkins para previsão de vazões médias mensais, *XII Simpósio Brasileiro de Recursos Hídricos*, pp. 143–149.
- Ballini, R., França, E. H., Soares, S. e Andrade, M. G. (1997a). Relação entre modelos auto-regressivos e a configuração da rede neural para previsão de séries temporais estacionárias, *IV Simpósio Brasileiro de Redes Neurais*, pp. 18–22.

- Ballini, R., Figueiredo, M., Soares, S., Andrade, M. e Gomide, F. (2000). A seasonal streamflow forecasting model using neurofuzzy network, *Information, Uncertainty and Fusion*, B. Bouchon-Meunier and R. R. Yager and L. Zadeh, Eds., Kluwer Academic Publishers, pp. 257-276.
- Ballini, R., Soares, S. e Andrade, M. G. (1998). An adaptive neural fuzzy network model for seasonal streamflow forecasting, *Proceedings of the IV Brazilian Symposium on Neural Networks-IEEE Computer Society*, pp. 215-219.
- Ballini, R., Soares, S. e Andrade, M. G. (1999a). A comparison between neural fuzzy network and bayesian periodic autoregressive models to seasonal streamflow series forecasting, *Proceedings of the International Congress on Computational Intelligent Methods and Applications - CIMA-ISFL*, pp. 74-80.
- Ballini, R., Soares, S. e Andrade, M. G. (1999b). Comparison of seasonal streamflow forecasting models, *Proceedings of the International Conference on Intelligent System Application to Power Systems - ISAP'99*, pp. 417-421.
- Ballini, R., Soares, S. e Andrade, M. G. (1999c). Seasonal streamflow forecasting via a neural fuzzy system, *Proceedings of International Federation of Automatic Control - IFAC'99*, Vol. k, pp. 81-86.
- Box, G. E. P. e Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco.
- Box, G. E. P., Jenkins, G. M. e Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, Holden Day, 3th Edição.
- Box, G. E. P. e Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*, Addison-Wesley, New York.
- Braga, B. P. F. (1983). Geração e previsão de vazões através de modelos ARMA e ARIMA, *Curso de Engenharia Hidrológica - Hidrologia Operacional*, Vol. 2, pp. Q.1-Q.59.
- Buckley, A. e Hayashi, Y. (1994). Fuzzy neural networks: a survey, *Fuzzy Sets and Systems* 66: 41-49.
- Caminhas, W. M., Tavares, H. M. F. e Gomide, F. (1996). Pattern classification using neurofuzzy and neural networks, *Proceedings Information Processing and Management of Uncertainty on Knowledge - Based Systems - IPMU'96*, Vol. 1, pp. 175-180.
- Caminhas, W., Tavares, H., Gomide, F. e Pedrycz, W. (1999). Fuzzy set based neural networks: Structure, learning and application, *Journal of Advanced Computational Intelligence* 3(3): 151-157.
- Casella, G. e George, E. J. (1992). Explaining Gibbs sampler, *The American Statistician* 46(3): 167-174.

- Castro, J. (1995). Fuzzy logic controllers are universal approximators, *IEEE Transactions on Systems, Man, and Cybernetics* **25**(4): 629–635.
- Castro, J. e Delgado, M. (1996). Fuzzy systems with defuzzification are universal approximators, *IEEE Transactions on Systems, Man, and Cybernetics* **26**(1): 149–152.
- CEPEL (1983). *Centro de Pesquisas de Energia Elétrica. Modelo de Séries Hidrológicas, Manual de Metodologia., Relatório Técnico 113-81-A.*
- Chang, H. (1995). *Model Determination Using Predictive Distributions*, Ph.D. Thesis, University of Connecticut.
- Contrell, M., Girard, B., Girard, Y., Manguéas, M. e Muller, C. (1995). Neural modeling for time series: A statistical stepwise method for weight elimination, *IEEE Transaction on Neural Networks* **6**(6): 1355–1364.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function, *Mathematics of Control, Signals, and Systems*, Vol. 2.
- Durbin, J. (1960). The fitting of time series models, *Revue Institut International Sattistique* **28**: 233–244.
- El-Sharkawi, M. A., Oh, S., Marksm, R. J., Damborg, M. J. e Brace, C. M. (1991). Short term electric load forecasting using an adapttive trained layered perceptron, in Proc. of the First Forum on Application of Neural Networks to Power Systems, pp. 3–6.
- Faria Neto, A., Alves da Silva, A. P. e Torres, G. L. (1998). Sistemas neuro-fuzzy para modelagem dinâmica de carga?, *XII Congresso Brasileiro de Automática*, Vol. 4, Uberlândia, pp. 1327–1332.
- Figueiredo, M. F. (1997). *Redes Neurais Nebulosas Aplicadas em Problemas de Modelagem e Controle Autônomo*, Tese de Doutorado, Faculdade de Engenharia Elétrica e de Computação, Unicamp.
- Figueiredo, M. e Gomide, F. (1997). Adaptive neuro fuzzy modelling, *Proceedings of the 6th IEEE Conference on Fuzzy Systems*, Barcelona – Espanha, pp. 1567–1572.
- Figueiredo, M. e Gomide, F. (1998). Fuzzy system design using neurofuzzy networks, *Proceedings of the 7th Information Processing and Management of Uncertainty on Knowledge - Based Systems - IPMU'98*, Vol. 2, pp. 1416–1422.
- Figueiredo, M., Gomide, F. e Pedrycz, W. (1995). A fuzzy neural network: Structure and learning, *Fuzzy Logic and Its Applications, Information Sciences and Intelligent Systems*, Z. Bien and K. Min, Eds., Kluwer Academic Publishers, Holanda, pp. 177–186.

- Fortunato, L. A. M., Neto, T. A. A., Albuquerque, J. C. R. e Ferreira, C. (1990). *Introdução ao Planejamento da Expansão e Operação de Sistemas de Produção de Energia Elétrica*, Editora Universitária, Niterói-RJ.
- Francelin, R., Ballini, R. e Andrade, M. G. (1996). Back-propagation and Box & Jenkins approaches to streamflow forecasting, *Latin-Iberian-American Congress on Operations Research and System Engineering - CLAIO, Simpósio Brasileiro de Pesquisa Operacional - SBPO*, Rio de Janeiro -RJ, pp. 1307-1312.
- Funahashi, K. (1989). On the approximate realization of continuous mappings by neural networks, *Neural Networks* 2: 183-192.
- GCOI (1983). *Grupo Coordenador da Operação Interligada*, Manual do Sistema PREVAZ, Eletrobrás.
- Geva, S. e Sitte, J. (1991). Adaptive nearest neighbor pattern classification, *IEEE Transactions on Neural Networks* 2(2).
- Gomide, F., Figueiredo, M. e Pedrycz, W. (1995). A neural fuzzy network: Structure and learning, *Fuzzy Logic and Its Applications, Information Sciences and Intelligent Systems*, Bien, Z. and Min, K., Kluwer Academic Publishers, Netherlands, pp. 177-186.
- Grossberg, S. (1967). Nonlinear difference-differential equations in prediction and learning theory, *Proceedings of the National Academy of Sciences of the U.S.A.* 58: 1329-1334.
- Grossberg, S. (1968). A prediction theory for some nonlinear functional-difference equations, *Journal of Mathematical Analysis and Applications* 21: 643-694.
- Gupta, M. M. e Qi, J. (1992). On fuzzy neuron models, *Fuzzy Logic for the Management of Uncertainty*, L. Zadeh and J. Kacprzyk, Eds, Wiley, New York, pp. 479-491.
- Haykin, S. (1994). *Neural Networks - A Comprehensive Foundation*, IEEE Press, New York.
- Hebb, D. (1949). *The Organization Behavior: A Neuropsychological Theory*, John Wiley & Sons, New York.
- Hertz, J., Krogh, A. e Palmer, R. (1991). *Introduction to the Theory of Neural Computation*, Reading, MA: Addison-Wesley.
- Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences of the U.S.A.*, Vol. 79, pp. 2554-2558.
- Hoptroff, R. G., Bramson, M. J. e Hall, T. J. (1991). Forecasting economic turning points with neural networks, Vol. 1, in Proc. of the IEEE/International Joint Conference of Neural Networks, pp. 347-352.

- Hutter, C. (1998). *Uma Abordagem Bayesiana para Modelos Auto-Regressivos Periódicos*, Tese de Mestrado, Instituto de Ciências Matemáticas e de Computação, USP-São Carlos.
- Hwang, J. L. e Monn, S. (1991). Temporal difference method for multi-step prediction: Application to power load forecasting, in Proc. of the First Forum on Application of Neural Networks to Power Systems, pp. 41-45.
- Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation, *Neural Networks* 1: 295-307.
- Jones, R. H. e Brelsford, W. M. (1967). Time series with periodic structure, *Biometrika* 54(3,4): 403-408.
- Kacprzyk, J. e Fedrizzi, M. (1992). *Fuzzy regression Analysis. Studies in Fuzziness*, Vol. 1, Physica-Verlag, Heidelberg.
- Kadowaki, M., Soares, S. e Andrade, M. G. (1997). Previsão de vazões mensais utilizando redes neurais multicamadas com algoritmo backpropagation, *IV Simpósio Brasileiro de Redes Neurais*.
- Keller, J. M. e Hunt, D. (1985). Incorporating fuzzy membership functions into the perceptron algorithm, *IEEE Trans. Pattern Anal. Mach. Intell.* 7: 693-699.
- Kramer, A. e Sangiovanni-Vincentelli, A. (1989). Efficient parallel learning algorithms for neural networks, *Advances in Neural Information Processing Systems 1*, D.S. Touretzky, ed., San Mateo, CA: Morgan Kaufmann, pp. 40-48.
- Kuo, R. J. e Xue, K. C. (1999). Fuzzy neural networks with application to sales forecasting, *Fuzzy Sets and Systems* 108: 123-143.
- Lachtermacher, G. e Fuller, J. D. (1995). Backpropagation in time-series forecasting, *Journal of Forecasting* 14(4): 381-393.
- Lapedes, A. e Farber, R. (1987). *Nonlinear Signal Processing Using Neural Networks: Prediction and System Modelling*, Technical Report LA-Ur-&-2662, Los Alamos National Laboratory.
- Lee, C. C. (1990). Fuzzy logic control in controller - part I and II, *IEEE Transactions on Systems, Man and Cybernetics* 20(2): 404-435.
- Lee, C. C. e Lin, C. (1991). Neural-network-base fuzzy logic control and decision system, *IEEE Transactions on Systems, Man and Cybernetics* 40(12): 1320-1336.
- Lee, S. C. e Lee, E. T. (1974). Fuzzy sets and neural networks, *Journal of Cybernetics* 4: 83-103.
- Lee, S. C. e Lee, E. T. (1975). Fuzzy neural networks, *Mathematic Biosciense* 423: 151-177.

- Lee, Y., Oh, S. e Kim, M. (1991). The effect of initial weights on premature saturation in back-propagation learning, *International Joint Conference on Neural Networks*, Vol. 1, pp. 765–770.
- Little, W. e Shaw, G. (1975). A statistical theory of short and long term memory, *Behavioral Biology* 14: 115–133.
- Liu, B., Xu, J. e Tan, Y. (1999). Neuro-Fuzzy Network with Dynamic Adjustment in Forecasting, *Proceedings of International Federation of Automatic Control - IFAC'99*, Vol. k, pp. 103–108.
- Liu, K., Subbarayan, S., Shoults, R. R., Manry, M. T., Kwan, C., Lewis, F. L. e Naccarino, J. (1996). Comparison of very short-term load forecasting techniques, *IEEE Transactions on Power Systems* 11(2): 877–882.
- Maier, H. R. e Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, *Environmental Modelling & Software* 15: 101–124.
- Mason, J. C., Teme'me, A. e Price, R. K. (1996). A neural network model of rainfall-runoff using radial basis functions, *Journal of Hydraulic Research* 34(4): 537–294.
- McCulloch, W. e Pitts, W. (1943). A logical calculus of the ideas imminent in nervous activity, *Bulletin of Mathematical Biophysics* 5: 115–133.
- Minsky, M. L. (1954). *Theory of neural-analog reinforcement systems and its application to the brain-model problem*, Ph.D. Thesis, Princeton University, Princeton, NJ.
- Minsky, M. L. e Papert, S. A. (1969). *Perceptrons*, MIT Press, Cambridge, MA.
- Morettin, P. A. e Toloí, C. M. C. (1985). *Previsão de Séries Temporais*, Editora Atual, São Paulo.
- Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning, *Neural Networks*, Vol. 6, pp. 525–533.
- Ng, G. W. (1997). *Application of Neural Networks to Adaptive Control of Nonlinear Systems*, John Wiley & Sons Inc, Control System Centre, UMIST, UK.
- Oliveira, J. R. T. (1998). *Uso do Amostrador de Gibbs e Metropolis-Hastings em Análise Bayesiana de Modelos AR*, Tese de Mestrado, Instituto de Ciências Matemáticas e de Computação, USP-São Carlos.
- Pagano, M. (1978). On periodic and multiple autoregressions, *The Annals of Statistics* 6: 1310–1317.
- Pedrycz, W. (1993). *Fuzzy Control and Fuzzy System*, John Wiley & Sons, Inc., New York, EUA.

- Pedrycz, W. e Gomide, F. (1998). *An Introduction to Fuzzy Sets: Analysis and Design*, MIT Press, Cambridge, MA.
- Pedrycz, W., Jam, P. C. F. e Rocha, A. F. (1995). Distributed fuzzy system modeling, *IEEE Transactions on Systems, Man and Cybernetics* **25**(5).
- Peng, T. M., Hubele, H. F. e Karady, G. G. (1992). Advancement in the application of neural networks for short-term load forecasting, Vol. 7, in Proc. of IEEE/Power Engineering Society Summer Meeting, 91 SM 451-5 PWRs, and in Transactions on Power Systems.
- Pereira, B. B., Pais, M. B. Z. e Sales, P. R. H. (1986). *Análise Espectral de Séries Temporais - Uma Introdução para Engenharia, Economia e Estatística*, Arte Final Leasing Editoria LTDA., Rio de Janeiro-RJ- Brasil.
- Quenouille, M. H. (1949). Approximate tests of correlation in time series, *Journal Royal Statistic Society* **B11**: 64-68.
- Ridmiller, M. (1994). Advanced supervised learning in multi-layer perceptrons - from back-propagation to adaptive learning algorithms, *Neural Networks* **5**.
- Roseblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review* **65**: 386-408.
- Rumelhart, D., Hinton, G. E. e Williams, R. (1986a). Learning representations by back-propagating errors, *Nature (London)* **323**: 533-536.
- Rumelhart, D., Hinton, G. e Williams, R. (1986b). Learning internal representation by error propagation, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, MIT: Press, Cambridge, M.A.
- Salas, J. D., Boes, D. C. e Smith, R. A. (1982). Estimation for ARMA models with seasonal parameters, *Water Resources Research* **18**(4): 1006-1010.
- Salas, J. D., Delleur, J. W., Yevjevich, V. e Lane, W. L. (1980). Applied modeling of hydrologic time series, *Water Resources Publications* .
- Sales, P. R. H., de Bragança Pereira, B. e Vieira, A. M. (1998). Procedimentos lineares para identificação e estimação dos parâmetros de modelos, *Revista Brasileira de Estatística - RBEs* **59**(212): 25-51.
- Sales, P. R. H. e Vieira, A. M. (1984). Modelagem de vazões médias mensais tendo em vista a obtenção de previsão segundo a metodologia Yevjevich/Box & Jenkins, *Seminário Internacional de Hidrologia e Hidráulica*, Assunção, Paraguai.
- Shamseldin, A. Y. (1997). Application of a neural network technique to rainfall-runoff modeling, *Journal of Hydrology* **199**: 272-294.

- Sharda, R. e Patil, R. (1990). Neural networks as forecasting experts: An empirical test, Vol. 1, in International Joint Conference on Neural Networks, Washington, DC., pp. 441–494.
- Silva, L. N. C. (1998). *Análise e Síntese de Estratégias de Aprendizado para Redes Neurais*, Tese de Mestrado, Faculdade de Engenharia Elétrica e de Computação, Unicamp.
- Soares, S. (1987). Planejamento da Operação de Sistemas Hidrotérmicos, *Controle e Automação-Revista Brasileira de Automática* 1(2): 122–123.
- Srinivasan, D., Liew, A. C. e Chen, J. S. P. (1991). Short term forecasting using neural network approach, in Proc. of the First Forum on Application of Neural Networks to Power Systems, pp. 12–16.
- Stone, M. (1974). Cross-validation choice assessment of statistical predictions, *Journal of the Royal Statistical Society*, Vol. B36, pp. 111–133.
- Tang, Z., de Almeida, C. e Fishwick, P. A. (1991). Time series forecasting using neural networks vs. Box-Jenkins methodology, *Simulation* 57(5): 303–310.
- Tierney, L., Kass, R. E. e Kadane, J. B. (1986). Fully exponential Laplace approximations for expectations and variances of nonpositive functions, *Journal of the American Statistical Association* 84: 710–716.
- Tong, H. e Lin, K. S. (1980). Threshold autoregressive, limit cycles and cyclical data, *Journal of the Royal Statistical Society* 42 (serie B)(3): 245–292.
- Vecchia, A. V. (1985). Maximum likelihood estimation for periodic autoregressive-moving average models, *Technometrics* 27(4): 375–384.
- Vieira, A. M., de Bragança Pereira, B. e Sales, P. R. H. (1998). Estimação conjunta dos parâmetros de um modelo estocástico multivariado contemporâneo periódico autoregressivo - par(p), *Revista Brasileira de Recursos Hídricos* 3(4): 5–17.
- von Neumann, J. (1956). Probabilistic logics and the synthesis of reliable organisms from unreliable components, *Automata Studies*, Princeton University Press, Princeton, NJ, pp. 43–98.
- Von Zuben, F. J. V. (1996). *Modelos Paramétricos e Não-Paramétricos de Redes Neurais Artificiais e Aplicações*, Tese de Doutorado, Faculdade de Engenharia Elétrica e de Computação, Unicamp.
- Walker, G. (1931). On periodicity in series of related terms, *Proceedings Royal Society* A131: 518–532.
- Wang, L. X. e Mendel, J. M. (1992). Back-propagation fuzzy system as non-linear dynamic system identifiers, *Proceedings of FUZZ-IEEE'92*, pp. 1409–1418.

- Weigend, A. S., Rumelhart, D. E. e Huberman, B. A. (1990). Predicting the future: A connectionist approach, *International Journal Neural System* 1: 3.
- Weigend, A. S., Rumelhart, D. E. e Huberman, B. A. (1991a). Generalization by weight-elimination application to forecasting, *Advances in Neural Information Processing Systems* 3, pp. 875–882.
- Weigend, A. S., Rumelhart, D. E. e Huberman, B. A. (1991b). Generalization by weight-elimination applied to currency exchange rate prediction, *Proceedings of the IEEE/International Joint Conference of Neural Networks*, Vol. I, pp. 837–841.
- Werbos, P. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences*, Ph.D. Thesis, Harvard University, Cambridge, MA.
- Widrow, B. (1962). Generalization and information storage in networks of adaline ‘neurons’, *Self-Organizing Systems (M.C. Yovitz, G.T. Jacobi, and G.D. Goldstein, eds.)* pp. 435–461.
- Widrow, B. e Hoff, M. (1960). Adaptive switching circuits, *IRE WESCON Convention Record*, pp. 96–104.
- Willshaw, D. e von der Malsburg, C. (1976). How pattern neural connections can be set up by self-organization, *Proceedings of the Royal Society of London, Series B*, Vol. 194, pp. 431–445.
- Yager, R. e Filev, D. (1994). *Essentials of Fuzzy Modeling and Control*, Wiley Interscience, New York.
- Yamakawa, T. e Tomoda, S. (1989). A fuzzy neuron and its application to pattern recognition, *Proceedings of the 3rd International Fuzzy Systems Association*, pp. 30–38.
- Yamakawa, T., Uchino, E., Miki, T. e Kusanagi (1992). A neo fuzzy neuron and its applications to system identification and predictions to system behavior, *Proceedings of the 2nd International Conference on Fuzzy Logic and Neural Networks*, pp. 477–483.
- Yule, G. (1927). On a method of investigation periodicities in disturbed series, with special reference to Wölfer’s sunspot numbers, *Philosophical Transactions Royal Society* 267(A226): 267–298.
- Zadeh, L. (1965). Fuzzy Sets, *Information Control* 8: 338–353.
- Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems* 1: 3–28.
- Zealand, C. M., Burn, D. H. e Simonovic, S. P. (1999). Short term streamflow forecasting using artificial neural networks, *Journal of Hydrology* 214: 32–48.
- Zimmermann, H. (1987). *Fuzzy Sets, Decision Making and Expert Systems*, Kluwer-Nihoft, Boston.

Apêndice A

Fundamentos de Sistemas Nebulosos

Uma característica da lógica clássica é o raciocínio “preto e branco”: um fato é verdadeiro ou não, sendo que cada conjunto contido no universo de discurso está associado a uma função de pertinência ou função característica, podendo assumir os valores 1, se o elemento do universo pertence ao conjunto, ou 0, caso contrário.

Freqüentemente, a classe de objetos do mundo real não tem precisamente definido um critério de pertinência (Zadeh, 1965). Por exemplo, a classe de “vazões mensais altas” depende do mês associado ao valor de vazão a ser analisado e do que significa um valor “alto” para este determinado mês e ano considerado. Nestas circunstâncias, a propriedade de pertinência de um elemento é realizada através de um grau de pertinência que pode estar associado a um número real no intervalo $[0, 1]$. Este fato conduz a um raciocínio impreciso para tratar as classes.

Em 1965, Lotfi Zadeh desenvolveu a teoria de conjuntos nebulosos como uma alternativa para tratar as incertezas e imprecisões. A lógica nebulosa estende os conceitos da lógica booleana clássica para os números reais. Ao contrário da lógica clássica, que tem somente dois valores, verdadeiro ou falso, a lógica nebulosa é multivalorada, onde os valores são conjuntos nebulosos. A lógica nebulosa é a lógica que trata de modelos de raciocínio aproximado. Em essência, o que a lógica nebulosa oferece é uma metodologia para tratar palavras ao invés de números, o que nenhum outro método faz.

DEFINIÇÃO A.1 Um conjunto nebuloso \mathcal{A} é caracterizado por uma função de pertinência (ou função característica) $f_{\mathcal{A}}(x)$, a qual associa a cada elemento de um domínio, espaço ou um universo de discurso X , um número real no intervalo unitário $[0, 1]$, com o valor de $f_{\mathcal{A}}(x)$ em x representando o grau de pertinência de x em A (Zadeh, 1965).

Assim, um conjunto nebuloso é definido como uma coleção de objetos com valores associados entre 0 (exclusão completa) e 1 (associação completa). Os valores associados expressam o grau com o qual cada objeto é compatível com as características ou propriedades que são distintas para a coleção.

As funções de pertinência associadas aos conjuntos nebulosos dependem não somente do conceito a ser representado, mas também do contexto no qual são usadas. As funções podem ter diferentes formas e algumas propriedades, como continuidade, podem ser exigidas (Pedrycz e Gomide, 1998). Duas formas de função de pertinência, geralmente utilizadas, são as funções triangulares e Gaussianas, descritas como segue.

1. Função Triangular

Neste caso, as funções de pertinência são definidas através dos parâmetros de dispersão ρ_{min} e ρ_{max} e o parâmetro c que define o centro ou valor modal da função (figura A.1), podendo ser representada da seguinte forma:

$$f_A(x) = \begin{cases} 0, & x \leq \rho_{min} \\ \frac{x - \rho_{min}}{c - \rho_{min}}, & x \in (\rho_{min}, c] \\ \frac{\rho_{max} - x}{\rho_{max} - c}, & x \in (c, \rho_{max}] \\ 0, & x \geq \rho_{max} \end{cases} \quad (A.1)$$

2. Função Gaussiana

As funções de pertinência Gaussianas são definidas pelos valores modais c (figura A.2), e podem ser representadas da seguinte forma:

$$f_A(x) = \frac{1}{1 + k(x - c)^2} \quad (A.2)$$

onde $k > 1$ é uma constante.

Como as funções de pertinência são representações equivalentes dos conjuntos nebulosos, as operações básicas de interseção, união e complemento são representadas tomando o mínimo, máximo e complemento das correspondentes funções características. Estas operações são chamadas operações nebulosas padrões.

Na teoria de conjuntos nebulosos, as normas triangulares têm um papel fundamental para fornecer os modelos genéricos para as operações de interseção e união, devendo possuir as propriedades de comutatividade, associatividade e monotonicidade e satisfazer

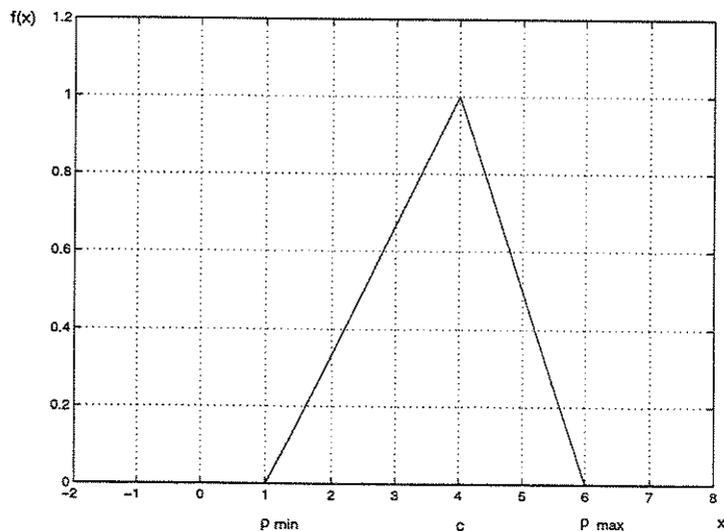


Figura A.1: Função de Pertinência Triangular.

as condições de contorno. Portanto, as normas triangulares formam classes gerais de interseção e união respectivamente caracterizadas pelas chamadas normas e co-normas. As normas fuzzy são também chamadas de t-normas e as co-normas de s-normas (Pedrycz e Gomide, 1998).

DEFINIÇÃO A.2 A t-norma é uma operação binária $t : [0, 1]^2 \rightarrow [0, 1]$, satisfazendo as seguintes propriedades:

1. Comutativa: $x t y = y t x$;
2. Associativa: $x t (y t z) = (x t y) t z$;
3. Monotônica: Se $x \leq y$ e $w \leq z$ então $x t w \leq y t z$;
4. Condições de Contorno: $0 t x = x$; $1 t x = x$.

DEFINIÇÃO A.3 Uma s-norma, também conhecido como co-norma triangular, é uma operação binária $s : [0, 1]^2 \rightarrow [0, 1]$, satisfazendo as propriedades:

1. Comutativa: $x s y = y s x$;
2. Associativa: $x s (y s z) = (x s y) s z$;

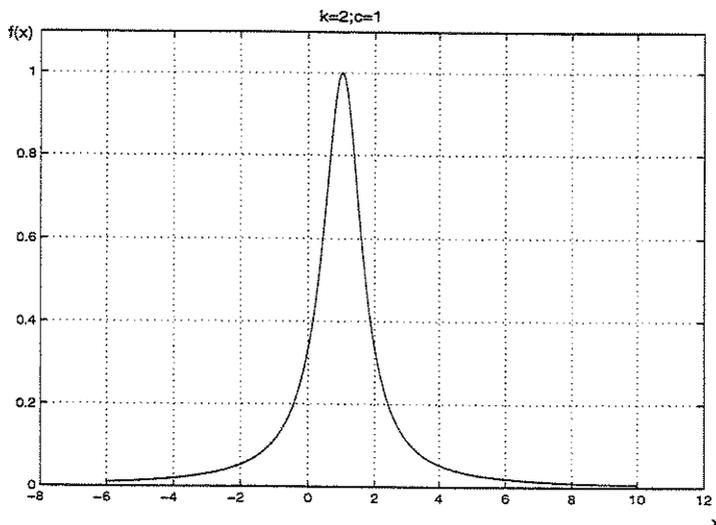


Figura A.2: Função de Pertinência Gaussiana.

3. Monotônica: Se $x \leq y$ e $w \leq z$ então $x s w \leq y t z$;
4. Condições de Contorno: $x s 0 = x$; $x t 1 = 1$.

Pelas propriedades descritas, pode-se notar que o operador \min (\wedge) é uma t-norma e o operador \max (\vee) é uma s-norma. Em (Pedrycz e Gomide, 1998) são apresentados outros exemplos de normas triangulares.

As operações sobre conjuntos nebulosos podem ser interpretadas como conectivos lógicos e as t-normas e s-normas podem ser identificadas como conectivos lógicos conjuntivos (AND) e disjuntivos (OR), respectivamente.

Uma base de regras nebulosas do tipo se-então é formada a partir dos conjuntos nebulosos e das variáveis associadas a estes conjuntos. Assim, realizando um mecanismo de inferência nebulosa é possível obter conclusões para fatos a partir de uma base de regras.

DEFINIÇÃO A.4 Um sistema nebuloso é composto por uma base de regras e um mecanismo de inferência nebulosa. Um sistema nebuloso faz o mapeamento $U \rightarrow V$ de um espaço de entrada para um espaço de saída, onde $U = U_1 \times \dots \times U_n \subset \mathfrak{R}^n$, $V \subset \mathfrak{R}$. A base de regras é constituída de M regras do tipo:

$R^{(l)} : \text{Se } x_1 \text{ é } F_1^l \text{ E } x_2 \text{ é } F_2^l \text{ E } \dots x_n \text{ é } F_n^l \text{ Então } y \text{ é } g^l$

O mecanismo de defuzzificação possui a mesma finalidade que o mecanismo de fuzzificação, só que no sentido inverso. Existem vários métodos para gerar as saída do sistema nebuloso (Pedrycz e Gomide, 1998). O modelo de rede neural nebulosa estudado neste trabalho utiliza o método do centro de gravidade (Figueiredo, 1997).

Os conjuntos nebulosos podem ser comparados de várias maneiras. A comparação dos conjuntos nebulosos é realizada, neste trabalho, através da medida de possibilidade, definida como segue (Pedrycz e Gomide, 1998).

DEFINIÇÃO A.5 A medida de possibilidade (Zadeh, 1978) de um conjunto nebuloso A com relação a um conjunto nebuloso B , denotado por $Poss(A, B)$ é definida como:

$$Poss(A, B) = \sup_{x \in X} [\min(A(x), B(x))]$$

UNICAMP
BIBLIOTECA CENTRAL
SEÇÃO CIRCULANTE