



UNIVERSIDADE ESTADUAL DE CAMPINAS  
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO  
DEP. DE ENGENHARIA DE COMPUTAÇÃO E AUTOMAÇÃO INDUSTRIAL

# **Redes Neurais, Metodologias de Agrupamento e Combinação de Previsores Aplicados à Previsão de Vazões Naturais**

Dissertação de Mestrado

Autor: **Marina Hirota Magalhães**

Orientador: **Prof. Dr. Fernando Gomide**

Co-orientador: **Prof<sup>a</sup>. Dr<sup>a</sup>. Rosângela Ballini**

Banca Examinadora: **Prof. Dr. Pyramo Pires da Costa Júnior**

**Prof. Dr. Secundino Soares Filho**

**Prof. Dr. Wagner Caradori do Amaral**

Dissertação submetida à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, como pré-requisito parcial para a obtenção do Título de Mestre em Engenharia Elétrica.

Novembro de 2004

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

M27r Magalhães, Marina Hirota  
Redes neurais, metodologias de agrupamento e combinação de previsores aplicados à previsão de vazões naturais / Marina Hirota Magalhães.--Campinas, SP: [s.n.], 2004.

Orientadores: Fernando Gomide, Rosangela Ballini.  
Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Sistemas difusos – Processamento de dados. 2. Redes neurais (Computação). 3. Análise de séries temporais – Processamento de dados. 4. Previsão hidrológica. I. Gomide, Fernando. II. Ballini, Rosangela. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

# Resumo

Planejamento de sistemas hidroelétricos possui um alto grau de complexidade e dificuldade, uma vez que envolve características de produção não lineares e depende de muitas variáveis. Um das variáveis chave é a vazão natural. Os valores de vazões devem ser previstos com acurácia, uma vez que esses valores influenciam significativamente na produção de energia. Atualmente, no setor de geração hidroelétrica, a previsão de vazões é baseada na metodologia de Box & Jenkins. Este trabalho propõe um modelo de previsão baseado em agrupamento nebuloso como alternativa para a previsão de vazões naturais médias mensais. O modelo utiliza o algoritmo de agrupamento *fuzzy c-means* para explorar a estrutura dos dados históricos, e procedimentos de mediana e reconhecimento de padrões para capturar similaridades na tendência das séries.

Ainda, este trabalho sugere um modelo que combina previsões geradas por um conjunto de métodos individuais de previsão, de uma maneira simples, mas efetiva. Utiliza-se, como combinador, uma rede neural treinada com o algoritmo do gradiente. O objetivo é combinar as previsões geradas por diferentes modelos na tentativa de capturar as contribuições das características de previsão mais importantes de cada previsor individual. Esse método também é aplicado à previsão de séries de vazões naturais médias mensais escolhendo-se, como modelos individuais, aqueles que obtiveram melhor desempenho para uma dada série.

Resultados experimentais com dados reais de vazão sugerem que o modelo preditivo baseado em agrupamento nebuloso obtém um desempenho superior, quando comparado com a metodologia atual de previsão de vazões adotada pelo setor hidroelétrico, e, ainda, com uma rede neural nebulosa, um modelo não linear. Além disso, o modelo de combinação alcança um desempenho superior que os modelos de previsão individuais, pois apresentam erros de previsão menores.

# Abstract

Planning of hydroelectric systems is a complex and difficult task once it involves non-linear production characteristics and depends on numerous variables. A key variable is the natural streamflow. Streamflow values covering the entire planning period must be accurately forecasted because they strongly influence energy production. Currently, streamflow prediction using Box & Jenkins methodology prevails in the electric power industry. This work proposes a fuzzy prediction model based on fuzzy clustering as an alternative for natural average monthly streamflow forecast. The model uses fuzzy c-means clustering to explore past data structure, and a median and pattern recognition procedures to capture similarities between streamflow history and data used for prediction.

In addition, this work suggests a linear approach to combine forecasts generated by a set of individual forecasting models in a simple and effective way. We use, as a combiner, a neural network trained with the gradient descent algorithm. The aim is to combine the forecasts generated by the different forecasting models as an attempt to capture the contributions of the most important prediction features of each individual model at each prediction step. The approach is also used for streamflow time series prediction choosing, as individual forecasting models, the most promising predictive methods.

Experimental results with actual data suggest that the predictive clustering approach performs globally better than the current streamflow forecasting methodology adopted by many hydroelectric systems worldwide, and a fuzzy neural network, a non-linear prediction model. The combination approach, with lower prediction errors, performs better than each of the individual forecasting models.



# Agradecimentos

À Deus pela luz que me auxiliou a chegar até aqui.

À minha família pelo amor e apoio.

Aos professores Fernando Gomide e Rosangela Ballini pela orientação, confiança e pelo exemplo de profissionalismo e seriedade.

Aos professores Wagner Amaral, Fernando Von Zuben, Leandro Nunes e Secundino Soares pelo auxílio durante o desenvolvimento do trabalho.

Ao amigo, amante e companheiro Tiago Vasconcellos pela compreensão, alegria, amor, carinho e apoio de todos os momentos.

Às amigas Cris e Gerusa por terem compartilhado comigo a minha trajetória Unicamp até hoje.

Aos amigos Igor, Giselle, Ivette, Thyago e Ricardo pela ajuda nos momentos de dúvida.

Aos amigos Rodrigo, Tiago, Cassiano, Thais, Duda e Virgínia por estarem sempre por perto...

À Bila por me mostrar que a vida pode ser leve e divertida.

Às amigas do time de volleyball da Unicamp pelos momentos de lazer e alegria.

Aos amigos do LCA e do COSE pelo companheirismo do dia-a-dia.

Aos professores do IMECC-UNICAMP, especialmente Cheti, Joni, Edmundo, Jayme, Filidor e Letelier por continuarem me auxiliando em minhas pesquisas.

A todos que eu não citei, mas que direta ou indiretamente passaram ou estão em minha vida e que por isso me ensinaram ou ainda me ensinam a ser uma pessoa melhor. Muito obrigada.

“O amor é a vida acontecendo no momento: sem passado, sem futuro, presente puro, eternidade numa bolha de sabão... Ponho-me a brincar com a vida e uma estranha metamorfose acontece: deixo de ser velho.

Sou criança de novo...”

Rubem Alves

# Sumário

<b>SUMÁRIO</b>	<b>vi</b>
<b>LISTA DE FIGURAS</b>	<b>ix</b>
<b>LISTA DE TABELAS</b>	<b>xii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação e Relevância . . . . .	1
1.2 Descrição do Problema de Previsão de Vazões . . . . .	2
1.3 Objetivos e Organização do Trabalho . . . . .	3
<b>2 Modelos Clássicos para Previsão de Séries Temporais</b>	<b>5</b>
2.1 Introdução . . . . .	5
2.2 Conceitos Básicos de Processos Estocásticos . . . . .	6
2.3 Modelos Lineares Estacionários . . . . .	8
2.3.1 Modelos Médias Móveis . . . . .	10
2.3.2 Modelos Autoregressivos . . . . .	11
2.3.3 Modelos Autoregressivos Médias Móveis . . . . .	11
2.4 Modelos Periódicos Autoregressivos Médias Móveis (PARMA) . . . . .	12
2.5 Ajuste de Modelos PARMA para Vazões Naturais . . . . .	15
2.6 Resumo . . . . .	19

---

<b>3</b>	<b>Redes Neurais Artificiais</b>	<b>21</b>
3.1	Introdução . . . . .	21
3.2	Um Breve Histórico da Área de RNAs e sua Aplicação em Previsão . . . . .	22
3.3	Componentes e Treinamento de uma RNA . . . . .	24
3.3.1	O Neurônio Artificial e as Funções de Ativação . . . . .	24
3.3.2	Treinamento Supervisionado . . . . .	26
3.3.3	Treinamento Não Supervisionado . . . . .	27
3.4	Redes Neurais Multicamadas . . . . .	27
3.4.1	Aplicação da Rede MLP para Previsão de Vazões . . . . .	30
3.5	Redes Neurais Nebulosas . . . . .	33
3.5.1	Aplicação da Rede Neural Nebulosa para Previsão de Vazões . . . . .	42
3.6	Resumo . . . . .	44
<b>4</b>	<b>Agrupamento de Dados em Previsão</b>	<b>45</b>
4.1	Introdução . . . . .	45
4.2	O Algoritmo <i>Fuzzy C-Means</i> (FCM) . . . . .	46
4.3	Modelo de Previsão de Séries Temporais Baseado em Agrupamento Nebuloso	48
4.3.1	Procedimento de Mediana de Dados (PMed) . . . . .	50
4.3.2	Procedimento de Reconhecimento de Padrões (PRP) . . . . .	51
4.4	Aplicação do Modelo baseado em Agrupamento FCM para Previsão de Vazões	53
4.5	Resumo . . . . .	61
<b>5</b>	<b>Combinação de Previsores</b>	<b>63</b>
5.1	Introdução . . . . .	63
5.2	Modelo de Combinação baseado em Redes Neurais . . . . .	65
5.3	Aplicação do Modelo de Combinação para Previsão de Vazões . . . . .	67
5.4	Resumo . . . . .	69

---

<b>6</b>	<b>Comparação e Discussão de Resultados</b>	<b>71</b>
6.1	Introdução . . . . .	71
6.2	Aplicação de Modelos para o Posto de Sobradinho . . . . .	73
6.2.1	Análise de Erros . . . . .	73
6.3	Aplicação de Modelos para o Posto de Furnas . . . . .	76
6.3.1	Análise de Erros . . . . .	82
6.4	Aplicação de Modelos para o Posto de Emborcação . . . . .	84
6.4.1	Análise de Erros . . . . .	89
6.5	Resumo . . . . .	91
<b>7</b>	<b>Conclusões</b>	<b>93</b>
7.1	Trabalhos Futuros . . . . .	95
<b>A</b>	<b>Tabelas de Erros Mensais de Previsão</b>	<b>97</b>
	<b>Referências Bibliográficas</b>	<b>102</b>

# Lista de Figuras

2.1	Série temporal como saída de um filtro linear. . . . .	9
2.2	Histórico de vazões: Sobradinho no período de 1931 à 1998. . . . .	15
2.3	Série estacionária das vazões de Sobradinho. . . . .	16
2.4	Histograma da série estacionária para a série de vazões de Sobradinho. . .	17
2.5	Previsão de vazões do posto de Sobradinho utilizando o modelo PARMA para o período de 1991 a 1998. . . . .	19
3.1	Modelo de neurônio artificial. . . . .	24
3.2	Arquitetura de uma rede MLP. . . . .	28
3.3	Previsão de vazões do posto de Sobradinho utilizando o modelo MLP( $m$ ) para o período de 1991 a 1998. . . . .	32
3.4	Modelo geral de neurônio nebuloso. . . . .	35
3.5	Arquitetura da rede neural nebulosa. . . . .	36
3.6	Previsão de vazões do posto de Sobradinho utilizando o modelo de RNN para o período de 1991 a 1998. . . . .	43
4.1	Grupos obtidos através do algoritmo FCM para PMed. Os círculos são os centros dos grupos e cada forma geométrica representa um grupo. . . . .	54
4.2	Previsão de vazões para setembro utilizando o PMed. As linhas contínuas e tracejadas representam os valores reais e previstos, respectivamente. . . . .	55

---

4.3	Grupos obtidos através do algoritmo FCM para o PRP. Os círculos são os centros dos grupos e cada forma geométrica representa um grupo. . . . .	56
4.4	Tendência da curva de vazões para Setembro. As linhas contínuas e tracejadas representam, respectivamente, os valores de $p_q$ e $p_j$ . . . . .	57
4.5	Previsão de vazões para setembro utilizando o PRP. As linhas contínuas e tracejadas representam os valores reais e previstos, respectivamente. . . . .	58
4.6	Previsão de vazões utilizando o modelo baseado em agrupamento com PMed para o período entre 1991 e 1998 do posto de Sobradinho. . . . .	60
4.7	Previsão de vazões utilizando o modelo baseado em agrupamento com PRP para o período entre 1991 e 1998 do posto de Sobradinho. . . . .	60
5.1	Modelo de combinação. . . . .	65
5.2	Agregador neural. . . . .	66
5.3	Previsão de vazões utilizando o modelo de combinação neural para o período entre 1991 e 1998 do posto de Sobradinho. . . . .	68
6.1	Erros mensais referentes ao posto de Sobradinho. . . . .	74
6.2	Previsão de vazões utilizando o modelo baseado em agrupamento com PMed para o período entre 1991 e 1998 do posto de Furnas. . . . .	79
6.3	Previsão de vazões utilizando o modelo baseado em agrupamento com PRP para o período entre 1991 e 1998 do posto de Furnas. . . . .	80
6.4	Previsão de vazões utilizando o modelo de combinação de previsores para o período entre 1991 e 1998 do posto de Furnas. . . . .	81
6.5	Erros mensais referentes ao posto de Furnas. . . . .	82
6.6	Previsão de vazões utilizando o modelo baseado em agrupamento com PMed para o período entre 1991 e 1998 do posto de Emborcação. . . . .	86
6.7	Previsão de vazões utilizando o modelo baseado em agrupamento com PRP para o período entre 1991 e 1998 do posto de Emborcação. . . . .	87

---

6.8	Previsão de vazões utilizando o modelo de combinação de previsores para o período entre 1991 e 1998 do posto de Emborcação. . . . .	89
6.9	Erros mensais referentes ao posto de Emborcação. . . . .	90

## Lista de Tabelas

2.1	Médias e desvios padrão mensais da série de vazões de Sobradinho. . . . .	16
2.2	Valores do critério BIC e da variância para o mês de junho para o posto de Sobradinho. . . . .	18
2.3	Estimativas dos parâmetros utilizando o método EMV para a série de vazões de Sobradinho. . . . .	18
3.1	Topologias das redes MLP escolhidas para cada mês. . . . .	31
3.2	Topologias das RNN escolhidas para cada mês. . . . .	42
3.3	Parâmetros da aprendizagem. . . . .	43
4.1	Valores reais normalizados e aproximações $v_9$ dadas pelo procedimento PRP.	58
4.2	Características para o modelo MPAN com PMed e PRP. . . . .	59
5.1	Pesos resultantes da combinação neural para o posto de Sobradinho. . . . .	68
6.1	Erros globais de previsão para o posto de Sobradinho. . . . .	75
6.2	Estimativas dos parâmetros utilizando o EMV: série de vazões de Furnas. .	76
6.3	Valores do critério BIC e da variância para o mês de setembro para o posto de Furnas. . . . .	77
6.4	Topologias das redes MLP para cada mês: posto de Furnas. . . . .	77
6.5	Topologias das RNN para cada mês: posto de Furnas. . . . .	78
6.6	Características para o modelo MPAN com PRP e PMed: Furnas. . . . .	79
6.7	Pesos resultantes da combinação neural para o posto de Furnas. . . . .	81
6.8	Erros globais de previsão para o posto de Furnas. . . . .	84
6.9	Estimativas dos parâmetros utilizando o EMV: série de vazões de Emborcação.	85

---

6.10	Valores do critério BIC e da variância para o mês de julho para o posto de Emborcação. . . . .	85
6.11	Topologias das redes MLP para cada mês: posto de Emborcação. . . . .	86
6.12	Topologias das RNN para cada mês: posto de Emborcação. . . . .	86
6.13	Características para o modelo MPAN com PRP e PMed: Emborcação. . . . .	87
6.14	Pesos resultantes da combinação neural para o posto de Emborcação. . . . .	88
6.15	Erros globais de previsão para o posto de Emborcação. . . . .	91
A.1	Erros mensais de previsão dos modelos PARMA, MLP e RNN para o posto de Sobradinho. . . . .	97
A.2	Erros mensais de previsão dos modelos MPAN-PMed, MPAN-PRP e CP para o posto de Sobradinho. . . . .	98
A.3	Coefficientes de correlação mensais ( $\rho$ ) para o posto de Sobradinho. . . . .	98
A.4	Erros mensais de previsão dos modelos PARMA, MLP e RNN para o posto de Furnas. . . . .	99
A.5	Erros mensais de previsão dos modelos MPAN-PMed, MPAN-PRP e CP para o posto de Furnas. . . . .	99
A.6	Coefficientes de correlação mensais ( $\rho$ ) para o posto de Furnas. . . . .	100
A.7	Erros mensais de previsão dos modelos PARMA, MLP e RNN para o posto de Emborcação. . . . .	100
A.8	Erros mensais de previsão dos modelos MPAN-PMed, MPAN-PRP e CP para o posto de Emborcação. . . . .	101
A.9	Coefficientes de correlação mensais ( $\rho$ ) para o posto de Emborcação. . . . .	101



# Capítulo 1

## Introdução

### 1.1 Motivação e Relevância

Há, atualmente, uma vasta literatura que descreve a previsão de séries temporais em diversas áreas, com exemplos que abrangem desde aplicações em medicina, medindo a irregularidade das batidas do coração; passam pelas pesquisas em hidrologia, prevenindo variáveis de recursos hídricos; e chegam até as aplicações em economia, observando o comportamento da volatilidade do mercado financeiro.

Muitos modelos são baseados na metodologia Box & Jenkins, a qual assume relações lineares entre as variáveis do sistema (Box et al., 1994). Entretanto, séries temporais, que descrevem uma seqüência real de valores, geralmente possuem um comportamento não linear e, nesse caso, as redes neurais artificiais (RNAs) e variantes têm sido largamente adotadas para prever valores futuros com base em valores passados (Gershenfeld e Weigend, 1994). Além disso, sistemas nebulosos vêm sendo combinados às RNAs com o objetivo de explorar a capacidade de representação e o poder de aprendizagem dessas abordagens, respectivamente (Pedrycz e Gomide, 1998).

Por outro lado, o agrupamento de dados, reconhecido como um dos paradigmas mais importantes para o reconhecimento e classificação de padrões, é empregado em problemas de previsão, com o objetivo de explorar a estrutura de dados históricos da série temporal. Este trabalho propõe um modelo de previsão baseado em agrupamento nebuloso de dados que, ao contrário dos métodos já desenvolvidos na literatura (Geva, 1999), (Shalizi et al., 2004), utiliza procedimentos de reconhecimento de padrões para

capturar similaridades entre o histórico de vazões e os dados utilizados para a previsão. A contribuição do modelo sugerido também inclui menor complexidade e menor tempo de execução do que os métodos baseados em redes neurais e redes neurais nebulosas.

Complementarmente, baseando-se na idéia de combinar previsores, para aprimorar o desempenho dos sistemas de previsão individuais, propõe-se um modelo linear simples, porém efetivo e pouco custoso para explorar as principais características de modelos promissores.

Os modelos propostos foram aplicados à previsão **um passo à frente** de séries temporais de vazões **naturais médias mensais**, um problema de grande importância no contexto do planejamento energético brasileiro.

## 1.2 Descrição do Problema de Previsão de Vazões

O planejamento da operação de sistemas de produção de energia elétrica visa determinar uma política de operação para usinas hidroelétricas e termoelétricas e para os intercâmbios, inclusive internacionais, do sistema que atenda à demanda de energia elétrica de forma econômica e confiável (Ballini, 2000).

Atividades de planejamento e operação de sistemas de energia e recursos hídricos requerem a utilização de muitas variáveis de entrada. No caso do planejamento de sistemas de energia brasileiro, predominantemente hidroelétrico, as previsões de vazões de curto e longo prazos são necessárias para propósitos de simulação, otimização e tomada de decisões.

Particularmente no Brasil, muitos sistemas hidroelétricos envolvem regiões distintas, com características geográficas bastante diferentes, com uma coleta de dados hidrométricos que abrange redes de aquisição de dados distintas e largamente espaçadas. Isso resulta numa incerteza considerável da informação hidrológica disponível. Além disso, a relação não-linear entre o fluxo de entrada e a saída desafia os modelos de previsão de vazões consideravelmente. Outra dificuldade dessa aplicação é a natureza não-estacionária e sazonal da série devido aos períodos intercalados de seca e humidade ao longo do ano (Maier e Dandy, 2000).

Atualmente, o planejamento energético brasileiro adota modelos baseados na

metodologia Box & Jenkins (Box et al., 1994) para a previsão de vazões, o que pode comprometer a qualidade dos resultados, implicando cenários menos confiáveis para a operação do sistema hidroelétrico. Abordagens capazes de capturar as incertezas, não-linearidades e padrões existentes ao longo da série de vazões, tais como modelos baseados em agrupamento nebuloso de dados, aparecem como alternativas promissoras.

### **1.3 Objetivos e Organização do Trabalho**

O objetivo deste trabalho se divide no desenvolvimento de duas abordagens de previsão. A primeira trata modelos de previsão de séries temporais baseados em agrupamento nebuloso de dados, que capacitam o sistema explorar a estrutura dos elementos anteriores da série, isto é, a estrutura dos dados históricos disponíveis.

A segunda abordagem se refere à combinação de modelos preditivos que apresentem desempenho individual promissor em determinados pontos da série. Dessa forma, o combinador deve ser capaz de determinar quais modelos locais são relevantes para uma dada previsão. O combinador construído, nesta dissertação, é uma rede neural composta por apenas uma camada de processamento, constituída por um neurônio linear, com treinamento via gradiente descendente. Trata-se, portanto, de um combinador linear.

O trabalho é dividido em sete capítulos. Após esta introdução, nos Capítulos 2 e 3, apresentam-se os principais modelos de previsão atualmente utilizados para a previsão de vazões, com os quais os modelos propostos serão comparados posteriormente. No Capítulo 2, apresenta-se conceitos básicos de processos estocásticos, os modelos de Box & Jenkins e a aplicação destes modelos à previsão de vazões. Por outro lado, no Capítulo 3, os modelos de redes neurais multicamadas (*MultiLayer Perceptron - MLP*) com aprendizagem supervisionada via algoritmo de retropropagação (*backpropagation*) e uma classe de redes neurais nebulosas com aprendizagem construtiva e competitiva são abordados.

A seguir, nos Capítulos 4 e 5 os modelos propostos são apresentados. No Capítulo 4, os principais trabalhos relacionando agrupamento de dados e previsão são discutidos e os modelos baseados em agrupamento nebuloso sugeridos são descritos e aplicados à previsão das séries de vazões. Analogamente, no Capítulo 5, um breve histórico na

área de combinação de previsores é apresentado e o modelo de combinação neural de previsores é detalhado com a respectiva aplicação à série de vazões.

No Capítulo 6, as metodologias são comparadas através da análise de erros de previsão para valores de vazões naturais médias mensais coletadas de três postos distintos, localizados em diferentes regiões do Brasil. Finalmente, no Capítulo 7 conclui-se o trabalho resumindo as principais contribuições e pontos de partida para desenvolvimentos futuros.

## Capítulo 2

# Modelos Clássicos para Previsão de Séries Temporais

Os modelos clássicos para análise e previsão de séries temporais são ferramentas bastante conhecidas na literatura. Tais modelos podem ser divididos em dois grandes grupos: métodos paramétricos e não-paramétricos. Em se tratando do contexto de modelagem paramétrica, as metodologias propostas por Box & Jenkins (Box et al., 1994) têm sido largamente utilizadas para previsões nas mais diversas áreas, como por exemplo, economia, finanças, meteorologia e hidrologia.

As séries de vazões apresentam características estatísticas uma vez que não podem ser determinadas por uma equação matemática exata, sendo, portanto, processos estocásticos. Assim, os modelos abordados neste capítulo apresentam características estocásticas e são baseados na metodologia de Box & Jenkins.

### 2.1 Introdução

De acordo com Box, Jenkins & Reinsel (Box et al., 1994):

**Definição 1** *Série Temporal é um conjunto de observações  $\{x_t, t \in T \subset \mathbb{R}_+\}$ ,  $T = \{t \mid 1 \leq t \leq N\}$  geradas sequencialmente no tempo. Se o conjunto é contínuo, a série é dita ser contínua e se o conjunto é discreto, a série é dita ser discreta.*

Se os valores futuros de uma série temporal são exatamente determinados por funções matemáticas, a série é dita ser determinística. Se os valores são descritos em termos de uma distribuição de probabilidade, a série temporal é não-determinística, ou, simplesmente, uma série temporal estocástica.

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

Neste trabalho, serão consideradas séries de vazões naturais afluentes médias mensais cujos dados foram observados somente para valores inteiros de  $t$ , ou seja,  $t \in \mathbb{Z}_+$ , onde  $\mathbb{Z}_+$  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 seguir, são apresentados alguns conceitos de processos estocásticos que ajudarão a compreender a descrição dos modelos utilizados para previsão das séries de vazões.

## 2.2 Conceitos Básicos de Processos Estocásticos

**Definição 2** *Seja  $\mathbb{Z}_+$  o conjunto dos números inteiros positivos. Um processo estocástico é uma família  $\{x_t, t \in \mathbb{Z}_+\}$ , tal que, para cada  $t \in \mathbb{Z}_+$ ,  $x_t$  é uma variável aleatória.*

Logo, uma série temporal a ser analisada é considerada como um resultado particular, produzido por um mecanismo de probabilidade intrínseco de um processo estocástico. Em outras palavras, uma observação é uma realização de uma variável aleatória  $x_t$  com função densidade de probabilidade  $p(x_t)$ . Da mesma forma, uma variável aleatória  $N$ -dimensional com função densidade de probabilidade conjunta  $p(x_1, x_2, \dots, x_N)$ ,  $t = 1, 2, \dots, N$ , pode descrever uma série temporal (Box et al., 1994).

**Definição 3** *Um processo estocástico é estritamente estacionário se suas propriedades não são afetadas pela mudança na origem do tempo, isto é, se a distribuição de probabilidade conjunta associada às  $k$  observações  $x_1, x_2, \dots, x_k$ ,  $1 < k < N$ , é a mesma associada*

às  $k$  observações  $x_{1+h}, x_{2+h}, \dots, x_{k+h}$ , para qualquer  $h$ . Em outras palavras, a escolha da origem no eixo  $t$  não afeta nenhum momento do processo.

Na prática, é impossível conhecer todas as distribuições conjuntas de  $x_1, x_2, \dots, x_k$ ,  $1 < k < N$ , e o processo fica restrito ao conceito de processo fracamente estacionário (ou estacionário de segunda ordem), que é aquele que obedece as seguintes condições:

- A média de  $x_t$  é constante, isto é,  $E[x_t] = \mu(t) = \mu$ , para todo  $t$ ;
- A variância de  $x_t$  é constante, isto é,  $Var[x_t] = \sigma^2(t) = \sigma^2$ , para todo  $t$ ;
- $Cov[x_t, x_{t+k}] = \gamma(k)$ , ou seja, é uma função que só depende de  $k$ .

A maioria das séries temporais contém componentes sazonais e/ou tendências, ou seja, possuem características não-estacionárias. Para uma série temporal sem estacionariedade sazonal, como o caso da série de vazões, uma transformação estatística é necessária para torná-la estacionária. Essa transformação nos dados é feita definindo-se uma nova variável:

$$z_t = \frac{x_t - \mu}{\sigma} \quad (2.1)$$

onde,  $\mu$  é a média dada por:

$$\mu = E[x_t] = \int_{-\infty}^{+\infty} x_t p(x_t) dx_t \quad (2.2)$$

e  $\sigma = \sqrt{\sigma^2}$  é o desvio padrão. Este é a raiz quadrada da variância a qual é dada por:

$$\sigma^2 = E[x_t - \mu]^2 = \int_{-\infty}^{+\infty} (x_t - \mu)^2 p(x_t) dx_t \quad (2.3)$$

Devido à estacionariedade induzida por (2.1),  $p(z_t)$  se reduz a  $p(z)$ , uma vez que a distribuição de probabilidade para esta nova variável é a mesma para todo tempo  $t$ .

Considerando que o processo é discretizado no tempo, a forma da distribuição de probabilidade,  $p(z)$ , pode ser descrita de acordo com os dados observados  $z_1, z_2, \dots, z_N$ . Assim, a média e a variância estimadas por:

$$\bar{z} = \hat{\mu} = \frac{1}{N} \sum_{t=1}^N z_t \quad (2.4)$$

$$\text{Var}[z] = \widehat{\sigma^2} = \frac{1}{N-1} \sum_{t=1}^N (z_t - \bar{z})^2 \quad (2.5)$$

são denominadas média e variância amostral, respectivamente.

A transformação (2.1) é aplicada para que a classe de modelos lineares estacionários possam ser empregados, uma vez que estes exigem esforços estatísticos e matemáticos menores dos que os apresentados pelos métodos não-estacionários.

Os modelos lineares estacionários estudados neste trabalho são desenvolvidos em três estágios (Box et al., 1994):

- **Identificação** – utilizar os dados para obter informações sobre a série e, assim, sugerir uma classe de modelos.
- **Estimação** – utilizar os dados para realizar inferências sobre os parâmetros, adequando-os ao modelo proposto.
- **Verificação** – aplicar o modelo ajustado aos dados e verificar seu grau de adequação, buscando contínua melhora através do ajuste dos parâmetros.

Assim, dependendo do desempenho do modelo testado na fase de verificação, outros modelos da classe pré-definida no primeiro estágio serão escolhidos para a mesma avaliação, de acordo com os três estágios descritos acima.

## 2.3 Modelos Lineares Estacionários

A classe de modelos lineares estacionários é composta por aqueles que podem ser conceitualmente detalhados e facilmente implementados (Gershenfeld e Weigend, 1994). Isso porque o processo estocástico, gerador das séries temporais, pode ser gerado por um filtro linear, cuja função de transferência é  $\Psi(B)$ , como mostra a Figura 2.1. As entradas  $a_t$  do sistema são valores aleatórios independentes cuja função de distribuição de probabilidade é normal com média igual a 0, variância constante e sem covariância entre valores referentes a dois instantes distintos, isto é:

$$E[a_t] = 0, \quad (2.6)$$

$$\text{Var}[a_t] = E[a_t^2] = \sigma_a^2 \quad (2.7)$$

$$\gamma_k = \text{cov}[a_t, a_{t+k}] = 0, \quad k \neq 0 \quad (2.8)$$

Esta seqüência de valores aleatórios é chamada ruído branco (Box et al., 1994). Assim, um possível modelo para uma série temporal é  $z_t = \Psi(B)a_t$ ,  $t = 1, \dots, N$ , onde  $a_t$  é um ruído branco não correlacionado com  $\Psi(B)$ , que é uma determinada função do tempo.

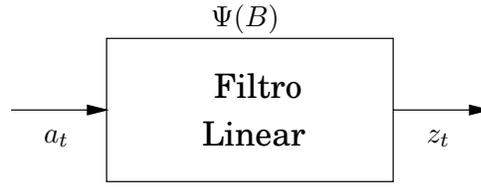


Figura 2.1: Série temporal como saída de um filtro linear.

A partir dessa abordagem, segue a formalização matemática:

$$z_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} = \Psi(B)a_t, \quad (2.9)$$

onde  $B$  é o operador de atraso, definido como:

$$Bz_t = z_{t-1}; \quad B^m z_t = z_{t-m}, \quad (2.10)$$

e a função de transferência  $\Psi(B)$  é definida como:

$$\Psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots = \sum_{j=0}^{\infty} \psi_j B^j; \quad \psi_0 = 1 \quad (2.11)$$

A condição de estacionariedade do modelo linear (2.9) está associada à série  $\Psi(B)$ . Em (Box et al., 1994), é demonstrado que para um processo linear, as condições de estacionariedade dependem da convergência da série  $\Psi(B)$ , isto é, para  $\sum_{j=1}^{\infty} |\psi_j| < \infty$ , a qual é verificada se  $|B| \leq 1$ , isto é, para valores dentro do círculo unitário (Ballini, 2000).

O modelo definido pela equação (2.9), também pode ser descrito como uma aproximação linear de valores anteriores  $z_{t-1}, z_{t-2}, \dots$  mais um ruído branco  $a_t$ , isto é,

$$\begin{aligned} z_t &= \pi_1 z_{t-1} + \pi_2 z_{t-2} + \dots + a_t \\ &= \sum_{j=1}^{\infty} \pi_j z_{t-j} + a_t \end{aligned} \quad (2.12)$$

Logo,

$$a_t = z_t - \sum_{j=1}^{\infty} \pi_j z_{t-j} = \Pi(B)z_t \quad (2.13)$$

Combinando as equações (2.9) e (2.13)

$$a_t = \Pi(B)\Psi(B)a_t \quad (2.14)$$

Simplificando,

$$\Pi(B) = \Psi(B)^{-1} \quad (2.15)$$

Assim como os parâmetros  $\psi_j$  estão relacionados à condição de estacionariedade, os parâmetros  $\pi_j$  estão relacionados à condição de invertibilidade. Analogamente, o modelo linear é inversível se a série  $\Pi(B)$  convergir, para  $|B| \leq 1$  (Box et al., 1994).

Entre os modelos lineares estacionários, os mais usados são os processos *médias móveis* (MA), *autoregressivos* (AR) e *autoregressivos médias móveis* (ARMA). A seguir, esses modelos são sucintamente descritos.

### 2.3.1 Modelos Médias Móveis

Seguindo o modelo descrito em (2.9), dada uma entrada  $a_t$  e uma função de transferência  $\Theta(B)$ , deseja-se gerar outra série  $\{z_t\}$ , do seguinte modo<sup>1</sup>:

$$z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} = a_t + \sum_{k=1}^q \theta_k a_{t-k} = \Theta(B)a_t \quad (2.16)$$

Assim, o valor atual de  $z_t$  é influenciado pelos valores atual e passados de  $a_t$ .

Essa equação descreve um filtro de convolução, isto é, a nova série  $z_t$  é gerada por um filtro de ordem  $q$  com os coeficientes  $\theta_i, i = 1, \dots, q$ , da série  $a_t$ . Essa relação é denominada *modelo médias móveis de ordem q* ou MA( $q$ ). Além desse nome, esse modelo é também conhecido como um *filtro de resposta ao impulso finito*, ou FIR, uma vez que esse filtro opera em laço aberto sem realimentação de forma que, a transformação realizada por ele é feita apenas com as entradas externas  $a_{t-k}, k = 1, \dots, q$  que lhe são fornecidas. Isso quer dizer que é garantido que a saída tenda a zero após  $q$  passos, assim que o sinal de entrada se anule (Gershenfeld e Weigend, 1994).

<sup>1</sup>A notação foi trocada de  $\Psi(B)$  para  $\Theta(B)$  para particularizar o modelo MA.

### 2.3.2 Modelos Autoregressivos

Diferentemente dos modelos  $MA(q)$ , os modelos autoregressivos são realimentados pelos dados  $z_{t-k}, k = 1, \dots, p$ , da série. Em outras palavras, a saída  $z_t$  pode ser continuamente processada mesmo depois que a entrada  $a_t$  cessa, uma vez que o modelo utiliza além da entrada externa, a memória interna do sistema. Formalmente,

$$z_t = a_t + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} = a_t + \sum_{k=1}^p \phi_k z_{t-k} \quad (2.17)$$

onde  $p$  é a ordem do modelo autoregressivo. Este modelo também é conhecido como *filtro de resposta ao impulso infinita* (IIR), uma vez que as saídas não são apenas uma transformação das entradas externas  $a_t$  mas também de uma dinâmica interna que envolve os valores  $z_{t-k}, k = 1, \dots, p$ . A expressão (2.17) define um *modelo autoregressivo de ordem  $p$* , ou  $AR(p)$ .

### 2.3.3 Modelos Autoregressivos Médias Móveis

Após um breve conhecimento dos conceitos envolvidos nos processos autoregressivos (AR) e médias móveis (MA), a idéia dos modelos *autoregressivos médias móveis* (ARMA) é facilmente compreendida, uma vez que este é uma combinação dos modelos previamente descritos nas seções 2.3.1 e 2.3.2. Matematicamente, o modelo ARMA é descrito como:

$$z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.18)$$

Em termos do operador de atraso  $B$ :

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) z_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \quad (2.19)$$

ou simplesmente,

$$\Phi(B) z_t = \Theta(B) a_t \quad (2.20)$$

onde,  $\Phi(B)$  e  $\Theta(B)$  são polinômios de graus  $p$  e  $q$ , respectivamente.

Para a modelagem de séries temporais que apresentam variação sazonal em sua estrutura, classes de processos periódicos, conhecidos como modelos periódicos, foram inicialmente propostas por (Jones e Brelsford, 1967) e posteriormente estudadas

por (Vecchia, 1985) e (Anderson e Vecchia, 1993). Esses modelos podem ser combinados aos processos  $AR(p)$  e  $ARMA(p, q)$ , originando, respectivamente, os modelos periódicos autoregressivos ( $PAR(p_m)$ ) e periódicos autoregressivos médias móveis ( $PARMA(p_m, q_m)$ ). A seguir, serão descritas as características principais desses modelos.

## 2.4 Modelos Periódicos Autoregressivos Médias Móveis (PARMA)

Os modelos periódicos autoregressivos médias móveis são extensões dos métodos  $ARMA(p, q)$ , com  $p \neq 0, q = 0$ , ( $PAR(p_m)$ ) ou  $p \neq 0, q \neq 0$ , ( $PARMA(p_m, q_m)$ ) e são destinados às séries temporais que apresentam variação em sua estrutura (Vecchia, 1985).

Seja uma série temporal sazonal  $\{x_t\}, t = 1, \dots, N$ , com período  $s$  e número total de observações  $N$ . Em outras palavras,  $m = 1, 2, \dots, s$  observações constituem um conjunto de dados que determina um período e  $r = 1, \dots, N$ , o índice de cada um desses conjuntos. Logo, o índice de tempo  $t$  pode ser escrito como função de  $r$  e  $m$ , isto é,  $t = t(r, m) = (r - 1)s + m$ .

Neste trabalho, são consideradas séries de vazões mensais com período de 12 meses, ou seja,  $s = 12$ ,  $r = 1, 2, \dots, n$  é o número de anos e  $m = 1, 2, \dots, 12$  representa os meses. Assim,  $x_{t(r, m)}$  é a observação da série  $x_t$  no ano  $r$  e mês  $m$ .

Nesse ponto, vale ressaltar que o conceito de estacionariedade é extensível para o caso periódico, conforme a definição 3. Dessa forma, o processo é periodicamente estacionário se a distribuição de probabilidade conjunta associada às  $k$  observações  $x_{1(r, m)}, x_{2(r, m)}, \dots, x_{k(r, m)}$ ,  $1 < k < N$  é a mesma associada às  $k$  observações  $x_{1+h(r, m)}, x_{2+h(r, m)}, \dots, x_{k+h(r, m)}$ , para qualquer  $h$ . Para eliminar o componente sazonal da série e torná-la periodicamente estacionária, realiza-se a seguinte mudança de variável para a padronização dos dados:

$$z_{t(r, m)} = \frac{x_{t(r, m)} - \mu_m}{\sigma_m} \quad (2.21)$$

onde  $\mu_m$  é a média e  $\sigma_m$  o desvio padrão do mês  $m$ .

Considerando a periodicidade da série, a idéia é ajustar  $s$  modelos  $ARMA(p_m, q_m)$ ,  $m = 1, \dots, s$ , ou seja, um modelo  $ARMA(p_m, q_m)$  para cada mês  $m$  do ano. Matematicamente,

mente, tais modelos são representados da seguinte forma:

$$z_{t(r,m)} = a_{t(r,m)} + \sum_{i=1}^{p_m} \phi_{i,m} z_{t(r,m)-i} - \sum_{j=1}^{q_m} \theta_{j,m} a_{t(r,m)-j} \quad (2.22)$$

onde  $\phi_{i,m}$  e  $\theta_{j,m}$  são funções periódicas de período  $s$ ,  $a_{t(r,m)}$  é um ruído branco com distribuição  $N(0, \tau_m^{-1})$ ,  $\tau_m^{-1} = \sigma_m^2$ ,  $\tau_m > 0$  e  $z_{t(r,m)}$  representa a série padronizada.

O próximo passo é estimar os parâmetros  $\phi_{i,m}$ ,  $\theta_{j,m}$  e  $\tau_m$  de modo que seus valores reflitam a relação existente entre os dados a serem previstos. Há diversas metodologias para essa estimação, tais como, o método dos momentos proposto por (Pagano, 1978), o método de inferência bayesiana proposto por (Salas et al., 1980) e o método de máxima verossimilhança (Vecchia, 1985). Neste trabalho, utilizou-se o método de máxima verossimilhança.

O método de máxima verossimilhança consiste em adotar, como estimativas dos parâmetros, os valores que maximizam a probabilidade (no caso da variável aleatória ser discreta) ou a densidade de probabilidade (no caso de variável contínua) da amostra observada ter sido obtida. Para obter estimadores de máxima verossimilhança, é necessário conhecer a distribuição da variável em estudo (Hoffmann, 1998).

Considerando a função densidade como uma gaussiana, o objetivo é maximizar a função de verossimilhança dada por:

$$L_m(\Phi_m, \Theta_m, \tau_m | Z_m) = \tau_m^{\frac{n_m}{2}} \exp \left\{ -\frac{\tau_m}{2} (Z_m - X_m \Phi_m + Y_m \Theta_m)^t (Z_m - X_m \Phi_m + Y_m \Theta_m) \right\} \quad (2.23)$$

onde,

$$n_m = \begin{cases} n - 1, & m = 1, \dots, p_m \\ n, & m = p_m + s; s = 1, 2, \dots, 12 \end{cases}$$

$$Z_m = \begin{pmatrix} z_{t(1,m)} \\ z_{t(2,m)} \\ \vdots \\ z_{t(n,m)} \end{pmatrix}_{n \times 1} ; \Phi_m = \begin{pmatrix} \phi_{1,m} \\ \phi_{2,m} \\ \vdots \\ \phi_{p_m,m} \end{pmatrix}_{p_m \times 1} ; \Theta_m = \begin{pmatrix} \theta_{1,m} \\ \theta_{2,m} \\ \vdots \\ \theta_{q_m,m} \end{pmatrix}_{q_m \times 1}$$

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

$$Y_m = \begin{pmatrix} a_{t(1,m)-1} & \cdots & a_{t(1,m)-q_m} \\ a_{t(2,m)-1} & \cdots & a_{t(2,m)-q_m} \\ \vdots & \ddots & \vdots \\ a_{t(n,m)-1} & \cdots & a_{t(n,m)-q_m} \end{pmatrix}_{n \times q_m}$$

Os parâmetros  $\Phi_m$ ,  $\Theta_m$  e  $\tau_m$  podem ser estimados de forma que a equação (2.23) seja maximizada. O procedimento segue tomando-se o logaritmo neperiano ( $\ln$ ) de ambos os membros da equação (2.23):

$$l_m(\Phi_m, \Theta_m, \tau_m | Z_m) = \frac{n_m}{2} \ln(\tau_m) - \frac{\tau_m}{2} (Z_m - X_m \Phi_m - Y_m \Theta_m)^t (Z_m - X_m \Phi_m + Y_m \Theta_m). \quad (2.24)$$

Derivando parcialmente com relação aos parâmetros  $\Phi_m$ ,  $\Theta_m$  e  $\tau_m$  e igualando a zero, os estimadores de máxima verossimilhança são dados por:

$$\hat{\Phi}_m = (X_m^t X_m)^{-1} (X_m^t Z_m) (X_m^t Y_m \hat{\Theta}_m); \quad (2.25)$$

$$\hat{\Theta}_m = (Y_m^t Y_m)^{-1} (Y_m^t Z_m) (X_m^t Y_m \hat{\Phi}_m); \quad (2.26)$$

$$\hat{\tau}_m^{-1} = \frac{1}{n_m} (Z_m - X_m \hat{\Phi}_m + Y_m \hat{\Theta}_m)^t (Z_m - X_m \hat{\Phi}_m + Y_m \hat{\Theta}_m). \quad (2.27)$$

A ordem do modelo pode ser selecionada através de diversos critérios propostos na literatura, entre os quais o mais explorado é o *Bayesian Information Criterium* (BIC), proposto por (Schwarz, 1978), dado pela equação:

$$BIC = N \log \hat{\sigma}_m^2 + M \log N, \quad (2.28)$$

onde  $M = p_m + q_m + 1$ . Esse critério incorpora a escolha de parcimônia sugerida por Box & Jenkins, isto é, utilizar um modelo com o menor número de parâmetros possível, penalizando caso ocorra o contrário (Francelin et al., 1997).

A partir de uma ordem inicial, calcula-se os estimadores de máxima verossimilhança, repetindo esse procedimento para ordens superiores. Através da minimização do critério de seleção BIC, determina-se, então, o modelo PARMA( $p_m, q_m$ ) mais parcimonioso a ser usado para cada mês.

## 2.5 Ajuste de Modelos PARMA para Vazões Naturais

Os dados utilizados consistem em uma série histórica de vazão natural afluyente média mensal que abrange o período de janeiro de 1931 a dezembro de 1998 (Figura 2.2), colhida em postos de medição de vazão, onde existe aproveitamento hidroelétrico. Esse conjunto de dados compõe um banco de vazões históricas mantido pelo ONS.

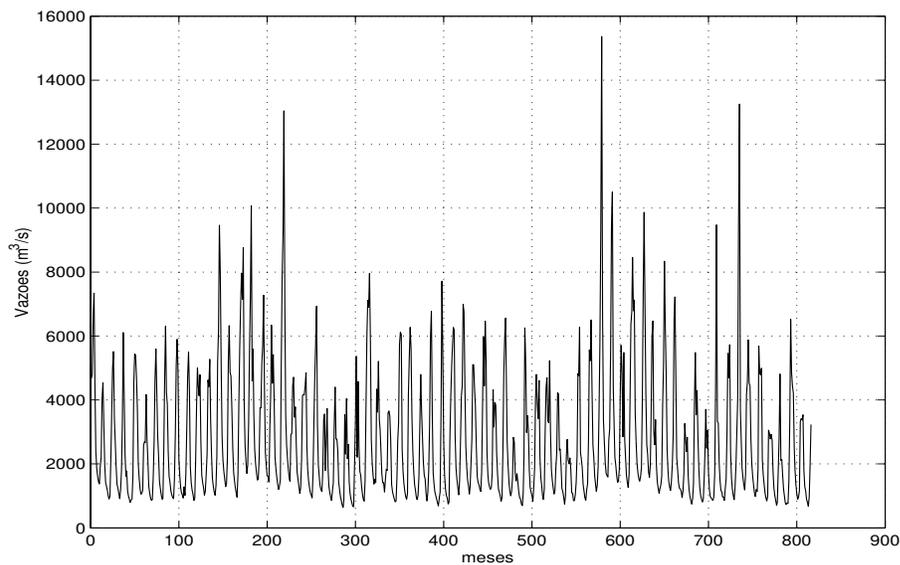


Figura 2.2: Histórico de vazões: Sobradinho no período de 1931 à 1998.

Para avaliar o desempenho do modelo descrito neste capítulo foram selecionados os últimos 8 anos (96 meses) do histórico de vazão, correspondente ao período de 1991 a 1998. Os parâmetros do modelo PARMA foram ajustados no período de janeiro de 1931 a dezembro de 1990.

A série de vazão natural média mensal tem como característica o comportamento periódico, em que os dados oscilam entre um valor mínimo, que ocorre nos meses secos, e um máximo, que ocorre nos meses de cheia, com uma variação sazonal, apresentando um período aproximado de 12 meses. Este comportamento pode ser observado na Figura 2.3 e na Tabela 2.1, em que são apresentados o desvio padrão e a média da série de vazão para o posto de Sobradinho, localizado na bacia do rio São Francisco, região nordeste do Brasil.

Para retirar o componente sazonal, a série é padronizada conforme a equação (2.21),

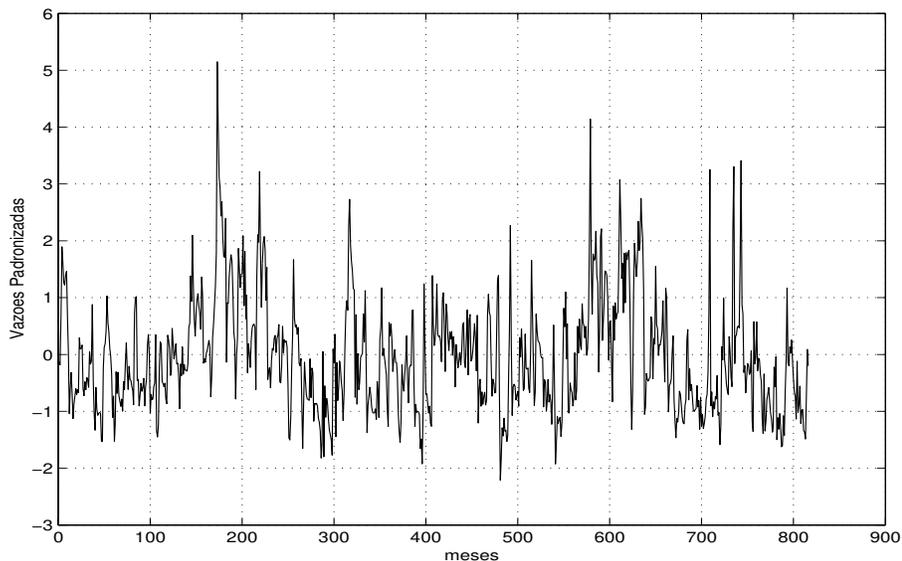


Figura 2.3: Série estacionária das vazões de Sobradinho.

Tabela 2.1: Médias e desvios padrão mensais da série de vazões de Sobradinho.

Meses	Média ( $\mu_m$ )	Desvio Padrão ( $\sigma_m$ )
Janeiro	4822,6	1407,3
Fevereiro	5119,4	2067,1
Março	4948,0	2611,4
Abril	3911,6	1738,9
Mai	2430,6	1168,9
Junho	1673,5	535,6
Julho	1376,9	374,7
Agosto	1178,1	293,2
Setembro	1041,6	258,3
Outubro	1164,9	297,9
Novembro	1899,1	703,2
Dezembro	3438,3	1178,9

tornando-se estacionária (Figura 2.3).

É importante ressaltar que, considerando o ajuste de um modelo  $PARMA(p_m, q_m)$  para previsão da série, são ajustados 12 diferentes modelos  $ARMA(p, q)$ , um para cada

mês do ano, de acordo com a equação (2.22).

A inferência dos parâmetros  $\Phi_m = \{\phi_{1,m} \cdots \phi_{p_m,m}\}$ ,  $\Theta_m = \{\theta_{1,m} \cdots \theta_{q_m,m}\}$  e  $\tau_m$  é realizada considerando os estimadores de máxima verossimilhança (EMV) descritos na seção (2.4). Para o caso das séries de vazões, utilizou-se a distribuição normal devido à análise do histograma representado na Figura 2.4. Porém deve-se notar que a aproximação pode não ser a mais adequada, já que a distribuição apresenta uma certa assimetria. Isso certamente acarretará em algumas dificuldades e perda das propriedades assintóticas nas estimativas dos parâmetros para ajuste dos modelos de séries temporais.

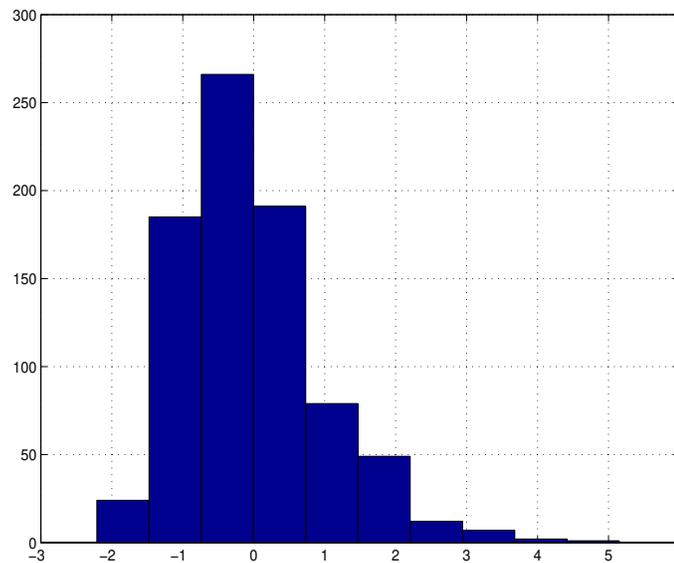


Figura 2.4: Histograma da série estacionária para a série de vazões de Sobradinho.

As ordens  $p_m$  e  $q_m$  do modelo PARMA foram escolhidas de acordo com o critério BIC definido na equação (2.28), que leva em consideração os números de parâmetros  $M$  e de dados  $N$ , além da variância  $\sigma_m^2 = \tau_m^{-1}$ ,  $\tau_m > 0$ , onde  $\tau_m$  é dado na Tabela 2.3. Considerando as ordens máximas  $p_m = 5$  e  $q_m = 1^2$ , verificou-se que para  $m = 6$  (junho), as ordens  $p_m = 3$  e  $q_m = 0$  tornam o modelo mais parcimonioso, segundo valores observados na Tabela 2.2.

Ainda, vale notar, na Tabela 2.2 que, embora o critério BIC e a variância ( $\sigma^2$ ) tenham valores inferiores para as ordens  $p_m = 4$  e  $p_m = 5$ , escolheu-se a ordem  $p_m = 3$ .

<sup>2</sup>Não há grandes vantagens em aumentar a ordem  $q_m$ , uma vez que o desempenho permanece constante e o esforço computacional é maior.

Tabela 2.2: Valores do critério BIC e da variância para o mês de junho para o posto de Sobradinho.

$p_m$	1		2		3		4		5	
	$\sigma^2$	BIC								
0	0,14	-134,87	0,10	-156,49	0,07	-174,47	0,06	-175,33	0,06	-176,43
1	0,10	-156,49	0,97	3,86	0,25	-82,73	0,09	-150,63	0,26	-72,32

Isso porque as diferenças tanto do BIC quanto da variância são baixas entre a ordem 3 e as ordens 4 e 5. Dessa forma, é razoável considerar uma ordem menor, neste caso, 3.

Utilizando a metodologia descrita acima, as estimativas foram obtidas e são mostradas na Tabela (2.3).

Assim, a ordem dos modelos mensais AR varia entre 1 e 3, com as maiores ordens ocorrendo nos meses mais secos. Além disso, para o mês de julho é estabelecida ordem 1 para a modelagem MA.

Tabela 2.3: Estimativas dos parâmetros utilizando o método EMV para a série de vazões de Sobradinho.

Meses	$\phi_{1,m}$	$\phi_{2,m}$	$\phi_{3,m}$	$\theta_{1,m}$	$\tau_m$	$(p_m, q_m)$
Janeiro	0,6062	-	-	-	1,7240	(1,0)
Fevereiro	0,4372	-	-	-	1,4020	(1,0)
Março	0,7153	-	-	-	2,3556	(1,0)
Abril	0,6534	-	-	-	2,1941	(1,0)
Mai	0,8004	-	-	-	3,2048	(1,0)
Junho	0,6541	0,2045	0,2193	-	16,6742	(3,0)
Julho	1,3363	-0,3687	-	-0,4307	19,4301	(2,1)
Agosto	0,9736	-	-	-	26,3190	(1,0)
Setembro	0,9399	-	-	-	10,3242	(1,0)
Outubro	0,8156	-	-	-	3,2251	(1,0)
Novembro	0,6167	-	-	-	1,8211	(1,0)
Dezembro	0,6165	-	-	-	1,8829	(1,0)

A seguir, a validação do modelo é feita através do conjunto de dados do período de 1991-1998. A Figura (2.5) mostra a previsão um passo à frente, dada pelo modelo PARMA ajustado. Nota-se que a performance da aproximação é pouco efetiva nos picos

da curva de previsão que correspondem ao período úmido, em que, normalmente, há grande ocorrência de chuvas. Por outro lado, a modelagem nos “vales”, que representam os meses de seca, acompanha melhor a curva real, embora ainda seja um pouco deficiente.

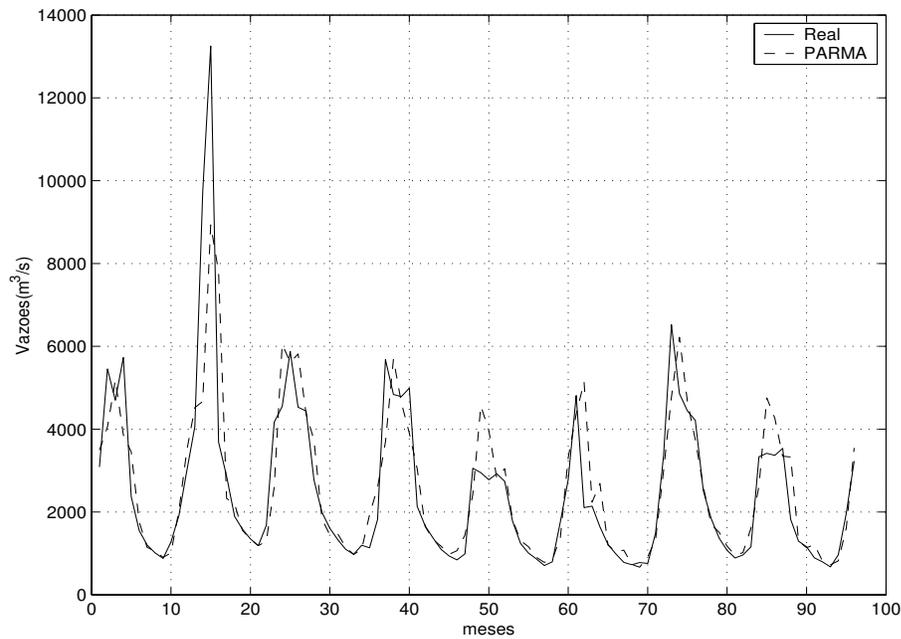


Figura 2.5: Previsão de vazões do posto de Sobradinho utilizando o modelo PARMA para o período de 1991 a 1998.

## 2.6 Resumo

Inicialmente, apresentou-se algumas definições básicas de processos estocásticos a fim de auxiliar o entendimento sobre os modelos de previsão de séries temporais. Baseando-se nesses conceitos, modelos lineares estacionários foram descritos seguindo a metodologia Box & Jenkins.

A estrutura dessa metodologia é composta por três estágios principais: identificação, estimação e validação. Assim, o modelo de séries temporais PARMA foi identificado para a série de vazões mensais de Sobradinho, e através do método de máxima verossimilhança os parâmetros foram estimados.

O critério para verificar a ordem com que o modelo se mostra mais parcimonioso foi o *Bayesian Information Criterium* (BIC), de acordo com o valor calculado pela sua equação. Esse valor leva em consideração a variância do modelo, o número de parâmetros envolvidos e o número de dados da série. A ordem que possui menor valor BIC, corresponde ao modelo PARMA( $p_m, q_m$ ) que melhor ajusta os dados.

## Capítulo 3

# Redes Neurais Artificiais

Redes Neurais Artificiais (RNAs) são poderosas ferramentas para tratar uma grande diversidade de problemas, tais como classificação e reconhecimento de padrões, processamento de sinais, aproximação de funções e previsão de séries temporais.

Neste capítulo é discutida a aplicação de duas estruturas de RNAs para a previsão de vazões naturais médias mensais, um passo à frente. Inicialmente, a previsão é feita utilizando uma rede neural multicamadas, seguida de uma rede híbrida que agrega a representatividade dos sistemas nebulosos.

### 3.1 Introdução

As RNAs constituem sistemas de processamento distribuído paralelo que, por serem baseados no funcionamento do sistema nervoso biológico, possuem propensão natural de armazenar conhecimento experimental (Haykin, 1994). Assim como o cérebro humano é capaz de transmitir e/ou armazenar informações através de impulsos nervosos entre os neurônios, as RNAs possuem conexões inter-neurais, representadas por pesos sinápticos, com função de armazenar e processar o conhecimento.

Do ponto de vista estrutural, as RNAs podem ser classificadas como redes neurais estáticas (não recorrentes) ou como redes neurais recorrentes (dinâmicas) (Haykin, 1994). A principal diferença entre as duas estruturas é a presença ou não de conexões que realimentem os neurônios presentes na arquitetura da rede. Quando não existe dependência das saídas atuais com seus valores passados, as redes são estáticas. Entre

essas estruturas, se destacam as rede neurais multicamadas (*Multi Layer Perceptron - MLP*) por serem amplamente utilizadas na literatura.

Por outro lado, as redes recorrentes são capazes de representar uma grande variedade de comportamentos dinâmicos, uma vez que a presença de realimentação de informação permite a criação de representações internas e dispositivos de memória capazes de processar e armazenar informações temporais e seqüenciais (Von Zuben, 1996).

Além disso, uma das tendências mais recentes é utilizar conjuntamente as potencialidades das RNAs com as de outras teorias, resultando em modelos híbridos. Entre outras, uma das fusões mais bem sucedidas combina tanto o poder de aprendizagem das redes neurais quanto a capacidade de representação dos sistemas nebulosos. Essa combinação é denominada rede neural nebulosa, ou simplesmente rede neurofuzzy.

### **3.2 Um Breve Histórico da Área de RNAs e sua Aplicação em Previsão**

O primeiro modelo matemático para um neurônio biológico foi proposto por McCulloch e Pitts em 1943 (McCulloch e Pitts, 1943) e considera a saída do neurônio uma resposta binária a um estímulo fornecido. Apesar de apresentar grande poder computacional, faltavam-lhe características adaptativas e de aprendizagem, imprescindíveis em modelos de sistemas inteligentes (Ballini, 2000).

Após a publicação do trabalho de McCulloch e Pitts, muitos foram os avanços na área de redes neurais durante as décadas subseqüentes por pesquisadores de áreas como a neurofisiologia (Hebb, 1949) e a engenharia (Rosenblatt, 1958), introduzindo e aperfeiçoando paradigmas de aprendizagem. Em meados da década de 60, (Hu, 1964) sugere o primeiro modelo de RNA para previsão de séries temporais, utilizado em dados meteorológicos.

Apesar do aparente sucesso dos modelos desenvolvidos, em 1969, Minsky e Papert (Minsky, 1969) mostraram as limitações da rede perceptron com uma única camada para problemas não linearmente separáveis, o que levou a um declínio nas investigações de metodologias relacionadas à área de redes neurais.

Entretanto, alguns trabalhos importantes foram publicados durante essa década,

destacando-se os mapas auto-organizáveis (Kohonen, 1982). No início da década de 80, o interesse na teoria das RNAs foi reativado após a publicação dos estudos de John Hopfield (Hopfield, 1982) baseados no princípio físico de armazenamento de informação em configurações dinamicamente estáveis, sendo este um dos primeiros modelos a introduzir dinâmica em RNAs (Von Zuben, 1996).

Em 1986, através da publicação do livro *Parallel Distributed Processing* (Rumelhart e McClelland, 1986), o algoritmo de retropropagação (*backpropagation*), previamente proposto por Werbos (Werbos, 1974), foi divulgado e, atualmente, é o algoritmo mais popular para o treinamento de RNAs estáticas multicamadas. A partir disso, pesquisadores de diversas áreas aumentaram seu interesse em aplicações de RNAs, particularmente no campo de previsão de séries temporais, foco deste trabalho. Neste contexto, (Weigend et al., 1990), (Weigend et al., 1992) conduziram um trabalho sobre estruturas neurais para previsão de série temporais reais.

Além de considerarem as não linearidades presentes entre as entradas e as saídas, as RNAs possuem características bastante atraentes (Zhang et al., 1998):

- Não requerem *a priori* informações sobre o modelo pois aprendem a partir de exemplos e, assim, capturam relações funcionais entre os dados mesmo se tais relações sejam dificilmente descritíveis. Para que uma RNA possa aprender e garantir um bom desempenho no resultado, o conjunto de dados disponíveis deve ser suficientemente grande. Nesse sentido, estes modelos podem ser tratados como um *método estatístico multivariado não-paramétrico e não-linear*.
- Têm a capacidade de generalizar. Após a fase de aprendizagem, quando já possuem algum conhecimento sobre as características do conjunto de dados, as RNAs conseguem inferir na população restante mesmo se os dados amostrais contêm ruído nas informações.
- São aproximadores universais, isto é, podem aproximar uma função contínua, definida sob um domínio compacto, com a precisão desejada.

Diversos estudos comparando o desempenho dos modelos de RNAs e das clássicas abordagens propostas por Box e Jenkins (Box et al., 1994) para previsão de séries temporais vêm sendo realizados. Entre outros trabalhos, destacam-se (Tang et al., 1991), (Tang e Fishwick, 1993) e (Gershenfeld e Weigend, 1994).

A seguir, os parâmetros que compõem a estrutura de uma RNA e métodos de aprendizagem são descritos detalhadamente.

### 3.3 Componentes e Treinamento de uma RNA

O funcionamento e o desempenho de uma RNA dependem diretamente da escolha dos componentes adotados em sua estrutura e dos métodos adotados em seu processo de aprendizagem.

#### 3.3.1 O Neurônio Artificial e as Funções de Ativação

Inicialmente, é necessário compreender a unidade básica de processamento de uma RNA, denominada neurônio artificial. Essa unidade processa suas entradas através da soma ponderada e de uma função de ativação. A função de ativação é o primeiro componente a ser escolhido na estrutura de um neurônio para uma dada aplicação.

A Figura 3.1 mostra o modelo genérico de um neurônio artificial onde  $f(\cdot)$  é a função de ativação do neurônio.

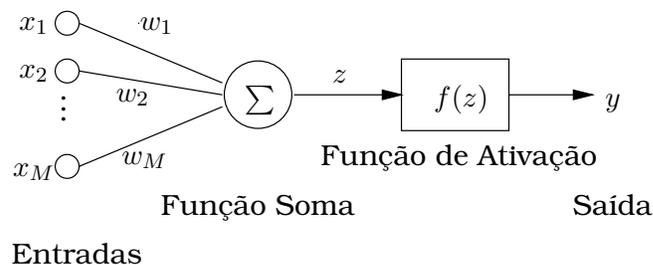


Figura 3.1: Modelo de neurônio artificial.

Matematicamente, a saída do neurônio é dada por:

$$y = f\left(\sum_{i=1}^M x_i w_i\right) = f(z) \quad (3.1)$$

sendo  $x_i$ ,  $i = 1, \dots, M$ , as entradas e  $w_i$  os pesos sinápticos correspondentes.

Por exemplo, para modelo de neurônio proposto por McCulloch e Pitts, a função

de ativação  $f(z)$  é uma função degrau da seguinte forma:

$$f(z) = \begin{cases} 1, & \text{se } z > a, \\ 0, & \text{se } z \leq a. \end{cases} \quad (3.2)$$

A função de ativação pode ser definida principalmente por outros três tipos de funções: linear, logística ou tangente hiperbólica (Haykin, 1994). A primeira é descrita da seguinte forma:

$$f(z) = \begin{cases} 0, & \text{se } z \leq \alpha, \\ \frac{z - \alpha}{\beta - \alpha} & \text{se } \alpha < z \leq \beta, \\ 1, & \text{se } z > \beta. \end{cases} \quad (3.3)$$

com,  $\alpha$  e  $\beta$  previamente definidos. Já as funções logística e tangente hiperbólica são representadas respectivamente como:

$$f(z) = \frac{1}{1 + \exp^{-\gamma z}} \quad (3.4)$$

$$f(z) = \frac{\exp^{\gamma z} - \exp^{-\gamma z}}{\exp^{\gamma z} + \exp^{-\gamma z}} \quad (3.5)$$

sendo,  $\gamma$  o ponto de inflexão de uma função sigmoide. Note que, as funções sigmoideais (3.4) e (3.5) possuem suas imagens em intervalos diferentes, isto é, para a função logística  $f(z) \in [0, 1]$ , enquanto que, para a tangente hiperbólica  $f(z) \in [-1, 1]$ .

A organização dos neurônios de uma RNA é feita em camadas, sendo a primeira, denominada camada de entrada e a última, camada de saída. O número de camadas intermediárias e o número de neurônios em cada uma delas são escolhidos empiricamente de acordo com a maior ou menor necessidade de se processar as entradas da rede, isto é, os números de camadas e de neurônios aumentam, caso o conjunto de dados exija um processamento mais refinado para capturar características não lineares.

Os neurônios recebem as informações das entradas e propagam-nas através dos pesos  $w_i$ ,  $i = 1, \dots, M$ , procedimento que simula o processamento sináptico. Os pesos sinápticos são parâmetros adaptáveis ajustados via processo de treinamento (Haykin, 1994). Esse processo pode ser classificado, principalmente, como treinamento supervisionado, treinamento não supervisionado ou treinamento por reforço. Neste trabalho, dois tipos de processos de treinamento (alternativamente, aprendizagem) são utilizados, o supervisionado e o não supervisionado. Ambos serão descritos com mais detalhes a seguir.

### 3.3.2 Treinamento Supervisionado

O treinamento supervisionado caracteriza-se pela disponibilidade de conhecimento ou informação sobre um sistema na forma de padrões entrada-saída (Haykin, 1994). Aplicando um algoritmo de treinamento supervisionado, uma RNA adquire conhecimento ou informação relevante sobre um problema de forma análoga àquela utilizada pelo ser humano e outros animais, ou seja, a partir de exemplos e de experiência (Von Zuben, 1996). Contudo, essa experiência é baseada apenas nos exemplos fornecidos pelo conjunto de treinamento. Caso apareça uma situação nova e particular, a RNA pode apresentar dificuldades na modelagem, diminuindo o desempenho.

Um processo supervisionado segue basicamente o seguinte esquema. Considerando um conjunto de treinamento composto por padrões entrada-saída desejada; para uma dada entrada, a objetivo é ajustar os parâmetros da rede através da comparação entre a saída desejada da rede e a saída correspondente. O ajuste é feito com base no sinal de erro, calculado pela diferença entre a saída desejada (conjunto de treinamento) e a saída da rede.

O algoritmo de treinamento mais difundido e utilizado na literatura é o de retropropagação (*backpropagation*) (Rumelhart e McClelland, 1986). Esse método consiste em duas fases. Na primeira, conhecida como fase de propagação direta (*forward*), as entradas são apresentadas e propagadas através da rede, camada a camada, calculando a saída de cada neurônio. Durante essa fase os pesos são fixos e a saída calculada é comparada com a saída desejada, resultando em um erro para cada unidade de saída. Na segunda fase, o erro calculado é propagado da camada de saída para a camada de entrada, fase de propagação reversa (*backward*), e os pesos são ajustados de acordo com a regra de correção do erro, originando o termo retropropagação do erro. A segunda fase pode ser formulada como um problema de otimização não linear irrestrita, tendo os pesos sinápticos como variáveis de decisão e o objetivo é minimizar a somatória dos erros quadráticos de todos os neurônios de saída. No algoritmo de retropropagação tradicional, para resolver o problema de otimização, adota-se o método do gradiente descendente, que considera apenas condições de primeira ordem a partir da função objetivo.

Muitas metodologias de treinamento foram desenvolvidas com base no algoritmo de retropropagação a fim de aumentar o desempenho ou a velocidade de convergência

(Jacobs, 1988).

### 3.3.3 Treinamento Não Supervisionado

Nos métodos não supervisionados a aprendizagem se processa sem que as saídas corretas sejam fornecidas. Esses métodos capacitam as redes neurais extrair correlações presentes nos dados de entrada. Após a aprendizagem, a rede neural torna-se capaz de classificar suas entradas de acordo com as correlações aprendidas (Figueiredo, 1997).

Neste trabalho, esse tipo de aprendizagem é aplicado no treinamento da estrutura neurofuzzy, proposta em (Figueiredo, 1997), para o ajuste dos parâmetros da segunda camada da rede. O método utilizado durante essa etapa é baseado em métodos competitivos utilizados pelas redes de Kohonen, ou seja, na aprendizagem, após receberem um sinal de entrada, os neurônios competem entre si e vence aquele que apresentar a menor distância entre os correspondentes pesos sinápticos e as entradas.

## 3.4 Redes Neurais Multicamadas

A arquitetura da rede MLP (Figura 3.2) é composta por neurônios conectados, seguindo uma formação em camadas. Essas conexões são responsáveis por propagar as entradas  $x_i$ ,  $i = 1, \dots, M$ . As unidades de entrada, que compõem essa camada, têm o objetivo de difundir o sinal inicial sem nenhuma modificação para a segunda camada. Na Figura 3.2, o  $j$ -ésimo neurônio da camada  $k$ ,  $k = 1, \dots, K$  é representado por  $p_j^k$  e está conectado ao neurônio  $i$  da camada anterior com o peso sináptico  $w_{ji}^k$  correspondente. O sinal de entrada é processado ao longo da rede, e após percorrer todas as suas camadas, produz uma resposta gerada pelos neurônios da camada de saída. Essa resposta é representada pelo sinal  $y_j$ ,  $j = 1, \dots, M_K$ .

Assim, a arquitetura de uma MLP apresenta três características distintas (Ballini, 2000):

- Variações quanto ao número de camadas e neurônios intermediários da rede;
- Tipo de conexão determinada pelas sinapses (estática ou recorrente);
- Função de ativação.

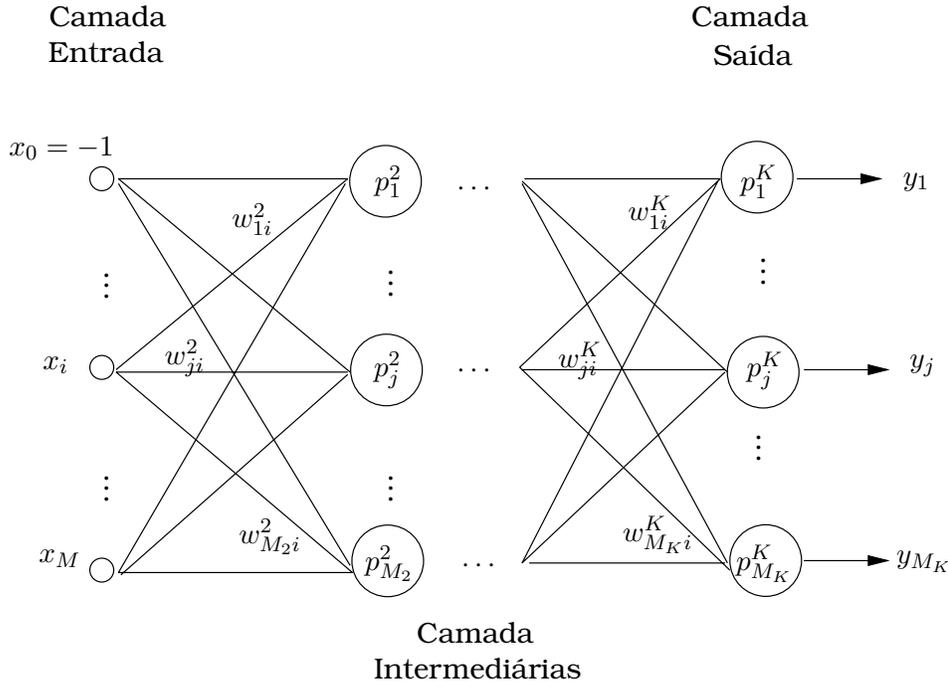


Figura 3.2: Arquitetura de uma rede MLP.

Como foi dito na seção 3.3.2, o método mais popular para o treinamento supervisionado é o de retropropagação. A extração do conhecimento é realizada a partir de um conjunto de  $N$  padrões entrada-saída  $\{(\mathbf{x}^1, \mathbf{d}^1), \dots, (\mathbf{x}^l, \mathbf{d}^l), \dots, (\mathbf{x}^N, \mathbf{d}^N)\}$ , com  $\mathbf{x}^l = (x_1^l \dots x_i^l \dots x_M^l)$  e  $\mathbf{d}^l = (d_1^l \dots d_j^l \dots d_{M_K}^l)$ . Na fase de propagação direta (*forward*), o sinal de entrada  $\mathbf{x}^l$  é propagado camada a camada e a respectiva saída  $\mathbf{y}^l = (y_1^l \dots y_j^l \dots y_{M_K}^l)$  é comparada à saída desejada  $\mathbf{d}^l$ . Assim, para cada neurônio  $p_j^K$ ,  $j = 1, \dots, M_K$  da camada de saída, na  $t$ -ésima apresentação de um padrão  $(\mathbf{x}^l, \mathbf{d}^l)$ , define-se um erro associado:

$$e_j^l(t) = d_j^l(t) - y_j^l(t) \quad (3.6)$$

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

$$\epsilon^l(t) = \frac{1}{2} \sum_{j=1}^{M_K} (e_j^l(t))^2 \quad (3.7)$$

Considerando o número total  $N$  de pares entrada-saída contidos no conjunto de treinamento, o erro quadrático médio é obtido pela soma dos erros correspondentes a

cada padrão, normalizado com relação ao número de padrões  $N$ :

$$E = \frac{1}{N} \sum_{l=1}^N \epsilon^l \quad (3.8)$$

O objetivo do processo de aprendizagem é ajustar os parâmetros livres (pesos) da rede para minimizar  $E$ . Para a derivação do algoritmo de retropropagação, é considerado um método de treinamento em que os pesos são ajustados padrão-a-padrão, ou seja, o ajuste dos pesos é realizado a partir do erro calculado para cada entrada apresentada à rede, conforme (3.7).

Para a resolução do problema de otimização do erro (3.8), adota-se o algoritmo de aproximação estocástica. Esta aproximação é aplicada em problemas em que o ambiente é desconhecido, como, por exemplo, no caso do treinamento de retropropagação, em que a função a ser minimizada é definida por amostragem e não tem uma equação conhecida. Dessa forma, os pesos seguem uma trajetória aleatória ao minimizar uma estimativa instantânea da função  $E$  ((3.8)). Conseqüentemente, o vetor gradiente é “aleatório” e sua acurácia aumenta à medida que o número de iterações  $t$  tende ao infinito (Haykin, 1994). A correção do erro é calculada para cada camada  $k$ , composta por  $M_k$  neurônios da seguinte forma<sup>1</sup>:

$$\Delta w_{ji}^k(t) = -\alpha \frac{\partial \epsilon(t)}{\partial w_{ji}^k(t)} \quad (3.9)$$

sendo  $\alpha$  a taxa de aprendizagem.

Para aumentar a velocidade de aprendizagem sem aumentar a taxa de aprendizagem  $\alpha$  e tornar a rede instável, utiliza-se um método simples que modifica a regra (3.9), adicionando um termo *momentum*, como proposto por (Rumelhart et al., 1986):

$$\Delta w_{ji}^k(t) = -\alpha \frac{\partial \epsilon(t)}{\partial w_{ji}^k(t)} + \beta \Delta w_{ji}^k(t-1) \quad (3.10)$$

ou seja, uma generalização de (3.9) para  $\beta \neq 0$ . O termo *momentum* relaciona as alterações dos pesos sinápticos na iteração  $t$  com as alterações realizadas na iteração  $t-1$ . 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 (Ballini, 2000).

<sup>1</sup>A partir deste ponto, o superíndice  $l$ , referente ao  $l$ -ésimo padrão, será omitido para simplificar a notação.

Explicitando a função  $\epsilon(t)$  e aplicando sucessivamente a regra da cadeia, calcula-se a derivada parcial  $\partial\epsilon(t)/\partial w_{ji}^k$ . Dessa forma, o ajuste dos pesos sinápticos é dado por:

$$w_{ji}^k(t+1) = w_{ji}^k(t) + \Delta w_{ji}^k(t) \quad (3.11)$$

A aprendizagem termina, caso a rede tenha alcançado o índice de desempenho estabelecido, isto é,  $\|E\| \leq \delta_{max}$ , sendo  $\|\cdot\|$  a norma Euclidiana e  $\delta_{max}$  um limiar suficientemente pequeno. Alternativamente, a aprendizagem termina caso o número de iterações ultrapasse um valor máximo especificado (Haykin, 1994).

A seguir, serão apresentadas as características topológicas de uma rede MLP quando aplicada ao problema de previsão de vazões.

### 3.4.1 Aplicação da Rede MLP para Previsão de Vazões

Nesta seção, as redes MLP são aplicadas à previsão de séries de vazão afluente. Analogamente ao Capítulo 2, os dados utilizados caracterizam o posto de Sobradinho e o período de teste é definido entre os anos de 1991 e 1998. Os dados são inicialmente normalizados no intervalo  $[0, 1]$ . Além disso, uma rede MLP é ajustada para cada mês do ano, resultando em 12 modelos  $MLP(m)$ ,  $m = 1, \dots, 12$ , cada um com uma arquitetura diferente.

A topologia da rede possui a primeira camada com  $M$  entradas, uma camada intermediária com  $M_1$  neurônios sigmoidais (logística) e a camada de saída com um neurônio linear. Logo, é necessário estabelecer os valores de  $M$ ,  $M_1$ , a taxa de aprendizagem  $\alpha$  e o termo *momentum*  $\beta$ . Para determinar esses parâmetros que compõe a topologia da rede, utiliza-se a técnica estatística de validação cruzada (Stone, 1974).

O procedimento inicial da técnica de validação cruzada consiste em particionar o conjunto de treinamento em dois subconjuntos distintos (Ballini, 2000):

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

A seguir, a partir do subconjunto de treinamento, estima-se os parâmetros da rede e utilizando esses parâmetros no subconjunto de validação, verifica-se o desempenho de várias topologias da rede. A melhor topologia é escolhida de acordo com erros apresentados durante a validação. Quanto menores os erros, melhor é o desempenho da rede testada.

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.

A Tabela 3.1 mostra os valores desses parâmetros para cada mês.

Tabela 3.1: Topologias das redes MLP escolhidas para cada mês.

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
$M/M_1$	2/7	6/31	1/14	3/28	3/6	6/9	2/18	2/23	2/30	3/33	3/30	1/32
$\alpha$	0,25	0,5	0,5	0,75	0,5	0,5	0,25	0,75	0,75	0,5	0,75	0,75
$\beta$	0,25	0,5	0,75	0,5	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,5

A Figura 3.3 mostra a previsão um passo à frente, para o período de 1991 e 1998. Observa-se resultados satisfatórios para os períodos de seca, representados pelas regiões dos “vales” do gráfico; por outro lado, nas regiões de “picos”, a curva prevista não acompanha bem a curva real. Isso demonstra, neste caso, a dificuldade do modelo  $MLP(m)$ , com camada de saída linear, em prever vazões em períodos úmidos ou de cheia, em contraste com o bom desempenho nos períodos de seca.

A seguir, a classe de redes neurais nebulosas utilizada neste trabalho será de-

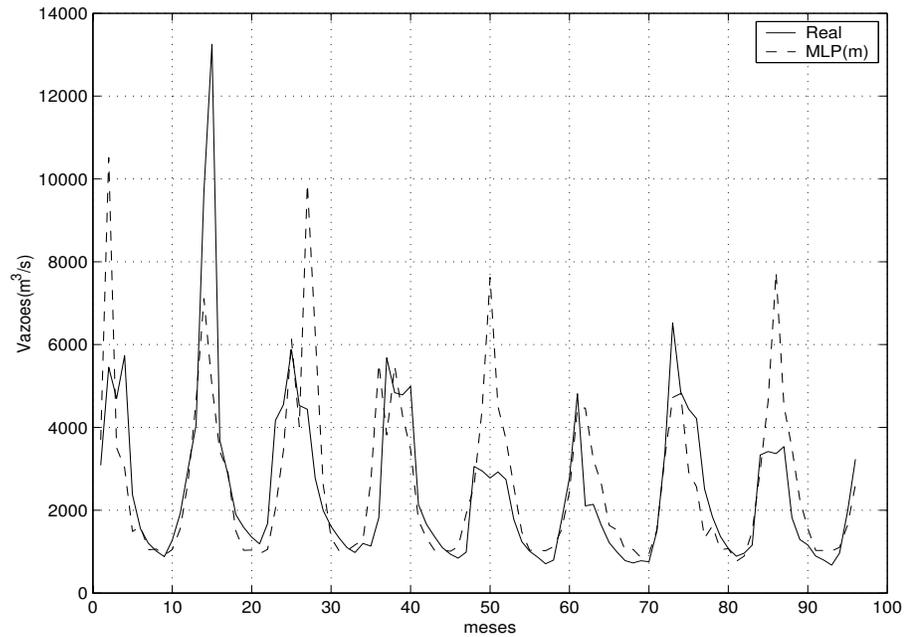


Figura 3.3: Previsão de vazões do posto de Sobradinho utilizando o modelo MLP(m) para o período de 1991 a 1998.

talhada e aplicada à série de vazões de Sobradinho, analogamente ao que foi realizado nesta seção.

### 3.5 Redes Neurais Nebulosas

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 o conhecimento e inferem conclusões a partir dos padrões apresentados (Pedrycz et al., 1995). Essencialmente, estes sistemas são capazes de processar o conhecimento de forma compreensível e de manipular 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 esses paradigmas poderiam ser unificados, resultando em sistemas de características mais completas no que se refere às suas capacidades inteligentes. O uso concomitante das duas abordagens resulta numa compensação das deficiências de cada uma. De um lado, as redes neurais são limitadas 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 (Ballini, 2000). O resultado dessa simbiose resultou nas redes neurais nebulosas (Pedrycz e Gomide, 1998).

Em se tratando da previsão de séries temporais, as redes neurais nebulosas, ou simplesmente, redes neurofuzzy, têm se mostrado bastante eficientes e promissoras (Luna et al., 2003). Por esse motivo, diversos algoritmos de aprendizagem vêm sendo propostos com o objetivo de aprimorar o funcionamento e o desempenho desse tipo de rede para os mais diversos tipos de séries temporais.

Em particular, uma classe de redes neurais nebulosas com aprendizagem construtiva e competitiva é considerada nesse trabalho. Esse modelo foi desenvolvido por (Figueiredo e Gomide, 1999) e aplicado à previsão de vazões em (Ballini, 2000). A topologia é caracterizada, essencialmente, por duas propriedades: o mapeamento das regras nebulosas que compõem a estrutura da rede é direto e o processamento neural deve equivaler totalmente a um mecanismo de inferência nebulosa. Logo, essa aproximação possui natureza implicitamente dual, ou seja, pode ser vista tanto como uma rede neural quanto como um sistema de inferência baseado em regras nebulosas (Figueiredo et al., 2004).

### Arquitetura da Rede Neural Nebulosa

A estrutura da rede neural nebulosa (RNN) processa um conjunto de regras nebulosas da forma “**Se** um conjunto de condições é satisfeito **Então** um conjunto de conseqüentes é inferido” e compõem uma base de regras, que pode ser definida 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**  $x_1 \text{ é } A_1^Q \text{ e } \dots \text{ e } x_M \text{ é } A_M^Q$  **Então**  $y \text{ é } w^Q$

---

saída:  $y = y(\mathbf{x})$

sendo  $x_j$ ,  $j = 1, \dots, M$  variáveis nebulosas do espaço de entrada;  $A_j$  e  $A_j^i$  conjuntos nebulosos,  $i = 1, \dots, Q$  o índice que denota a regra nebulosa; e  $w^i$  e  $y$  os valores dos conseqüentes das regras e variável do espaço de saída, respectivamente.

Seja o grau de pertinência da entrada  $x_k$  em relação ao conjunto nebuloso  $A_j^i$  dado por  $A_j^i(x_k) = a_{jk}^i$ , se  $x_k \in I_k = (x_I, x_F]$ , com  $I_k$ ,  $k = 1, \dots, Q$  representando o  $k$ -ésimo intervalo de discretização em relação à entrada  $x_k$ . 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), em que a saída  $y$  é determinada através dos seguintes passos:

1. *Comparação*: Para cada regra  $i$  e para cada antecedente  $j$ , computar a medida de possibilidade  $P_j^i$  entre os conjuntos nebulosos  $A_j$  e  $A_j^i$ :

$$P_j^i(\mathbf{x}) = S_k\{T(a_{jk}, a_{jk}^i)\} \quad (3.12)$$

com  $S$  e  $T$  denotando operadores lógicos, correspondentes a  $s$ -norma e  $t$ -norma, respectivamente com  $S$  sendo aplicada para todo intervalo de discretização  $k$ , e  $\mathbf{x} = (x_1, \dots, x_M)'$  representando a entrada da rede.

2. *Agregação dos Antecedentes*: Para cada regra  $i$ , agregar os valores de cada um dos antecedentes, resultando no nível de ativação  $H^i$  definido por:

$$H^i(\mathbf{x}) = T_j\{P_j^i(\mathbf{x})\} \quad (3.13)$$

3. *Agregação de Regras*: Computar a saída  $y$  da seguinte forma:

$$y(\mathbf{x}) = \frac{\sum_{i=1}^Q H^i(\mathbf{x})w^i}{\sum_{i=1}^Q H^i(\mathbf{x})} \quad (3.14)$$

A unidade básica de processamento é denominada neurônio nebuloso, cujo modelo geral é mostrado na Figura 3.4. Matematicamente, a saída desse neurônio é dada por:

$$y(\mathbf{x}) = \varphi(\phi(\psi(\mathbf{x}))) = \varphi(\phi(\psi_1(x_1), \dots, \psi_M(x_M))) \quad (3.15)$$

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

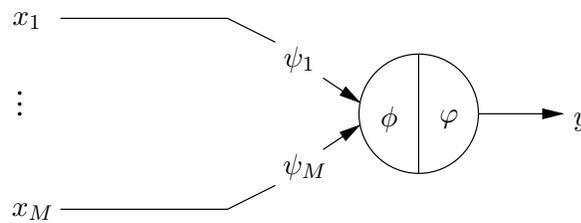


Figura 3.4: Modelo geral de neurônio nebuloso.

A arquitetura da rede nebulosa é estática e apresenta cinco camadas, conforme mostra a Figura 3.5. A primeira camada é dividida em  $M$  grupos de neurônios, cada um associado a uma variável de entrada. Cada neurônio  $k$  do  $j$ -ésimo grupo está associado a um intervalo de discretização do espaço de entradas cuja saída  $a_{jk} \in \{0, 1\}$  de modo que, para o  $k$ -ésimo intervalo  $I_k = (x_I, x_F]$ ,  $a_{jk}$  é definido como:

$$a_{jk} = \varphi(x_j) \begin{cases} 1, & \text{se } x_j \in (x_{jI}, x_{jF}], \\ 0, & \text{caso contrário} \end{cases} \quad (3.16)$$

ou seja, para a primeira camada, atribui-se a função identidade para  $\psi$  e  $\phi$  na equação (3.15).

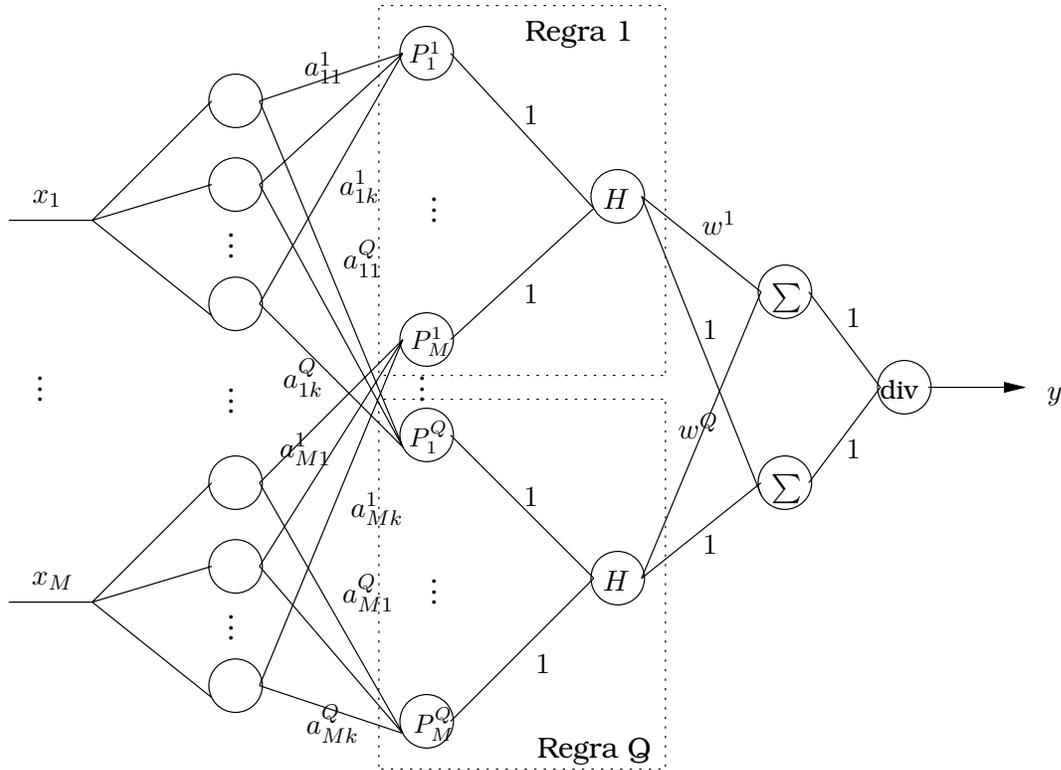


Figura 3.5: Arquitetura da rede neural nebulosa.

A segunda camada é composta por  $Q$  grupos de regras cada qual com  $M$  neurônios (número de antecedentes de cada regra). Esta camada é responsável pelo primeiro passo do processo de inferência, denominado *comparação* (Ballini, 2000). Considerando o  $j$ -ésimo neurônio do  $i$ -ésimo grupo de regras, a medida de possibilidade  $P_j^i$  é calculada, de acordo com a equação (3.12), assumindo  $\psi = t$ -norma,  $\phi = s$ -norma e  $\varphi =$  identidade no neurônio geral representado por (3.15).

A terceira camada, por sua vez realiza a *agregação dos antecedentes* de cada grupo de regras  $i$  da camada anterior (equação 3.13). Neste caso, considera-se, no neurônio geral, a função  $\phi = t$ -norma e as demais,  $\psi$  e  $\varphi$  como funções identidade.

A quarta camada processa os valores do numerador e do denominador de (3.14), através de dois neurônios, cujo operador de agregação associado é a soma algébrica. A diferença entre esses dois neurônios é o peso atribuído às conexões sinápticas. Um desses neurônios se conecta aos  $i, i = 1, \dots, Q$ , neurônios da terceira camada com pesos

$w^i$ . Dessa forma, considerando que  $\psi$  = produto,  $\phi$  = soma algébrica e  $\varphi$  = identidade, as saídas desse neurônio são os numeradores de (3.14). Por outro lado, o valor atribuído aos pesos sinápticos que conectam o outro neurônio da quarta camada aos neurônios da camada anterior é unitário. Logo, tomando os operadores  $\psi$  e  $\varphi$  como funções identidade e  $\phi$  = soma algébrica, obtêm-se os denominadores de (3.14).

Finalmente, o processamento realizado na quinta camada equivale ao passo 3 do processo de inferência, denominado *agregação de regras*, uma vez que esta última camada possui apenas um neurônio que calcula a razão entre as saídas dos dois neurônios da camada anterior fornecendo a saída da rede neural.

### Método de Aprendizagem

A estratégia de aprendizagem na rede neural nebulosa se divide em duas fases (Ballini, 2000). Na primeira, a rede aprende os parâmetros que determinam os conjuntos nebulosos,  $A_j^i(\cdot)$ , dos antecedentes das regras. Considerando que, os conjuntos nebulosos são representados por funções de pertinência,  $F_j^i(\cdot)$ , simétricas e do mesmo tipo, os parâmetros que definem essas funções são os valores modais  $c_j^i$  e as dispersões  $r^i$ . Desse modo, é possível ajustar os pesos sinápticos  $a_{jk}^i$ , da segunda camada, uma vez que:

$$a_{jk}^i = F_j^i(x_j), \quad \text{tal que } x_j \in I_k = [x_I, x_F] \quad (3.17)$$

Essa fase é realizada por um treinamento não supervisionado, baseado nas redes tipo Kohonen, o que implica a existência de competição entre os neurônios.

Por outro lado, na segunda fase, desenvolve-se a aprendizagem dos consequentes das regras, de modo que os pesos,  $w^i$  a serem ajustados, são pertencentes à quarta camada (Figura 3.5). O treinamento, neste caso, é supervisionado e baseado no método do gradiente. Logo, analogamente ao treinamento por retropropagação apresentado na seção 3.4, é fornecido um conjunto de pares entrada-saída desejada,  $(\mathbf{x}^l, d^l)$ , com  $\mathbf{x}^l = (x_1^l, \dots, x_M^l)$ ,  $l = 1, 2, \dots$ <sup>2</sup>.

Em suma, o método de aprendizagem, composto pelas duas fases descritas, consiste em ajustar os valores modais, as dispersões e os pesos, minimizando, a cada

<sup>2</sup>Daqui em diante, o superíndice  $l$  será omitido para simplificar a notação.

iteração  $t$ , o erro quadrático<sup>3</sup>:

$$E(\mathbf{x}, t) = E(\mathbf{x}(t)) = \frac{1}{2}(d(t) - y(t))^2 \quad (3.18)$$

sendo  $y(t)$  é a saída da rede. Aplicando o método do gradiente, o ajuste dos pesos é dado por:

$$w^i(t+1) = w^i(t) - \alpha \frac{\partial E(\mathbf{x}(t))}{\partial w^i(t)} \quad (3.19)$$

A derivada  $\partial E(\mathbf{x}(t))/\partial w^i(t)$  é calculada pela regra da cadeia aplicada à equação (3.18), resultando em:

$$w^i(t+1) = w^i(t) + \alpha \left( \frac{(d(t) - y(t))H^i(\mathbf{x}(t))}{\sum_{h=1}^Q H^h(\mathbf{x}(t))} \right) \quad (3.20)$$

(Figueiredo, 1997) introduz uma metodologia de aprendizagem construtiva e competitiva *off-line* e outra *on-line*. Neste trabalho, utiliza-se apenas a última para a previsão de vazões, uma vez que em (Ballini, 2000) a rede nebulosa com aprendizagem *on-line* tem desempenho superior ao aprendizado *off-line*.

Uma das razões para que a aprendizagem *on-line* forneça resultado superior, é o fato de que a estrutura da rede se adapte durante o treinamento. Essa modificação ocorre no número de regras  $Q$ , nos valores modais e de dispersão e nos pesos. Isso significa que sempre que a rede receber um novo padrão de entrada, os grupos de neurônios vão competir.

Os fatores que caracterizam a estrutura da RNN são definidos como:

1. Fator de aprendizagem  $\varepsilon^i(t) \in (0,1)$  como uma função monotônica decrescente definida neste trabalho da seguinte maneira:

$$\varepsilon^i(t) = \frac{\beta}{\eta^i(t) + 1} \quad (3.21)$$

para o ajuste dos valores modais  $c_j^i(t)$  das funções de pertinência  $F_j^i(\cdot)$ . O parâmetro  $\beta \in (0,1]$  é constante, enquanto  $\eta^i(l)$  é variável e representa o número de vezes que o valor modal  $c_j^i$  é ajustado;

---

<sup>3</sup>Notar que o erro é, em cada iteração  $t$ , função de  $\mathbf{x}$ , uma vez que  $y$  depende de  $\mathbf{x}$ . Conforme (3.14), a notação  $E(\mathbf{x}(t))$  denota este fato, e não que  $\mathbf{x}$  é uma função de  $t$ .

2. Fator de aprendizagem constante  $\alpha \in (0, 1)$  para ajustar os conseqüentes das regras nebulosas  $w_i(t)$ ;
3. Fator de redução  $\gamma \in (0, 1)$  para o ajuste da dispersão dos conjuntos nebulosos.

A inicialização do algoritmo de aprendizagem (em  $t = 0$ ), é feita de acordo com o seguinte procedimento (Ballini, 2000):

- Fazer o número inicial de regras  $Q(0) = 0$ .
- Definir as funções de pertinência  $F_j^i(\cdot)$  correspondentes às regras nebulosas.
- Escolher valores para  $\alpha$ ,  $\beta$  e  $\gamma$ .
- Definir o valor inicial da dispersão das funções de pertinência,  $\theta$ .

Ao apresentar o par  $(\mathbf{x}, d)$  à RNN, a rede codifica a primeira regra nebulosa. A seguir o valor  $Q(t)$  é atualizado e os valores modais definidos como sendo cada componente do vetor de entrada e o conseqüente como a saída desejada correspondente. Esta etapa pode ser resumida da seguinte forma:

**para**  $i = 1$  **faça**:

$$c_j^i(t+1) = x_j(t), \quad j = 1, \dots, M \quad (3.22)$$

$$r^i(t+1) = \theta, \quad (3.23)$$

$$w^i(t+1) = d(t) \quad (3.24)$$

$$\eta^i(t+1) = 0 \quad (3.25)$$

A partir desta etapa, duas situações são possíveis (Figueiredo et al., 2004):

1.  $\exists i \mid H^i(\mathbf{x}(t)) \neq 0$ . O grupo  $v$  vence a competição se

$$H^v(\mathbf{x}(t)) \geq H^i(\mathbf{x}(t)), \quad \forall i \in \{1, \dots, Q(t)\} \quad (3.26)$$

Há então dois casos:

- Caso o desempenho atinja o nível desejado, ou seja,  $\|d(t) - y(t)\| \leq \delta$ , os parâmetros são ajustados da seguinte forma:

$$c_j^v(t+1) = c_j^v(t) + \varepsilon^v(t)(x_j(t) - c_j^v(t)), \quad j = 1, \dots, M \quad (3.27)$$

$$c_j^i(t+1) = c_j^i(t), \quad i \neq v \quad (3.28)$$

$$\eta^v(t+1) = \eta^v(t) + 1 \quad (3.29)$$

$$\eta^i(t+1) = \eta^i(t), \quad i \neq v \quad (3.30)$$

$$w^i(t+1) = w^i(t) + \alpha \left( \frac{(d(t) - y(t))H^i(\mathbf{x}(t))}{\sum_{h=1}^{Q(t)} H^h(\mathbf{x}(t))} \right) \quad (3.31)$$

$$r^i(t+1) = r^i(t), \quad i = 1, \dots, Q(t) \quad (3.32)$$

$$Q(t+1) = Q(t) \quad (3.33)$$

- Caso o desempenho não atinja o nível desejado, ou seja,  $\|d(t) - y(t)\| > \delta$ , um novo grupo de neurônios é adicionado à rede, isto é  $Q(t+1) = Q(t) + 1$ , e os parâmetros são ajustados da seguinte maneira:

**para**  $i = 1$  **até**  $Q(t)$  **faça**

**se**  $H^i(\mathbf{x}(t)) \neq 0$  **então**

$$r^i(t+1) = \gamma r^i(t) \quad (3.34)$$

**fim se**

$$c_j^i(t+1) = c_j^i(t) \quad (3.35)$$

$$w^i(t+1) = w^i(t) \quad (3.36)$$

$$\eta^v(t+1) = \eta^v(t) + 1 \quad (3.37)$$

$$\eta^i(t+1) = \eta^i(t), \quad i \neq v \quad (3.38)$$

**fim para**

**para**  $i = Q(t+1)$  **faça**

$$c_j^i(t+1) = x_j(t) \quad (3.39)$$

$$r^i(t+1) = D(\mathbf{c}^v(t), \mathbf{x}(t)) \quad (3.40)$$

$$w^i(t+1) = d(t) \quad (3.41)$$

$$\eta^i(t+1) = 0 \quad (3.42)$$

**fim para**

com,  $0 < \gamma < 1$  é um fator de redução e  $D(\mathbf{c}^v(t), \mathbf{x}(t))$  é a distância Euclidiana entre o padrão de entrada  $\mathbf{x}(t)$  e  $\mathbf{c}^v(t) = (c_1^v, \dots, c_M^v)'$ .

2.  $H^i(t)(\mathbf{x}(t)) = 0, \forall i \in \{1, \dots, Q(t)\}$ . Neste caso, um novo grupo deve ser adicionado à rede e  $D(\mathbf{c}^i(t), \mathbf{x}(t))$  é calculado entre o padrão de entrada  $\mathbf{x}(t)$  e  $\mathbf{c}^i(t) = (c_1^i, \dots, c_M^i)'$ . O vencedor  $v$  é tal que  $D(\mathbf{c}^v(t), \mathbf{x}(t)) \leq D(\mathbf{c}^i(t), \mathbf{x}(t)), \forall i \in \{1, \dots, Q(t)\}$ . Nenhuma atualização é feita nos valores modais, nas dispersões, nos pesos e no valor de  $\eta^i$ . Para o novo grupo  $Q(t+1)$ , os parâmetros são inicializados como:

$$c_j^i(t+1) = x_j(t) \quad (3.43)$$

$$r^i(t+1) = D(\mathbf{c}^v(t), \mathbf{x}(t)) \quad (3.44)$$

$$w^i(t+1) = d(t) \quad (3.45)$$

$$\eta^i(t+1) = 0 \quad (3.46)$$

Para um número *finito* de padrões de entrada  $N$ , o treinamento termina quando os seguintes critérios são satisfeitos:

$$e_c = \max_j |c_j(t+1) - c_j(t)| \leq \delta_c \quad (3.47)$$

$$e_y = \frac{1}{N} \sum_{l=1}^N (d^l(t) - y^l(t))^2 \leq \delta_y \quad (3.48)$$

sendo  $y^l$  a saída correspondente à entrada  $x^l(t)$  apresentada na iteração  $t$  e  $\delta_c$  e  $\delta_y$  limites previamente selecionados.

Na próxima seção, a rede neural nebulosa descrita será aplicada para previsão um passo à frente para a série de vazões médias mensais.

### 3.5.1 Aplicação da Rede Neural Nebulosa para Previsão de Vazões

Nesta seção, a rede neural nebulosa (RNN) apresentada é aplicada à previsão de vazões. Assim como nos capítulos anteriores, são construídos 12 modelos diferentes de RNN, um para cada mês do ano (Ballini et al., 2000). Os dados utilizados são referentes ao posto de Sobradinho entre 1931 e 1998, sendo que o período de 1991 a 1998 foi escolhido para testar o desempenho dos modelos.

Os dados foram normalizados para o intervalo  $[0, 1]$ . Esta normalização é realizada para simplificar a entrada na rede pois, desta forma, pode-se considerar que a primeira camada da RNN contém  $M$  grupos, cada um contendo um neurônio (Ballini, 2000). A Tabela 3.2 mostra o número de entradas, definido, neste caso, com os mesmos valores da rede MLP.

Tabela 3.2: Topologias das RNN escolhidas para cada mês.

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
$M$	2	6	1	3	3	6	2	2	2	3	3	1
Regras	49	48	47	50	41	46	44	47	44	48	47	46

As funções de pertinência das regras foram selecionadas como funções Gaussianas, definidas como:

$$F_j^i(x_j) = \begin{cases} 0, & \text{se } |x_j - c_j^i| > 2r^i \\ \exp\left(\frac{-|x_j - c_j^i|}{r^i}\right), & \text{caso contrário} \end{cases} \quad (3.49)$$

e os valores dos parâmetros da aprendizagem, os quais são escolhidos experimentalmente, são mostrados na Tabela 3.3.

Tabela 3.3: Parâmetros da aprendizagem.

$\alpha$	$\beta$	$\gamma$	$\Theta$	$\delta_c$	$\delta_y$	$\delta$
0.5	1.0	0.9	2.0	0.01	0.01	0.09

Os operadores de agregação *t-norma* e *s-norma*, para a segunda camada, foram adotados como produto e máximo, respectivamente. Para a terceira camada, o operador *t-norma* escolhido foi o produto.

Durante a aprendizagem, a RNN define o número de regras que melhor representa o universo de discurso. Para o posto de Sobradinho, o número de regras para cada mês é mostrado na Tabela 3.2. O número de entradas variou de 1 a 6, dependendo do mês. Depois do treinamento, o número de regras nebulosas ajustadas, para cada mês, variou de 41 a 50. O maior número de regras nebulosas ocorre nos meses úmidos do ano porque durante esses períodos, os dados de vazões apresentam grande variância.

Após o treinamento, a arquitetura da rede está ajustada para a previsão dos dados de teste (1991-1998), que resulta na curva de previsões de vazões mostrada na Figura 3.6.

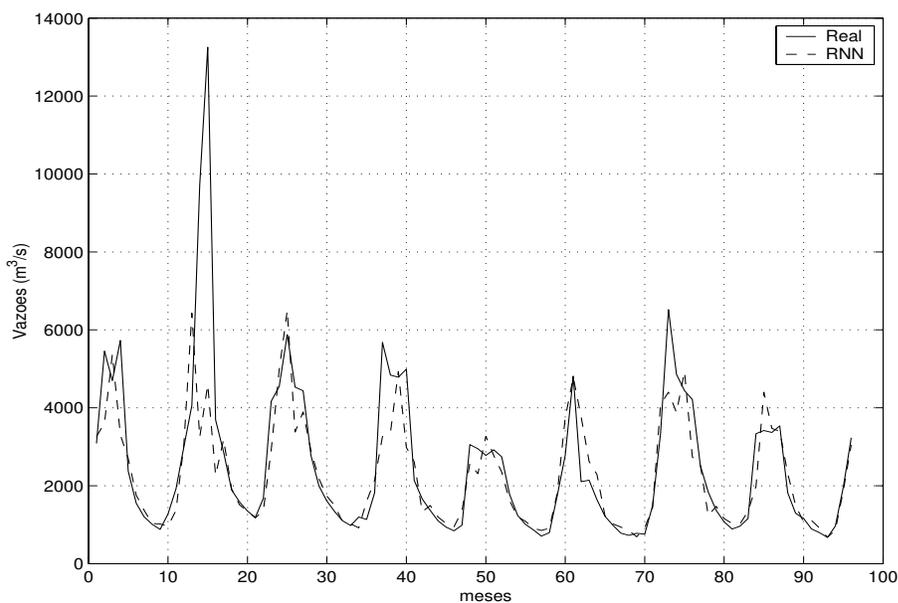


Figura 3.6: Previsão de vazões do posto de Sobradinho utilizando o modelo de RNN para o período de 1991 a 1998.

### **3.6 Resumo**

Redes neurais artificiais são largamente utilizadas na literatura para a solução de uma grande variedade de problemas. Com o avanço dos métodos de aprendizagem e das arquiteturas, as RNAs têm se mostrado uma ferramenta bastante promissora, em particular, para o problema de previsão de séries temporais.

Neste capítulo, apresentou-se, dois tipos de RNAs. Inicialmente, foi introduzida a topologia genérica das redes MLP com o treinamento por retropropagação. O outro modelo combina o poder dos sistemas nebulosos de extrair conhecimento à capacidade de aprendizado das redes neurais, resultando na chamada rede neural nebulosa. Adotou-se uma arquitetura de rede adaptativa com uma metodologia de aprendizagem construtiva e competitiva.

Ambos os modelos de rede foram aplicados ao problema de previsão de vazões médias mensais um passo à frente para a usina hidroelétrica de Sobradinho no período de 1991 a 1998.

## Capítulo 4

# Agrupamento de Dados em Previsão

Além das técnicas apresentadas nos Capítulos 2 e 3, há, alternativamente, modelos baseados em agrupamentos de dados utilizados para previsão de séries temporais. Essencialmente, a idéia de aplicar algoritmos de agrupamento em previsão consiste em organizar os elementos da série em grupos com características similares.

Apesar do agrupamento de dados ser um conceito comum em modelos estatísticos, existem também algoritmos de agrupamento baseados na teoria de conjuntos nebulosos. Esses algoritmos constituem generalizações dos algoritmos clássicos e são utilizados, nesse trabalho, como base para o desenvolvimento de modelos destinados à previsão.

### 4.1 Introdução

O objetivo dos algoritmos de agrupamento é identificar duas ou mais coleções de dados que formam uma estrutura de grupo e que possam ser associadas a classes. Em sistemas cujos dados contêm um grande número de atributos, o algoritmo de agrupamento, ao particionar a base de dados em um conjunto de grupos, também proporciona um mecanismo de sumarização e compactação de informação (Cardoso, 2003).

Atualmente, algoritmos de agrupamento de dados estão sendo utilizados em diversas áreas, como economia, biologia e medicina, entre outras. Em previsão de séries temporais, os algoritmos de agrupamento são utilizados como componentes de metodologias híbridas. Geralmente, essas metodologias híbridas são desenvolvidas em duas fases. Na primeira, através dos algoritmos de agrupamento, o objetivo é organizar os

dados em grupos com características similares, estabelecendo padrões e capturando comportamentos semelhantes ao longo da série. Entre outros algoritmos, destacam-se um método clássico de detecção de pontos mutáveis (Oh e Han, 2001) e uma variação hierárquica do algoritmo *fuzzy c-means* (Geva, 1999). Supondo que os dados estão rotulados, a segunda fase consiste em aplicar métodos, como os descritos nos Capítulos 2 e 3, a cada uma das diferentes classes pré-estabelecidas. A principal diferença entre os métodos existentes reside na escolha do método de agrupamento.

A metodologia descrita e aplicada neste capítulo utiliza o algoritmo de agrupamento *fuzzy c-means* e, diferentemente dos modelos citados, faz aproximações utilizando reconhecimento de padrões ou mediana dos dados para realizar a previsão ao invés de modelar cada grupo com uma metodologia de previsão diferente (Magalhães, Ballini, Gonçalves e Gomide, 2004), (Magalhães, Ballini, Soares e Gomide, 2004).

A aplicação de algoritmos de agrupamento em previsão de séries de vazão é promissora considerando que as características dos dados dessas séries incluem relações entrada-saída não-lineares, incertezas devido à esparsidade da rede de coleção de dados hidrométricos e ainda natureza não-estacionária causada por períodos úmidos e secos intervalados ao longo do ano.

## 4.2 O Algoritmo *Fuzzy C-Means* (FCM)

O algoritmo de agrupamento de dados nebuloso *fuzzy c-means* (Bezdek, 1981) baseia-se em um modelo não linear de otimização que agrupa dados com características similares de acordo com um procedimento iterativo de minimização de uma função objetivo que representa um critério de partição, ponderado pelos graus de pertinência dos dados aos respectivos grupos.

O algoritmo FCM utiliza as seguintes notações. Seja um conjunto finito de padrões  $P = \{p_1, p_2, \dots, p_N\} \subset \mathbb{R}^s$ , onde  $\mathbb{R}^s$  é um espaço Euclidiano  $s$ -dimensional;  $c$  é o número de grupos, com  $2 \leq c \leq N$ ;  $U = [\mu_{ij}]$  é a matriz de pertinência, onde  $\mu_{ij}$ ,  $1 \leq i \leq c$  e  $1 \leq j \leq N$  denota o grau de pertinência do padrão  $p_j$  ao grupo  $i$ . Uma  $c$ -partição de  $P$  é

definida por:

$$M_{fc} = \left\{ U \mid \mu_{ij} \in [0, 1], \forall i, j; \sum_{i=1}^c \mu_{ij} = 1, \forall j; 0 < \sum_{j=1}^N \mu_{ij} < N, \forall i \right\} \quad (4.1)$$

O algoritmo FCM procura agrupar os dados minimizando a seguinte função:

$$J_m(U, V) = \sum_{j=1}^N \sum_{i=1}^c \mu_{ij}^m d_{ij}^2, \quad U \in M_{fc}, \quad 1 < m < \infty \quad (4.2)$$

sendo  $d_{ij} = \|p_j - \nu_i\|_A$  a distância entre  $p_j$  e  $\nu_i$  dada por uma norma induzida por um produto interno, isto é,  $\|x\|_A = x^T A x$ , com  $A$  uma matriz  $s \times s$  definida positiva;  $V = \{\nu_1, \nu_2, \dots, \nu_c\}$  é um conjunto de centros de grupo em que  $\nu_i$  é denominado o  $i$ -ésimo grupo; e  $m$  é o fator que define o grau de nebulosidade da partição nebulosa do sistema.

A seguir, os passos básicos do algoritmo FCM:

---

#### Algoritmo 1 FCM

---

Dado  $P$ , escolher o número de grupos  $1 < c < N$ , o parâmetro  $m > 1$ , o critério de parada  $\epsilon > 0$  e o número máximo de iterações  $lmax$ .

**1.** Inicializar  $U^{(0)}$  e o contador de iterações  $l = 1$ .

**2.** Calcular os  $c$  centros de grupos  $\{\nu_1^{(l)}, \nu_2^{(l)}, \dots, \nu_c^{(l)}\}$  utilizando  $U^{(l)}$ , com a equação:

$$\nu_i^{(l)} = \frac{\sum_{j=1}^N (\mu_{ij})^m p_j}{\sum_{j=1}^N (\mu_{ij})^m}, \quad i = 1, 2, \dots, c$$

**3.** Utilizando  $\nu_i^{(l)}$ , atualizar  $U^{(l-1)}$  com o seguinte procedimento:

Para  $j = 1$  até  $N$

Se  $\|p_j - \nu_i^{(l)}\|^2 > 0$

$$\mu_{ij}^l = \left[ \sum_{k=1}^c \left( \frac{\|p_j - \nu_i\|^2}{\|p_j - \nu_k\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}$$

Se  $\|p_j - \nu_i^{(l)}\|^2 = 0$

$$\mu_{ij}^l = 1, \quad 1 \leq i \leq c$$

**4.** Calcular  $\Delta = \|U^{(l)} - U^{(l-1)}\| = \max_{i,j} |\mu_{ij}^{(l)} - \mu_{ij}^{(l-1)}|$

Se  $\Delta > \epsilon$  ou  $l < lmax$

$l = l + 1$  e voltar ao passo **2**

Senão parar.

---

### 4.3 Modelo de Previsão de Séries Temporais Baseado em Agrupamento Nebuloso

Esta seção detalha o modelo de previsão baseado em agrupamento nebuloso (MPAN) proposto nesta dissertação. O método de previsão é composto por duas fases. A primeira é referente ao agrupamento de padrões e a segunda à classificação de novos padrões e previsão.

Neste trabalho, considera-se a previsão um passo à frente, isto é, dadas  $p_{t-1} \in \mathfrak{R}$ ,  $t = 1, \dots$ , amostras de uma série temporal, o objetivo é estimar o valor de  $p_t$ , utilizando informação de um conjunto de valores passados de  $p_t$ .

Seja um padrão genérico  $(p + 1)$  dimensional construído pela concatenação de  $p$  valores passados amostrais de uma série temporal da seguinte maneira:

$$p_j = [p_{t-p}^j \quad p_{t-p+1}^j \quad \cdots \quad p_{t-1}^j \quad p_t^j] \quad (4.3)$$

com  $j = 1, 2, \dots, N$  o número de  $(p + 1)$ -padrões<sup>1</sup>.

Os elementos  $p_{t-k}^j$ ,  $k = 1, \dots, p$  dos  $(p + 1)$ -padrões são escolhidos dependendo da informação requerida para uma previsão acurada. Neste caso, adota-se  $p = l$ ,  $l \in \{1, 2, \dots, L\}$  valores passados da série temporal de vazões, obtendo:

$$p_j = [v_{t-l}^j \quad v_{t-l+1}^j \quad \cdots \quad v_{t-1}^j \quad v_t^j] \quad (4.4)$$

O  $(p + 1)$ -padrão de dados  $p_j$  também pode ser composto, não só pelos  $l$  valores anteriores  $v_{t-k}$ ,  $k = 1, \dots, l$ , como também pelas  $l - 1$  inclinações correspondentes, isto é, pelas primeiras diferenças  $(v_{t-k+1} - v_{t-k})$ ,  $k = 2, \dots, l$ . Assim,  $p = 2l - 1$  e tem-se que:

$$p_j = [(v_{t-l+1}^j - v_{t-l}^j) \cdots (v_{t-1}^j - v_{t-2}^j) \quad v_{t-l}^j \cdots v_{t-1}^j \quad v_t^j] \quad (4.5)$$

Claramente, diferenças de ordem superior podem ser adotadas para definir os  $(p + 1)$ -padrões de dados.

O algoritmo FCM, descrito na seção 4.2, é aplicado, na primeira fase do modelo MPAN, aos  $(p + 1)$ -padrões de dados definidos por (4.4), embora a expressão (4.5) possa,

<sup>1</sup>A notação  $(p + 1)$ -padrões é adotada pelo fato do padrão possuir dimensão  $(p + 1)$ .

de forma análoga, ser considerada para o agrupamento. Como resultado, obtém-se a matriz de centros de grupos dada por:

$$C = \begin{bmatrix} \nu_{11} & \nu_{12} & \cdots & \nu_{1(p+1)} \\ \nu_{21} & \nu_{22} & \cdots & \nu_{2(p+1)} \\ \vdots & \vdots & \ddots & \vdots \\ \nu_{i1} & \nu_{i2} & \cdots & \nu_{i(p+1)} \\ \vdots & \vdots & \ddots & \vdots \\ \nu_{c1} & \nu_{c2} & \cdots & \nu_{c(p+1)} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_i \\ \vdots \\ \nu_c \end{bmatrix} \quad (4.6)$$

com  $\nu_i \in \mathbb{R}^{p+1}$ . O número apropriado de centros  $c$  é escolhido empiricamente. Isto é, o modelo é testado para  $c$  variando de 1 a 10 e aquele que apresentar o melhor desempenho, ou seja, os menores erros de previsão, indica a melhor escolha do parâmetro  $c$ . Esses centros de grupo serão utilizados no próximo estágio, para a classificação de novos padrões de entrada.

A segunda fase classifica novos  $(p+1)$ -padrões  $p_q$ , denominados  $(p+1)$ -padrões de previsão, de acordo com os grupos nebulosos (4.6) obtidos na primeira fase, e calcula as previsões. O  $(p+1)$ -padrão de previsão tem como último componente  $(p+1)$ , o valor a ser previsto. Os demais  $p$  elementos compõem o  $(p+1)$ -padrão de previsão de forma similar à composição dos  $(p+1)$ -padrões de dados definidos em (4.4). Novamente, tomando  $p = l$  e os  $t - k$ ,  $k = 1, \dots, l$  valores imediatamente anteriores, tem-se que:

$$p_q = [v_{t-l}^q \quad v_{t-l+1}^q \quad \cdots \quad v_{t-1}^q \quad v_t] \quad (4.7)$$

ou, juntamente com as  $l - 1$  inclinações correspondentes,

$$p_q = [(v_{t-l+1}^q - v_{t-l}^q) \cdots (v_{t-1}^q - v_{t-2}^q) \quad v_{t-l}^q \cdots v_{t-1}^q \quad v_t] \quad (4.8)$$

A classificação dos  $(p+1)$ -padrões de previsão (4.7) é feita utilizando a seguinte equação do algoritmo FCM:

$$u_{iq} = \left[ \sum_{k=1}^c \left( \frac{\|p_q - \nu_i\|^2}{\|p_q - \nu_k\|^2} \right)^{\frac{1}{m-1}} \right]^{-1} \quad (4.9)$$

sendo  $u_{iq}$  o grau de pertinência do  $(p+1)$ -padrão de previsão  $p_q$  ao grupo  $i$ . Assim, o valor da previsão  $\hat{v}_t$  é computado como segue:

$$\hat{v}_t = u_{1q}\nu_{1(p+1)} + u_{2q}\nu_{2(p+1)} + \cdots + u_{cq}\nu_{c(p+1)} = \sum_{i=1}^c u_{iq}\nu_{i(p+1)} \quad (4.10)$$

Note que como o último componente do  $(p + 1)$ -padrão de previsão  $p_q$  é o valor a ser previsto, deve haver um mecanismo que substitua  $v_t$  para a segunda fase, em que há classificação. Isso significa que o valor de  $v_t$  em  $p_q$  deve ser substituído por uma aproximação adequada. Neste trabalho, introduz-se dois procedimentos para se obter tais aproximações. No primeiro, a aproximação é dada pela mediana dos valores de  $v_t^j$ ,  $j = 1, \dots, N$ . O segundo, denominado reconhecimento de padrões, utiliza os  $p$  primeiros componentes dos  $(p + 1)$ -padrões de previsão  $p_q$  para escolher o mais próximo entre os  $N$   $(p + 1)$ -padrões de dados  $p_j$ .

A seguir, ambos esquemas serão descritos com maiores detalhes.

#### 4.3.1 Procedimento de Mediana de Dados (PMed)

Existem diversos valores estatísticos, tais como média, moda e mediana, de se representar um conjunto de dados. A escolha da mediana como aproximação de  $v_t$  é adotada nesse trabalho pelo fato dessa medida corresponder à metade de uma distribuição de frequência, isto é, a mediana depende de uma posição, o que faz com que seu valor seja menos sensível a valores extremos.

A mediana é calculada a partir da ordenação ascendente dos elementos de um conjunto de dados. Então, o valor da mediana corresponde ao valor do elemento localizado no meio do conjunto ordenado. Assim, se o número de dados de um conjunto é ímpar, a mediana é o valor central; caso o número de dados seja par, a mediana é, por convenção, a média aritmética dos dois valores centrais.

A partir do conceito descrito acima, a construção dos  $(p + 1)$ -padrões de previsão para a segunda fase é feita a partir da mediana de  $v_t^j$ ,  $j = 1, \dots, N$ . Formalmente:

$$p_q = [v_{t-l}^q \quad v_{t-l+1}^q \quad \cdots \quad v_{t-1}^q \quad md(v_t^j)] \quad (4.11)$$

Em suma, o modelo MPAN com a aproximação pela mediana segue os passos mostrados no Algoritmo 2.

**Algoritmo 2** MPAN-PMed

Dada uma série temporal, escolher o valor de  $l$  dependendo da informação requerida para uma previsão acurada.

1. Construir os  $(p + 1)$ -padrões de dados  $p_j$  e de previsão  $p_q$  com  $p = l$  ou  $p = 2l - 1$ .
2. Escolher o número de grupos  $c$  tal que  $1 < c < N$  e aplicar o algoritmo FCM descrito na seção 4.2, para obter a matriz de centros de grupos  $C = [\nu_1 \cdots \nu_i \cdots \nu_c]^T$ .
3. Calcular a mediana de  $v_t^j, j = 1, \dots, N$ .
4. Nos  $(p + 1)$ -padrões de previsão  $p_q$ , substituir o valor de  $v_t$  pela aproximação fornecida pelo valor da mediana.
5. Utilizando  $C$ , classificar cada um dos  $(p + 1)$ -padrões de previsão  $p_q$  através da equação:

$$u_{iq} = \left[ \sum_{k=1}^c \left( \frac{\|p_q - \nu_i\|^2}{\|p_q - \nu_k\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}$$

6. Computar o valor das previsões através da expressão:

$$\hat{v}_t = \sum_{i=1}^c u_{iq} \nu_{i(p+1)}$$

**4.3.2 Procedimento de Reconhecimento de Padrões (PRP)**

A idéia principal deste procedimento é incluir informações sobre a tendência do  $(p + 1)$ -padrão de previsão para selecionar o valor que melhor aproxima o valor real  $v_t$ . A tendência do  $(p + 1)$ -padrão de previsão é definida tanto pela concatenação dos  $p = l$  valores de  $v_{t-k}^q, k = 1, \dots, l$  quanto pela concatenação dos  $p = 2l - 1$  valores  $v_{t-k}^q$  com suas respectivas inclinações.

A diferença dos padrões definidos para o PRP está na exclusão dos valores  $v_t^j$  e  $v_t$ , o que reduz para  $p$  a dimensão desses padrões. Dessa forma, esses vetores de tendência são denominados  $p$ -padrões de dados ou de previsão, seguindo a notação adotada anteriormente. Assim, os  $p$ -padrões de dados podem ser tanto:

$$p_j = [v_{t-l}^j \quad v_{t-l+1}^j \quad \cdots \quad v_{t-1}^j] \quad (4.12)$$

quanto:

$$p_j = [(v_{t-l+1}^j - v_{t-l}^j) \cdots (v_{t-1}^j - v_{t-2}^j) \quad v_{t-l}^j \cdots v_{t-1}^j] \quad (4.13)$$

e, similarmente, os  $p$ -padrões de previsão correspondentes podem ser:

$$p_q = [v_{t-l}^q \quad v_{t-l+1}^q \quad \cdots \quad v_{t-1}^q] \quad (4.14)$$

ou

$$p_q = [(v_{t-l+1}^q - v_{t-l}^q) \cdots (v_{t-1}^q - v_{t-2}^q) \quad v_{t-l}^q \cdots v_{t-1}^q] \quad (4.15)$$

assumindo  $p = l$  ou  $p = 2l - 1$ , respectivamente.

A aproximação é obtida pelo valor de  $v_t^j$  do  $p$ -padrão de dados que mais se aproxima do  $p$ -padrão de previsão  $p_q$ . Mais precisamente, o procedimento de reconhecimento de padrões (PRP) compara os  $N$  valores dos  $p$ -padrões de dados com um  $p$ -padrão de previsão utilizando a norma Euclidiana como medida de distância. Logo, o valor  $v_t^j$  do  $p$ -padrão de dados, cuja norma é a menor, fornece a aproximação para  $v_t$ . Portanto, matematicamente, o  $p$ -padrão de dados mais próximo,  $p_j^*$ , é tal que:

$$\|p_q - p_j^*\| = \min_{1 \leq j \leq N} \|p_q - p_j\| \quad (4.16)$$

e o valor  $v_t^j$  imediatamente posterior ao último componente de  $p_j^*$  estabelece a aproximação para  $v_t$ .

Os passos básicos do modelo MPAN com a aproximação de  $v_t$  dada pelo PRP estão descritos no Algoritmo 3.

**Algoritmo 3** MPAN-PRP

Dada uma série temporal, escolher o valor de  $l$  dependendo da informação requerida para uma previsão acurada.

1. Construir os  $(p + 1)$ -padrões de dados  $p_j$  e de previsão  $p_q$  com  $p = l$  ou  $p = 2l - 1$ .
2. Escolher o número de grupos  $c$  tal que  $1 < c < N$  e aplicar o algoritmo FCM descrito na seção 4.2, para obter a matriz de centros de grupos  $C = [\nu_1 : \nu_c]$ .
3. Escolher experimentalmente o valor de  $l$  e a necessidade da inclusão das inclinações na construção dos  $p$ -padrões de dados e de previsão.
4. Para cada  $p$ -padrão de previsão  $p_q$ , computar o  $p$ -padrão de dados  $p_j^*$  tal que:

$$\|p_q - p_j^*\| = \min_{1 \leq j \leq N} \|p_q - p_j\|$$

adotando-se  $\|\cdot\|$  como a norma Euclidiana.

5. Tomar o valor de  $v_t^j$  imediatamente posterior ao último elemento de  $p_j^*$  como aproximação para  $v_t$ .
6. Utilizando  $C$ , classificar cada um dos  $(p + 1)$ -padrões de previsão  $p_q$  através da equação:

$$u_{iq} = \left[ \sum_{k=1}^c \left( \frac{\|p_q - \nu_i\|^2}{\|p_q - \nu_k\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}$$

7. Computar o valor das previsões através da expressão:

$$\hat{v}_t = \sum_{i=1}^c u_{iq} \nu_{i(p+1)}$$

## 4.4 Aplicação do Modelo baseado em Agrupamento FCM para Previsão de Vazões

A metodologia de previsão descrita na seção 4.3 visa prever o valor da série um passo à frente e está baseada na organização das amostras das séries através do algoritmo FCM descrito na seção 4.2. Analogamente aos Capítulos 2 e 3, o modelo baseado no agrupamento FCM será utilizado para prever valores de vazões do Posto de Sobradinho.

Para cada mês, foram disponibilizados 68 anos (1931 a 1998) de dados históricos de vazão, os quais foram normalizados no intervalo  $[-1, 1]$ . Considera-se os primeiros 60 anos para o agrupamento e os 8 anos restantes para o teste.

Um modelo MPAN é desenvolvido e utilizado para cada mês  $t$ ,  $t = 1, \dots, 12$ . Des-

sa forma, de acordo com (4.4) e (4.7), os  $(p + 1)$ -padrões de dados e de previsão têm, respectivamente, a seguinte forma genérica:

$$p_j = [v_{t-l}^j \quad v_{t-l+1}^j \quad \cdots \quad v_{t-1}^j \quad v_t^j], \quad j = 1, \dots, 60 \quad (4.17)$$

$$p_q = [v_{t-l}^q \quad v_{t-l+1}^q \quad \cdots \quad v_{t-1}^q \quad v_t], \quad q = 61, \dots, 68 \quad (4.18)$$

Inicialmente, para melhor compreender os procedimentos de medianas (PMed) e reconhecimento de padrões (PRP), descritos nas seções 4.3.1 e 4.3.2, considera-se, como exemplo, as vazões do mês de setembro, ou seja,  $t = 9$ . O número de características relevantes é definido como  $l = 1$  na equação (4.17). Assim, para o PMed, os  $(p + 1)$ -padrões de dados, para o mês 9 no ano  $j$ , têm a forma:

$$p_9^j = [v_8^j \quad v_9^j] \quad (4.19)$$

Estes padrões são construídos e agrupados. Para o conjunto de dados considerado, o número de grupos foi escolhido experimentalmente como  $c = 4$ . Os grupos são mostrados na Figura 4.1.

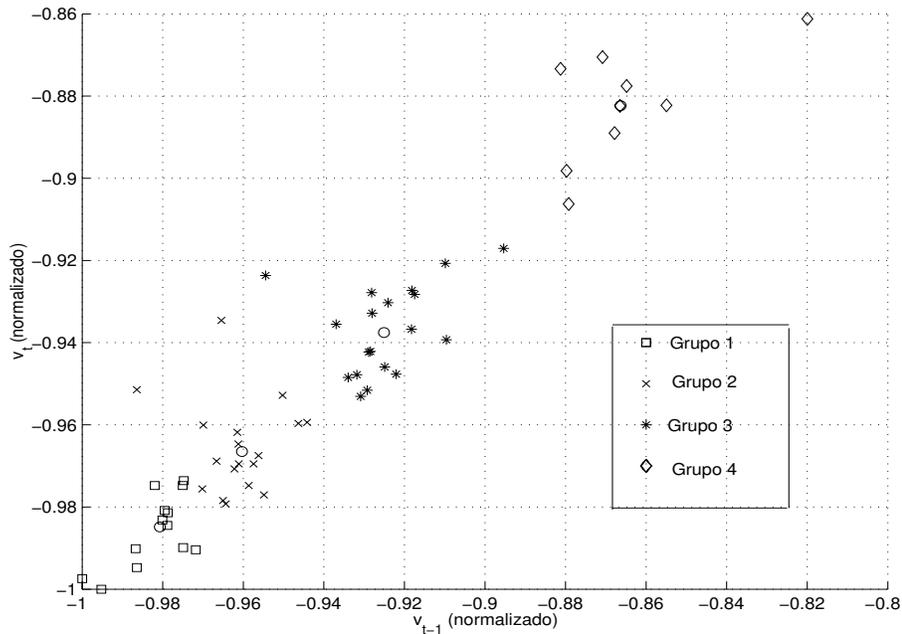


Figura 4.1: Grupos obtidos através do algoritmo FCM para PMed. Os círculos são os centros dos grupos e cada forma geométrica representa um grupo.

Para a classificação, os  $(p + 1)$ -padrões de previsão são definidos, segundo (4.18), com  $l = 1$  e  $t = 9$ , como:

$$p_9^q = [v_8^q \quad v_9] \quad (4.20)$$

Neste caso, a mediana dos  $v_9^j$ ,  $j = 1, \dots, 60$  é calculada e adotada como aproximação para a vazão referente ao mês de setembro de 1991 ( $q = 61$ ), representada por  $v_9$ , do padrão de previsão (4.20). Note que para o procedimento PMed, a aproximação para  $v_9$ , para todo  $q$  é a mesma pois, neste exemplo, ela é calculada de acordo com um banco de dados fixo composto por 60 dados, ou seja,  $v_9 = med(v_9^j) = -0,9528$ , para  $q = 61, \dots, 68$ .

Após a classificação, utiliza-se os centros de grupos  $\nu_i$ ,  $i = 1, \dots, c$  encontrados na primeira fase para computar a previsão para  $\hat{v}_9$  através da equação (4.10):

$$\hat{v}_9 = u_{1,61}\nu_{1,2} + u_{2,61}\nu_{2,2} + u_{3,61}\nu_{3,2} + u_{4,61}\nu_{4,2} = \sum_{i=1}^4 u_{i,61}\nu_{i,2} \quad (4.21)$$

isto é,  $c = 4$ ,  $q = 61$  e  $p = 1$ .

Esse esquema é repetido para prever os valores de vazões para setembro de 1992 a 1998, ou seja, para  $q = 62, \dots, 68$ . Na Figura 4.2, observa-se a previsão das vazões de setembro para os 8 anos de teste.

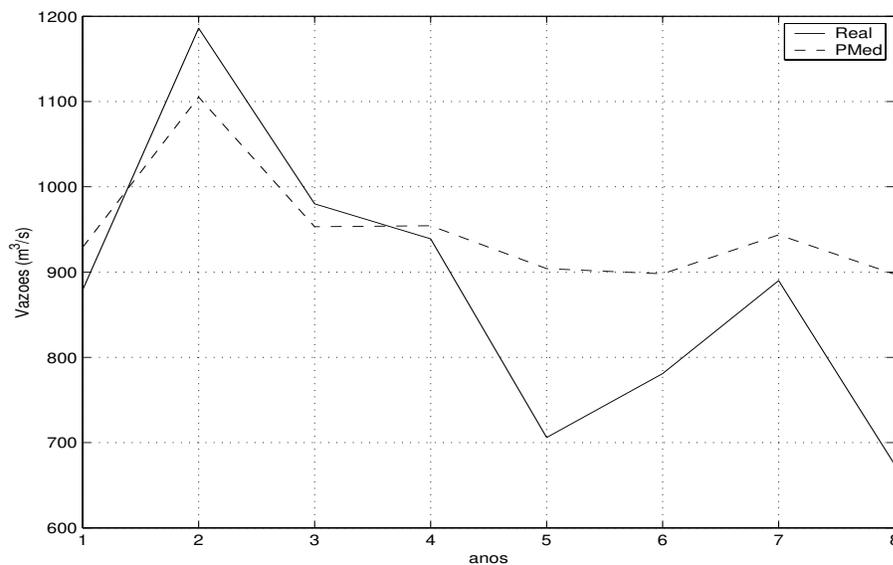


Figura 4.2: Previsão de vazões para setembro utilizando o PMed. As linhas contínuas e tracejadas representam os valores reais e previstos, respectivamente.

Para o esquema PRP, os  $(p + 1)$ -padrões de dados  $p_9^j$  definidos em (4.19) são agrupados como no procedimento PMed. Entretanto, o número de centros de grupo, escolhido empiricamente, é  $c = 8$ . A Figura 4.3 mostra esses grupos.

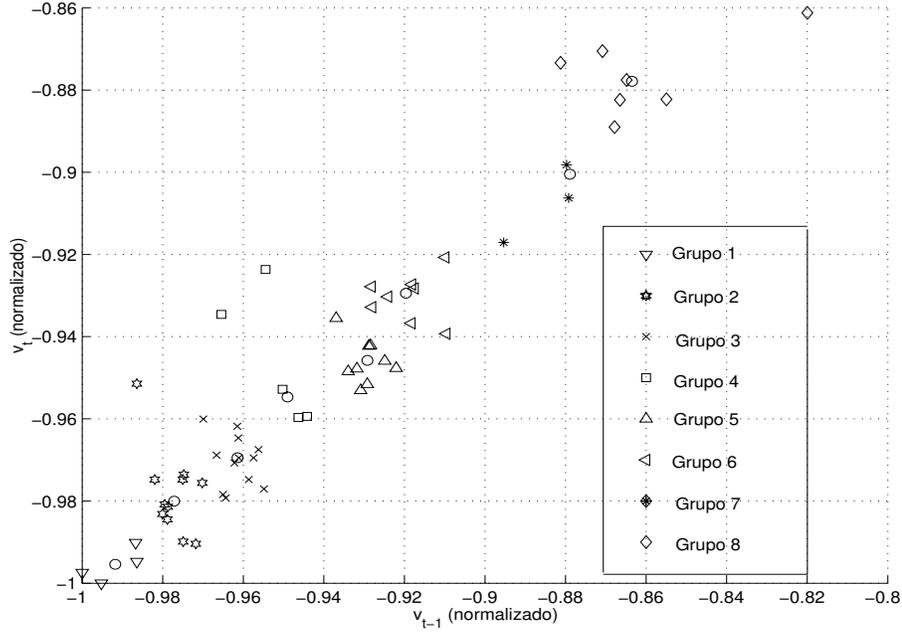


Figura 4.3: Grupos obtidos através do algoritmo FCM para o PRP. Os círculos são os centros dos grupos e cada forma geométrica representa um grupo.

Para a classificação, considera-se através de experimentos, que o número de características relevantes para os  $p$ -padrões de dados e os  $p$ -padrões de previsão é  $l = 2$ . Logo, esses padrões têm, respectivamente, a forma (segundo (4.13) e (4.15)):

$$p_j = [(v_8^j - v_7^j) \quad v_7^j \quad v_8^j] \quad (4.22)$$

$$p_q = [(v_8^q - v_7^q) \quad v_7^q \quad v_8^q] \quad (4.23)$$

As inclinações foram consideradas para a previsão de vazões de setembro porque foi observado que essas inclinações auxiliam na acurácia das previsões.

Logo, para prever a vazão de setembro de 1991 ( $q = 61$ ), utiliza-se os dados de vazão de julho e agosto de 1991 para construir o  $p$ -padrão de previsão correspondente e a equação (4.16) para encontrar o  $p$ -padrão de dados mais próximo dentre os 60

$p$ -padrões de dados. A Figura 4.4 mostra as vazões reais de julho ( $v_7^q$ ), agosto ( $v_8^q$ ) e setembro ( $v_9$ ) contidas nos  $p$ -padrões de previsão e o  $p$ -padrão de dados  $p_j^*$  obtido. Assim, a aproximação para  $v_9$  no  $p$ -padrão de previsão é adotada como o valor imediatamente posterior a  $v_8^*$  do  $p$ -padrão de dados  $p_j^*$ .

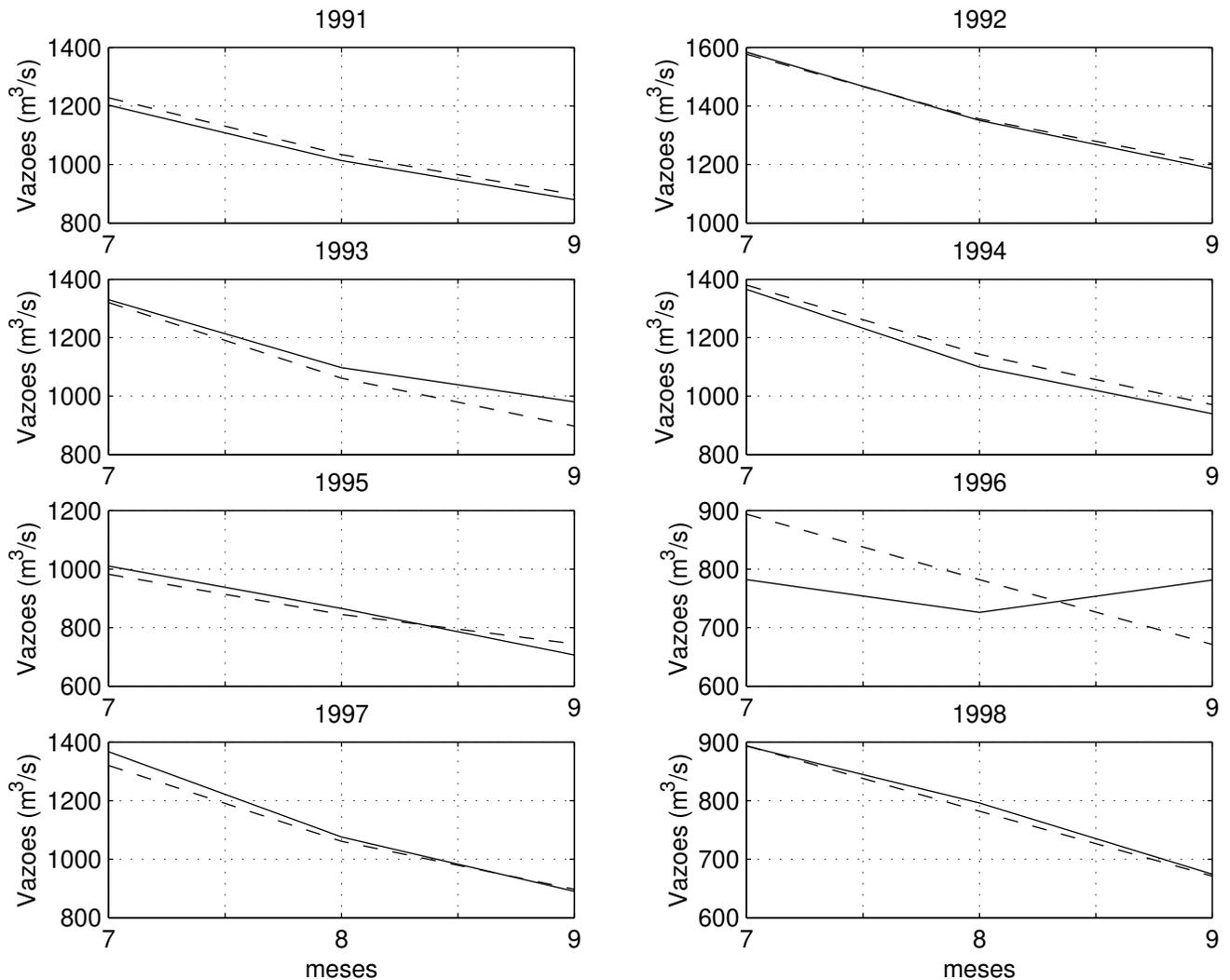


Figura 4.4: Tendência da curva de vazões para Setembro. As linhas contínuas e tracejadas representam, respectivamente, os valores de  $p_q$  e  $p_j$ .

Uma vez obtida a aproximação  $v_9$ ,  $q = 61$ , o  $p$ -padrão de previsão pode ser classificado e utilizando (4.10), a previsão é calculada para a vazão de setembro de 1991, da

seguinte forma:

$$\hat{v}_9 = u_{1,61}\nu_{1,2} + u_{2,61}\nu_{2,2} + \dots + u_{8,61}\nu_{8,2} = \sum_{i=1}^8 u_{i,61}\nu_{i,2} \quad (4.24)$$

Os passos descritos acima são, então, repetidos para prever as vazões restantes de setembro, ou seja, para  $q = 62, \dots, 68$ . A Figura 4.4 mostra a comparação das tendências para os anos 1992, ..., 1998. Observando essas tendências e a Tabela 4.1, nota-se que as aproximações têm valores bastante próximos dos reais, com exceção do ano de 1996.

Tabela 4.1: Valores reais normalizados e aproximações  $v_9$  dadas pelo procedimento PRP.

	1991	1992	1993	1994	1995	1996	1997	1998
Real	-0,9719	-0,9305	-0,9583	-0,9639	-0,9953	-0,9852	-0,9705	-0,9996
Aproximação	-0,9696	-0,9282	-0,9695	-0,9597	-0,9902	-1,0000	-0,9695	-1,0000

Depois de aplicado o procedimento para todos os anos de teste, as previsões para setembro podem ser observadas na Figura 4.5.

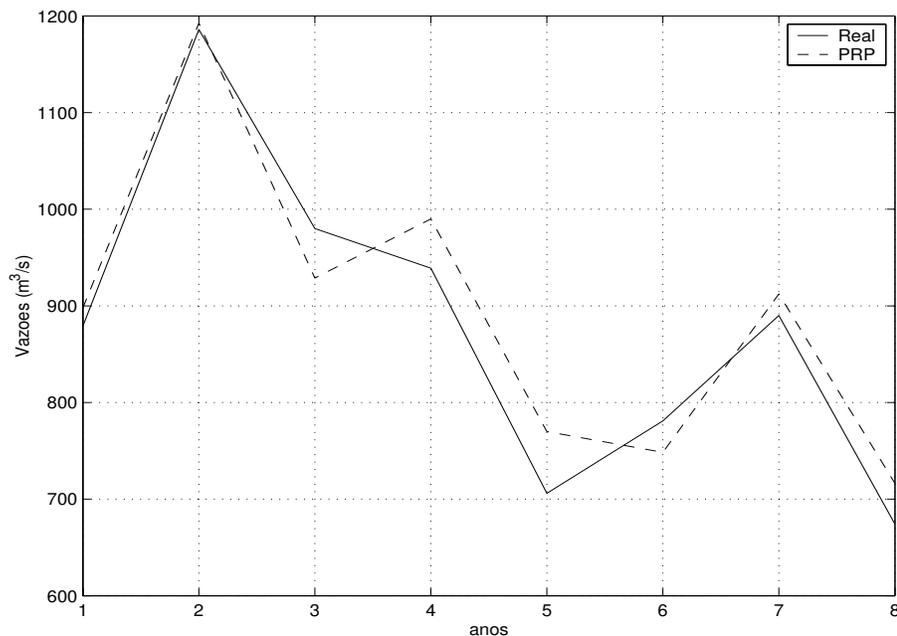


Figura 4.5: Previsão de vazões para setembro utilizando o PRP. As linhas contínuas e tracejadas representam os valores reais e previstos, respectivamente.

A Tabela 4.2 possui os dados que caracterizam os procedimentos PMed e PRP, para cada mês do ano. Pode-se observar que o número de centros de grupos,  $c$ , varia de acordo com o procedimento empregado. Além disso, é necessário estabelecer mais dois parâmetros: o número de características  $l$  para os  $p$ -padrões de dados e de previsão, para o esquema PRP e a inclusão ou não das inclinações. Esses parâmetros são obtidos experimentalmente para todos os meses.

Tabela 4.2: Características para o modelo MPAN com PMed e PRP.

Meses	PRP			PMed
	$c$	$l$ (para os $p$ -padrões)	Inclusão da Inclinação	$c$
Jan	2	3	não	2
Fev	3	3	não	6
Mar	4	3	sim	3
Abr	7	3	não	5
Mai	8	3	não	8
Jun	10	2	não	9
Jul	8	1	não	5
Ago	8	1	não	4
Set	8	2	sim	4
Out	5	2	não	2
Nov	4	2	não	2
Dez	3	2	não	2

Em particular, para o posto de Sobradinho, observa-se que a inclinação é agregada aos  $p$ -padrões de dados e de previsão apenas nos meses de março e setembro. Possivelmente, isso se deve ao fato desses meses representarem as transições entre períodos úmidos e secos para esse reservatório. Por exemplo, os valores de vazão de março são antecidos de valores altos correspondentes aos meses de dezembro, janeiro e fevereiro e procedidos de valores mais baixos, em decréscimo (meses de abril, maio e junho). Assim, a inclinação que possuía valores positivos começa a apresentar valores negativos e por isso influencia na escolha das aproximações de  $v_t$ . O mês de setembro é caracterizado da mesma forma com o diferencial de apresentar como precedentes, valores de vazão baixos relativos aos meses de julho e agosto, e posteriores altos correspondentes aos meses de outubro e novembro.

As Figuras 4.6 e 4.7 mostram os resultados de previsão de vazões após aplicação dos esquemas PMed e PRP, respectivamente, para todos os meses  $t = 1, \dots, 12$  de todos os anos de teste  $q = 61, \dots, 68$ .

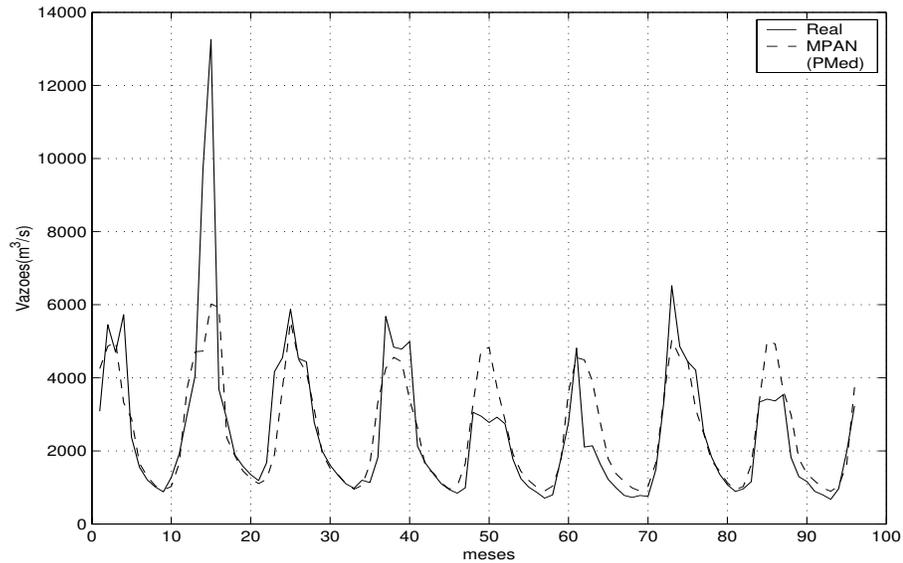


Figura 4.6: Previsão de vazões utilizando o modelo baseado em agrupamento com PMed para o período entre 1991 e 1998 do posto de Sobradinho.

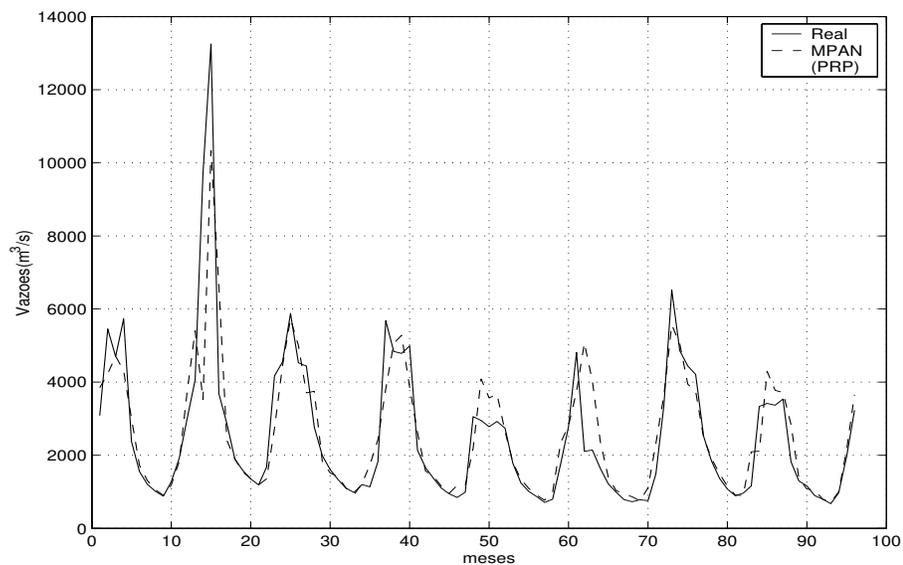


Figura 4.7: Previsão de vazões utilizando o modelo baseado em agrupamento com PRP para o período entre 1991 e 1998 do posto de Sobradinho.

## 4.5 Resumo

Neste capítulo apresentou-se um modelo baseado em agrupamento nebuloso de dados destinado à previsão de séries temporais. Esse método é dividido em 2 estágios. No primeiro, o algoritmo *fuzzy c-means* (FCM) é aplicado aos padrões de dados  $p_j$ . O objetivo é organizar padrões em grupos com características similares presentes nos dados. Como resultado, obtêm-se os representantes desses grupos chamados centros de grupo.

O próximo passo é utilizar os centros de grupo obtidos no primeiro estágio para classificar novos padrões ou padrões de previsão,  $p_q$ . Considerando que o procedimento envolve grandezas nebulosas, o resultado dessa classificação é uma matriz de graus de pertinência dos dados em relação aos centros de grupo. Através dessa matriz e dos centros de grupo, a previsão é feita como uma combinação linear dos centros ponderados pelos graus de pertinência do padrão de previsão a estes centros.

O modelo foi aplicado à série de vazões do posto de Sobradinho. Analogamente aos capítulos anteriores, foi aplicado um modelo de agrupamento para cada mês do ano. Os resultados gráficos foram significativamente superiores utilizando o procedimento PRP, devido a sua capacidade de capturar a tendência da curva de vazões (os resultados analíticos computados pelos erros de previsão serão mostrados no Capítulo 6).

No próximo capítulo descreve-se o outro modelo de previsão proposto neste trabalho. Esse modelo combina as previsões dos métodos mais promissores a fim de obter resultados que explorem as potencialidades de cada modelo individual.



# Capítulo 5

## Combinação de Previsores

Assim como as metodologias descritas nos capítulos anteriores, diversas outras foram propostas e aperfeiçoadas por pesquisadores a fim de atingir desempenho superior no campo de previsão de séries temporais.

Dessa forma, nas últimas décadas, surgiu a idéia de combinar os resultados dos modelos de previsão mais eficientes, com o objetivo de agregar as características relevantes de cada um para uma determinada aplicação.

### 5.1 Introdução

Agregação ou combinação é um campo de pesquisa extenso na área estatística. Na modelagem de sistemas, agregação significa combinar classes de modelos para representar mais efetivamente um sistema, segundo atribuição adequada de pesos para cada um desses modelos (Magalhães, Ballini, Molck e Gomide, 2004). A principal diferença entre os métodos de combinação de previsores é a maneira como esses pesos são calculados. Por isso, atualmente, existem duas grandes classes de combinações, as lineares e as não-lineares.

Mecanismos típicos de combinação linear de dados incluem operadores de agregação como mínimo, máximo e média. Entretanto, o operador média ignora a habilidade e a relevância dos previsores individuais, uma vez que pesos iguais são atribuídos para cada previsão local.

A idéia de combinar os resultados dos modelos de previsão mais eficientes, de forma menos intuitiva, foi proposta originalmente por Bates e Granger (Bates e Granger, 1969), sugerindo um esquema de combinação linear, cujos coeficientes ou pesos são calculados utilizando uma matriz de covariância de erros de previsão individuais.

Desde então, muitos estudos e testes empíricos vêm mostrando que os modelos de combinação linear de previsores aprimoram o desempenho das previsões individuais (Clemen, 1989), (Armstrong, 1989), (Makridakis, 1989). Esses modelos foram extensivamente explorados por pesquisadores e compõem um vasto conteúdo na literatura que abrange desde modelos de combinação com média simples até modelos não lineares mais complexos, como os baseados em inteligência computacional (Menezes et al., 2000). Nesse contexto, pode-se destacar pesquisas recentes na estimação dos pesos através do uso de redes neurais artificiais (Donaldson e Kamstra, 1996), redes neurais evolutivas (Harrald e Kamstra, 1997) e redes neurais nebulosas (Jingrong, 2000), (Jingrong, 2002). Um outro exemplo, que utiliza um paradigma da inteligência computacional, é um método de combinação não linear baseado nos sistemas nebulosos Takagi-Sugeno (Fiordaliso, 1998).

Em geral, o desempenho dessas metodologias de combinação não lineares é computado pela comparação com modelos de combinação lineares para diversas aplicações, com ênfase em séries econômicas. Essa comparação geralmente é favorável à utilização de modelos de combinação não lineares, como as redes neurais artificiais (Khotanzad et al., 2000) e as aproximações neurais nebulosas (Palit e Popovic, 2000).

Além das combinações de modelos, outras aproximações, como as arquiteturas *ensembles* (Sharkey, 1999), redes neurais modulares (Petridis e Kehagias, 1998) e mistura de especialistas, são desenvolvidas com base, essencialmente, na mesma idéia das combinações, isto é, com o objetivo de capturar as características mais importantes de previsores promissores.

Neste capítulo, propõe-se um modelo simples, porém efetivo para a combinação linear de previsões, com uma estrutura neural composta por uma única camada com um neurônio linear. A seguir, o modelo será detalhado e posteriormente aplicado à previsão de séries temporais.

## 5.2 Modelo de Combinação baseado em Redes Neurais

De uma forma geral, o mecanismo de combinação é dividido em duas fases. Na primeira,  $r$  modelos locais ou individuais são selecionados de acordo com os **erros globais de previsão** fornecidos por cada um deles. Assim, dado um conjunto de dados, esses modelos são aplicados e posteriormente, as previsões individuais obtidas são coletadas e combinadas utilizando um operador de combinação (Figura 5.1) treinado para explorar a capacidade de previsão de cada previsor na previsão final.

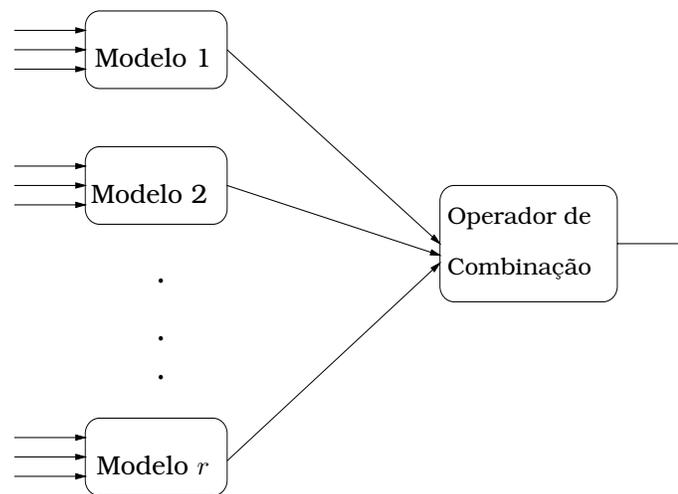


Figura 5.1: Modelo de combinação.

Considerando a primeira etapa, os modelos de previsão locais podem ser lineares, como os do Capítulo 2, baseados em redes neurais, como os do Capítulo 3, baseados em agrupamento de dados, como no Capítulo 4, ou ainda, qualquer previsor qualificado como apropriado para o problema de interesse.

Na segunda etapa, para combinar os resultados individuais de diferentes modelos de previsão, sugere-se, neste trabalho, um método de combinação baseado nas redes neurais artificiais, porém mais simples. O operador de combinação é uma rede neural composta por um neurônio linear, conforme mostra a Figura 5.2, com  $\hat{y}_j, j = 1, \dots, r$  denotando a saída de cada modelo individual.

Note que se trata de um caso especial de rede neural, em que o processamento é realizado através de apenas um neurônio linear em uma única camada, ou seja, trata-se de um combinador linear. Portanto, o procedimento de aprendizagem é realizado da

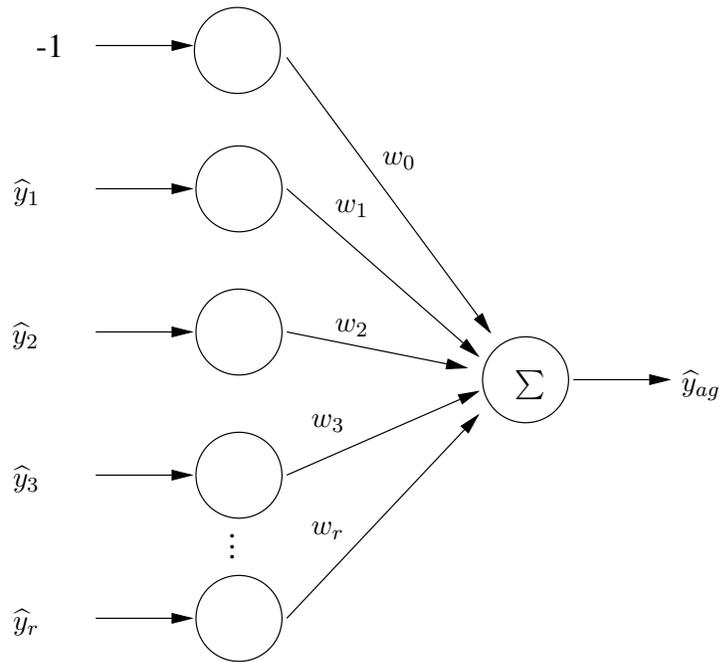


Figura 5.2: Agregador neural.

mesma forma que na seção 3.4, através do método do gradiente descendente, em que os pesos da rede,  $w_i, i = 1, 2, \dots, r$ , são ajustados visando minimizar o erro quadrático médio de previsão, isto é:

$$\min E = \frac{1}{N} \sum_{l=1}^N \epsilon^l(t) \quad (5.1)$$

sendo,  $N$  o número de padrões e  $\epsilon^l(t)$  o erro quadrático de previsão de cada padrão de entrada  $l$  na iteração  $t$

$$\epsilon^l(t) = \frac{1}{2} (d^l(t) - \hat{y}_{ag}^l(t))^2 \quad (5.2)$$

com  $d^l(t)$  a saída desejada e  $\hat{y}_{ag}^l(t)$  a saída do combinador dada por:

$$\hat{y}_{ag}^l(t) = \sum_{i=0}^r w_i(t) \hat{y}_i^l(t) \quad (5.3)$$

Analogamente à seção 3.4, a atualização dos pesos é feita pelo método do gradiente da seguinte forma:

$$w_i(t) = w_i(t) + \Delta w_i(t) \quad (5.4)$$

o qual,

$$\Delta w_i(t) = -\alpha \frac{\partial \epsilon^l(t)}{\partial w_i(t)} \quad (5.5)$$

com  $\alpha$  representando a taxa de aprendizagem. Logo, utilizando a regra da cadeia, derivando (5.2) em relação a  $\widehat{y}_{ag}^l(t)$ , (5.3) em relação a  $w_i(t)$ , considerando o neurônio linear e substituindo em (3.9):

$$\Delta w_i(t) = -\alpha(d^l(t) - \widehat{y}_{ag}^l(t))\widehat{y}_i^l(t) \quad (5.6)$$

Os  $r$  modelos individuais que são as entradas do combinador neural são definidas a partir do desempenho de cada modelo para uma determinada aplicação.

A seguir, o agregador neural será utilizado para combinar modelos locais destinados às séries de vazões.

### 5.3 Aplicação do Modelo de Combinação para Previsão de Vazões

Nesta seção, o combinador neural é aplicado à série de vazões de Sobradinho, cujos dados compreendem o período entre 1931 e 1998. Adota-se como conjunto de teste as vazões de 1991 a 1998, analogamente aos Capítulos anteriores.

A escolha dos modelos foi feita de acordo com o desempenho apresentado por cada método local para este posto. Assim, a combinação foi feita entre três modelos: o periódico autoregressivo médias móveis (PARMA - Capítulo 2), a rede neural nebulosa (RNN - Capítulo 3) e o baseado em agrupamento nebuloso com procedimento de reconhecimento de padrões (MPAN-PRP - Capítulo 4).

Para utilizar o modelo de combinação proposto, os dados foram normalizados no intervalo  $[-1, 1]$  e os pesos inicializados no intervalo  $[-0, 01; 0, 01]$ . A taxa de aprendizagem foi fixa e dada por  $\alpha = 0,01$ , resultando numa busca lenta mas capaz de minimizar a função objetivo, dada em (5.1). Os pesos resultantes da combinação neural estão dispostos na Tabela 5.1. Observa-se que os pesos que ponderam as entradas referentes ao modelo RNN têm valores mais altos que os demais, nos meses mais úmidos (janeiro, fevereiro, março, abril, outubro, novembro e dezembro). Por outro lado, nos meses mais secos, há um equilíbrio entre os valores dos pesos, para este caso especificamente.

Tabela 5.1: Pesos resultantes da combinação neural para o posto de Sobradinho.

<b>Meses</b>	$w_1$ (PARMA)	$w_2$ (RNN)	$w_3$ (PRP)
Jan	0,1626	0,6921	0,3434
Fev	0,0556	0,6557	0,4326
Mar	0,1622	0,4599	0,4189
Abr	0,1180	0,6224	0,4065
Mai	0,2238	0,3329	0,3292
Jun	0,2757	0,2875	0,2761
Jul	0,2521	0,3102	0,4091
Ago	0,3153	0,3300	0,2348
Set	0,2621	0,4274	0,3088
Out	0,1557	0,6653	0,3264
Nov	0,1955	0,5313	0,4548
Dez	0,0992	0,6371	0,4032

A Figura 5.3 mostra os resultados da combinação neural.

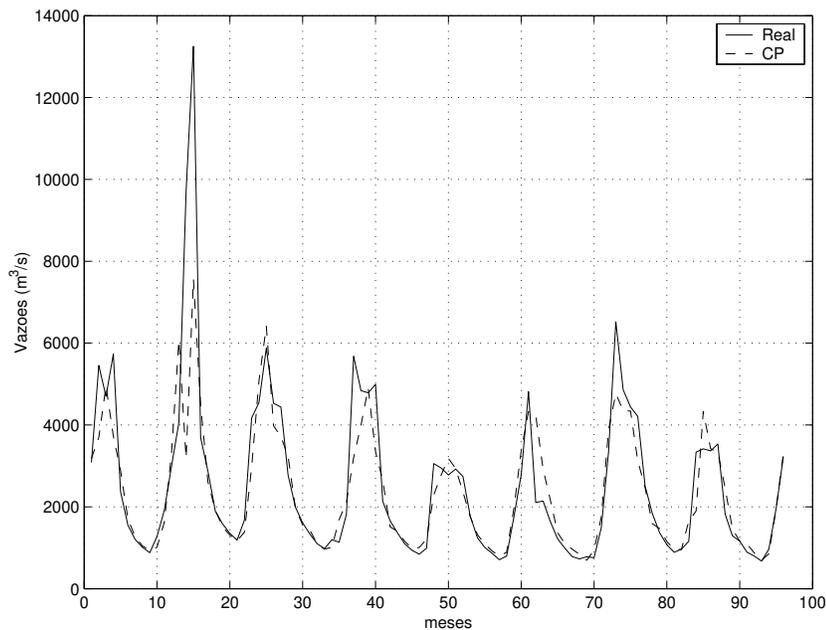


Figura 5.3: Previsão de vazões utilizando o modelo de combinação neural para o período entre 1991 e 1998 do posto de Sobradinho.

É possível observar que as regiões de “pico”, que apresentam maior dificuldade em serem previstas, seguem de forma satisfatória a curva de valores reais, com exceção do segundo período de cheia. Isso porque nenhum dos modelos individuais apresentou um bom desempenho na modelagem desse “pico”, uma vez que este, por ser um evento anômalo, não apresenta uma tendência semelhante ao histórico de dados.

## **5.4 Resumo**

Neste Capítulo um modelo de combinação linear de previsores foi proposto para aprimorar a previsão de séries temporais. Trata-se de um modelo constituído em duas fases. Na primeira, os modelos individuais são escolhidos de acordo com o respectivo desempenho na previsão de uma determinada série.

Na segunda, uma rede neural simples composta por apenas uma camada com um neurônio linear responsável, através do processamento das previsões locais (entradas), pela combinação. A rede é treinada pelo método do gradiente descendente a fim de ajustar os pesos adequadamente.

O modelo de combinação linear é aplicado para a previsão do período de 1991-1998 da série de vazões de Sobradinho. Os resultados mostram que a combinação fornece uma previsão mais próxima da curva real que as previsões resultantes dos modelos individuais. Isso será verificado no próximo Capítulo, em que os modelos serão comparados com base em critérios quantitativos de erros.



# Capítulo 6

## Comparação e Discussão de Resultados

### 6.1 Introdução

Os modelos, descritos nos capítulos anteriores, foram aplicados a três séries de vazões pertencentes a três bacias hidrográficas distintas do Brasil. A primeira é referente ao reservatório da usina hidroelétrica de Sobradinho que está localizada na região nordeste do Brasil, faz parte da bacia do Rio São Francisco e pertence à Companhia Hidroelétrica do São Francisco (CHESF). A seguir, utilizou-se a série de vazões referente ao reservatório de Furnas, localizado na bacia do Rio Grande, região sudeste do Brasil e pertencente à Companhia Furnas Centrais Elétricas S.A.. Finalmente, os modelos foram aplicados às vazões do reservatório de Emborcação, localizado na bacia do rio Paranaíba, também na região sudeste do Brasil. A usina hidroelétrica de Emborcação pertence à empresa Centrais Elétricas de Minas Gerais (CEMIG).

Os dados de vazões oscilam entre valores mínimos e máximos ao longo da variação sazonal no período de 12 meses. Assim, a sazonalidade das vazões mensais sugere o uso de 12 modelos diferentes, um para cada mês do ano.

A comparação foi feita com base na análise dos erros fornecidos pelas previsões de cada modelo, sendo computados tanto os erros mensais, considerando os 12 modelos distintos ajustados para cada mês, quanto os globais, relativos a toda série. Neste trabalho, foram considerados para avaliação dos erros mensais e globais, a raiz do erro quadrático médio (REQM), o erro percentual médio (EPM) e o erro percentual máximo (EPmax). Para a comparação global, adotou-se ainda o erro absoluto médio (EAM), o

coeficiente de correlação  $\rho$  e a variância ( $\sigma^2$ ). Esses critérios são definidos como:

$$REQM = \left( \frac{1}{P} \sum_{k=1}^P (d_k - \hat{y}_k)^2 \right)^{\frac{1}{2}} \quad (6.1)$$

$$EAM = \frac{1}{P} \sum_{k=1}^P |d_k - \hat{y}_k| \quad (6.2)$$

$$EPM = \frac{100}{P} \sum_{k=1}^P \frac{|d_k - \hat{y}_k|}{d_k} \quad (6.3)$$

$$EPmax = \max \left( 100 \left| \frac{d_k - \hat{y}_k}{d_k} \right| \right) \quad (6.4)$$

$$\rho = \frac{\sum_{k=1}^P (d_k - \bar{d})(\hat{y}_k - \bar{y})}{\left( \sum_{k=1}^P (d_k - \bar{d})^2 (\hat{y}_k - \bar{y})^2 \right)^{\frac{1}{2}}} \quad (6.5)$$

$$\sigma^2 = \frac{1}{P-1} \sum_{k=1}^N (\hat{y}_k - \bar{y})^2 \quad (6.6)$$

sendo  $P$  o número de previsões,  $d_k$  a saída real,  $\hat{y}_k$  a saída do modelo e  $\bar{d}$  e  $\bar{y}$  suas respectivas médias. Por exemplo, se os critérios de erros fossem computados para as previsões de 8 anos do mês de Janeiro,  $P = 8$ ,  $\bar{d}$  seria a média dos valores reais dos 8 anos previstos e  $\bar{y}$  seria a média das previsões dos 8 anos.

Em particular, o coeficiente de correlação mede o quanto as previsões estão correlacionadas com as vazões reais. Assim,  $-1 < \rho < 1$  e quanto mais próximo de um 1 estiver o valor de  $\rho$ , mais acurada é a previsão. Por outro lado, um coeficiente negativo indica uma correlação inversa, ou seja, significa uma previsão deficiente.

A seguir, a aplicação dos modelos às séries descritas será exposta e a análise dos erros detalhada.

## 6.2 Aplicação de Modelos para o Posto de Sobradinho

Ao longo deste trabalho, os dados de vazão do posto de Sobradinho foram utilizados para exemplificar a aplicação dos modelos descritos para uma série temporal. Assim, a análise e o ajuste dos modelos clássicos de séries temporais (PARMA), as redes neurais artificiais (MLP) e nebulosas (RNN), os modelos baseados em agrupamento nebuloso (MPAN-PMed, MPAN-PRP) e o método de combinação de previsores (CP), para esse posto foram apresentados nos Capítulos 2, 3, 4 e 5, respectivamente. Vale ressaltar que foram realizadas previsões de vazão um passo à frente para o período de 1991 e 1998.

### 6.2.1 Análise de Erros

Inicialmente, foram testados os modelos já conhecidos na literatura de previsão de séries temporais. Os coeficientes  $\phi_{i,m}$  e  $\theta_{j,m}$ ,  $i = 1, \dots, p_m$ ,  $j = 1, \dots, q_m$ , e as ordens  $p_m$  e  $q_m$  dos modelos AR e MA, respectivamente, são mostrados na Tabela 2.3 do Capítulo 2. Da mesma forma, as características das topologias utilizadas nas redes MLP e RNN e os parâmetros do treinamento são sumarizados, respectivamente, nas Tabelas 3.1, 3.2 e 3.3 do Capítulo 3.

Os modelos propostos neste trabalho também foram aplicados à série de vazões de Sobradinho, considerando as características dadas pela Tabela 4.2, Capítulo 4, e os parâmetros de aprendizagem definidos no Capítulo 5.

Em geral, os períodos úmidos (outubro a março) apresentam maior variabilidade nas vazões, o que prejudica a modelagem da série e conseqüentemente as respectivas previsões. A Figura 6.1 mostra essa dificuldade através da curva dos erros mensais (REQM, EPM e EPmax) das previsões obtidas pelos 6 modelos descritos anteriormente (as tabelas referentes aos erros mensais de previsão estão dispostas para consulta no Apêndice A). É notável uma diminuição desses erros durante os períodos de seca (abril a setembro) e o respectivo aumento durante os períodos úmidos.

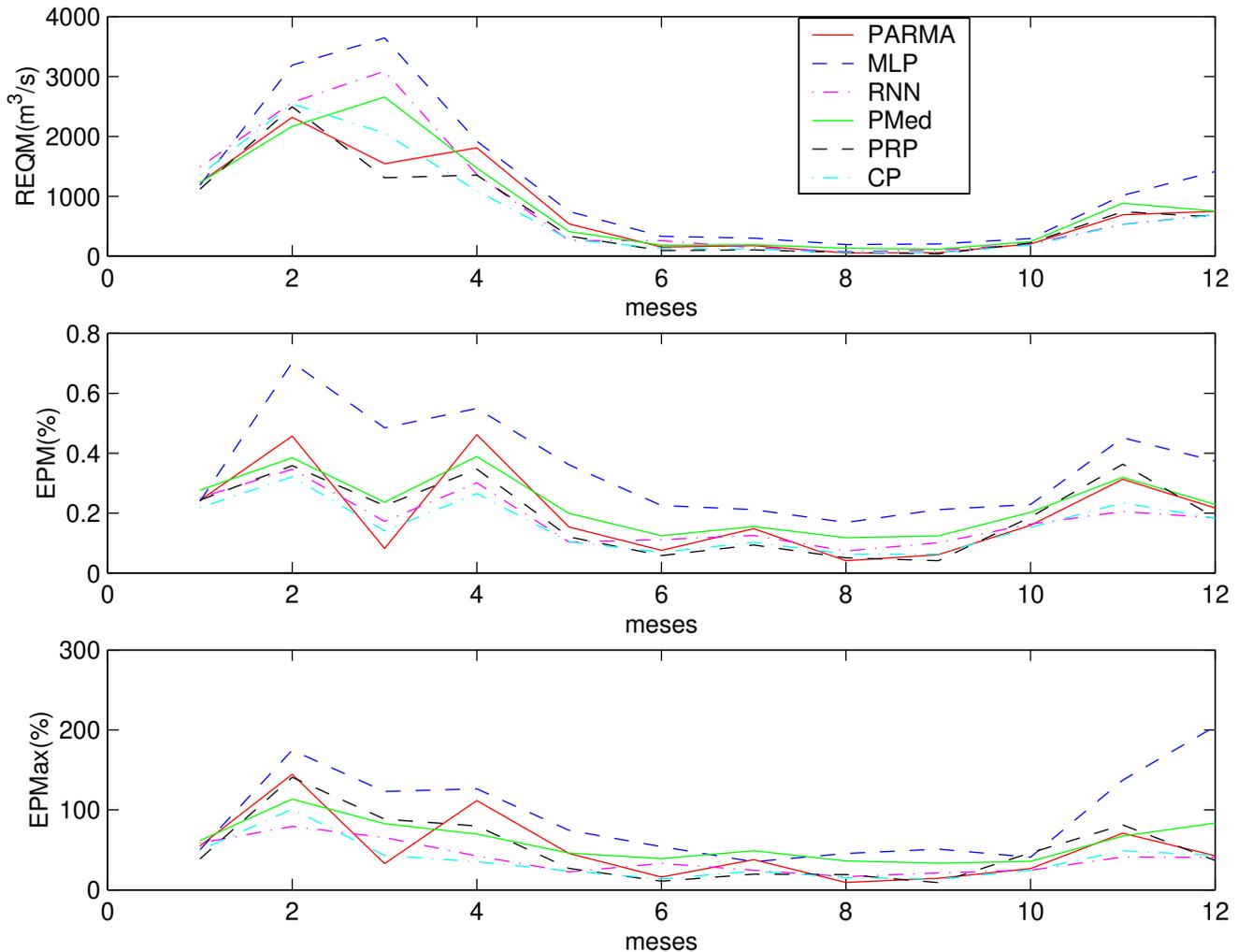


Figura 6.1: Erros mensais referentes ao posto de Sobradinho.

Ainda na Figura 6.1, com relação ao critério REQM, observa-se uma equivalência dos modelos durante as previsões do período seco. Entretanto, é possível notar para esse período que a MLP apresenta erros superiores, enquanto o MPAN-PRP e o modelo CP apresentam erros mensais inferiores aos demais modelos. Nos meses úmidos, os modelos propostos, MPAN-PRP e CP, novamente fornecem resultados satisfatórios quando comparados aos modelos já existentes, embora em janeiro e em março haja uma sutil superioridade da MLP e do PARMA, respectivamente.

Os critérios EPM e EPmax mostram com maior clareza as diferenças entre os modelos, principalmente nos meses secos. Observa-se que, para este caso, a MLP forne-

ceu resultados inferiores aos demais modelos. O modelo MPAN-PMed apresentou erros mensais ligeiramente superiores aos demais e fica comprovada a eficiência dos modelos MPAN-PRP e CP.

Vale ressaltar que embora não forneça erros de previsão inferiores para todos os meses, é notável que o modelo CP aproveita as características favoráveis dos modelos mais promissores para cada mês, como por exemplo, nos meses de julho e junho (MPAN-PRP) ou novembro (RNN).

Com relação aos erros de previsão globais, o modelo proposto MPAN-PMed apresentou desempenho inferior aos modelos PARMA e RNN (Tabela 6.1). Isso pode ser explicado pela grande variabilidade das vazões para o posto de Sobradinho, a qual não foi modelada adequadamente pela mediana dos dados, mesmo considerando a baixa sensibilidade desta medida a valores extremos.

Tabela 6.1: Erros globais de previsão para o posto de Sobradinho.

Métodos	<b>REQM</b> ( $m^3/s$ )	<b>EAM</b> ( $m^3/s$ )	<b>EPM</b> (%)	<b>EPmax</b> (%)	$\rho$	$\sigma^2$ ( $\times 10^6$ )
PARMA	1079,30	593,83	20,09	144,27	0,84	1,17
MLP	1462,80	820,24	31,31	149,70	0,69	2,16
RNN	1330,40	606,37	17,80	79,31	0,76	1,72
MPAN-PMed	1191,60	622,60	22,81	113,21	0,80	1,44
MPAN-PRP	1005,00	537,10	18,93	140,75	0,86	1,02
CP	1103,10	520,13	15,97	100,94	0,85	1,20

Ainda, a Tabela 6.1 mostra que o modelo MPAN-PRP obteve um resultado superior aos métodos clássico, neural (MLP) e MPAN-PMed. Além disso, o modelo MPAN-PRP apresentou erros globais (REQM e EAM) mais baixos e o coeficiente de correlação consideravelmente maior que a abordagem neural nebulosa (RNN). O EPM apontou um desempenho superior da RNN, com relação aos modelos individuais, decorrente da avaliação dos erros mensais de maio e novembro. A variância mais baixa é dada pelas previsões do MPAN-PRP, o que sugere que a dispersão dos erros de previsão com relação à média é pequena para este modelo implicando na equivalência da inferioridade do REQM.

Finalmente, os erros globais fornecidos pelas previsões do combinador linear fo-

ram significativamente inferiores aos dos modelos individuais, com exceção do REQM que se apresentou mais elevado do que nos modelos PARMA e MPAN-PRP (Tabela 6.1). Vale notar que, considerando os modelos individuais agregados PARMA, RNN e MPAN-PRP, a variância foi reduzida para o modelo CP, embora não tenha sido inferior à variância dos modelos PARMA e MPAN-PRP.

### 6.3 Aplicação de Modelos para o Posto de Furnas

Nesta seção, os modelos descritos nesse trabalho foram aplicados à previsão de vazões um passo à frente do posto de Furnas. Da mesma forma que na seção anterior, o conjunto de dados para o ajuste dos métodos foi adotado entre os anos de 1931 e 1990 enquanto os testes são feitos para as vazões do período de 1991 a 1998.

Inicialmente, o modelo PARMA foi ajustado utilizando o método de máxima verossimilhança e forneceu os coeficientes estimados  $\phi_{i,m}$ ,  $i = 1, \dots, p_m$ ,  $\theta_{j,m}$ ,  $j = 1, \dots, q_m$  e variância  $\tau_m$ , conforme mostra a Tabela 6.2. As ordens dos modelos AR ( $p_m$ ) e MA ( $q_m$ ) são determinadas segundo o critério BIC (equação (2.28)), cujo objetivo é encontrar o modelo mais parcimonioso possível.

Tabela 6.2: Estimativas dos parâmetros utilizando o EMV: série de vazões de Furnas.

Meses	$\phi_{1,m}$	$\phi_{2,m}$	$\phi_{3,m}$	$\phi_{4,m}$	$\theta_{1,m}$	$\tau_m$	$(p_m, q_m)$
Janeiro	0,5983	-	-	-	-	1,7149	(1,0)
Fevereiro	0,5312	-	-	-	-	1,7600	(1,0)
Março	0,4862	0,2404	-	-	-	1,9434	(2,0)
Abril	0,5400	0,1364	0,2688	-	-	3,5924	(3,0)
Maió	0,5439	0,1746	0,3130	-	-	6,5685	(3,0)
Junho	0,6877	0,2206	-	-	-	4,9054	(2,0)
Julho	0,6028	0,3769	-	-	-	10,9397	(2,0)
Agosto	0,9711	-0,0267	-	-	-	10,5399	(2,0)
Setembro	0,6058	0,1866	0,6371	-0,5793	-	5,9990	(4,0)
Outubro	0,9248	-	-	-	-0,6464	3,9039	(1,1)
Novembro	0,3531	-	-	-	-	1,2950	(1,0)
Dezembro	1,9013	-	-	-	-1,8872	2,0620	(1,1)

Segundo o critério BIC, a ordem do modelo AR variou de 1 a 4, com as maiores ordens ocorrendo nos meses de seca. Com relação à ordem do modelo MA, a parcimônia foi alcançada com a ordem 1 para os meses de outubro e dezembro. Vale ressaltar que, no mês de setembro, a ordem do modelo AR,  $p_m = 4$ , embora elevada, foi escolhida conforme os valores do critério BIC e da variância do erro ( $\sigma^2$ ) mostrados na Tabela 6.3.

Tabela 6.3: Valores do critério BIC e da variância para o mês de setembro para o posto de Furnas.

$p_m$ $q_m$	1		2		3		4		5	
	$\sigma^2$	BIC	$\sigma^2$	BIC	$\sigma^2$	BIC	$\sigma^2$	BIC	$\sigma^2$	BIC
0	0,31	-82,69	0,29	-82,26	0,26	-86,45	0,19	-100,73	0,19	-97,52
1	0,29	-82,26	0,32	-70,60	0,26	-80,68	0,51	-31,01	0,26	-72,76

Em seguida, para cada mês, uma rede MLP, com uma única camada intermediária, foi treinada e testada para um conjunto de topologias diferentes. Para determinar a estrutura da rede foi utilizada a técnica estatística de validação cruzada descrita na seção 3.4.1, Capítulo 3. A Tabela 6.4 apresenta a estrutura da rede para cada mês, ou seja, o número de entradas  $M$ , o número de neurônios intermediários  $M_1$ , a taxa de aprendizagem  $\alpha$  e o termo *momentum*  $\beta$ .

Tabela 6.4: Topologias das redes MLP para cada mês: posto de Furnas.

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
$M/M_1$	4/5	3/10	3/20	3/10	1/23	3/23	6/22	6/24	2/19	3/17	4/22	3/22
$\alpha$	0,5	0,25	0,75	0,5	0,75	0,5	0,75	0,75	0,25	0,25	0,5	0,75
$\beta$	0,75	0,25	0,75	0,75	0,75	0,75	0,25	0,75	0,75	0,25	0,25	0,75

Um outro modelo para comparação foi a rede neural nebulosa (RNN). Como descrito na seção 3.5, Capítulo 3, durante o treinamento da RNN, há um mecanismo de inserção de regras nebulosas que adiciona as regras à topologia da rede, de acordo com o espaço de entradas do problema. Dessa maneira, ao final da aprendizagem, a estrutura da RNN fornece o número de regras utilizadas, conforme mostra a Tabela 6.5. Note que o número de entradas  $M$  foi o mesmo adotado para a rede MLP.

Tabela 6.5: Topologias das RNN para cada mês: posto de Furnas.

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
$M$	4	3	3	3	1	3	6	6	2	3	4	3
Regras	44	44	34	34	30	18	17	23	20	24	12	43

O número de entradas para a RNN variou entre 3 e 6, dependendo do mês. O número de regras ou grupos de neurônios na camada intermediária variou entre 17 e 44, com o maior número de regras ocorrendo para meses úmidos já que nesses meses as vazões apresentam altas variâncias.

Embora o número de entradas  $M$  seja o mesmo para ambas redes, para a rede MLP, deve-se definir o número  $M_1$  de neurônios intermediários clássicos. Os parâmetros de aprendizagem da rede MLP, também escolhidos experimentalmente, variam de um mês para outro, enquanto que para a RNN, os parâmetros  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\Theta$ ,  $\delta_c$ ,  $\delta_y$  e  $\delta$  são definidos da mesma forma que para o posto de Sobradinho (Tabela 3.3).

Os modelos de previsão baseados em agrupamento nebuloso (MPAN), propostos nesse trabalho, também são caracterizados por parâmetros específicos. Tanto para o procedimento de reconhecimento de padrões (PRP), quanto para o de mediana de dados (PMed), é necessário escolher experimentalmente o parâmetro  $c$  que define o número de centros dos grupos. Além disso, o PRP utiliza  $l$  valores anteriores de vazões e as respectivas diferenças ou inclinações para construir os  $p$ -padrões de dados/previsão. De acordo com a Tabela 6.6, as inclinações foram relevantes nos meses de fevereiro, março e outubro, o que pode estar indicando, para este caso, uma possível mudança do período úmido para o seco em fevereiro e março, em contraste com a mudança do período seco para o úmido em outubro.

Vale ressaltar que o número de centros  $c$  definido para o modelo MPAN-PMed é diferente do definido para o modelo MPAN-PRP devido ao caráter experimental da escolha do parâmetro  $c$  e às características intrínsecas de cada método.

As previsões resultantes dos modelos MPAN-PMed e MPAN-PRP são mostradas, respectivamente, nas Figuras 6.2 e 6.3.

Tabela 6.6: Características para o modelo MPAN com PRP e PMed: Furnas.

Meses	PRP			PMed
	$c$	$l$ (para os $p$ -padrões)	Inclusão da Inclinação	$c$
Jan	2	2	não	10
Fev	2	2	sim	5
Mar	4	2	sim	5
Abr	5	3	não	3
Mai	3	2	não	3
Jun	3	2	não	3
Jul	3	2	não	3
Ago	3	1	não	2
Set	4	1	não	3
Out	8	3	sim	6
Nov	6	1	não	3
Dez	7	3	não	3

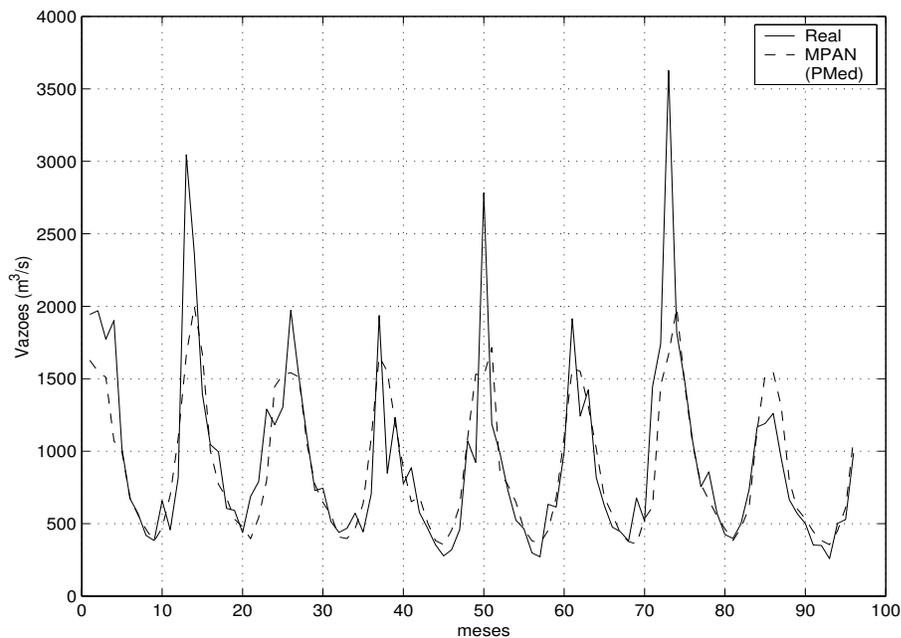


Figura 6.2: Previsão de vazões utilizando o modelo baseado em agrupamento com PMed para o período entre 1991 e 1998 do posto de Furnas.

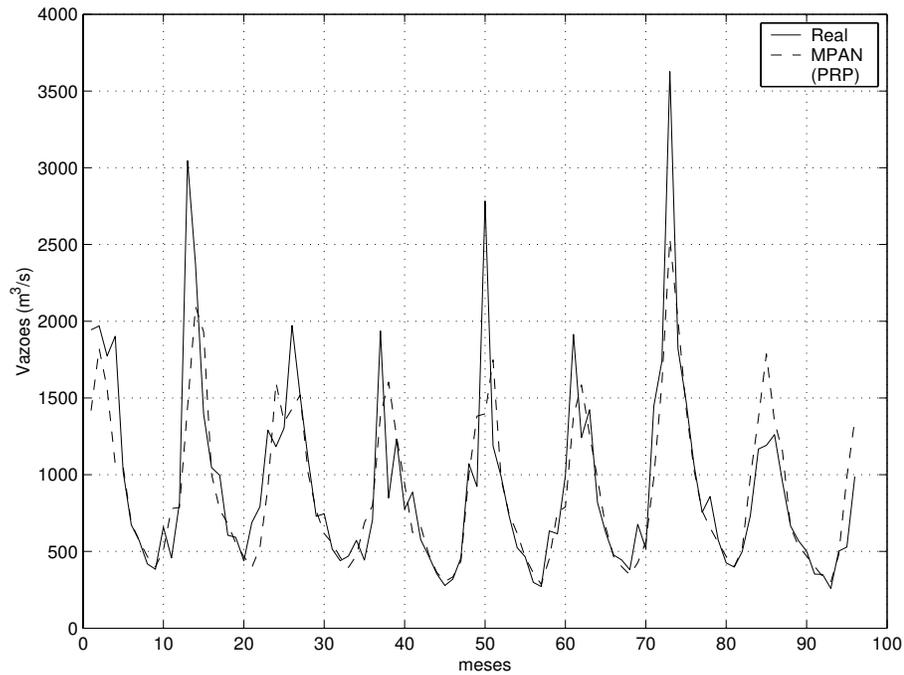


Figura 6.3: Previsão de vazões utilizando o modelo baseado em agrupamento com PRP para o período entre 1991 e 1998 do posto de Furnas.

Ainda, a combinação linear de previsões (CP) descrita no Capítulo 5 foi aplicada à série de vazões de Furnas agregando as características dos modelos RNN, MPAN-PMed e MPAN-PRP. A taxa de aprendizagem empregada foi  $\alpha = 0,01$ . Os pesos resultantes da combinação neural utilizada estão dispostos na Tabela 6.7 e mostram que, para este caso, há um equilíbrio entre os valores dos pesos para os meses de janeiro, junho, julho, setembro e outubro, enquanto que para os demais meses, com exceção de novembro, as entradas relativas à RNN são ponderadas por um peso maior que os dados aos demais modelos.

A Figura 6.4 mostra a curva prevista. Graficamente, é possível observar que as previsões do modelo CP no quarto e sexto picos acompanham melhor as não linearidades da curva real quando comparadas às curvas produzidas pelos modelos individuais MPAN-PMed (Figura 6.2) e MPAN-PRP (Figura 6.3).

Tabela 6.7: Pesos resultantes da combinação neural para o posto de Furnas.

Meses	$w_1$ (RNN)	$w_2$ (PMed)	$w_3$ (PRP)
Jan	0,4505	0,3508	0,2701
Fev	0,6616	0,3714	0,1958
Mar	0,5358	0,3556	0,2061
Abr	0,6351	0,1938	0,3387
Mai	0,6966	0,2999	0,2352
Jun	0,2938	0,2595	0,17498
Jul	0,3698	0,2830	0,2786
Ago	0,5779	0,1893	0,1758
Set	0,2960	0,2018	0,3348
Out	0,2355	0,2719	0,2977
Nov	0,0861	0,3964	0,4796
Dez	0,5180	0,1159	0,3863

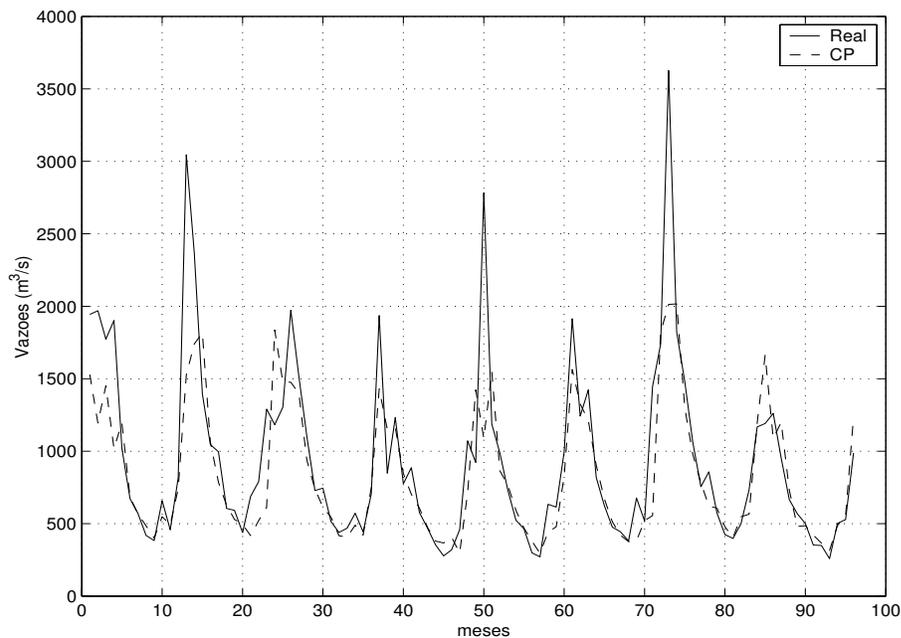


Figura 6.4: Previsão de vazões utilizando o modelo de combinação de previsores para o período entre 1991 e 1998 do posto de Furnas.

A seguir, os erros associados às previsões são analisados e comparados para avaliar o desempenho das metodologias propostas em relação às já desenvolvidas na literatura.

### 6.3.1 Análise de Erros

A análise de erros é feita de forma análoga à seção 6.2.1, considerando os mesmos critérios de avaliação de desempenho. Os erros mensais são sumarizados na Figura 6.5.

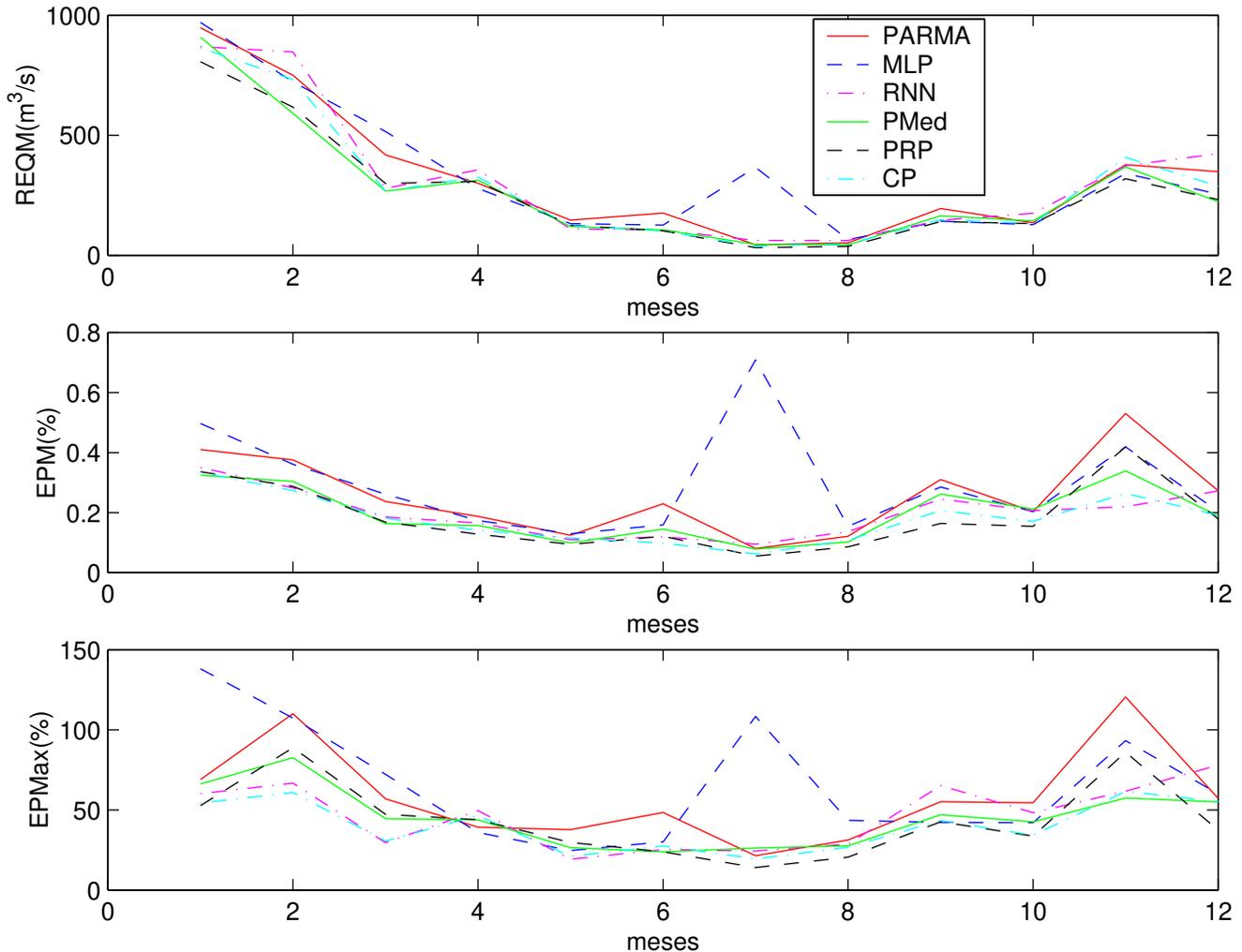


Figura 6.5: Erros mensais referentes ao posto de Furnas.

Observa-se que o critério REQM apresenta erros, cujos valores são inferiores aos apresentados para a usina de Sobradinho. Isso porque a variabilidade das vazões de Furnas é menor que as de Sobradinho. Entretanto, o desempenho da MLP é novamente inferior, excetuando-se o mês de setembro.

Por outro lado, analisando os critérios EPM e EPmax, em que é mais clara a diferença entre os modelos, os erros de previsão fornecidos pelo modelo PARMA foram, em geral, superiores aos demais modelos, com exceção do mês de agosto.

A RNN fornece erros de previsão equivalente aos modelos PMed e PRP para todos os meses com ênfase no mês de novembro em que apresenta o melhor resultado entre os modelos. Ainda, é notável que para esta usina, o modelo MPAN-PMed obtém resultados satisfatórios, contrariamente ao caso anterior, da usina de Sobradinho.

Comparando-se os erros mensais dos dois modelos MPAN, observa-se uma similaridade nos resultados com relação à maioria dos meses úmidos (janeiro, fevereiro, março e dezembro) e uma superioridade do procedimento PRP nos meses de transição (abril e outubro) e, em geral nos meses de seca (maio, julho, agosto e setembro).

Novamente, embora os erros mensais do modelo CP não apresentem valores inferiores em comparação com todos os modelos individuais, é visível a “herança” da RNN no mês de novembro e do MPAN-PRP nos meses de julho, setembro, outubro e dezembro, uma vez que nestes meses os erros do modelo CP se aproximam dos erros individuais da RNN e do MPAN-PRP, respectivamente.

Os erros globais dispostos na Tabela 6.8 seguem a mesma tendência dos erros mensais, isto é, os modelos propostos apresentam, para o posto de Furnas, resultados superiores aos fornecidos pelos modelos PARMA, rede MLP e RNN para a maioria dos critérios de erro.

Vale observar ainda que, considerando os modelos MPAN, embora os critérios REQM e EAM apresentem valores mais altos no modelo CP, há significativa redução no erros percentuais, EPM e EPmax, e da variância ( $\sigma^2$ ), e similaridade dos coeficientes de correlação  $\rho$ .

Tabela 6.8: Erros globais de previsão para o posto de Furnas.

Métodos	<b>REQM</b> ( $m^3/s$ )	<b>EAM</b> ( $m^3/s$ )	<b>EPM</b> (%)	<b>EPmax</b> (%)	$\rho$	$\sigma^2$ ( $\times 10^5$ )
PARMA	418,28	260,41	25,67	120,44	0,75	1,77
MLP	427,33	271,17	29,62	138,04	0,75	1,83
RNN	415,77	234,28	19,88	77,95	0,77	1,64
MPAN-PMed	365,61	209,03	19,77	82,60	0,82	1,32
MPAN-PRP	345,61	200,82	18,26	88,88	0,84	1,19
CP	384,58	214,66	17,89	61,60	0,81	1,39

## 6.4 Aplicação de Modelos para o Posto de Emborcação

Seguindo procedimento análogo, inicialmente, os parâmetros dos modelos são ajustados para a série de vazões de Emborcação.

A Tabela 6.9 mostra as estimativas dos parâmetros  $\phi_i$ ,  $i = 1, \dots, p_m$  e  $\theta_j$ ,  $j = 1, \dots, q_m$  dadas pelo modelo PARMA. A fim de encontrar os modelos de forma parcimoniosa, as ordens dos modelos AR ( $p_m$ ) e MA ( $q_m$ ) são determinadas segundo o critério BIC (equação (2.28)).

Observa-se que a ordem dos modelos mensais AR variou de 1 a 3 (Tabela 6.9), com as maiores ordens atribuídas aos meses secos. Além disso, para abril e novembro, a ordem do modelo MA foi 1. Ainda, observa-se que, no mês de julho, a ordem do modelo AR,  $p_m = 2$ , embora distoante dos demais meses secos, foi escolhida de acordo com os valores do critério BIC e da variância do erro, mostrados na Tabela 6.10.

A arquitetura da rede MLP foi obtida através da técnica de validação cruzada. Esta técnica determinou o número de entradas ( $M$ ) e o número de neurônios intermediários ( $M_1$ ) para cada mês, como mostra a Tabela 6.11. Observa-se ainda que os valores de  $M_1$  oscilaram entre 3 e 33 e ambos  $\alpha$  e  $\beta$  assumem os valores 0,25; 0,50 ou 0,75.

Novamente, a RNN utilizou as mesmas entradas que a MLP e o número de grupos de neurônios ou regras que compõem as camadas intermediárias foram determinados pelo treinamento da rede. A Tabela 6.12 mostra esses números, que tiveram uma pequena variação mês a mês (30 a 39). Isso é um indício de que a série apresenta poucas

Tabela 6.9: Estimativas dos parâmetros utilizando o EMV: série de vazões de Emborcação.

Meses	$\phi_{1,m}$	$\phi_{2,m}$	$\phi_{3,m}$	$\theta_{1,m}$	$\tau_m$	$(p_m, q_m)$
Janeiro	0,4365	-	-	-	1,3700	(1,0)
Fevereiro	0,6482	-	-	-	2,1286	(1,0)
Março	0,4892	-	-	-	1,6175	(1,0)
Abril	1,0344	-	-	-0,4842	2,4886	(1,1)
Mai	0,6819	0,1248	0,2637	-	8,7813	(3,0)
Junho	0,7947	-0,0422	0,1888	-	5,3395	(3,0)
Julho	0,6991	0,2853	-	-	21,6255	(2,0)
Agosto	0,8279	-0,1681	0,2992	-	16,0279	(3,0)
Setembro	0,8248	-	-	-	5,0048	(1,0)
Outubro	0,5111	-	-	-	1,5248	(1,0)
Novembro	0,6553	-	-	0,0032	1,6751	(1,1)
Dezembro	0,5264	-	-	-	1,5864	(1,0)

variações nas vazões ao longo dos meses, ou seja, é bem comportada.

A seguir, os parâmetros dos modelos propostos MPAN foram definidos experimentalmente. A Tabela 6.13 detalha os valores desses parâmetros. Na fase de agrupamento de dados, o número de centros dos grupos  $c$  é definido e posteriormente, na fase de classificação com o procedimento PRP, é necessário definir o parâmetro  $l$  correspondente ao número de valores anteriores de vazões e ainda se suas respectivas diferenças ou inclinações são relevantes na construção dos  $p$ -padrões de dados/previsão. Para o posto de Emborcação, as inclinações aumentam o desempenho do modelo nos meses de abril,

Tabela 6.10: Valores do critério BIC e da variância para o mês de julho para o posto de Emborcação.

$p_m$	1		2		3		4		5	
	$\sigma^2$	BIC								
0	0,07	-180,54	0,05	-196,36	0,05	-192,17	0,05	-189,25	0,04	-191,54
1	0,05	-196,36	0,06	-186,54	0,06	-180,96	0,05	-181,46	0,06	-169,38

Tabela 6.11: Topologias das redes MLP para cada mês: posto de Emborcação.

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
$M/M_1$	3/32	5/3	3/6	1/27	3/30	3/29	1/33	1/31	2/3	3/33	1/3	3/3
$\alpha$	0,75	0,25	0,75	0,5	0,5	0,5	0,75	0,75	0,25	0,75	0,75	0,25
$\beta$	0,25	0,25	0,5	0,5	0,75	0,25	0,75	0,75	0,25	0,75	0,75	0,25

Tabela 6.12: Topologias das RNN para cada mês: posto de Emborcação.

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
$M$	3	5	3	1	3	3	1	1	2	3	1	3
Regras	34	31	33	30	37	38	34	35	35	38	39	38

maio, agosto, outubro, novembro e dezembro.

As previsões resultantes dos modelos MPAN-PMed e MPAN-PRP são mostradas, respectivamente, nas Figuras 6.6 e 6.7, respectivamente.

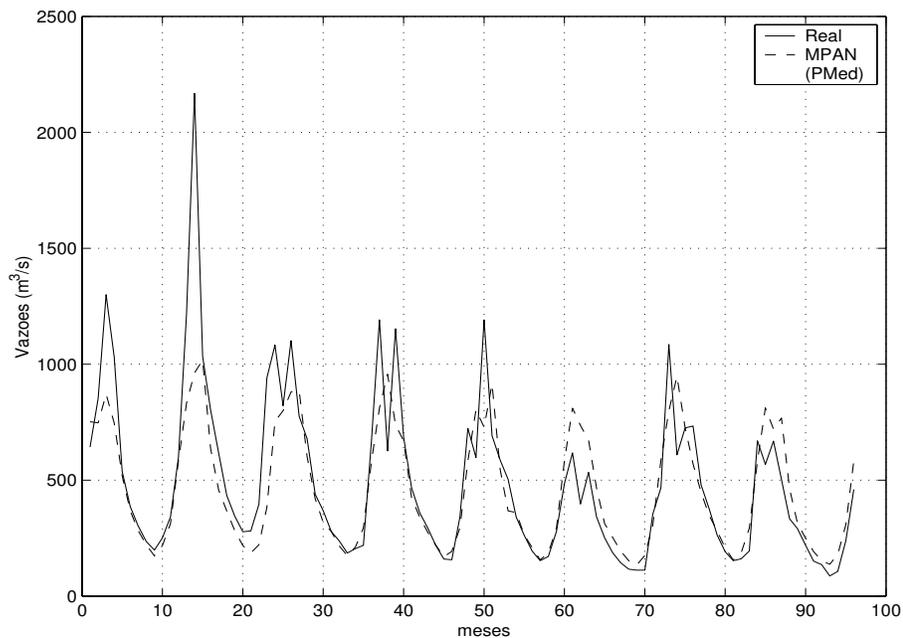


Figura 6.6: Previsão de vazões utilizando o modelo baseado em agrupamento com PMed para o período entre 1991 e 1998 do posto de Emborcação.

Tabela 6.13: Características para o modelo MPAN com PRP e PMed: Emborcação.

Meses	PRP			PMed
	$c$	$l$ (para os $p$ -padrões)	Inclusão da Inclinação	$c$
Jan	4	2	não	10
Fev	3	2	não	3
Mar	3	3	não	3
Abr	3	3	sim	3
Mai	6	2	sim	4
Jun	4	2	não	2
Jul	4	3	não	2
Ago	8	3	sim	2
Set	4	1	não	6
Out	2	2	sim	2
Nov	3	3	sim	2
Dez	2	3	sim	4

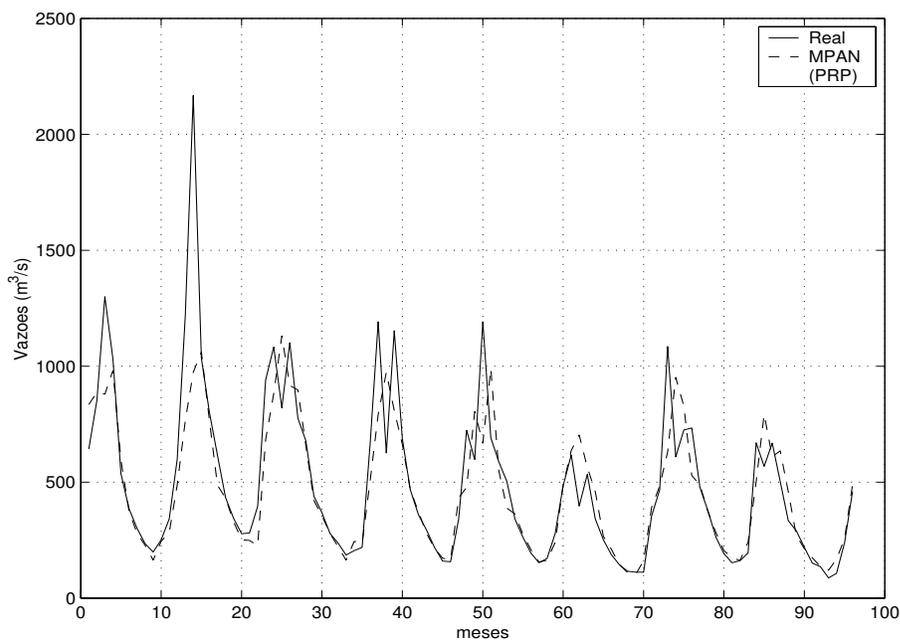


Figura 6.7: Previsão de vazões utilizando o modelo baseado em agrupamento com PRP para o período entre 1991 e 1998 do posto de Emborcação.

A combinação linear de previsões (CP) descrita no Capítulo 5 foi aplicada à série de vazões de Emborcação agregando as características dos modelos PARMA, MPAN-PMed e MPAN-PRP. A escolha desses modelos para a combinação de previsores foi feita experimentalmente, uma vez que apenas o modelo MPAN-PRP forneceu erros inferiores para a maioria dos critérios (Tabela 6.15). A taxa de aprendizagem empregada foi  $\alpha = 0,01$ . Os pesos resultantes da combinação são mostrados na Tabela 6.14.

Tabela 6.14: Pesos resultantes da combinação neural para o posto de Emborcação.

<b>Meses</b>	$w_1$ (PARMA)	$w_2$ (PMed)	$w_3$ (PRP)
Jan	-0,1501	1,1002	0,1675
Fev	-0,2451	0,6691	0,7773
Mar	-0,5444	0,8753	0,9658
Abr	-0,3599	0,7750	0,8105
Mai	0,6128	0,2031	0,2665
Jun	0,2450	0,1698	0,7645
Jul	0,0570	0,0712	1,0156
Ago	0,4367	0,1325	0,5306
Set	-0,5362	1,0205	0,6707
Out	0,1060	0,6486	0,7522
Nov	-0,0619	0,4984	0,9270
Dez	-0,1971	1,3191	0,02017

Observa-se que, neste caso, o modelo MPAN-PRP foi ponderado por valores mais altos que os atribuídos aos demais modelos, para a maioria dos meses. As exceções aparecem nos meses de janeiro, setembro e dezembro, em que atribui-se ao MPAN-PMed valores superiores, e no mês de maio, em que o PARMA tem peso maior. A Figura 6.8 mostra a curva prevista.

Na próxima seção, os erros serão analisados e comparados para verificar a acurácia dos modelos propostos.

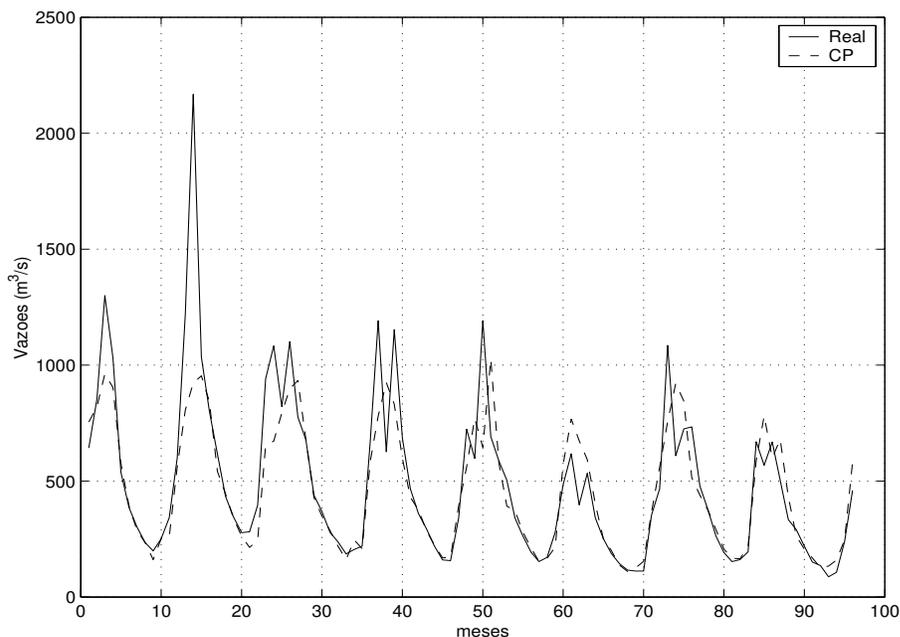


Figura 6.8: Previsão de vazões utilizando o modelo de combinação de previsores para o período entre 1991 e 1998 do posto de Emborcação.

#### 6.4.1 Análise de Erros

Inicialmente, os erros mensais foram computados. A Figura 6.9 detalha esses erros para os modelos descritos neste trabalho. Observa-se que a rede MLP obteve erros inferiores para a maioria dos meses, com exceção dos meses de maio e junho, em que o PARMA forneceu melhores resultados, e nos meses de janeiro e abril, o desempenho da RNN foi superior. Vale ressaltar que a rede MLP só obteve resultados satisfatórios para as vazões de Emborcação, mais um indício do bom comportamento dessa série.

Segundo a Figura 6.9, o modelo MPAN-PMed obteve desempenho superior à rede MLP apenas nos meses de janeiro, maio, junho, novembro e dezembro, considerando o critério EPM. Entretanto, o método PARMA supera esse modelo em maio e junho. Já o predictor MPAN-PRP fornece erros menores que a RNN em abril, maio, junho e agosto, além de uma significativa redução em outubro, novembro e dezembro, sendo bastante similar nos demais meses.

Para este caso, o combinador apresentou resultados bastante promissores, considerando que apenas em março e setembro os erros foram similares aos erros dos

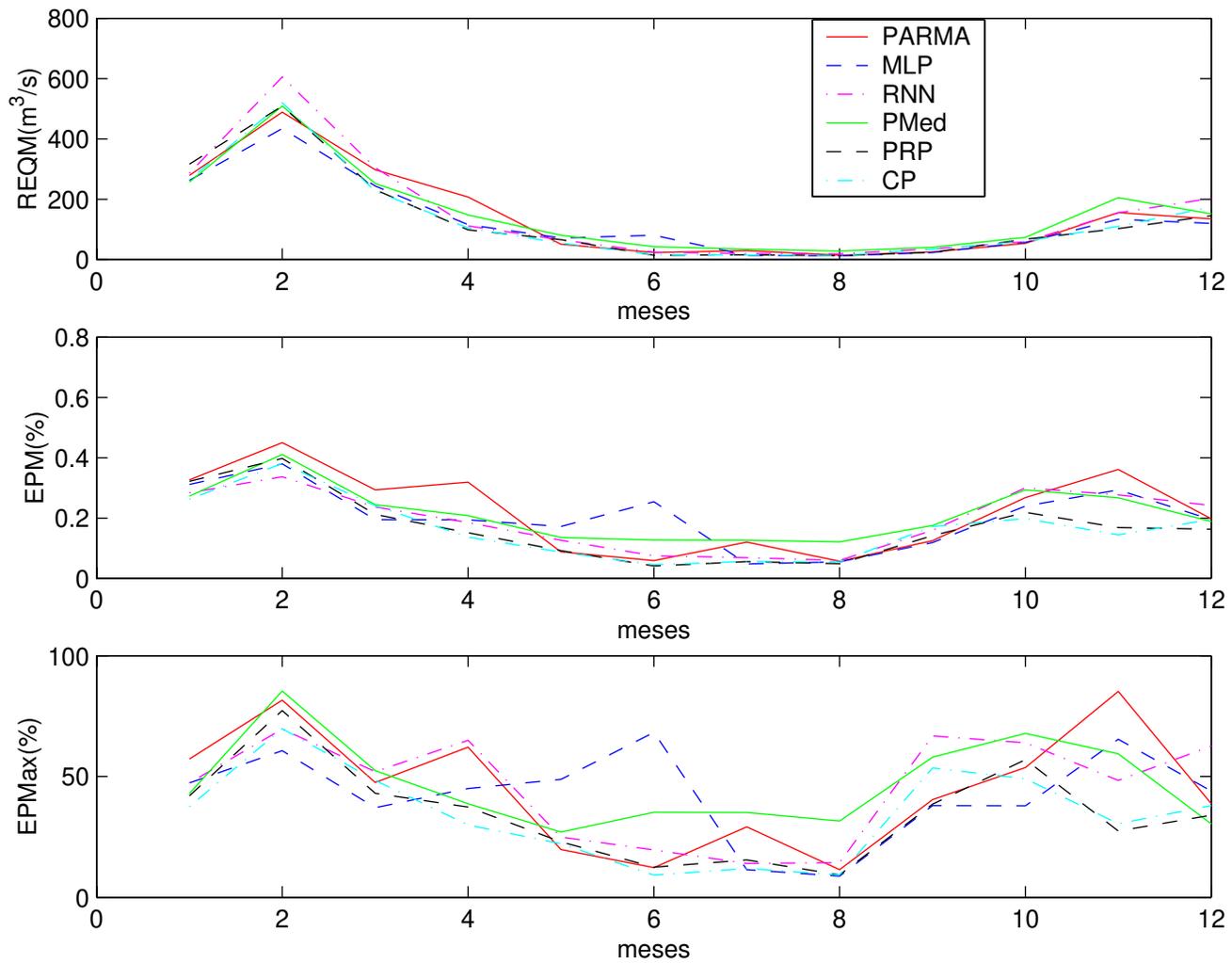


Figura 6.9: Erros mensais referentes ao posto de Emborcação.

modelos individuais.

Na Tabela 6.15, observa-se que no contexto dos modelos individuais, o modelo MPAN-PRP obtém os erros de previsão inferiores a todos os modelos, excetuando-se os critérios REQM e coeficiente de correlação, os quais foram mais altos que os fornecidos pela rede MLP.

Ainda, o modelo de combinação produz previsões que reduzem os erros dos modelos individuais e ainda aumentam o valor do coeficiente de correlação  $\rho$ , embora este ainda apresente um valor maior para as previsões dadas pela MLP. Considerando os modelos individuais (PARMA, MPAN-PMed e MPAN-PRP) utilizados, a variância das previsões do modelo CP também é reduzida, embora tenha sido superior à variância do modelo MLP .

Tabela 6.15: Erros globais de previsão para o posto de Emborcação.

Métodos	<b>REQM</b> ( $m^3/s$ )	<b>EAM</b> ( $m^3/s$ )	<b>EPM</b> (%)	<b>EPmax</b> (%)	$\rho$	$\sigma^2$ ( $\times 10^4$ )
PARMA	203,85	125,32	22,21	85,26	0,81	4,19
MLP	177,43	109,53	20,47	68,29	0,86	3,18
RNN	228,85	115,69	19,59	69,68	0,77	5,09
MPAN-PMed	202,48	117,29	21,44	85,38	0,83	4,04
MPAN-PRP	196,59	104,23	16,77	77,27	0,83	3,82
CP	194,26	102,44	16,46	69,85	0,84	3,72

## 6.5 Resumo

Neste Capítulo, tanto dos modelos individuais MPAN quanto do modelo de combinação de previsores propostos neste trabalho foram aplicados a três séries de vazões dos reservatórios das usinas de Sobradinho, Furnas e Emborcação.

Os resultados foram apresentados e o desempenho dos modelos foi avaliado comparando-se os erros de previsão dos modelos sugeridos e os fornecidos por três modelos de previsão da literatura: o modelo clássico PARMA, as redes MLP e uma rede neural nebulosa (RNN). Os critérios de erro adotados foram a raiz do erro quadrático

médio (REQM), o erro absoluto médio (EAM), o erro percentual médio (EPM), o erro percentual máximo (EPmax) e o coeficiente de correlação ( $\rho$ ).

Para os três postos, nota-se que o desempenho do modelo MPAN-PRP é, em geral, superior aos demais modelos individuais, sugerindo que o modelo consegue capturar a tendência das vazões e acompanhar melhor a curva real de previsões. Por outro lado, o modelo MPAN-PMed, apesar de fornecer resultados satisfatórios para Furnas e Emborcação, não apresentou um bom desempenho para o posto de Sobradinho. Isso pode ser explicado pela grande variância nas vazões deste reservatório, o que não pôde ser capturado pela mediana dos dados, embora este valor estatístico seja pouco sensível a valores extremos.

Ainda, o modelo de combinação linear de previsores forneceu resultados bastante promissores e conseguiu, em geral, explorar as características individuais de cada método de previsão, principalmente para os dados de Emborcação e Sobradinho. Isso sugere que, considerando a simplicidade do modelo proposto, se houver um aprimoramento em sua arquitetura ou o envolvimento de abordagens não lineares, a combinação pode se tornar uma ferramenta ainda mais eficiente.

# Capítulo 7

## Conclusões

O problema de previsão um passo à frente de série de vazões naturais médias mensais abordado, neste trabalho, é de extrema importância para o setor elétrico brasileiro. Isto porque essas previsões de vazões são uma das variáveis de entrada exigidas para o planejamento e operação do sistema hidroelétrico.

Atualmente, essas previsões são realizadas pelo setor, tendo como base, os modelos de séries temporais desenvolvidos por Box & Jenkins. Entretanto, considerando a periodicidade e a não-estacionariedade presentes na série de vazões, esses modelos nem sempre fornecem resultados satisfatórios. Por esse motivo, para obter previsões mais efetivas, modelos de redes neurais e redes neurais nebulosas foram explorados na literatura, atingindo um desempenho superior aos de Box & Jenkins.

Neste trabalho, métodos baseados em agrupamento nebuloso (MPAN) foram propostos para a previsão séries de vazões, um passo à frente. Estes métodos são compostos por 2 estágios. No primeiro, utiliza-se o algoritmo *fuzzy c-means* para agrupar padrões e obter a matriz de centros dos grupos. No segundo estágio o de classificação, os padrões de teste são rotulados de acordo com os centros determinados na primeira etapa. Dois procedimentos são utilizados para aproximar o valor real de um dos atributos de cada padrão de teste: um através de reconhecimento de padrões (PRP) e outro através das medianas (PMed) dos dados históricos. O objetivo dessa fase é encontrar os graus de pertinência associados aos padrões de teste, gerados a partir da matriz de centros previamente obtida. A previsão é determinada por uma combinação ponderada dos graus de pertinência e seus respectivos centros.

Os modelos propostos foram utilizados para a previsão de vazões de três diferentes postos: Sobradinho, Furnas e Emborcação, situados em regiões distintas do Brasil. Os resultados foram comparados aos fornecidos por modelos PARMA (baseados na metodologia Box & Jenkins), por redes neurais MLP e por uma classe de redes neurais nebulosas (RNN). De forma geral, o MPAN-PRP tem desempenho superior aos modelos RNN e PARMA e apresenta uma estrutura mais simples que a RNN, com menor complexidade e menor tempo de execução.

Além disso, propôs-se um método de combinação de previsores visando aprimorar ainda mais a previsão de séries de vazões. Esse método é baseado na idéia de capturar as características de previsão mais importantes de conjunto de previsores individuais. A agregação é dividida em duas etapas. Na primeira, seleciona-se os modelos individuais mais apropriados para domínio de previsão. A seguir os resultados são combinados por um agregador linear na forma de uma rede neural treinada pelo método do gradiente descendente.

Os modelos individuais considerados foram os previamente testados: PARMA, MLP, RNN, MPAN-PMed e MPAN-PRP. Para cada um dos postos, foram escolhidos os que obtiveram, individualmente, os melhores desempenhos medidos pelos erros globais de previsão fornecidos por cada um desses modelos. Embora apresente uma estrutura muito simples e exigências computacionais modestas, o modelo de combinação produziu as melhores previsões, quando comparado aos modelos individuais.

Vale ressaltar que o esforço computacional e o tempo de execução dos modelos propostos MPAN-PMed, MPAN-PRP e CP é bastante razoável em comparação aos modelos PARMA, MLP e RNN. Isso porque os procedimentos utilizados nos modelos MPAN, *fuzzy c-means*, mediana de dados e reconhecimento de padrões não exigem um grande esforço computacional e são executados rapidamente, assim como os procedimentos que compõem o PARMA. Contrariamente, uma MLP com um número grande de neurônios na camada intermediária demanda um tempo execução e um esforço computacional relativamente altos devido às formulações matemáticas envolvidas no processo de treinamento. A RNN, apesar de fornecer resultados num tempo menor que a MLP, exige um esforço computacional grande considerando que o número de regras nebulosas aumenta exponencialmente.

## 7.1 Trabalhos Futuros

O bom desempenho apresentado pelos modelos MPAN-PRP sugere a continuidade do estudo desses modelos, adotando ordens superiores na definição dos  $(p + 1)$  e/ou  $p$ -padrões de dados e de previsão, o que em outras palavras significa considerar, como atributos, mais valores anteriores da série. Paralelamente, deve se levar em conta que o aumento na dimensão do padrões acarreta um aumento na dimensão do problema de agrupamento nebuloso, o que afeta diretamente no desempenho do algoritmo FCM. Dessa forma, outros modelos de agrupamento, como as redes de Kohonen, podem ser testados e avaliados.

Ainda, aspectos teóricos dos modelos propostos MPAN devem ser estudados com objetivo de verificar a formulação do método como a solução de um problema de otimização.

No contexto de combinação de previsores, trabalhos futuros deverão considerar a análise de operadores de combinação não lineares, e a influência da eficiência dos previsores individuais na previsão por combinação.

Além disso, uma alternativa para melhorar a escolha dos previsores mais promissores seria considerar não só os erros globais de cada predictor individual, mas sim, os erros mensais. Assim, para cada mês, três modelos de previsão seriam escolhidos, não sendo necessariamente iguais para todos os meses.

No setor elétrico brasileiro, a previsão  $k$  passos à frente de vazões naturais médias mensais também se mostra interessante para o planejamento e operação a longo prazo. Dessa forma, novos testes deverão considerar, por exemplo, previsões com 6 ou 12 passos à frente.

Para melhorar a análise estatística dos resultados, uma alternativa seria utilizar o método de validação cruzada, isto é, dividir o conjunto de dados em  $n$  subconjuntos e através de uma combinação em pares treinamento(ajuste)/teste desses subconjuntos, computar os resultados dos modelos para cada par e obter a média desses resultados, o que fornecerá maior consistência nas previsões.

Ainda, para evitar problemas na análise estatística dos resultados, novos testes poderiam ser realizados utilizando dados de vazões geradas sinteticamente a partir do comportamento das séries históricas, uma vez que as séries utilizadas neste traba-

lho possuem um conjunto limitado de dados históricos. Além disso, o uso das séries sintéticas fornece mais informações para o ajuste ou treinamento e mais dados de teste para os modelos de previsão.

# Apêndice A

## Tabelas de Erros Mensais de Previsão

Este apêndice mostra as tabelas de erros mensais para todos os modelos descritos nesta dissertação, para os postos de Sobradinho, Furnas e Emborcação.

Tabela A.1: Erros mensais de previsão dos modelos PARMA, MLP e RNN para o posto de Sobradinho.

Meses	PARMA			MLP			RNN		
	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)
Jan	1226,72	24,24	54,73	1165,35	23,85	49,84	1489,08	25,13	58,33
Fev	2320,68	45,69	144,27	2184,22	35,60	105,32	2568,01	34,50	79,31
Mar	1542,68	8,18	32,60	3683,45	48,79	126,50	3088,56	17,18	65,35
Abr	1809,16	46,16	111,44	1607,78	43,51	66,30	1365,01	30,11	42,24
Mai	538,01	15,36	45,33	758,99	36,77	70,50	259,25	10,30	22,30
Jun	154,20	7,55	15,98	350,04	22,99	68,10	260,70	11,13	32,96
Jul	179,23	14,79	37,70	296,95	23,23	72,16	143,41	12,49	24,32
Ago	51,82	4,12	9,22	250,87	24,16	61,80	78,23	7,37	16,41
Set	59,81	6,11	14,32	183,17	17,52	43,20	100,73	10,10	20,98
Out	203,48	15,88	26,41	298,60	23,81	42,23	205,03	16,32	24,44
Nov	691,81	31,28	70,80	1048,89	43,72	131,20	532,62	20,54	40,87
Dez	750,45	21,68	42,60	1140,67	31,73	149,70	693,29	18,41	40,47

Tabela A.2: Erros mensais de previsão dos modelos MPAN-PMed, MPAN-PRP e CP para o posto de Sobradinho.

Meses	MPAN-PMed			MPAN-PRP			CP		
	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)
Jan	1225,77	27,59	61,48	1119,12	24,29	38,56	1353,13	21,84	49,58
Fev	2168,49	38,45	113,22	2493,78	35,81	140,75	2541,57	32,08	100,93
Mar	2656,77	23,65	82,43	1308,48	22,42	88,27	2052,74	14,03	42,98
Abr	1471,17	38,86	69,67	1354,63	34,67	79,60	1092,26	26,50	34,96
Mai	411,37	18,83	45,84	337,70	12,04	26,70	283,72	10,47	23,70
Jun	186,14	10,86	39,00	93,71	5,80	10,79	128,19	6,93	13,57
Jul	190,78	15,50	48,61	106,39	9,62	19,48	118,88	10,27	23,40
Ago	131,97	11,77	35,99	62,08	5,08	18,87	65,23	6,34	15,49
Set	120,20	12,38	33,13	38,35	4,09	8,82	59,98	6,28	13,09
Out	241,56	20,26	35,51	221,89	18,65	46,12	185,92	15,22	24,52
Nov	885,98	31,91	67,16	742,00	36,30	80,95	527,90	23,42	48,59
Dez	751,90	22,86	83,34	653,47	18,44	36,60	686,72	18,25	43,11

Tabela A.3: Coeficientes de correlação mensais ( $\rho$ ) para o posto de Sobradinho.

Meses	PARMA	MLP	RNN	MPAN-PMed	MPAN-PRP	CP
Jan	0,34	0,41	0,39	0,35	0,62	0,44
Fev	0,03	0,21	-0,36	0,10	0,20	-0,34
Mar	0,97	0,11	0,44	0,91	0,33	0,95
Abr	0,28	-0,20	0,80	0,24	0,63	0,67
Mai	0,74	0,08	0,96	0,75	0,70	0,89
Jun	0,91	-0,09	0,69	0,91	0,59	0,91
Jul	0,87	-0,03	0,95	0,96	0,91	0,97
Ago	0,98	-0,06	0,96	0,96	0,85	0,99
Set	0,93	0,04	0,87	0,92	0,79	0,93
Out	0,79	0,16	0,80	0,81	0,61	0,86
Nov	0,78	0,04	0,93	0,92	0,55	0,94
Dez	0,76	-0,46	0,69	0,46	0,77	0,72

Tabela A.4: Erros mensais de previsão dos modelos PARMA, MLP e RNN para o posto de Furnas.

Meses	PARMA			MLP			RNN		
	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)
Jan	946,95	41,00	68,90	1179,39	49,68	138,04	868,67	34,99	60,20
Fev	749,85	37,56	110,04	2162,22	36,10	107,73	846,24	28,26	66,72
Mar	417,91	23,69	56,92	3682,86	26,11	72,17	277,67	18,51	29,57
Abr	299,99	18,70	39,22	1513,69	17,33	35,85	355,38	16,55	49,59
Mai	145,91	12,47	37,66	756,35	12,91	24,57	109,59	10,91	19,24
Jun	175,69	22,91	48,37	372,46	15,92	30,09	104,55	11,88	25,20
Jul	43,80	8,05	21,38	283,25	70,85	108,34	61,82	9,51	24,37
Ago	51,29	12,11	31,22	366,52	15,44	43,50	61,31	13,59	28,67
Set	194,34	30,97	55,12	183,99	28,50	42,20	146,11	24,54	65,23
Out	134,76	20,33	54,52	296,52	20,27	42,06	174,94	20,62	48,27
Nov	376,94	52,99	120,44	1080,34	41,91	93,24	372,78	21,95	61,63
Dez	348,12	27,27	57,15	1149,73	20,38	60,03	422,90	27,23	77,96

Tabela A.5: Erros mensais de previsão dos modelos MPAN-PMed, MPAN-PRP e CP para o posto de Furnas.

Meses	MPAN-PMed			MPAN-PRP			CP		
	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)
Jan	907,05	32,45	66,19	805,40	33,60	52,67	863,63	33,41	54,33
Fev	590,77	30,40	82,60	616,75	28,73	88,88	731,69	27,35	60,92
Mar	266,58	16,30	44,56	298,71	16,82	47,26	266,39	18,12	30,55
Abr	312,74	15,70	43,71	307,67	12,79	43,81	325,29	14,07	46,46
Mai	119,00	9,92	26,41	124,77	9,40	29,71	122,81	11,53	21,05
Jun	105,45	14,50	23,83	101,03	12,10	23,86	100,36	9,85	27,51
Jul	45,63	7,84	26,10	31,10	5,50	13,97	37,61	6,24	19,45
Ago	42,64	10,18	27,61	36,31	8,60	20,49	47,85	10,80	26,71
Set	164,24	26,16	46,92	140,11	16,39	42,58	148,07	20,63	43,79
Out	142,92	21,10	42,61	131,51	15,42	33,66	131,80	17,06	34,15
Nov	368,00	33,86	57,46	318,93	41,91	85,94	407,47	26,40	61,59
Dez	222,70	18,85	54,93	231,83	17,87	37,46	287,52	19,15	55,22

Tabela A.6: Coeficientes de correlação mensais ( $\rho$ ) para o posto de Furnas.

Meses	PARMA	MLP	RNN	MPAN-PMed	MPAN-PRP	CP
Jan	0,31	0,27	0,48	0,86	0,62	0,67
Fev	0,00	0,00	0,17	0,26	0,20	0,21
Mar	0,41	0,39	0,20	0,24	0,33	0,26
Abr	0,54	0,65	0,49	0,63	0,63	0,66
Mai	0,63	0,66	0,79	0,68	0,70	0,77
Jun	0,37	-0,50	0,64	0,54	0,59	0,64
Jul	0,91	0,69	0,78	0,82	0,91	0,87
Ago	0,70	0,76	0,73	0,76	0,85	0,77
Set	0,20	0,49	0,47	0,47	0,79	0,66
Out	0,42	0,62	0,25	0,48	0,61	0,50
Nov	0,35	0,41	0,52	0,32	0,55	0,62
Dez	0,78	0,70	0,75	0,78	0,77	0,78

Tabela A.7: Erros mensais de previsão dos modelos PARMA, MLP e RNN para o posto de Emborcação.

Meses	PARMA			MLP			RNN		
	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)
Jan	279,27	32,69	57,23	262,34	31,12	47,33	286,27	28,35	47,33
Fev	488,75	44,99	81,57	433,90	38,00	60,64	605,17	33,70	69,68
Mar	297,92	29,37	47,49	243,97	19,48	37,04	304,22	23,69	52,00
Abr	207,36	31,87	62,11	115,28	19,41	45,00	111,22	18,52	64,98
Mai	50,61	8,78	19,78	71,11	62,90	48,82	65,25	12,68	24,90
Jun	23,17	5,89	12,27	80,14	25,41	68,29	24,46	7,50	19,64
Jul	29,24	12,02	29,19	14,01	4,82	11,39	19,56	6,85	13,99
Ago	13,81	5,70	11,42	11,98	5,51	87,61	18,64	5,94	14,31
Set	25,45	12,54	40,42	23,09	11,90	37,98	34,67	15,98	66,80
Out	53,47	26,81	53,73	57,14	24,00	37,89	58,81	30,07	63,94
Nov	155,78	36,13	85,26	133,78	29,33	65,43	154,39	27,67	48,39
Dez	134,71	19,73	38,74	119,82	19,48	43,97	202,87	24,18	62,57

Tabela A.8: Erros mensais de previsão dos modelos MPAN-PMed, MPAN-PRP e CP para o posto de Emborcação.

Meses	MPAN-PMed			MPAN-PRP			CP		
	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)	REQM ( $m^3/s$ )	EPM (%)	EPmax (%)
Jan	256,64	27,25	42,91	316,10	32,17	41,98	264,57	26,25	37,51
Fev	508,62	41,06	85,38	508,90	39,73	77,27	519,55	38,09	69,85
Mar	251,85	24,47	52,43	231,43	21,19	43,05	226,35	24,10	48,34
Abr	148,04	20,80	38,69	98,23	15,17	37,32	104,51	13,65	29,90
Mai	80,63	13,59	27,00	65,72	9,18	22,93	54,04	8,71	22,18
Jun	42,31	12,74	35,22	13,86	4,09	12,45	16,09	4,52	9,16
Jul	33,98	12,66	35,08	15,84	5,57	15,42	14,92	5,62	11,93
Ago	28,11	12,09	31,59	13,08	4,89	9,28	12,47	5,49	9,29
Set	40,08	17,58	57,97	24,14	14,15	38,59	33,93	17,19	53,52
Out	73,48	29,35	67,87	67,48	21,92	56,98	60,34	19,85	49,04
Nov	205,47	26,71	59,37	102,55	16,88	27,44	109,97	14,48	30,49
Dez	151,11	18,90	30,36	145,23	16,34	34,02	174,09	19,57	37,94

Tabela A.9: Coeficientes de correlação mensais ( $\rho$ ) para o posto de Emborcação.

Meses	PARMA	MLP	RNN	MPAN-PMed	MPAN-PRP	CP
Jan	0,06	0,08	-0,07	0,43	-0,11	0,39
Fev	0,39	0,65	0,95	0,35	0,32	0,31
Mar	0,31	0,49	0,15	0,38	0,53	0,56
Abr	0,52	0,87	0,87	0,94	0,90	0,89
Mai	0,95	0,86	0,86	0,83	0,85	0,93
Jun	0,96	0,74	0,98	0,93	0,99	0,98
Jul	0,96	0,97	0,96	0,90	0,97	0,97
Ago	0,96	0,97	0,96	0,90	0,97	0,97
Set	0,92	0,94	0,84	0,92	0,92	0,92
Out	0,97	0,88	0,86	0,88	0,69	0,79
Nov	0,77	0,95	0,83	0,96	0,94	0,95
Dez	0,82	0,85	0,82	0,86	0,87	0,84

## Referências Bibliográficas

- Anderson, P. e Vecchia, A. (1993). Asymptotic results for periodic autoregressive moving average processes, *Journal Time Series Analyses* **1**: 1–18.
- Armstrong, J. (1989). Combining forecasts: The end of the beginning or the beginning of the end?, *International Journal of Forecasting* **5**(4): 585–588.
- Ballini, R. (2000). *Análise e previsão de vazões utilizando modelos de séries temporais, redes neurais e redes neurais nebulosas*, Tese de Doutorado, Unicamp, Campinas, São Paulo, Brasil.
- Ballini, R., Figueiredo, M., Soares, S., Andrade, M. e Gomide, F. (2000). A seasonal streamflow forecasting model using neurofuzzy network, *Information, Uncertainty and Fusion*, Vol. 1, Kluwer Academic Publishers, Massachusetts, EUA, pp. 257–276.
- Bates, J. e Granger, C. (1969). The combination of forecasts, *Operational Research Quarterly* **20**: 451–468.
- Bezdek, J. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, EUA.
- Box, G., Jenkins, G. e Reinsel, G. C. (1994). *Time Series Analysis, Forecasting and Control*, 3rd edn, Holden Day, Oakland, California, EUA.
- Cardoso, G. (2003). *Modelo de previsão baseado em agrupamento e base de regras nebulosas*, Tese de Mestrado, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil.
- Clemen, R. (1989). Combining forecasts: A review and annotated bibliography, *International Journal of Forecasting* **5**(4): 559–583.

- Donaldson, R. e Kamstra, M. (1996). Forecast combining with neural networks, *Journal of Forecasting* **15**(1): 49–61.
- Figueiredo, M. (1997). *Redes neurais nebulosas aplicadas em problemas de modelagem e controle autônomo*, Tese de Doutorado, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil.
- Figueiredo, M., Ballini, R., Soares, S., Andrade, M. e Gomide, F. (2004). Learning algorithms for a class of neurofuzzy network and application, *IEEE Transactions on System, Man, and Cybernetics, Part C* **34**(3): 293–301.
- Figueiredo, M. e Gomide, F. (1999). Design of fuzzy systems using neurofuzzy networks, *IEEE Transactions on Neural Networks* **10**(4): 815–827.
- Fiordaliso, A. (1998). A nonlinear forecasts combination method based on Takagi-Sugeno fuzzy systems, *International Journal of Forecasting* **14**: 367–379.
- Francelin, R., Ballini, R. e Andrade, M. (1997). Back-propagation vs. Box and Jenkins model to streamflow forecasting, *Technical Report 38*, Instituto de Ciências Matemáticas de São Carlos - (ICMSC-USP), São Carlos, São Paulo, Brasil.
- Gershenfeld, N. e Weigend, A. (1994). The future of time series: learning and understanding, *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison-Wesley Publishing Company, Reading, Massachusetts, EUA.
- Geva, A. (1999). Non-stationary time series prediction using fuzzy clustering, *Proceedings of the International Conference of the North American Fuzzy Information Processing Society*, pp. 413–417.
- Harrald, P. e Kamstra, M. (1997). Evolving artificial neural networks to combine financial forecasts, *IEEE Transactions on Evolutionary Computation* **1**(1): 40–52.
- Haykin, S. (1994). *Neural Networks - A Comprehensive Foundation*, Macmillan College Publishing Company, New York, EUA.
- Hebb, D. (1949). *The Organization of Behavior*, N.Y.Wiley, New York, EUA.
- Hoffmann, R. (1998). *Estatística para Economistas*, 3a edn, Pioneira Thomson Learning, São Paulo, Brasil.

- Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences of the USA*, Vol. 79, pp. 2554–2558.
- Hu, M. (1964). *Application to the adaline system to weather forecasting*, Tese de Mestrado, Stanford Electronic Laboratories, Stanford, CA, EUA.
- Jacobs, R. (1988). Increased rates of convergence through learning rate adaptation, *Neural Networks* **1**: 295–307.
- Jingrong, D. (2000). Research on the method of nonlinear combining forecasts based on fuzzy-neural systems, *Proceedings of the 3<sup>rd</sup> World Congress on Intelligent Control and Automation*, Hefei, China, pp. 899–903.
- Jingrong, D. (2002). A nonlinear combining forecast method based on fuzzy neural network, *Proceedings of the First International Conference on Machine Learning and Cybernetics*, Beijing, China, pp. 2160–2164.
- Jones, R. e Brelford, W. (1967). Time series with periodic structure, *Biometrika* **54**(3,4): 403–408.
- Khotanzad, A., Elragal, H. e Lu, T. (2000). Combination of artificial neural-network forecasters for prediction of natural gas consumption, *IEEE Transactions on Neural Networks* **11**(2): 464–473.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps, *Biologica Cybernetica* **43**: 59–69.
- Luna, I., Magalhães, M., Soares, S., Ballini, R. e Gomide, F. (2003). Aplicação de sistemas neurofuzzy recorrentes a previsão de vazões, *Anais do Simpósio Brasileiro de Automação Inteligente*, Vol. 1, Bauru, São Paulo, Brasil, pp. 733–738.
- Magalhães, M., Ballini, R., Gonçalves, R. e Gomide, F. (2004). Predictive fuzzy clustering model for natural streamflow forecasting, *Proceedings of the IEEE International Conference on Fuzzy Systems*, Budapest, Hungary, pp. 390–394.
- Magalhães, M., Ballini, R., Molck, P. e Gomide, F. (2004). Combining forecasts for natural streamflow prediction, *Proceedings of the International Conference of the North American Fuzzy Information Processing Society*, Banff, Canadá.

- Magalhães, M., Ballini, R., Soares, S. e Gomide, F. (2004). Modelo preditivo baseado em agrupamento nebuloso – um aplicação para previsão de vazões, *Anais do XV Congresso Brasileiro de Automática*, Gramado, Rio Grande do Sul.
- Maier, H. e Dandy, G. (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.
- Makridakis, S. (1989). Why combining works?, *International Journal of Forecasting* **5**(4): 601–603.
- McCulloch, W. e Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics* **5**: 115–133.
- Menezes, L., Bunn, D. e Taylor, J. (2000). Review of guidelines for the use of combined forecasts, *European Journal of Operational Research* **120**: 190–204.
- Minsky, M. (1969). *Perceptrons*, MIT Press, Cambridge, MA, EUA.
- Oh, K. e Han, I. (2001). An intelligent clustering forecasting system based on change-point detection and artificial neural networks: Application to financial economics, *Proceedings of the 34th Hawaii International Conference on System Science*.
- Pagano, M. (1978). On periodic and multiple autoregressions, *The Annals of Statistics* **6**: 1310–1317.
- Palit, A. e Popovic, D. (2000). Nonlinear combination of forecasts using artificial neural network, fuzzy logic and neuro-fuzzy approaches, *Proceedings of the IEEE International Conference on Fuzzy Systems*, Vol. 2, pp. 566–571.
- Pedrycz, W. e Gomide, F. (1998). *An Introduction to Fuzzy Sets: Analysis and Design*, MIT Press, Cambridge, MA, EUA.
- Pedrycz, W., Jam, P. e Rocha, A. (1995). Distributed fuzzy system modeling, *IEEE Transactions on System, Man and Cybernetics* **25**(5).
- Petridis, V. e Kehagias, A. (1998). *Predictive Modular Neural Networks*, Kluwer Academic Publishers, Boston, EUA.
- Rosenblatt, 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 Williams, R. (1986). Learning representations by backpropagation errors, *Nature* **323**: 533–536.
- Rumelhart, D. e McClelland, J. (1986). *Parallel Distributed Processing*, Vol. 1, MIT Press, Cambridge, MA, EUA.
- Salas, J., Delleur, J. e Yevjevich, W. (1980). *Applied modeling of hydrologic time series*, Water Resources Publications, Littleton, Colorado, EUA.
- Schwarz, G. (1978). Estimating the dimension of a model, *Ann. Statist.* **6**(2): 461–468.
- Shalizi, C., Shalizi, K. e Crutchfield, J. (2004). An algorithm for pattern discovery in time series, *Journal of Machine Learning Research* . (in press).
- Sharkey, A. (1999). *Combining Artificial Neural Nets: Ensemble and Modular Multi-net Systems*, Springer-Verlag, London, UK.
- Stone, M. (1974). Cross-validation choice assessment of statistical predictions, *Journal of the Statistical Society* **B**(36): 111–133.
- Tang, Z., Almeida, C. e Fishwick, P. (1991). Time series forecasting using neural networks vs Box-Jenkins methodology, *Simulation* **57**(5).
- Tang, Z. e Fishwick, P. (1993). Feedforward neural nets as models for time series forecasting, *ORSA Journal on Computing* **5**(4): 374–385.
- Vecchia, A. (1985). Maximum likelihood estimation for periodic autoregressive-moving average models, *Technometrics* **27**(4): 375–384.
- Von Zuben, F. (1996). *Modelos Paramétricos e Não Paramétricos de Redes Neurais Artificiais e Aplicações*, Tese de Doutorado, FEEC - Unicamp, Campinas, São Paulo, Brasil.
- Weigend, A., Ruberman, B. e Rumelhart, D. (1990). Predicting the future: a connectionist approach, *International Journal of Neural Systems* **1**: 193–209.
- Weigend, A., Ruberman, B. e Rumelhart, D. (1992). Predicting sunspots and exchange rates with connectionist networks, in M. Casdagli e S. Eubank (eds), *Nonlinear Modeling and Forecasting*, Addison-Wesley, Redwood City, CA, EUA, pp. 395–432.

- 
- Werbos, P. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences*, Tese de Doutorado, Harvard University, Cambridge, MA, EUA.
- Yager, R. e Filev, D. (1994). *Essentials of Fuzzy Modeling and Control*, Wiley Interscience, New York, EUA.
- Zhang, Q., Patuwo, B. e Hu, M. (1998). Forecasting with artificial neural networks: the state of the art, *International Journal of Forecasting* **14**: 35–62.
- Zimmermann, H. (1987). *Fuzzy Sets, Decision Making and Expert Systems*, Kluwer-Nihoft, Boston, EUA.