



COPPE/UFRJ

MODELAGEM E PREVISÃO DE VAZÕES DE RIOS BASEADAS EM
CLIMATOLOGIA

Daniel Sica de Faria

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientadores: Mariane Rembold Petraglia

Jacques Szczupak

Rio de Janeiro
Dezembro de 2009

MODELAGEM E PREVISÃO DE VAZÕES DE RIOS BASEADAS EM
CLIMATOLOGIA

Daniel Sica de Faria

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA
(COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE
DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof^ª. Mariane Rembold Petraglia, Ph.D.

Prof. Jacques Szczupak, Ph.D.

Prof. José Gabriel Rodriguez Carneiro Gomes, Ph.D.

Prof. Marco Antonio Cetale Santos, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 2009

Faria, Daniel Sica de

Modelagem e Previsão de Vazões de Rios baseadas em Climatologia/Daniel Sica de Faria. – Rio de Janeiro: UFRJ/COPPE, 2009

XIV, 65: il.; 29,7

Orientadores: Mariane Rembold Petraglia e Jacques Szczupak

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia Elétrica, 2009

Referências Bibliográficas: p. 63 – 65

1. Previsão de Vazões. 2. Climatologia. 3. Redes Neurais. 4. Quantização Vetorial. I. Petraglia, Mariane Rembold, *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Agradecimentos

Durante esse trabalho tive o prazer de conviver com os professores orientadores, Mariane e Jacques, com a professora Leontina e com o amigo e pesquisador Luiz Henrique, aos quais sou eternamente grato por tornar possível um sonho. As discussões produtivas sobre técnicas de processamento de sinais e a conceitos climatológicos muito ajudaram para minha formação como pesquisador.

Agradeço também ao professor José Gabriel pelas discussões e aulas, que tiveram uma grande participação nas modelagens do problema.

Aos amigos da Engenho que sempre contribuíram e incentivaram esse trabalho.

Ao, mais que, amigo Fernando Tomaz e seus pais, Fernando e Rosângela, que sempre me apoiaram e muito ajudaram não só para o trabalho, mas para toda a formação pessoal.

Aos meus tios, Ana Maria e Mozart, e meu padrinho Marcus Vinícius, que desde muito cedo, quando tudo isso era apenas um sonho, me apoiaram e sempre estiveram presentes em todos os momentos.

À minha mãe, que sem dúvidas é a minha maior incentivadora e quem mais fez para tornar tudo isso uma realidade, ultrapassando todos os obstáculos em seu caminho e retirando os que estavam no meu. À Giulia, minha irmã tão querida, por todo seu apoio, carinho e incentivo. Ao meu pai por todo apoio.

A quem me faltam palavras para agradecer, minha avó.

E a Deus, por tudo e por essa extensa lista de pessoas a agradecer.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

MODELAGEM E PREVISÃO DE VAZÕES DE RIOS BASEADAS EM
CLIMATOLOGIA

Daniel Sica de Faria

Dezembro/2009

Orientadores: Mariane Rembold Petraglia

Jacques Szczupak

Programa: Engenharia Elétrica

Este trabalho desenvolve a identificação e a utilização de um modelo de previsão de vazões de rios, baseado em climatologia. O modelo em questão busca informações sobre o comportamento das vazões a partir de variáveis ditas explicativas, que se antecipam à vazão sob análise, transcendendo o clássico estudo de seu histórico. O objetivo principal do trabalho é identificar um modelo capaz de prever as vazões futuras com base em conjuntos de variáveis explicativas, fazendo para isto uso de conceitos de Redes Neurais e Quantização Vetorial em sua modelagem matemática. No aspecto climatológico são utilizados fenômenos onde ocorram teleconexões e/ou propagação global, como fundamentos para o mapeamento entre as variáveis climatológicas e a vazão. A previsão é feita utilizando a natural precedência das variáveis explicativas com respeito à vazão.

O trabalho mostra como identificar e utilizar os modelos não lineares capazes de explicar e prever as vazões em questão, sendo apresentados exemplos de rios brasileiros como suporte aos resultados da pesquisa.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of master of Science (M.Sc.)

MODELING AND PREDICTION OF RIVER WATER FLOWS BASED ON
CLIMATOLOGY

Daniel Sica de Faria

December/2009

Advisors: Mariane Rembold Petraglia

Jacques Szczupak

Department: Electric Engineering

This work introduces the identification and application of a river water flow prediction model based on climatology. This model searches for information on water flow behavior in explaining variables which anticipate the water flow under analysis, transcending the classical approach limited by its historical. The main goal of this work is to identify a model based on a set of explaining variables, capable of predicting the future water flow, using concepts of Neural Networks and Vector Quantization on its mathematical modeling. In terms of climatology, are used phenomena containing teleconnections and/or global propagation as fundamentals for mapping the climatological variables into water flow. The prediction is performed using the natural precedence of explaining variables with respect to the water flow.

This work shows how to identify and use non-linear models capable of explaining and predicting the water flow in question, presenting Brazilian rivers examples as support to research results.

Sumário

1	Introdução.....	1
1.1	Motivações.....	1
1.2	Estrutura do texto.....	2
2	O Modelo Físico da Vazão	4
2.1	Introdução	4
2.2	Teleconexões e Propagações Climatológicas	4
2.3	Previsão de Vazões	9
2.4	O Modelo Proposto.....	12
2.5	Conclusão.....	13
3	Revisão dos Conceitos para Modelagem.....	14
3.1	Introdução	14
3.2	Redes Neurais	14
3.2.1	Introdução.....	14
3.2.2	O Neurônio Artificial e Funções de Ativação	17
3.2.2.1	O Neurônio Artificial	17
3.2.2.2	Funções de Ativação	18
3.2.3	Perceptrons	21
3.2.3.1	Perceptrons de uma camada	21
3.2.3.2	Perceptrons de múltiplas camadas.....	24
3.2.4	Uma Rede Neural como um Combinador Linear	26

3.3	Quantizador Vetorial.....	27
3.3.1	Diagrama de Voronoi	28
3.3.2	O Algoritmo Generalizado de Lloyd.....	29
3.4	Conclusões	32
4	Identificação do Modelo do Preditor	33
4.1	Introdução	33
4.2	A identificação do modelo desconhecido	33
4.3	A identificação por Redes Neurais	37
4.4	A identificação por Quantização Vetorial	40
4.5	A identificação pela combinação dos métodos anteriores	44
4.6	Conclusões	47
5	Previsão da Vazão	48
5.1	Introdução	48
5.2	A previsão utilizando a Rede Neural	48
5.3	A Previsão pelo Quantizador Vetorial	50
5.4	Previsão por Combinação	52
5.5	Conclusões	54
6	Conclusões.....	56
6.1	Contribuições	57
6.2	Sugestões para trabalhos futuros.....	58
A	Apêndice.....	59
A.1	Regra da Cadeia	59
A.2	Derivada da função Tangente Hiperbólica	59
A.3	O Produto de Kronecker	59

A.4	Equacionamento de um perceptron de duas camadas.....	60
7	Referências Bibliográficas.....	63

Lista de Figuras

Figura 1.1 – Matriz Energética por combustível (Fonte: ONS)	1
Figura 1.2 – Curva de Energia Natural Afluente e Custo Marginal da Operação.....	2
Figura 2.1 – Pontos teleconectados. Azul: Latirude 24°S/Longitude 67,5°O e Verde: Latitude 30°S/Longitude 55°L.....	5
Figura 2.2 – Série de Temperatura do Ar (Azul) em Latitude 24°S/Longitude 67,5°O e de Taxa de Precipitação em (Verde) Latitude 30°S/longitude 55°L.....	5
Figura 2.3 – Propagação da anomalia da temperatura da superfície do mar representando o fenômeno El Niño entre os anos de 1982 e 1983.....	7
Figura 2.4 – Célula de Walker sob condições normais	8
Figura 2.5 – Deslocamento da célula de Walker em anos de Niño.....	8
Figura 2.6 – Séries de anomalia da taxa de precipitação. (a) Seca no Nordeste e (b) Cheia no Sul Brasileiro por conta do fenômeno do El Niño nos anos de 1982/83.....	9
Figura 2.7 – Sinal de Vazão.....	10
Figura 2.8 – Espectro de Frequência do sinal de Vazão.....	11
Figura 2.9 – Zoom no espectro de Frequência do sinal de Vazão.....	11
Figura 2.10 – Sistema Desconhecido	13
Figura 3.1 – Modelo de um neurônio	15
Figura 3.2 – Modelo esquemático de um neurônio artificial.....	17
Figura 3.3 – Deslocamento ocasionado pelo <i>bias</i>	18
Figura 3.4 – Funções de ativação	19
Figura 3.5 – Perceptron de uma camada com dois neurônios e quatro entradas.....	21

Figura 3.6 – Composição de um perceptron de múltiplas camadas - (a) e (b) perceptron de uma camadas e (c) perceptrons de duas camadas a partir de (a) e (b)..	25
Figura 3.7 – Diagrama de Voronoi.....	29
Figura 3.8 – Diagrama esquemático do GLA.....	31
Figura 4.1 – Processo de identificação do sistema desconhecido	34
Figura 4.2 – Divisão da série de vazão nos conjunto de treinamento, teste e validação	35
Figura 4.3 – Fasores de Tempo	35
Figura 4.4 – Coordenadas encontradas para o caso exemplo das próximas seções	36
Figura 4.5 – Processo de treinamento e validação da Rede Neural.....	38
Figura 4.6 – Identificação do modelo pela Rede Neural – em escala de tempo diária...	39
Figura 4.7 – Identificação do modelo pela Rede Neural – em escala de tempo mensal	39
Figura 4.8 – Erro entre a vazão original e a aproximação pelo VQ no conjunto de validação à cada quebra do dicionário	41
Figura 4.9 – Identificação do modelo pelo Quantizador Vetorial – em escala de tempo diária.....	43
Figura 4.10 – Identificação do modelo pelo Quantizador Vetorial - em escala de tempo mensal.....	43
Figura 4.11 – Modelo Completo de Identificação.....	44
Figura 4.12 – Processo de treinamento e validação do Combinador	45
Figura 4.13 – Identificação do modelo pela Combinação dos métodos anteriores – em escala de tempo diária	45
Figura 4.14 – Identificação do modelo pela Combinação dos métodos anteriores – em escala de tempo mensal	46
Figura 4.15 – Comparação dos Erros de Validação para todos os métodos.....	46
Figura 5.1 – Previsão pela Rede Neural – em escala de tempo diária.....	49

Figura 5.2 – Previsão pela Rede Neural – em escala de tempo mensal	50
Figura 5.3 – Previsão pelo Quantizador Vetorial – em escala de tempo diária.....	51
Figura 5.4 – Previsão pelo Quantizador Vetorial – em escala de tempo mensal	52
Figura 5.5 – Previsão pelo Combinação – em escala de tempo diária	53
Figura 5.6 – Previsão pelo Combinação – em escala de tempo mensal	53
Figura 5.7 – Comparação dos Erros de Teste para todos os métodos	54
Figura A.6.1 – Perceptron de duas camadas.....	60

Lista de Tabelas

Tabela 4.1 – Descrição das variáveis explicativas do caso exemplo das próximas seções	36
Tabela 4.2 – Comparação dos Erros de Validação para todos os métodos	47
Tabela 5.1 – Comparação dos Erros de Validação para todos os métodos	54

Lista de Abreviaturas

OLR – *Outgoing Longwave Radiation*

CMO – Custo Marginal de Operação

ENA – Energia Natural Afluente

GLA – *Generalized Lloyd Algorithm*

LMS – *Least Mean Square*

MSE – *Mean Square Error*

PCH – Pequenas Centrais Hidrelétricas

MLT – Média de Longo Termo

MLP – *Multi Layer Perceptron*

ONS – Operador Nacional do Sistema

1 Introdução

1.1 Motivações

A motivação inicial deste trabalho parte da forte dependência que a matriz energética brasileira apresenta com relação aos recursos hídricos do país (atualmente mais de 80% da oferta de energia se encontra em usinas hidrelétricas) de grande e pequeno porte (PCH). A Figura 1.1 mostra um gráfico da distribuição da matriz energética para cada uma das fontes utilizadas no país, de acordo com o ONS (Operador Nacional do Sistema).

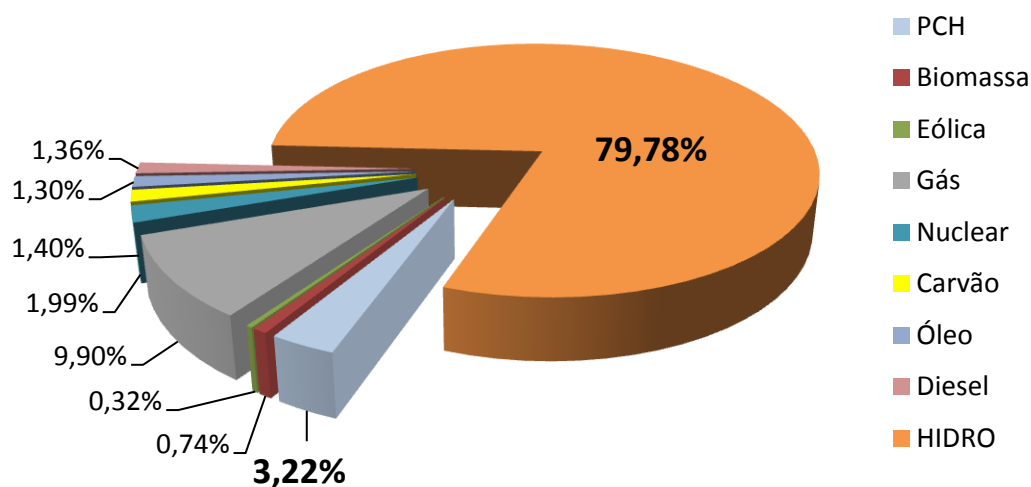


Figura 1.1 – Matriz Energética por combustível (Fonte: ONS)

Aliado a isso está a vulnerabilidade dos preços da energia elétrica. Uma previsão errada pode causar efeitos não somente no setor elétrico, mas também em outros, como por exemplo, na economia como um todo. Exemplos recentes, como o aumento do preço da energia no início de 2008 (Figura 1.2), que causou o fechamento de diversas empresas, e, pior ainda, o racionamento ocorrido em 2001 mostram os impactos que os resultados que uma previsão errada pode causar.

A Figura 1.2 (Fonte: ONS) apresenta um gráfico, no período de 27 de outubro de 2007 e 9 de maio de 2008, onde pode-se ver facilmente a correlação entre a curva de Energia Natural Afluente (ENA) em % MLT, que é o valor das vazões dos rios convertidas em energia e divididas pela média de longo termo, e a curva de Custo Marginal de Operação (CMO), que é o preço da energia para as empresas no mercado de curto prazo.

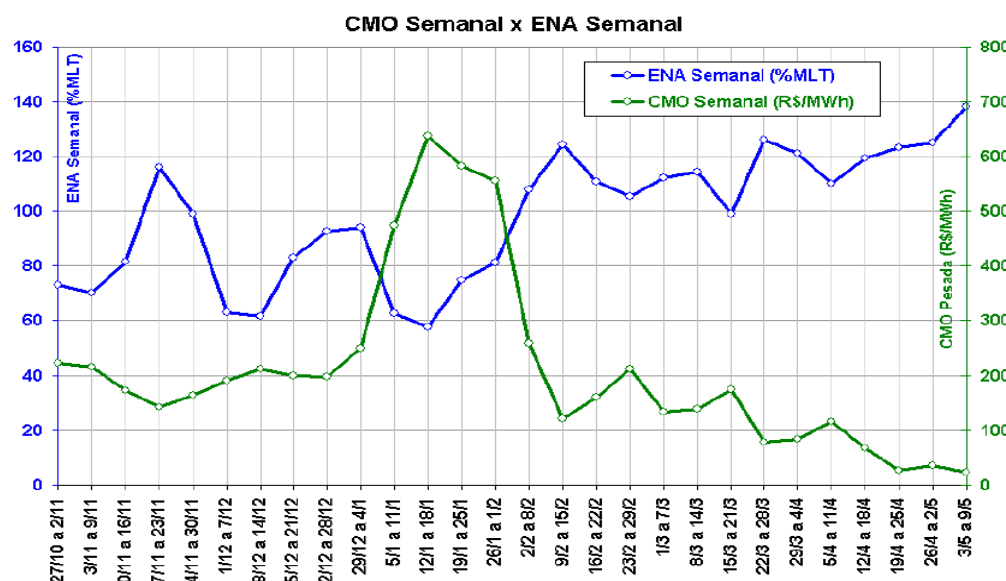


Figura 1.2 – Curva de Energia Natural Afluente e Custo Marginal da Operação

O que é proposto neste trabalho é uma nova metodologia de previsão, não mais baseada somente no histórico da série de vazão e nem na previsão de chuva (como são muitos dos modelos atualmente utilizados) e sim um modelo que produz a vazão através de suas variáveis, ditas explicativas.

1.2 Estrutura do Texto

O texto está organizado em seis capítulos, descritos a seguir.

Este capítulo apresenta as motivações e importância das previsões de vazão dos rios brasileiros para o setor elétrico do país.

O Capítulo 2 faz um panorama das teorias que fundamentam as relações existentes entre fenômenos que acontecem em lugares distantes no globo terrestre, conhecidos como teleconexões, e a propagação destes por todo o planeta; faz também um apanhado das metodologias de previsão de vazões mais utilizadas atualmente no setor elétrico. Por fim, introduz o modelo físico que mapeia as variáveis explicativas na vazão através de modelagens matemáticas, sendo esse modelo a base do trabalho aqui apresentado.

Já o Capítulo 3 faz uma revisão das ferramentas matemáticas, Redes Neurais e Quantização Vetorial, que serão utilizadas em processos de identificação. A Seção 3.2 deste capítulo trata de Redes Neurais, o porquê da sua utilização, a estrutura de uma rede, os parâmetros existentes e as suas aplicações mais comuns. Em seguida traz a descrição de um Quantizador Vetorial, incluindo os métodos utilizados para encontrar a matriz de modulação, dicionário, o algoritmo de otimização e suas aplicações.

Os Capítulos 4 e 5 são os mais específicos e importantes, pois tratam, respectivamente, da identificação do modelo e a previsão da vazão desejada. Os processos de identificação e de previsão são apresentados divididos em duas etapas. Primeiramente é feita a identificação/previsão utilizando uma Rede Neural e, em paralelo, um Quantizador Vetorial. Depois é feita a combinação dos resultados através de um combinador linear, que terá a função combinatória, onde será obtido o resultado final.

O Capítulo 6 apresenta as principais conclusões obtidas com esse trabalho, as principais contribuições e sugestão para trabalhos futuros.

Por fim, um apêndice, com algumas informações complementares ao texto, é incluído.

2 O Modelo Físico da Vazão

2.1 Introdução

Este capítulo aborda os conceitos de climatologia que embasam a sua utilização para a explicação e previsão da vazão, que é foco deste trabalho, apresenta alguns exemplos de modelos de previsão utilizados no setor elétrico Brasileiro e também discute o modelo que será utilizado para alcançar este objetivo. Os conceitos são aqui inicialmente apresentados sendo mais detalhados e exemplificados ao longo dos Capítulos 4 e 5.

2.2 Teleconexões e Propagações Climatológicas

As teleconexões são eventos climatológicos relacionados, que acontecem em diferentes localizações sobre o globo terrestre e que nos proporcionam o entendimento do sinal de um ponto baseado no sinal de outro que esteja teleconectado com o primeiro.

A Figura 2.1 mostra dois pontos distantes no globo terrestre (Azul: Latitude 24°S/Longitude 67,5°O e Verde: Latitude 30°S/Longitude 55°L) que estão teleconectados como podemos perceber observando a Figura 2.2, onde duas séries, de diferentes pontos do globo, de diferentes variáveis climatológicas (Temperatura do Ar e Taxa de Precipitação) apresentam um comportamento semelhante. As séries foram plotadas normalizadas devido à diferença entre as unidades e valores das variáveis, não alterando em nada a sua forma (Fonte de dados: Noaa - Physical Science Division, 2009).

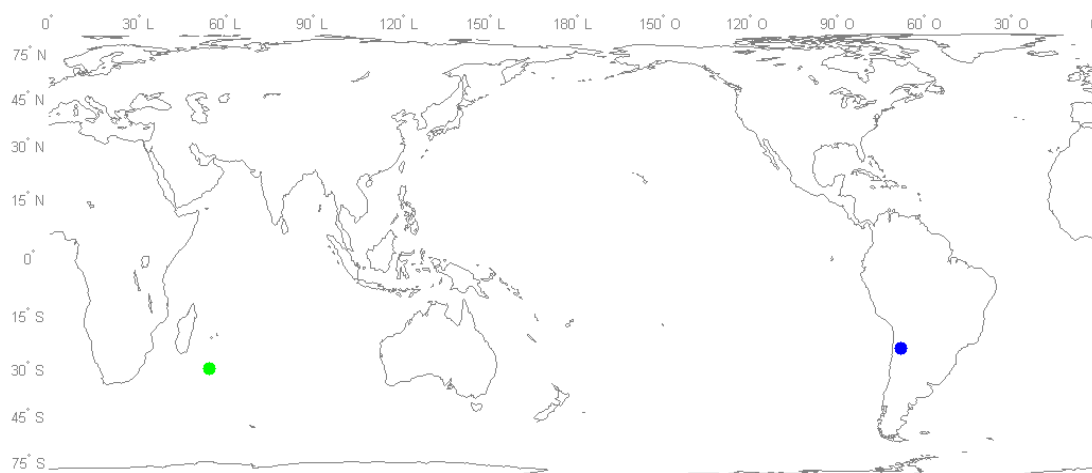


Figura 2.1 – Pontos teleconectados. Azul: Latitude 24°S/Longitude 67,5°O e Verde: Latitude 30°S/Longitude 55°L

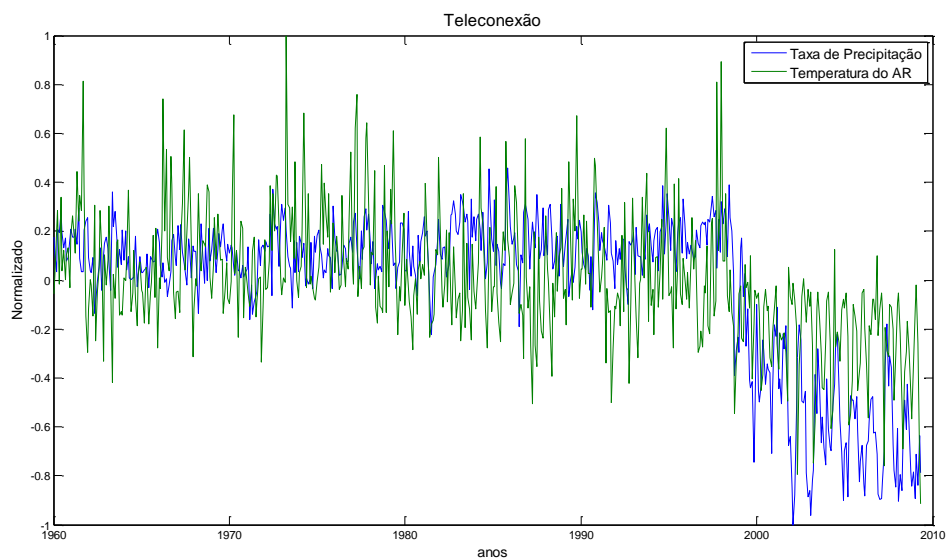


Figura 2.2 – Série de Temperatura do Ar (Azul) em Latitude 24°S/Longitude 67,5°O e de Taxa de Precipitação em (Verde) Latitude 30°S/longitude 55°L

Já as propagações climáticas são eventos relacionados à dinâmica global, ou seja, sendo um fenômeno não estático, ocorrendo em um ponto do globo, ele se propaga ao longo do mesmo, alcançando outras regiões do planeta.

Existem diversos fenômenos que se propagam ao longo do globo: um deles, talvez o mais conhecido de todos, é o fenômeno El Niño/La Niña (Ambrizzi, 2003), que

é caracterizado pela diferença de pressão atmosférica entre o Tahiti e Darwin, na Austrália, assim como pela alteração na temperatura da água do Oceano Pacífico – aquecimento no caso El Niño e resfriamento no caso La Niña – que ocorre de oeste para leste do oceano. A Figura 2.3 ilustra, utilizando a anomalia da temperatura da superfície do mar, a propagação do evento que ocorreu durante os anos de 1982 e 1983 (Fonte de dados: (Noaa - Physical Science Division, 2009)).

Estes eventos trazem consigo interconexões climatológicas que causam perturbações de secas e cheias em vários lugares (Drumond & Ambrizzi, 2003). Por exemplo, na América do Sul, no verão austral em anos de El Niño há geralmente aumento da precipitação na região nordeste da Argentina, parte do Uruguai e do Estado do Rio Grande do Sul. Enquanto isso, no Nordeste Brasileiro e no Mar do Caribe há em geral seca, devido ao deslocamento da célula de Walker (LabMet), que é uma célula de circulação com movimentos ascendentes no Pacífico central/ocidental e movimentos descendentes no oeste da América do Sul e com ventos de leste para oeste próximos à superfície (ventos alísios) e de oeste para leste em altos níveis da troposfera. A Figura 2.4 ilustra esse comportamento em anos sob condições normais, e a Figura 2.5 em anos de Niño, sobre o Norte da América do Sul.

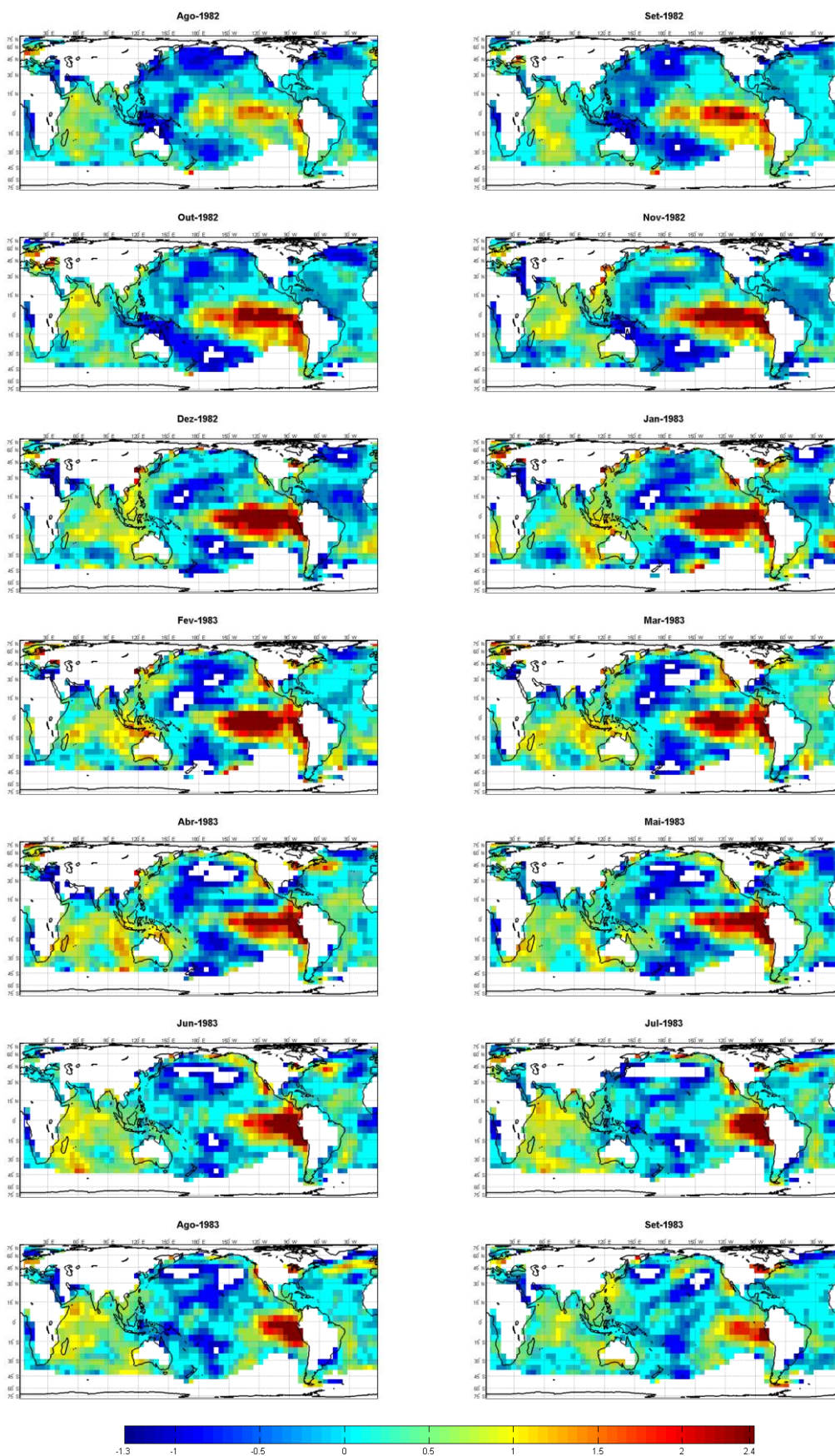


Figura 2.3 – Propagação da anomalia da temperatura da superfície do mar representando o fenômeno El Niño entre os anos de 1982 e 1983

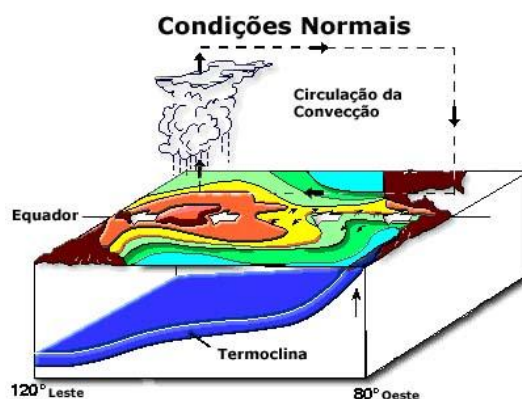


Figura 2.4 – Célula de Walker sob condições normais

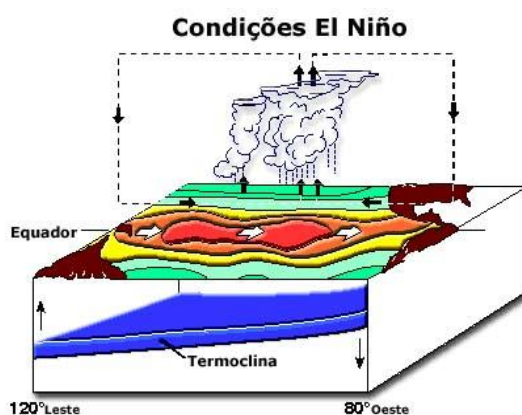


Figura 2.5 – Deslocamento da célula de Walker em anos de Niño

A Figura 2.6 traz dois gráficos que ilustram as perturbações de seca e cheia causadas pelo El Niño, através de sinais de anomalia de taxa de precipitação. No eixo horizontal temos a informação temporal, em escala mensal, entre os anos de 1975 e 1990 e no eixo vertical a anomalia da precipitação (diferença com relação a média para cada mês) em $kg/m^2/s$, que acontecem nos anos de Niño (por exemplo 1982/83). Podemos perceber a diferença nos valores para estes anos com relação aos demais. A Figura 2.6(a) ilustra, através da anomalia da taxa de precipitação, a seca no Nordeste Brasileiro e a Figura 2.6(b) a cheia na região Sul do Brasil neste período.

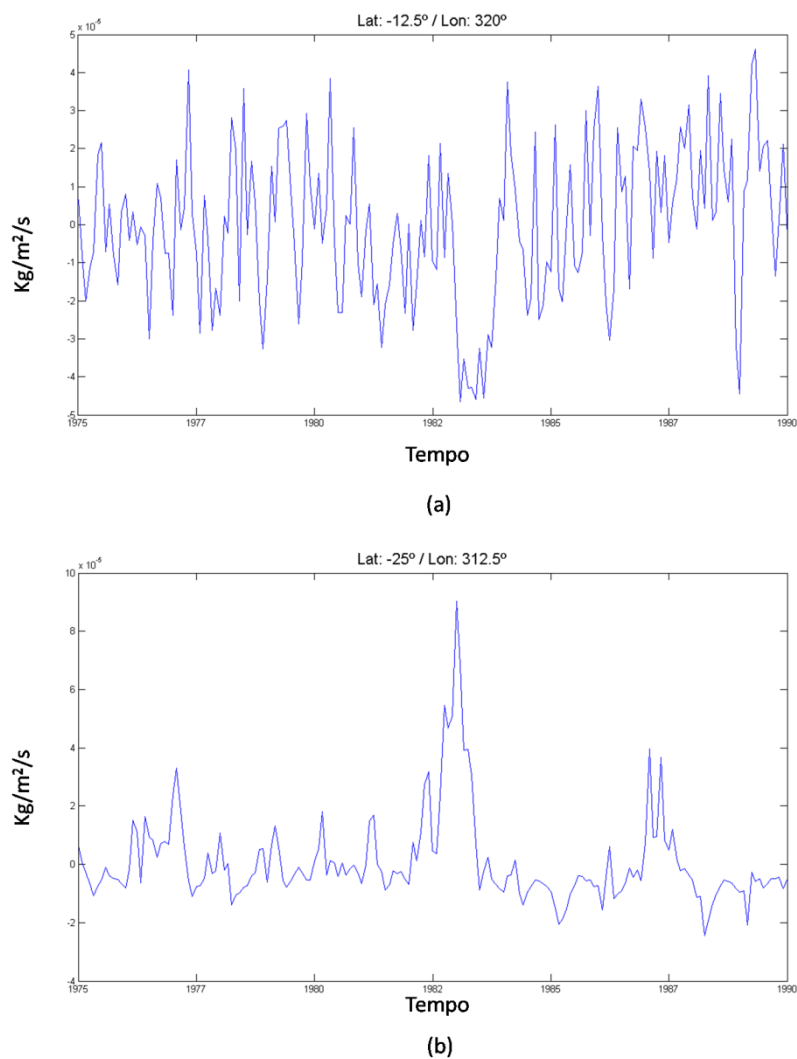


Figura 2.6 – Séries de anomalia da taxa de precipitação. (a) Seca no Nordeste e (b) Cheia no Sul Brasileiro por conta do fenômeno do El Niño nos anos de 1982/83

2.3 O sinal de Vazão

O sinal de vazão apresenta algumas características que devem ser analisadas com bastante cuidado antes de fazer a sua previsão. Apesar de sempre apresentar, ao longo do ano, um período com dados de baixo valor (seca) e um período de valores altos (cheia), pode existir um deslocamento, encurtamento ou alargamento e variações do montante de águas que fogem ao comportamento médio e a sazonalidade do sinal.

A Figura 2.7 ilustra um sinal de vazão, mais especificamente, o sinal de vazão do rio São Francisco, que será utilizado para estudo neste trabalho e está localizado nas regiões Sudeste e Nordeste do Brasil.

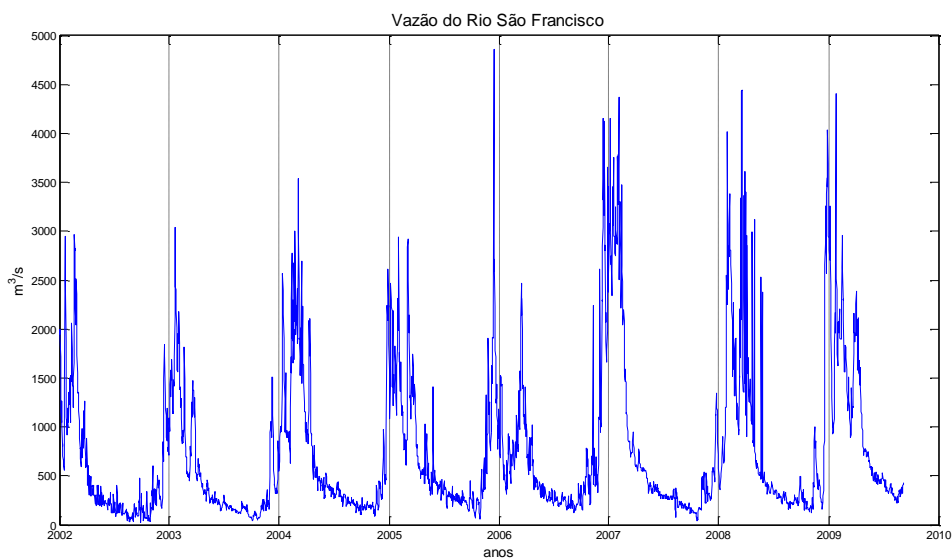


Figura 2.7 – Sinal de Vazão

A análise espectral do sinal, ilustrada na Figura 2.8, apresenta uma oscilação com período de um ano, que pode ser melhor observado na Figura 2.9. Fora a oscilação de um ano, a análise espectral não apresentou nenhum outro ciclo muito significativo, mesmo porque o intervalo de dados disponíveis contém apenas 8 anos.

Os dados utilizados iniciam-se somente em 2002 por questões relacionadas à dificuldade em se conseguir dados confiáveis para os anos anteriores.

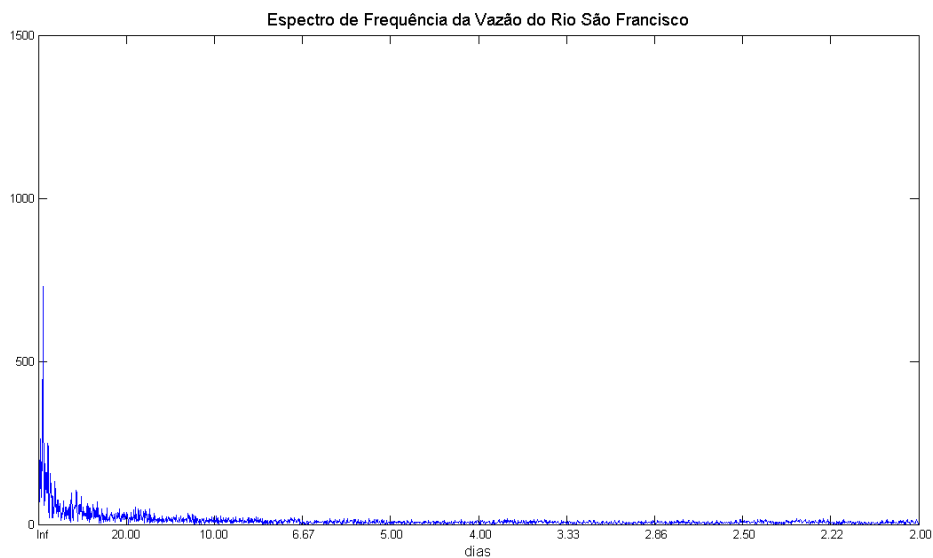


Figura 2.8 – Espectro de Frequência do sinal de Vazão

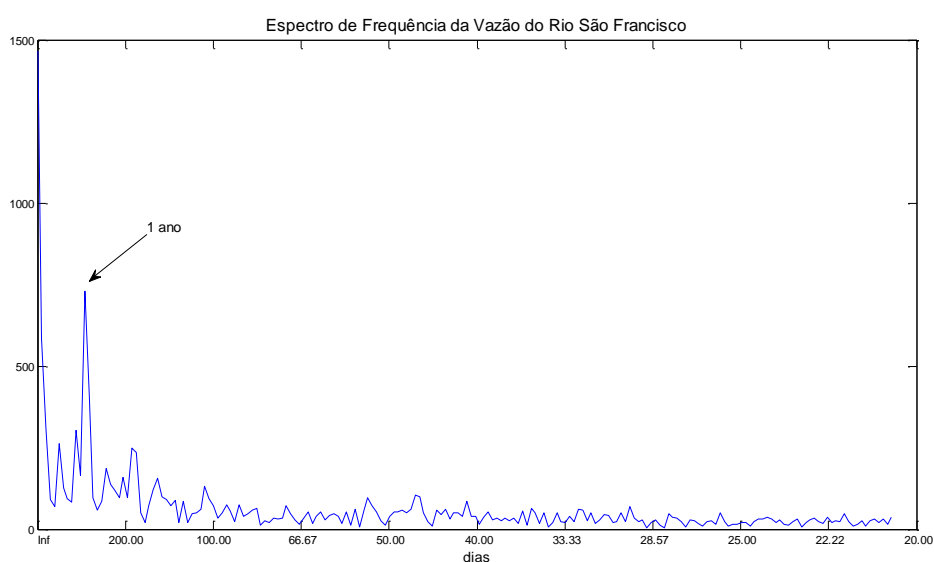


Figura 2.9 – Zoom no espectro de Frequência do sinal de Vazão

2.4 Previsão de Vazões

A previsão de vazões, na maior parte dos casos, se baseia em linhas matemáticas quase sempre calcadas na própria série histórica da vazão e tentando extrair daí informações que possam levar aos valores futuros da mesma. Nesta linha de raciocínio

estão, por exemplo, os preditores diretos e reversos (Costa, Maceira, & Damázio, 2007), os preditores estatísticos (Quintão, Souza, Costa, & Damázio, 2008), etc. O que há de comum entre eles é que a única informação disponível está no passado e sob o qual deverá ser desenhado o futuro. Em termos mais rasteiros, o futuro terá de “repetir” o passado.

Há também técnicas de predição baseadas em modelos físicos, como por exemplo, os que se baseiam na chuva para estabelecer a vazão do rio, dependendo, portanto, da previsão da própria chuva (Guilhon & Rocha, 2008).

A Seção 2.5 apresenta o conceito de previsão empregado neste trabalho. Um modelo matemático que descreve a formação da vazão a partir de um conjunto de variáveis físicas, que denotaremos simplesmente como variáveis explicativas, é proposto.

2.5 O Modelo Proposto

Temos como objetivo principal a identificação e previsão de vazões. Portanto o primeiro passo na estruturação do preditor de vazões é sua caracterização como um sistema até esse ponto desconhecido. Esse sistema é excitado por um conjunto de variáveis explicativas, capazes de produzir à sua saída a vazão desejada. A Figura 2.10 ilustra o processo em questão, apresentando o sistema desconhecido como uma caixa tendo como entradas as variáveis explicativas e como saída a série de vazão.

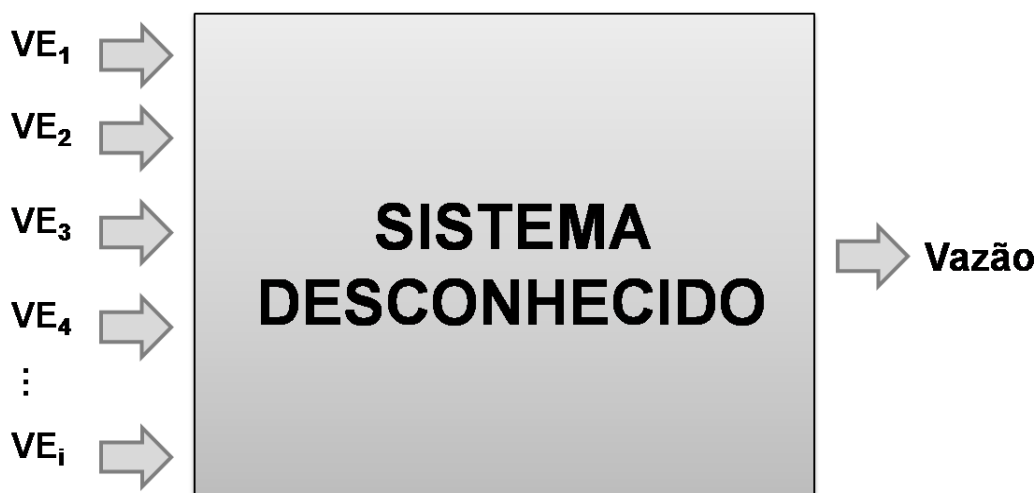


Figura 2.10 – Sistema Desconhecido

A variável explicativa corresponde a uma série de parâmetros físicos, recolhidos ao longo do globo terrestre. Essa técnica, embasada pelas teorias explicadas na Seção 2.2, utiliza variáveis explicativas contendo informações que possam caracterizar – explicar – eventos ocorridos na série de vazões considerada. São utilizados desde dados provindos diretamente de medidas de satélite (Noaa - Physical Science Division, 2009) até os originários de medidas pré-processadas (Macêdo, 2006).

Nos próximos capítulos serão descritas as metodologias matemáticas utilizadas e como é feita a identificação deste modelo, para que então seja possível utilizá-lo para fazer a previsão dos valores de vazão.

2.6 Conclusão

Esse capítulo apresentou algumas explicações climatológicas que servirão de suporte à modelagem aqui adotada como base para previsão de vazões. Apresentou as metodologias atualmente mais adotadas para a previsão de vazões no setor elétrico brasileiro. Introduziu um modelo físico, calcado em variáveis explicativas, que será utilizado neste trabalho no intuito de encontrar os valores futuros da vazão.

3 Revisão dos Conceitos para Modelagem

3.1 Introdução

Este capítulo faz uma breve apresentação das técnicas utilizadas para modelar o sistema desconhecido introduzido no capítulo anterior: Redes Neurais e Quantização Vetorial.

Na seção de redes neurais será apresentado o princípio de funcionamento de uma rede neural MLP (*Multi Layer Perceptron*), as funções de ativação utilizadas, o neurônio artificial e o seu aprendizado através de retro propagação do erro.

Na seção de Quantização Vetorial será apresentada a sua base de funcionamento, a criação do dicionário, o método de minimização de erros e a classificação pelo diagrama de Voronoi.

3.2 Redes Neurais

3.2.1 Introdução

As Redes Neurais Artificiais – que durante esse texto serão simplesmente chamadas de Redes Neurais – foram inspiradas pela sofisticada capacidade de cognição, já conhecida no cérebro humano, e pela alta capacidade de processamento numérico de um computador digital. A idéia seria levar ao computador a possibilidade de replicar a operação do cérebro humano, explorando sua alta capacidade em processar as informações, paralelamente e até mesmo de forma não-linear. Esta idéia esbarra muitas vezes na sofisticação do cérebro humano. Um exemplo disto é o reconhecimento de um

rostro familiar, que pode levar em torno de 100 a 200 ms para ser efetuado pelo cérebro e até dias por um computador convencional.

Outra grande vantagem para o cérebro humano é a sua habilidade em aprender com o passar do tempo, tornando-se mais experiente. Já o computador não usufrui desta dádiva, tendo usualmente um conhecimento limitado e estático.

Neste aspecto, as mais importantes estruturas do cérebro são os neurônios, ilustrados na Figura 3.1, onde se localizam as unidades de processamento de informação. Estes neurônios possuem enorme capacidade de adaptação, conforme ocorram mudanças no meio. Similarmente, as redes neurais fazem uso de neurônios artificiais e de sua capacidade de aprendizado para se ajustarem às necessidades dos modelos em que são empregadas.

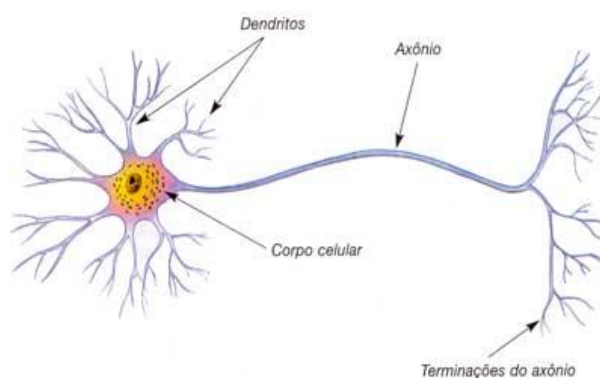


Figura 3.1 – Modelo de um neurônio

As Redes Neurais também fazem uso de ligações entre neurônios, atribuindo a cada uma delas um peso, contraparte dos pesos sinápticos, replicando a forma como o cérebro faz uso de ligações físicas para a transmissão de informação entre os receptores e os atuadores do corpo humano.

A adaptação da rede neural se dá por um processo denominado aprendizagem, que tem como objetivo modificar os pesos das conexões existentes entre os neurônios. Esse processo de aprendizagem se dá através de um mapeamento entre a entrada e a saída onde, para a mesma entrada, é minimizada uma medida da diferença entre a saída

desejada e o valor esperado da saída da rede. A exemplo do que se faz usualmente em eletrônica, essa medida permite que façamos o caminho inverso (saída – entrada), modificando progressivamente os valores dos pesos sinápticos da mesma; em outras palavras, tornando-a a cada iteração mais experiente.

Além do benefício de aprendizagem da rede neural, que já era conhecido e empregado nos filtros digitais adaptativos (Haykin, 1996), elas apresentam também propriedades de não linearidade através do uso de funções de ativação em cada um de seus neurônios. Estas funções serão apresentadas na Seção 3.2.2.2..

A definição de Redes Neurais dada por (Haykin, 1999) é:

“Uma rede neural é um processador maciçamente paralelamente distribuído, constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimentos experimentais e torná-lo disponível para o uso. Ela se assemelha ao cérebro em dois aspectos:

1. O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem.
2. Forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.”

Nas próximas seções serão apresentadas, com mais detalhes, as partes referentes ao aprendizado da rede neural, as funções de ativação mais usuais e sua topologia.

Apesar de existirem diversos tipos de redes neurais, como por exemplo, Redes Neurais Recorrentes, *Self-Organizing Maps* (Kohonen, 2001), Máquinas de Vetor Suporte (Vapnik, 1998), Árvores de decisão Redes Bayesiana (Duda, Hart, & Stork, 2000) e Redes de Função de Base Radial (Hassoun, 1995), neste texto será dado enfoque às redes neurais utilizadas para a modelagem do sistema desconhecido introduzido no capítulo anterior - as redes MLP (*Multi Layer Perceptron*).

3.2.2 O Neurônio Artificial e Funções de Ativação

3.2.2.1 O Neurônio Artificial

Um neurônio artificial - que a partir de agora será chamado somente neurônio, assim como o neurônio humano - é uma unidade de processamento de informação. Um diagrama esquemático ilustrando um neurônio é mostrado na Figura 3.2.

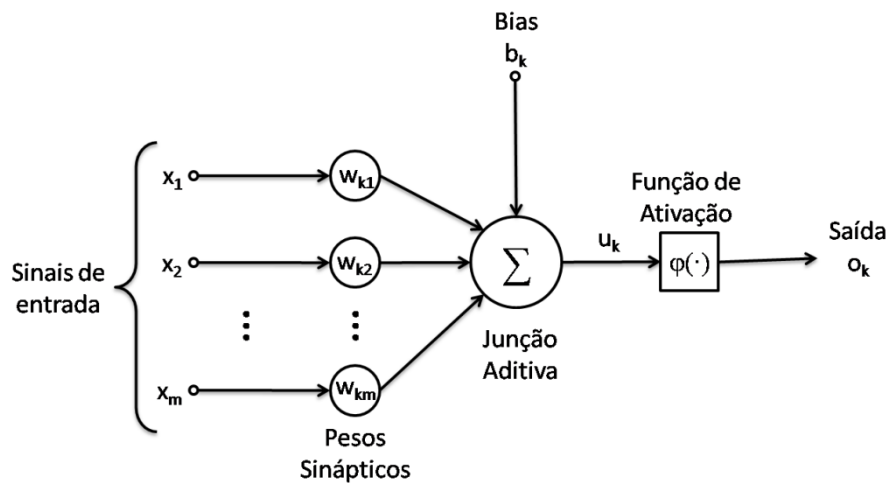


Figura 3.2 – Modelo esquemático de um neurônio artificial

Um neurônio apresenta um conjunto de pesos sinápticos que tem a função de ponderar cada uma de suas entradas. As entradas já ponderadas são aplicadas a um somador, encaminhando o somatório resultante a uma função de ativação, como detalhado na próxima seção.

Podemos modelar matematicamente um neurônio da seguinte maneira:

$$v_k = \sum_{j=1}^m w_{kj} x_j \quad (3.1)$$

$$o_k = \varphi(u_k) = \varphi(v_k + b_k) \quad (3.2)$$

onde

x_1, x_2, \dots, x_m são os sinais de entrada

$w_{k1}, w_{k2}, \dots, w_{km}$ são os pesos sinápticos do neurônio k

v_k é a saída da combinação das entradas ponderada pelos pesos sinápticos

b_k é o *bias* ou polarização

u_k é igual a v_k deslocado por b_k

$\varphi(\cdot)$ é a função de ativação

o_k é a saída do neurônio

Outra maneira de se modelar o neurônio é retirando o valor do sinal de *bias* e colocando em seu lugar um novo peso sináptico que tem como entrada sempre um valor +1. A função do *bias* é causar um deslocamento à entrada da função de ativação, como representado na Figura 3.3.

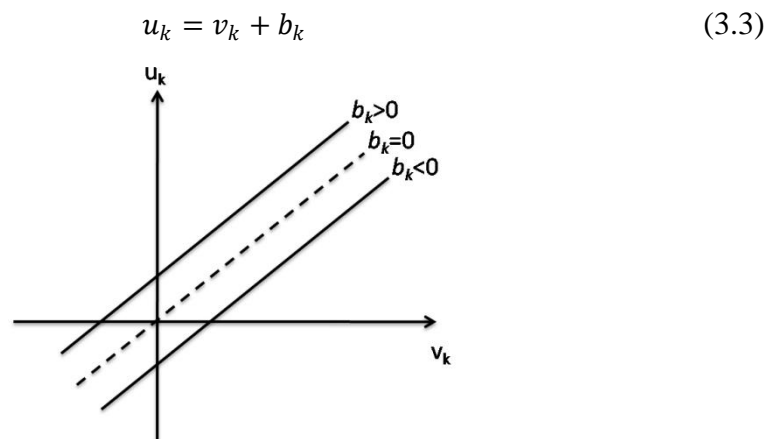


Figura 3.3 – Deslocamento ocasionado pelo *bias*

3.2.2.2 Funções de Ativação

As funções de ativação de uma rede neural são utilizadas após a saída do somatório, conforme a seção anterior, para manipular o resultado da ponderação das entradas em conjunto com o bias, permitindo restringir ou apenas modificar esses valores.

Os tipos mais usuais de funções de ativação são: Limiar, Linear por Partes, Sigmóide e Linear.

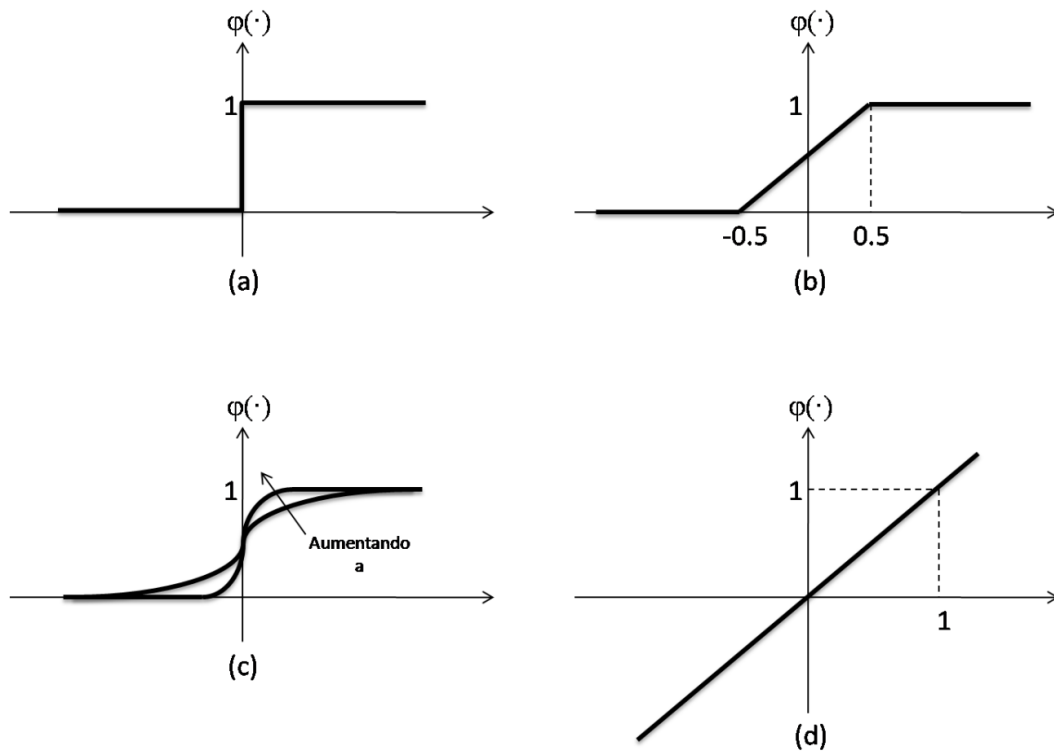


Figura 3.4 – Funções de ativação

(a) Limiar (b) Linear por Partes (c) Sigmóide (d) Linear

A função limiar é utilizada geralmente em redes neurais classificatórias, onde é interessante que a saída apresente somente dois valores: por exemplo 0 (zero) caso a saída do somatório seja negativa e 1 caso seja positiva, conforme ilustrado na Figura 3.4(a), ou seja,

$$\varphi(u) = \begin{cases} 0 & \text{se } u < 0 \\ 1 & \text{se } u \geq 0 \end{cases} \quad (3.4)$$

A função Linear por Partes é utilizada quando se deseja uma saída linear com uma saturação, ela apresenta uma saída linear quando o resultado do somatório está restrito a $(-0.5, 0.5)$ e satura para valores onde o módulo deste valor é maior ou igual à 0.5, conforme ilustrado na Figura 3.4(b), ou seja

$$\varphi(u) = \begin{cases} 1 & u \geq 0.5 \\ u & -0.5 > u > 0.5 \\ 0 & u \leq -0.5 \end{cases} \quad (3.5)$$

A função sigmóide é a função mais utilizada na construção de redes neurais, isso porque ela tem um perfeito balanceamento entre o comportamento linear e o não-linear. O exemplo mais comum é a função logística, definida como

$$\varphi(u) = \frac{1}{1 + \exp(-au)} \quad (3.6)$$

onde o parâmetro a indica a inclinação da função, conforme pode ser visto na Figura 3.4(c).

Outro exemplo muito comum, utilizando uma correspondente da função sigmóide é a função tangente hiperbólica, dada por

$$\varphi(u) = \tanh(u) \quad (3.7)$$

Podemos perceber fazendo uma comparação entre a Figura 3.4(a) e Figura 3.4(c) que a função sigmóide apresenta, assim como a função limiar, um limite quando seus valores tendem ao infinito, porém ela apresenta uma propriedade muito importante na teoria de redes neurais que é a diferenciabilidade; isso ficará mais claro na Seção 3.2.3.1 quando for explicado sobre a condição de optimalidade da rede.

A mais simples das funções de ativação é a função linear; na verdade é até um abuso de notação chamá-la de função de ativação, uma vez que os valores de saída do somatório não têm seus valores alterados. A Figura 3.4(d) ilustra a função linear, definida por

$$\varphi(u) = u \quad (3.8)$$

A função linear é geralmente utilizada quando não se deseja restringir o valor de saída, sendo seu uso muito comum na última camada de uma rede preditiva, onde os valores de saída podem estar fora do intervalo dos dados de treinamento e são minimamente distorcidos pelo aspecto linear da saída.

3.2.3 Perceptrons

3.2.3.1 Perceptrons de uma camada

Para facilitar o entendimento sobre redes neurais, nesta seção será introduzida uma rede com uma única camada e na seção seguinte o conceito será estendido para múltiplas camadas (Gomes, 2007).

O perceptron de uma camada pode ser facilmente associado a um filtro adaptativo e seu método de aprendizado ao algoritmo LMS (*Least Mean Square*) (Widrow & Hoff, 1960).

Como exemplo do modelo matemático, será utilizado um modelo de um perceptron, tendo uma camada com dois neurônios e quatro entradas, conforme mostrado na Figura 3.5.

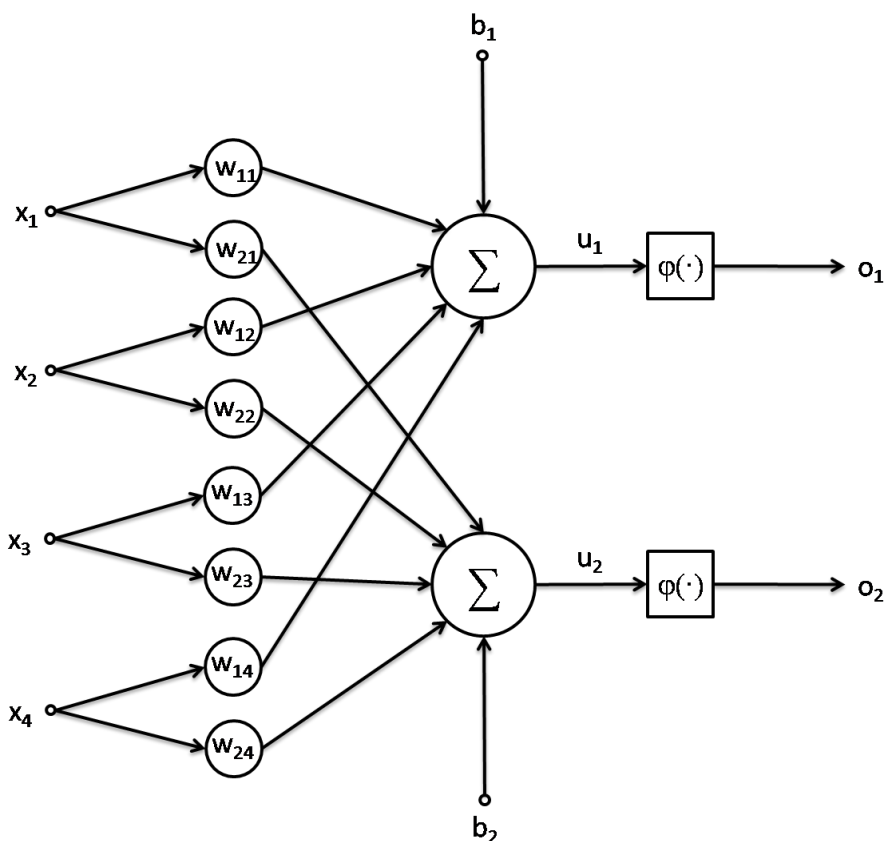


Figura 3.5 – Perceptron de uma camada com dois neurônios e quatro entradas

A formulação matemática pode então ser descrita por

$$u_i = \left(\sum_j w_{ij} x_j \right) + b_i \quad (3.9)$$

$$o_i = \varphi(u_i) \quad (3.10)$$

onde

$j = 1, \dots, M$ é o número de entradas (neste exemplo 4)

$i = 1, \dots, N$ é o número de neurônio (neste exemplo 2)

Caso a rede neural deva replicar um sistema com as mesmas quatro entradas e saída t_i , a medida de erro entre a saída da rede neural, o_i , e o real valor, t_i , é dada pela diferença entre esses valores, e a função custo total pelo somatório dos erros quadrados dividido por 2, ou seja,

$$e_i = t_i - o_i \quad (3.11)$$

$$J = \frac{\sum_i e_i^2}{2} \quad (3.12)$$

Como o valor do erro depende de o_i , que depende do *bias*, dos pesos e das entradas, a função custo será também função destas variáveis. Portanto a minimização da mesma será feita em função do *bias* e dos pesos, únicos controles sobre as entradas, levando a

$$(\mathbf{W}^*, \mathbf{b}^*) = \underset{(\mathbf{W}, \mathbf{b})}{\operatorname{argmin}} \bar{J} \quad (3.13)$$

Uma forma de fazer isso é através do gradiente desta função, já que este indica o caminho de crescimento da função. Neste caso, caminhando na direção contrária à do gradiente, estaremos minimizando o valor da função.

O gradiente da função custo é dado pela derivada parcial da mesma em função de seus parâmetros, pesos e *bias*. Para encontrarmos essas derivadas, será necessária a utilização da regra da cadeia, que pode ser encontrada no Apêndice A ou em (Malta, Pesco, & Lopes, 2002), resultando em

$$\frac{\partial J}{\partial b_i} = \frac{\partial J}{\partial e_i} \frac{\partial e_i}{\partial o_i} \frac{\partial o_i}{\partial u_i} \frac{\partial u_i}{\partial b_i} \quad (3.14)$$

$$\frac{\partial J}{\partial w_{ij}} = \frac{\partial J}{\partial e_i} \frac{\partial e_i}{\partial o_i} \frac{\partial o_i}{\partial u_i} \frac{\partial u_i}{\partial w_{ij}} \quad (3.15)$$

Essas derivadas podem ser resolvidas termo a termo. Para a resolução de $\frac{\partial o_i}{\partial u_i}$ será utilizada a função tangente hiperbólica ((3.7) como efeito de exemplificação, as demais derivadas não perdem suas propriedades devido a isso. Assim, temos

$$\frac{\partial J}{\partial e_i} = \frac{\partial \left(\frac{\sum_i e_i^2}{2} \right)}{\partial e_i} = e_i \quad (3.16)$$

$$\frac{\partial e_i}{\partial o_i} = \frac{\partial (t_i - o_i)}{\partial o_i} = -1 \quad (3.17)$$

$$\frac{\partial o_i}{\partial u_i} = \frac{\partial \varphi(u_i)}{\partial u_i} = 1 - o_i^2 \quad (3.18)$$

$$\frac{\partial u_i}{\partial b_i} = \frac{\partial [(\sum_j w_{ij} x_j) + b_i]}{\partial b_i} = 1 \quad (3.19)$$

$$\frac{\partial u_i}{\partial w_{ij}} = \frac{\partial [(\sum_j w_{ij} x_j) + b_i]}{\partial w_{ij}} = x_j \quad (3.20)$$

A resolução da equação ((3.18) pode ser encontrada com mais detalhes no Apêndice A. De posse das equações ((3.16)-((3.20) temos as derivadas da função custo para cada um de seus parâmetros:

$$\frac{\partial J}{\partial b_i} = -e_i(1 - o_i^2) \quad (3.21)$$

$$\frac{\partial J}{\partial w_{ij}} = -e_i(1 - o_i^2)x_j \quad (3.22)$$

Mediante isso podemos fazer a atualização dos valores dos pesos e dos *bias* da rede segundo as equações:

$$b_i(n+1) = b_i(n) - \eta \frac{\partial J}{\partial b_i} \quad (3.23)$$

$$w_{ij}(n+1) = w_{ij}(n) - \eta \frac{\partial J}{\partial w_{ij}} \quad (3.24)$$

onde

$\eta < 1$ é chamado de coeficiente de aprendizado e sua função é controlar o tamanho do passo a cada iteração, para que ao caminhar para o mínimo não o ultrapassemos.

3.2.3.2 Perceptrons de múltiplas camadas

O perceptron com múltiplas camadas é uma concatenação seqüencial de perceptrons de uma camada, ou seja, cada camada tem como entrada as saídas da camada anterior, com exceção da primeira que tem como entrada as próprias entradas da rede. A comparação do valor encontrado com o valor esperado é feita na saída da última camada; portanto, o erro se retropropaga por todas as camadas alterando o valor de todos os pesos e *biases*.

A Figura 3.6 ilustra o processo de composição de um perceptron de duas camadas a partir de dois perceptrons de uma camada. Esse processo pode ser repetido para a composição de perceptrons de mais camadas. É importante observar a notação dada depois da junção dos perceptrons. O índice entre parêntese, por exemplo $x_i^{(l)}$, indica a camada da rede à qual o parâmetro corresponde.

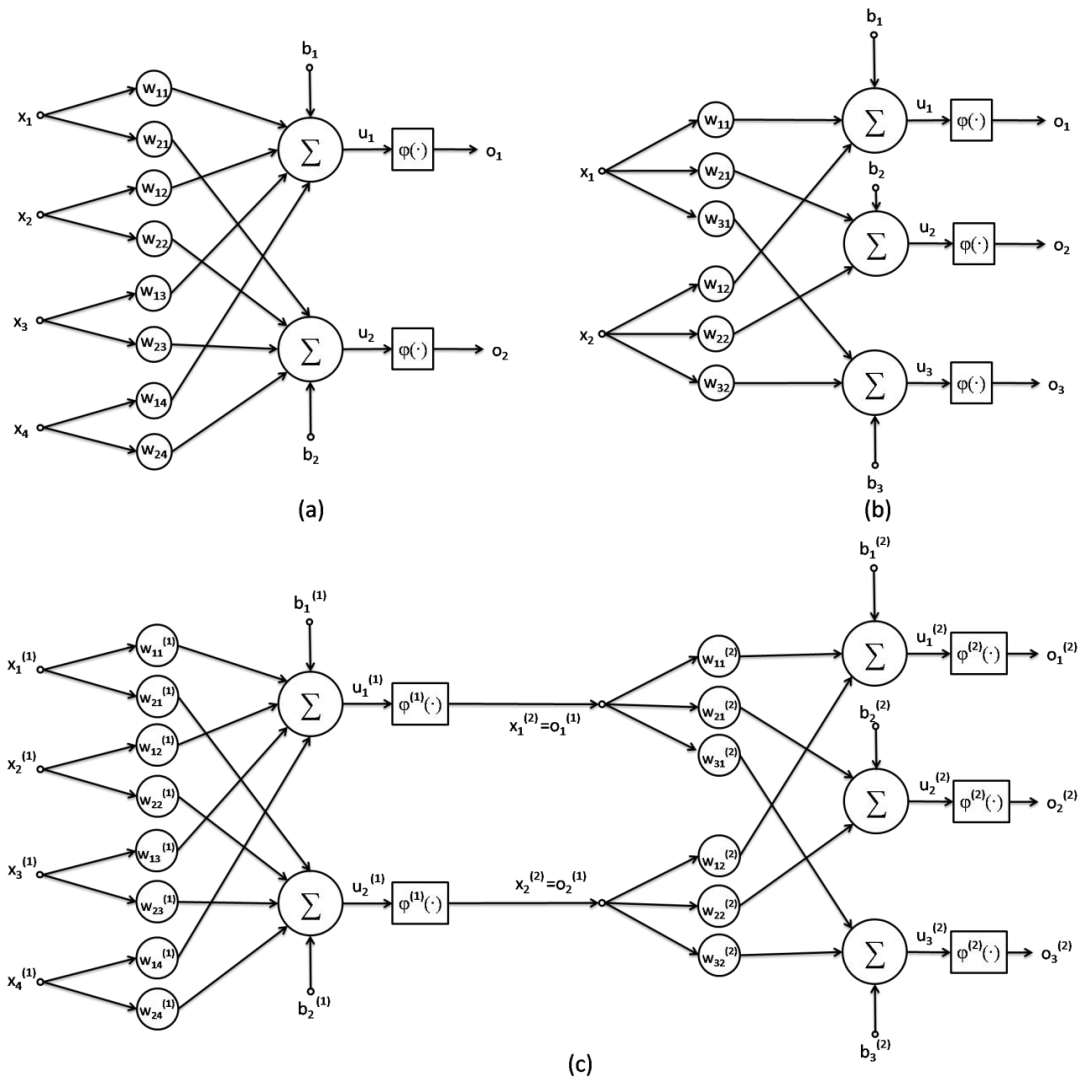


Figura 3.6 – Composição de um perceptron de múltiplas camadas - (a) e (b) perceptron de uma camadas e (c) perceptrons de duas camadas a partir de (a) e (b)

A Figura 3.6(c) leva ao seguinte conjunto de equações:

$$\mathbf{u}^{(l)} = \mathbf{W}^{(l)}\mathbf{x}^{(l)} + \mathbf{b}^{(l)} \quad (3.25)$$

$$\mathbf{o}^{(l)} = \varphi^{(l)}(\mathbf{u}^{(l)}) \quad (3.26)$$

$$\mathbf{x}^{(l+1)} = \mathbf{o}^{(l)} \quad (3.27)$$

$$\mathbf{e} = \mathbf{t} - \mathbf{o}^{(L)} \quad (3.28)$$

$$J = \frac{\mathbf{e}^T \mathbf{e}}{2} \quad (3.29)$$

onde

$l = 1, \dots, L$ é o número das camadas

L é o número de camadas

T é o operador transposto

A atualização dos pesos e bias da rede segue o seguinte conjunto de equações:

$$\frac{\partial J}{\partial \mathbf{b}^{(l)}} = -\left(\mathbf{I} - \text{diag}^2(\mathbf{o}^{(l)})\right) (\mathbf{W}^{(l+1)})^T \left(\mathbf{I} - \text{diag}^2(\mathbf{o}^{(l+1)})\right) \dots (\mathbf{W}^{(L)})^T \left(\mathbf{I} - \text{diag}^2(\mathbf{o}^{(L)})\right) \mathbf{e} \quad (3.30)$$

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{W}^{(l)}} &= -\mathbf{o}^{(l-1)} \odot \left(\mathbf{I} - \text{diag}^2(\mathbf{o}^{(l)})\right) (\mathbf{W}^{(l+1)})^T \left(\mathbf{I} - \text{diag}^2(\mathbf{o}^{(l+1)})\right) \dots (\mathbf{W}^{(L)})^T \left(\mathbf{I} - \text{diag}^2(\mathbf{o}^{(L)})\right) \mathbf{e} \\ &= -\mathbf{o}^{(l-1)} \odot \left[-\frac{\partial J}{\partial \mathbf{b}^{(l)}}\right] \end{aligned} \quad (3.31)$$

$$\mathbf{b}^{(l)}(n+1) = \mathbf{b}^{(l)}(n) - \eta \frac{\partial J}{\partial \mathbf{b}^{(l)}} \quad (3.32)$$

$$\mathbf{W}^{(l)}(n+1) = \mathbf{W}^{(l)}(n) - \eta \frac{\partial J}{\partial \mathbf{W}^{(l)}} \quad (3.33)$$

O operador \odot indica o produto de Kronecker, detalhado no conjunto de equações ((3.25)-(3.33) no Apêndice A. Assim como na seção anterior foi utilizada a função de ativação tangente hiperbólica para exemplificação.

3.2.4 Uma Rede Neural como um Combinador Linear

Uma rede neural pode ser utilizada para fazer a combinação linear das entradas caso sejam utilizadas somente funções de ativação lineares em seus neurônios. Com isso reduzimos a equação ((3.26) a

$$\mathbf{o}^{(l)} = \mathbf{u}^{(l)} \quad (3.34)$$

E utilizando ((3.25) e ((3.27) chegamos a:

$$\begin{aligned} \mathbf{u}^{(l+1)} &= \mathbf{W}^{(l+1)} \mathbf{x}^{(l+1)} + \mathbf{b}^{(l+1)} = \mathbf{W}^{(l+1)} \mathbf{o}^{(l)} + \mathbf{b}^{(l+1)} \\ &= \mathbf{W}^{(l+1)} \mathbf{u}^{(l)} + \mathbf{b}^{(l+1)} \\ &= \mathbf{W}^{(l+1)} [\mathbf{W}^{(l)} \mathbf{x}^{(l)} + \mathbf{b}^{(l)}] + \mathbf{b}^{(l+1)} \\ &= \mathbf{W}^{(l+1)} \mathbf{W}^{(l)} \mathbf{x}^{(l)} + \mathbf{W}^{(l+1)} \mathbf{b}^{(l)} + \mathbf{b}^{(l+1)} \end{aligned} \quad (3.35)$$

Fazendo $\mathbf{V}^{(l)} = \mathbf{W}^{(l+1)} \mathbf{W}^{(l)}$ e $\mathbf{c}^{(l)} = \mathbf{W}^{(l+1)} \mathbf{b}^{(l)} + \mathbf{b}^{(l+1)}$ e utilizando ((3.34)

chegamos a:

$$\mathbf{o}^{(l+1)} = \mathbf{V}^{(l)}\mathbf{x}^{(l)} + \mathbf{c}^{(l)} \quad (3.36)$$

que corresponde à redução de uma rede neural de múltiplas camadas a uma rede de uma única camada. Portanto o valor de saída se torna uma combinação linear dos valores de entrada, ponderados pelos pesos e somados a um *bias*.

3.3 Quantizador Vetorial

A metodologia de quantização vetorial (Gersho & Gray, 1992) é amplamente conhecida e empregada para a compressão de sinais com o intuito de reduzir o volume da informação que deverá ser transmitida ou armazenada, mas com um objetivo bem claro: possibilitar a recuperação do sinal com uma distorção/ruído de quantização admissível. Na quantização vetorial é necessária a criação de um dicionário para codificação deste sinal – um conjunto de vetores, em menor quantidade que os da informação original, que a represente de uma forma aproximada. O vetor representante é transmitido e ao final da transmissão decodificado, recuperando assim o sinal com o nível de perdas desejado. Para isso é necessária a otimização desta aproximação de forma a contrabalançar seus dois principais parâmetros: tamanho da informação a ser enviada e nível de ruído desejado.

A etapa mais importante de uma quantização vetorial é a criação do dicionário de codificação, pois é nela que será definido o nível de compressão do sinal e o nível de ruído inserido na informação. Existem diversas técnicas utilizadas para a criação deste dicionário que são amplamente empregadas em análises de clusterização e quantização vetorial. As mais usuais são: Poda (*Pruning*), Vizinheiro mais próximo (*Pairwise Nearest Neighbor - PNN*), Partição (*Splitting*), Algoritmo Generalizado de Lloyd (*Generalized Lloyd Algorithm – GLA*), Relaxação Estocástica (*Stochastic Relaxation*) e Clusterização Fuzzy (*Fuzzy Clustering*).

Um dicionário é formado por centróides, que são os vetores centrais da clusterização de uma determinada classe, que por sua vez são as divisões entre os vetores de entrada que mais se assemelham, (mais próximos pela métrica adotada).

Na próxima seção apresentaremos o diagrama de Voronoi e a técnica do GLA combinada com a Partição.

3.3.1 Diagrama de Voronoi

O diagrama de Voronoi é uma divisão do espaço em regiões de acordo com a distância a determinados pontos. Para um dado conjunto $A = [a_1, a_2, \dots, a_n]$ de n pontos, existe uma divisão em n regiões p , otimizando as distâncias, ou seja, as linhas que delimitam cada conjunto são encontradas de modo que os pontos pertencentes a este conjunto tenham uma distância menor em relação ao seu centróide do que em relação a qualquer outro centróide. Isso pode ser melhor entendido observando-se a Figura 3.7, que ilustra um diagrama de Voronoi para um conjunto de $n = 10$ pontos. O conjunto delimitado pelas linhas em vermelho é o conjunto relacionado ao ponto a_i formando uma região p_i . Os pontos dentro desta região apresentam menor distância à a_i do que a qualquer outro ponto do conjunto A . Ou seja, a delimitação encontrada pelo diagrama é a melhor possível em termos de distância.

Esse diagrama é utilizado para fazer a clusterização dos vetores de entrada de um quantizador vetorial em relação aos seus centróides. A próxima seção descreve isso mais detalhadamente.

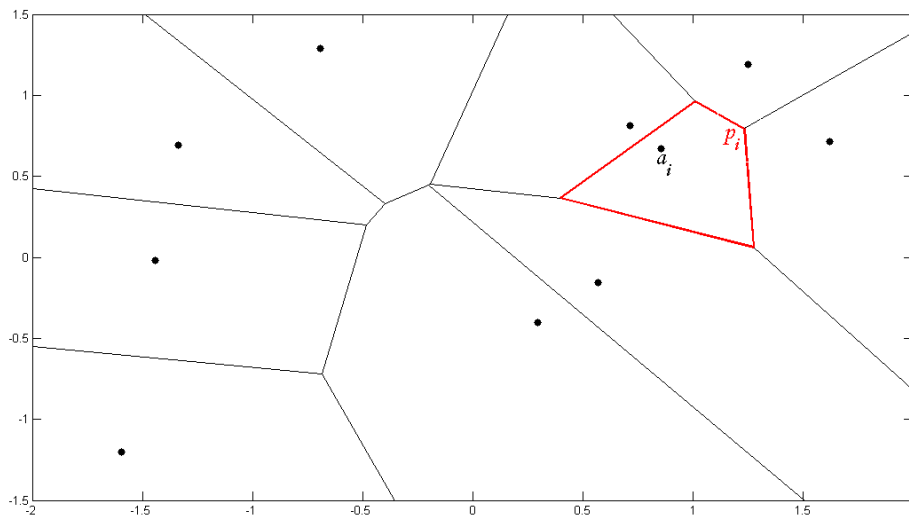


Figura 3.7 – Diagrama de Voronoi

3.3.2 O Algoritmo Generalizado de Lloyd

O algoritmo generalizado de Lloyd, ou simplesmente GLA (será utilizada a sigla em inglês devido à maioria das referências no assunto serem em inglês), é baseado na modificação do dicionário através de um processo de iterações. Em princípio assumimos que a função densidade de probabilidade dos vetores de entrada é conhecida, continuamente e uniformemente distribuída.

O objetivo principal do algoritmo é criar um dicionário que apresente a menor distorção possível. Ou seja, a clusterização dos vetores de entrada é otimizada de acordo com a métrica de distância utilizada.

Inicialmente temos um conjunto de dados de entrada \mathbf{X} e um dicionário inicial \mathbf{Y} .

$$\mathbf{X} \in \mathbb{R}^{M \times N} \quad (3.37)$$

$$\mathbf{x}(n) \in \mathbb{R}^M, n = 1, \dots, N \quad (3.38)$$

$$\mathbf{Y} \in \mathbb{R}^{M \times K} \quad (3.39)$$

$$\mathbf{y}(k) \in \mathbb{R}^M, n = 1, \dots, K \quad (3.40)$$

Cada iteração procura, para cada vetor de entrada, utilizando uma medida de distância $d(\mathbf{x}, \mathbf{y})$, como por exemplo, qual o centróide mais próximo e atribui a ele o

índice de indexação do centróide $\hat{\mathbf{x}}(n)$ como efeito de clusterização conforme as equações

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2 = \sum_{m=1}^M (x_m - y_m)^2 \quad (3.41)$$

$$\hat{\mathbf{x}}(n) = \beta \left(\alpha(\mathbf{x}(n)) \right) \quad (3.42)$$

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \min_{k \in \{1, 2, \dots, K\}} d(\mathbf{x}, \mathbf{y}_k) \quad (3.43)$$

Após analisar todos os centróides é calculada a distorção de todos os vetores de entrada em relação ao seu centróide, ou seja

$$D = \frac{1}{N} \sum_{n=1}^N d(\mathbf{x}(n), \hat{\mathbf{x}}(n)) = \frac{1}{N} \sum_{n=1}^N \|\mathbf{x}(n) - \hat{\mathbf{x}}(n)\|^2 = E[d(\mathbf{x}, \hat{\mathbf{x}})] \quad (3.44)$$

A clusterização dos vetores de entrada é feita, por exemplo, utilizando a condição do vizinho mais próximo, gerando um conjunto R_k dado por

$$R_k = \{\mathbf{x} \in \mathcal{T}: d(\mathbf{x}, \mathbf{y}_k) \leq d(\mathbf{x}, \mathbf{y}_j); \forall j \neq k\} \quad (3.45)$$

Ao final desse processo tem-se tantos conjuntos R_i quanto centróides, e a partir desse agrupamento podemos calcular os novos valores de centróides, ou seja

$$\mathbf{y}_k = \frac{1}{N_k} \sum_{\mathbf{x}(n) \in R_k} \mathbf{x}(n) \quad (3.46)$$

Esse processo minimiza a distâncias dos centróides aos vetores associados a ele.

Feito isso para todos os k conjuntos/classes o processo do GLA descrito acima é refeito gerando uma nova divisão das classes, com uma distorção menor ou igual à anterior.

A Figura 3.8 ilustra um diagrama esquemático do algoritmo, onde a condição de parada é definida pelo usuário. Pode ser usada, por exemplo, a diferença entre a distorção atual e a da iteração anterior; se for menor que certo valor pare, se não continue, ou então até que um certo número máximo de iterações seja alcançado.

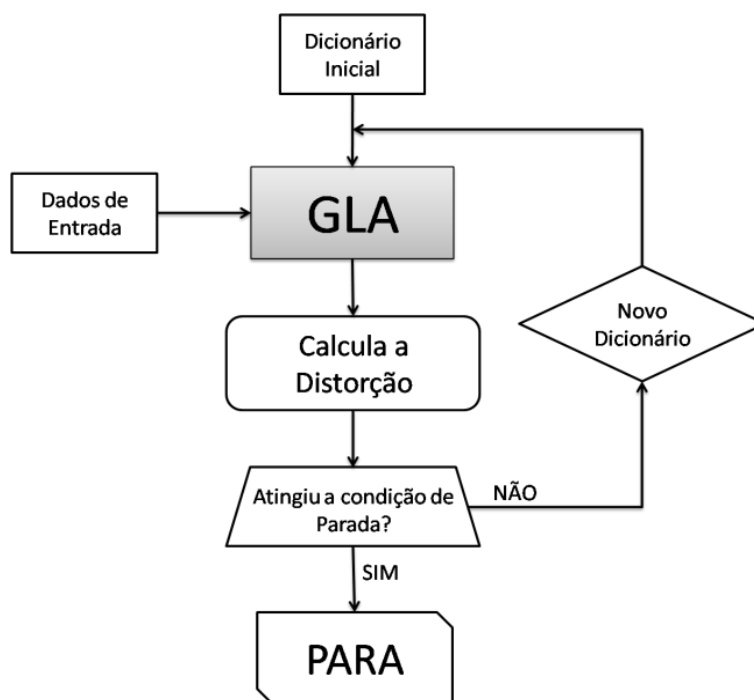


Figura 3.8 – Diagrama esquemático do GLA

O GLA em sua formulação padrão apresenta um dicionário inicial do mesmo tamanho (quantidade de centróides) que o dicionário final. Uma modificação muito usual é a criação do dicionário através de Partição, ou seja, o dicionário é inicializado com um número menor de centróides e então estes são quebrados, conforme o algoritmo vai evoluindo, até que ao final de todas as iterações o tamanho final do dicionário seja o desejado. Contudo, o algoritmo pode ser aplicado sem perda de generalidade alguma. Na verdade, até ocorrer uma quebra, o algoritmo é executado normalmente para um dicionário com um número de centróides. No instante da quebra, o algoritmo é interrompido e então se dá início a um novo processo com um novo dicionário, agora com um tamanho maior. Ou seja, a cada quebra executamos o algoritmo independente do anterior.

O processo de quebra é feito adicionando um pequeno vetor (ϵ) ao dicionário anterior, por exemplo, inicialmente, temos um único centróide (Y_0) que define uma única classe onde todos os vetores de entrada se encontram, após a partição, e antes de

executar o algoritmo, teremos $[Y_0 + \epsilon, Y_0]$. Podemos notar que esse novo dicionário conterá o dicionário anterior, ou seja, o passo seguinte apresentará uma perda menor ou igual à apresentada anteriormente. Esse processo é feito diversas vezes, até que o dicionário final seja do tamanho desejado ou um nível tolerável de ruído seja alcançado.

3.4 Conclusões

Esse capítulo fez uma breve revisão da literatura utilizada para a modelagem matemática do sistema desconhecido do Capítulo 2. Primeiramente foi detalhada a modelagem de uma rede neural, seus componentes e seus parâmetros, e a partir da Seção 3.3 foi abordada a metodologia de um quantizador vetorial, com detalhes nos tópicos utilizados neste trabalho.

4 Identificação do Modelo do Preditor

4.1 Introdução

Este capítulo mostrará como são feitas as identificações do modelo desconhecido, descrito em 2.5 e ilustrado na Figura 2.10, pelas ferramentas matemáticas apresentadas no Capítulo 3.

Na Seção 4.5 será feita a combinação dos resultados encontrados e a análise dos erros de cada um dos métodos.

4.2 A identificação do modelo desconhecido

O sistema desconhecido descrito em 2.5 e ilustrado na Figura 2.10 será identificado utilizando as ferramentas matemáticas descritas no Capítulo 3. As mesmas variáveis explicativas que, através do sistema desconhecido, são capazes de gerar a vazão como saída serão utilizadas em um sistema matemático com intuito de gerar uma estimativa da vazão. A Figura 4.1 ilustra o processo de identificação do sistema desconhecido.

Caso tenhamos como saída do sistema matemático uma réplica da vazão, ele representará exatamente o sistema desconhecido; mas como isso é, na prática, algo difícil de ser alcançado, o sistema matemático a ser utilizado é aquele que consegue gerar estimativas da vazão com erros toleráveis.

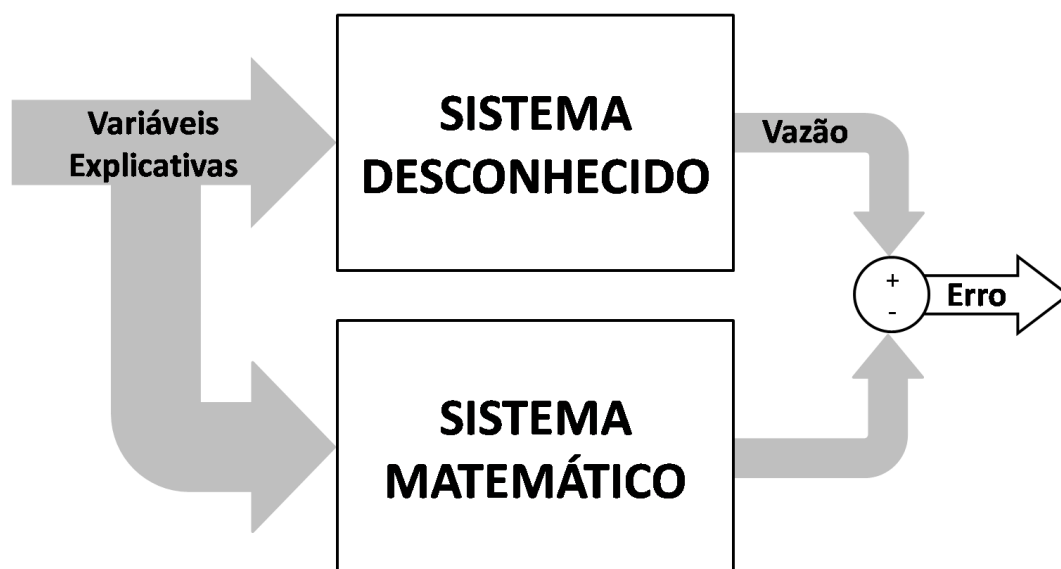


Figura 4.1 – Processo de identificação do sistema desconhecido

Para chegarmos a esse sistema matemático vamos dividir a série de vazão em três conjuntos de dados: a primeira parte será utilizada para o treinamento do modelo, a segunda para obtenção de uma medida de qualidade do treinamento e a terceira, que só será utilizada no próximo capítulo, para obtenção da medida de qualidade da previsão.

A divisão será feita da seguinte forma: a terceira parte (ou conjunto de teste) será o último ano da série, a segunda (ou conjunto de validação) será o penúltimo ano e o restante será a primeira (ou conjunto de treinamento). Essa divisão é feita de modo que sempre, independente da data final da série, os conjunto de validação e teste englobem períodos de seca e cheia dentro de seu ano de duração. A Figura 4.2 ilustra a divisão para série de vazão do rio São Francisco, localizado no Nordeste do Brasil.

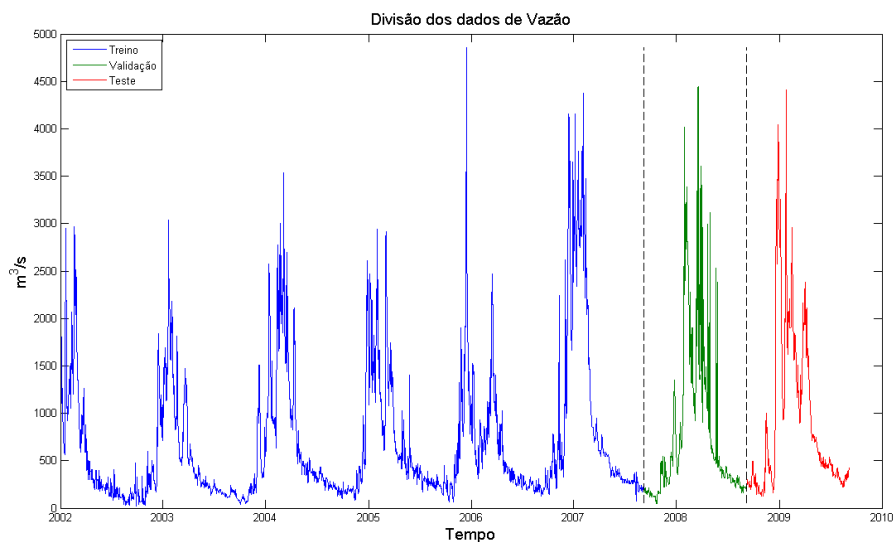


Figura 4.2 – Divisão da série de vazão nos conjunto de treinamento, teste e validação

Para auxiliar as variáveis explicativas, poderão ser, utilizados dois fasores de tempo (seno e cosseno com período de um ano) que introduzem no sistema uma informação temporal. A Figura 4.3 ilustra estes fasores para o mesmo intervalo apresentado na Figura 4.2.

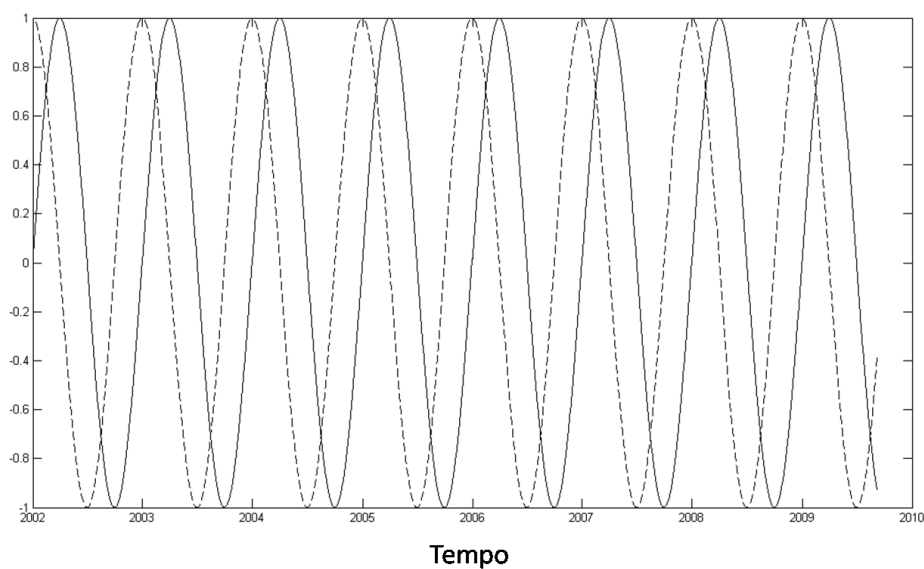


Figura 4.3 – Fasores de Tempo

A escolha das variáveis explicativas é um assunto que está sob forte estudo, não sendo abordado aqui por fugir aos objetivos deste trabalho. Para o caso exemplo que será mostrado nas próximas seções, as variáveis explicativas encontradas estão ilustradas na Figura 4.4 e descritas na Tabela 4.1.

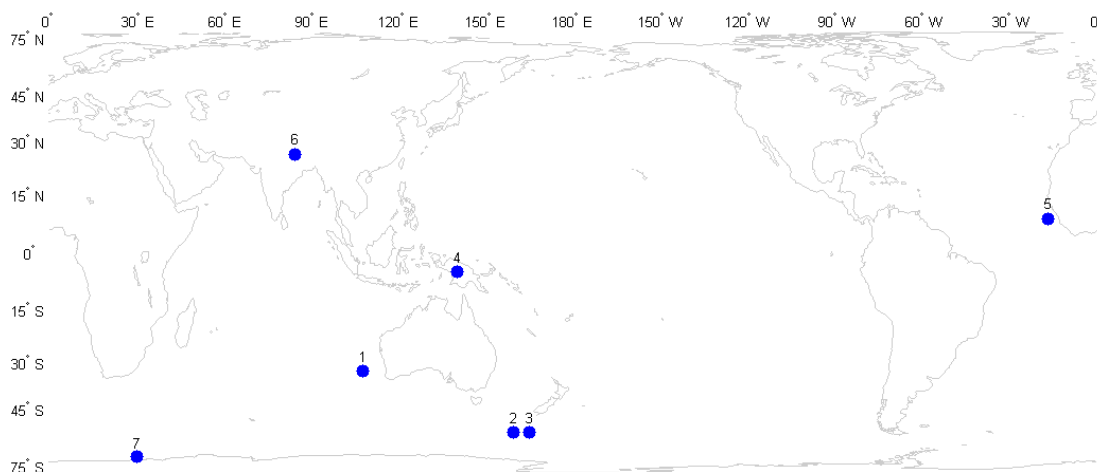


Figura 4.4 – Coordenadas encontradas para o caso exemplo das próximas seções

Tabela 4.1 – Descrição das variáveis explicativas do caso exemplo das próximas seções

	Variável Climatológica	Latitude	Longitude	Adiantamento (em dias)
1	Inovação OLR	32,5 S	107,5 L	200
2	Inovação Temperatura do Ar	54,3 S	159,4 L	200
3	Inovação Temperatura do Ar	54,3 S	165,0 L	200
4	Inovação OLR	5,0 S	140,0 L	195
5	Taxa de Precipitação	8,7 N	17,0 O	188
6	Taxa de Precipitação	26,1 N	84,4 L	188
7	Inovação Temperatura do Ar	67,6 S	30,0 L	190

Os dados de Inovação são obtidos através de um pré processamento de uma série, por exemplo OLR, como descrito em (Macêdo, 2006).

4.3 A identificação por Redes Neurais

A identificação do sistema desconhecido pela modelagem de redes neurais, descrita em 3.2, utiliza como entradas as variáveis explicativas e os fasores, e como saída a vazão histórica.

A modelagem de uma rede neural tem por construção o objetivo da minimização do erro, por exemplo, o erro médio quadrático. Mas a simples minimização do erro quadrático é insuficiente aos objetivos deste estudo. Sabe-se que um treinamento excessivo da rede ocasiona perda de generalização, ou seja, a rede fica restrita aos valores utilizados para treinamento, e a saída no intervalo de dados que não foram utilizados para o treinamento fica ruim. Para que isso não ocorra, é feita a divisão dos dados de entrada e de saída. O conjunto de dados de treinamento é utilizado para a otimização dos pesos e *bias* e o conjunto de validação é utilizado como medida de generalização a cada iteração da rede, mas em nenhum momento é utilizado para a otimização. O ponto de parada no treinamento da rede será o momento em que a curva de erro obtida pelos dados de validação pára de decrescer e começa a aumentar – isto é chamado na literatura de supertreinamento ou *overtraining*.

As curvas ilustradas na Figura 4.5 apresentam os erros de treinamento e validação da Rede Neural. O ponto onde ocorre o *overtraining* é o exato momento onde guardamos os valores dos pesos e *bias* da rede para as suas futuras simulações.

A topologia final da rede utilizada neste caso exemplo apresenta 4 camadas escondidas com 18, 20, 16 e 10 neurônios, todos com função de ativação tangente hiperbólica e uma camada de saída com um único neurônio com função linear.

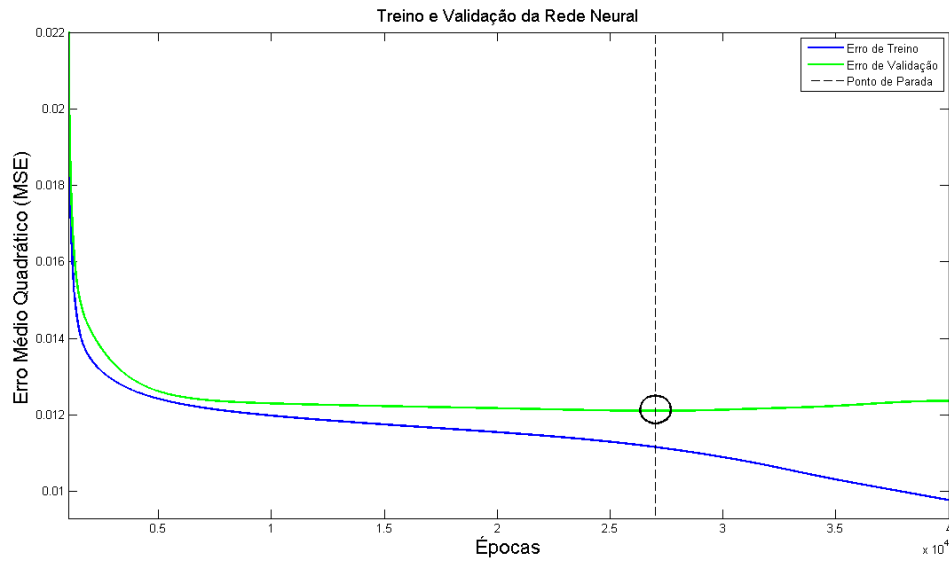


Figura 4.5 – Processo de treinamento e validação da Rede Neural

Com estes valores de pesos e *bias* temos uma Rede Neural treinada e pronta para a simulação do sistema desconhecido. Utilizando os valores de variáveis explicativas e fasores dos conjuntos de treino e validação temos os resultados apresentados nas Figura 4.6 e Figura 4.7. Podemos perceber que existe uma diferença entre os valores originais de vazão e os simulados pela rede, isso porque a identificação do sistema desconhecido apresenta um certo erro, que consideramos tolerável.

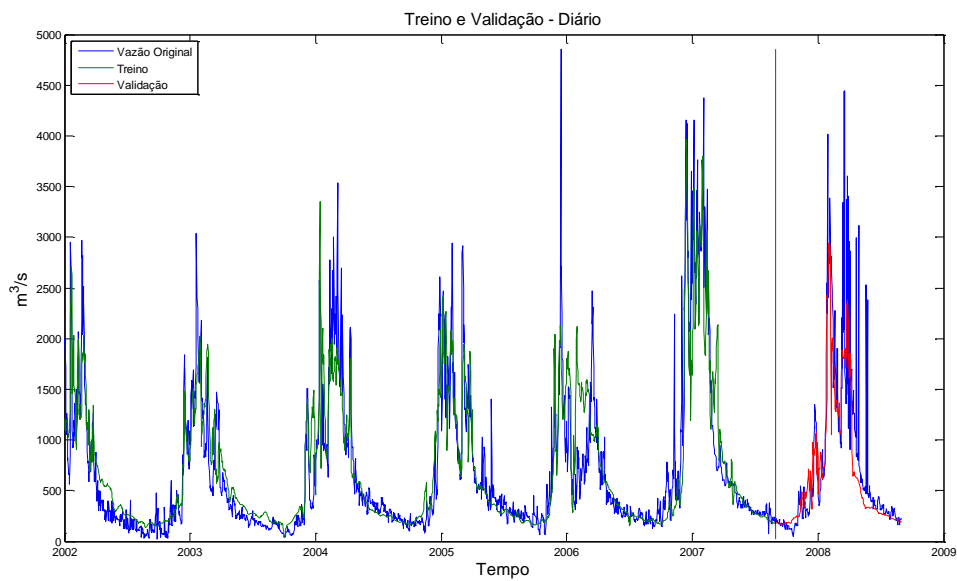


Figura 4.6 – Identificação do modelo pela Rede Neural – em escala de tempo diária

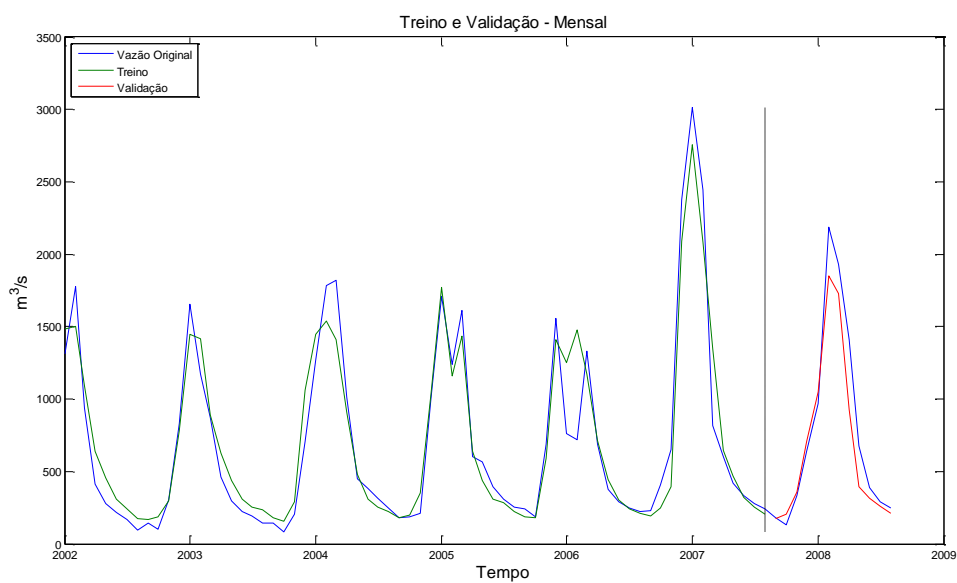


Figura 4.7 – Identificação do modelo pela Rede Neural – em escala de tempo mensal

4.4 A identificação por Quantização Vetorial

Para fazer a identificação do sistema desconhecido utilizando a modelagem matemática de quantização vetorial, descrita na Seção 3.3, é necessário que sejam criados vetores, um para cada instante de tempo, compostos pelas variáveis explicativas, pelos fasores de tempo e vazão, organizados na matriz:

$$\left[\begin{array}{cccc|ccc} VE_{11} & VE_{12} & \cdots & VE_{1N} & VE_{1(N+1)} & \cdots & VE_{1(N+M)} \\ VE_{21} & VE_{22} & \cdots & VE_{2N} & VE_{2(N+1)} & \cdots & VE_{2(N+M)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ VE_{I1} & VE_{I2} & \cdots & VE_{IN} & VE_{I(N+1)} & \cdots & VE_{I(N+M)} \\ F_{11} & F_{12} & \cdots & F_{1N} & F_{1(N+1)} & \cdots & F_{1(N+M)} \\ F_{21} & F_{22} & \cdots & F_{2N} & F_{2(N+1)} & \cdots & F_{2(N+M)} \\ V_1 & V_2 & \cdots & V_N & V_{N+1} & \cdots & V_{N+M} \end{array} \right] \quad (4.1)$$

Treinamento
Validação

onde

VE_{in} é o valor da variável explicativa i (para $i = 1, \dots, I$) no instante de tempo n (para $n = 1, \dots, N + M$)

F_{1n} e F_{2n} são os dois valores dos fasores no instante de tempo n (para $n = 1, \dots, N + M$)

V_n é o valor da vazão no instante de tempo n (para $n = 1, \dots, N + M$)

N é o número de instantes de tempo no conjunto de treino

M é o número de instantes de tempo no conjunto de teste

I é o número de variáveis explicativas utilizadas

O primeiro passo na identificação do modelo é a criação do dicionário de quantização. Esse dicionário é criado utilizando o algoritmo descrito na Seção 3.3.2 (GLA) de modo a classificar cada um dos vetores do conjunto de treinamento (somente os de treinamento), gerando um conjunto de centróides que possam reproduzir o dado de vazão com certo nível de distorção. Ou seja, será feito um mapeamento direto das variáveis explicativas na vazão.

O dicionário é criado de modo a otimizar a distorção entre os valores originais de vazão e a aproximação encontrada para a mesma no período de validação. A Figura 4.8 ilustra o erro de validação ao longo de um processo de treinamento. No eixo horizontal encontram-se marcadas as quebras, de acordo com o método de *splitting*, feitas no dicionário, e no eixo vertical o erro médio quadrático entre a vazão original e a aproximação dada pelo modelo ao longo do conjunto de dados de validação.

Da mesma forma que no treinamento da Rede Neural, o conjunto de validação é utilizado aqui para evitar o treinamento excessivo, ou seja, um dicionário muito grande. Ao observar a figura pode parecer que o aprendizado deveria ter sido interrompido após a segunda quebra (de 4 para 8 centróides), pois ocorreu um aumento do erro. Mas por experiência o modelo não deve ter uma quantidade muito pequena de centróides, o que pode ser comprovado pela figura, uma vez que os valores decresceram quando foram feitas mais quebras no dicionário. O modelo é escolhido para um número de centróides não muito pequeno e quando o erro de validação é o menor, neste caso, 128 centróides.

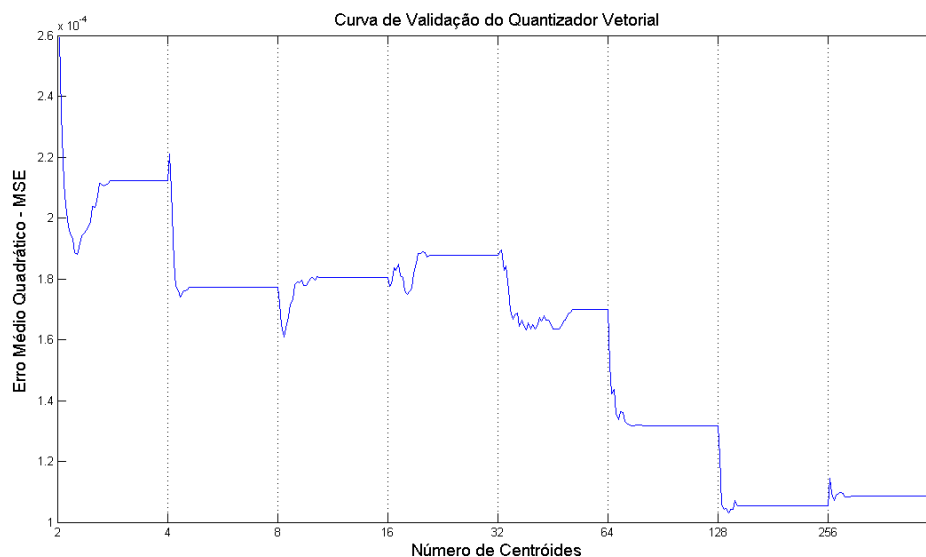


Figura 4.8 – Erro entre a vazão original e a aproximação pelo VQ no conjunto de validação à cada quebra do dicionário

Ou seja, o modelo é treinado utilizando o conjunto de dados de treinamento e otimizado pelo conjunto de validação, resultando em um dicionário que seja o mais adequado para os dois conjuntos de dados e que por construção deve ser o mais apropriado para o conjunto de teste, que será discutido no próximo capítulo.

Ao fim do treinamento teremos um dicionário, onde a última dimensão de cada centróide são os valores de vazão que poderemos ter. Com isso podemos encontrar a aproximação dada por esse dicionário ao longo do conjunto de treinamento e validação da seguinte forma:

1. Corta-se a última dimensão de cada um dos vetores formados pelas variáveis explicativas, fasores e vazão. Ou seja, os vetores descritos na equação ((4.1) serão agora formados somente por variáveis explicativas e fasores, conforme mostrado abaixo.

$$\left[\begin{array}{cccc|ccc} VE_{11} & VE_{12} & \dots & VE_{1N} & VE_{1(N+1)} & \dots & VE_{1(N+M)} \\ VE_{21} & VE_{22} & \dots & VE_{2N} & VE_{2(N+1)} & \dots & VE_{2(N+M)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ VE_{I1} & VE_{I2} & \dots & VE_{IN} & VE_{I(N+1)} & \dots & VE_{I(N+M)} \\ F_{11} & F_{12} & \dots & F_{1N} & F_{1(N+1)} & \dots & F_{1(N+M)} \\ F_{21} & F_{22} & \dots & F_{2N} & F_{2(N+1)} & \dots & F_{2(N+M)} \end{array} \right] \quad (4.2)$$

Treinamento *Validação*

2. Encontra-se, para cada um dos vetores, o centróide (também sem levar em conta o último elemento) mais próximo a ele.
3. Associa-se o valor do último elemento deste centróide mais próximo ao último elemento do vetor.

Depois de feito este processo para cada um dos vetores, os últimos elementos dos mesmos serão uma aproximação da vazão em cada instante e, então, podem ser comparados com o valor da vazão original.

As Figura 4.9 e Figura 4.10 ilustram a identificação do modelo pelo Quantizador Vetorial. A Figura 4.9 ilustra o processo em escala de tempo diária e a Figura 4.10

ilustra o mesmo processo após fazermos a média mensal dos dados diários. Nas duas figuras podemos perceber que a identificação apresenta um erro pequeno, ou seja, o sistema desconhecido descrito é agora um sistema conhecido que apresenta um erro tolerável.

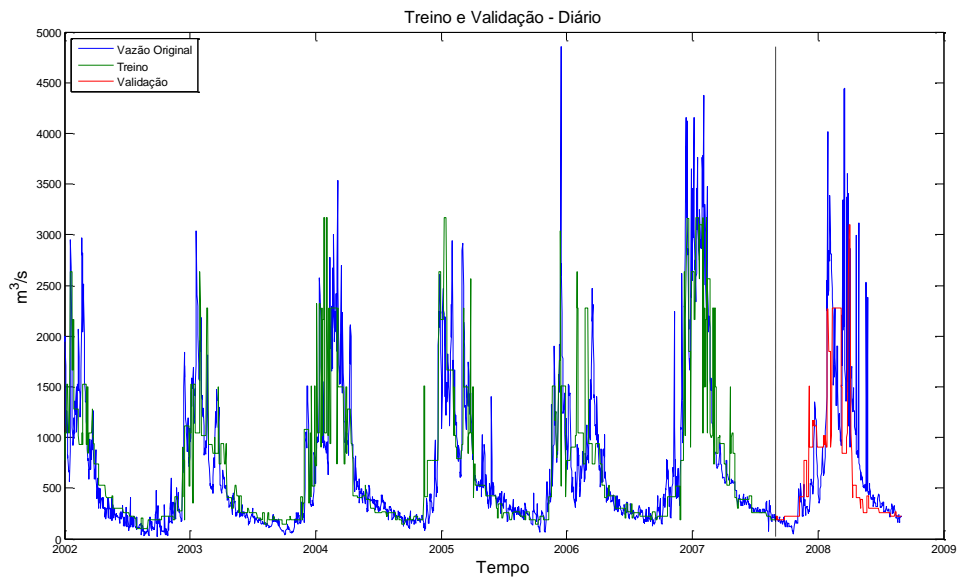


Figura 4.9 – Identificação do modelo pelo Quantizador Vetorial – em escala de tempo diária

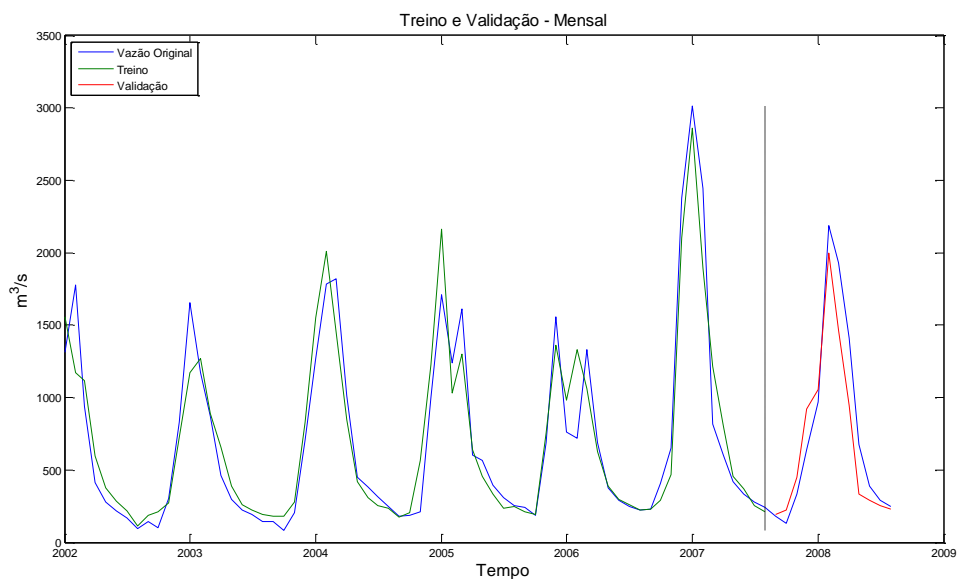


Figura 4.10 – Identificação do modelo pelo Quantizador Vetorial - em escala de tempo mensal

4.5 A identificação pela combinação dos métodos anteriores

O objetivo da combinação é unir o melhor dos dois métodos anteriores de modo a diminuir os erros. Logo, após concluirmos a identificação do modelo pelos métodos de Redes Neurais e Quantização Vetorial, descritos respectivamente nas Seções 4.3 e 4.4 podemos fazer uma combinação destes métodos.

A combinação será feita aqui utilizando-se uma nova Rede Neural, totalmente independente da primeira e puramente linear – matematicamente funcionará como um combinador linear – onde as entradas serão as saídas dos dois modelos para o conjunto de treino, série em verde das Figura 4.6 e Figura 4.9. Podem ser, também utilizados os fasores de tempo como entrada.

O modelo completo de identificação será composto por duas etapas de identificação: na 1ª Etapa é feita a identificação pela Rede Neural e, em paralelo, pelo Quantizador Vetorial; na segunda etapa as saídas destes dois modelos são utilizadas como entradas do Combinador, que tem como saída a resposta final de toda a identificação.

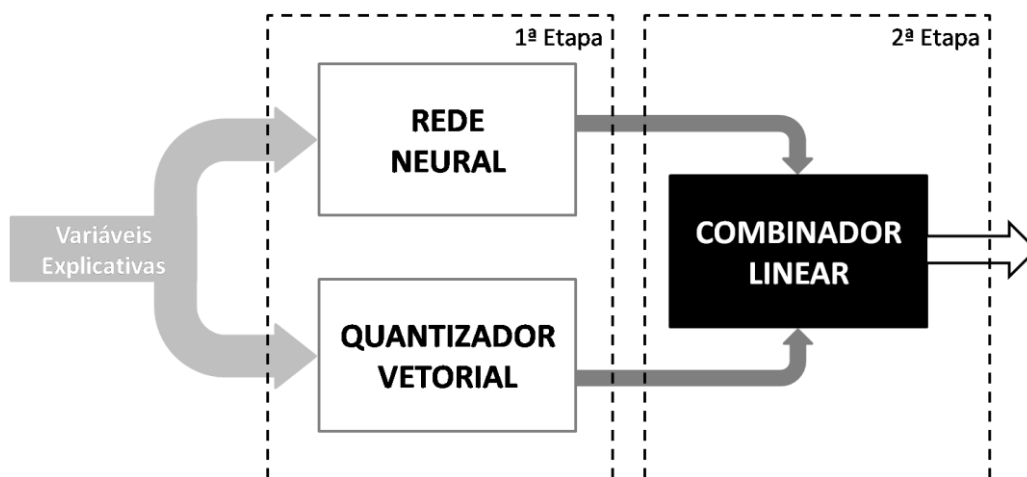


Figura 4.11 – Modelo Completo de Identificação

O processo de treinamento e validação desta Rede Neural segue o mesmo que foi feito para a primeira. Ou seja, o objetivo é minimizar o erro de treinamento, mas não

permitir que ocorra um super treinamento. A Figura 4.12 ilustra as curvas de erro de treinamento e validação do Combinador.

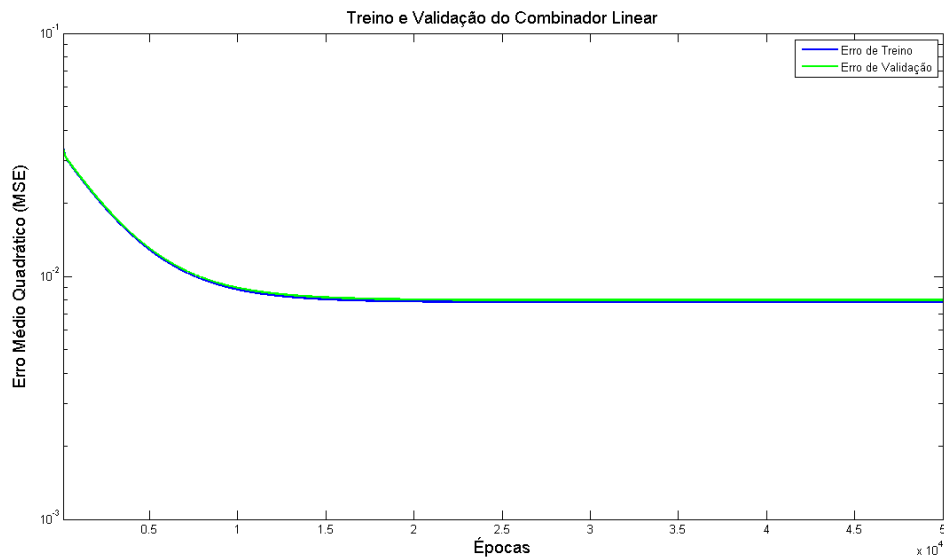


Figura 4.12 – Processo de treinamento e validação do Combinador

Os resultados encontrados como saída final da identificação estão ilustrados nas Figura 4.13 e Figura 4.14.

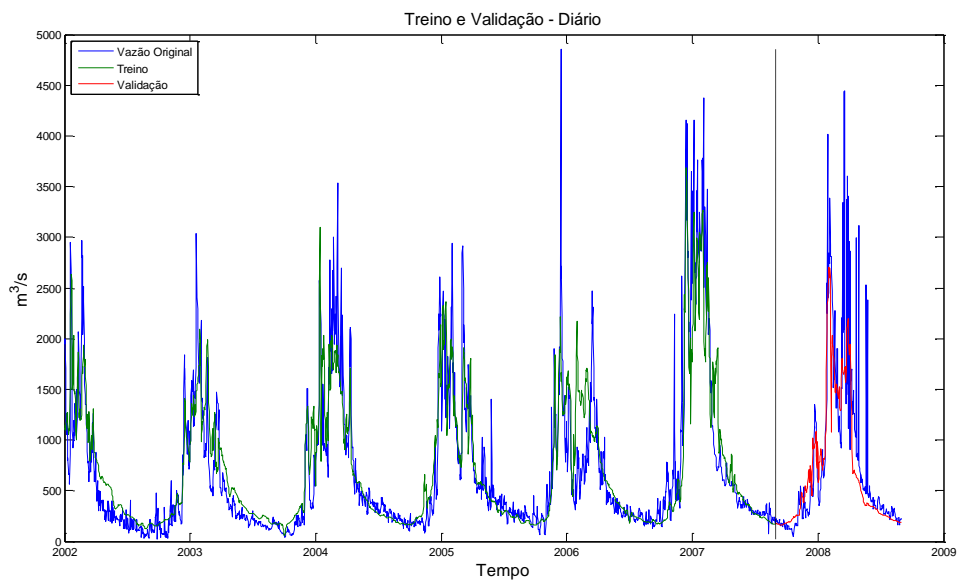


Figura 4.13 – Identificação do modelo pela Combinação dos métodos anteriores – em escala de tempo diária

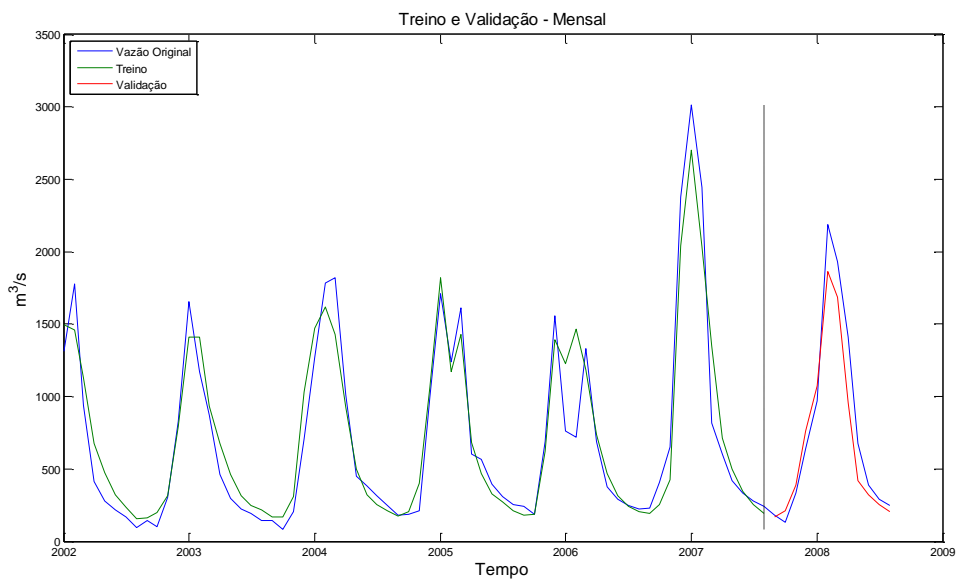


Figura 4.14 – Identificação do modelo pela Combinação dos métodos anteriores – em escala de tempo mensal

A Figura 4.15 e a Tabela 4.2 apresentam os erros médios quadráticos para o conjunto de validação de todos os métodos de identificação utilizados. Podemos perceber que em geral a combinação apresenta um erro menor, a cada mês, que um dos dois métodos que a compõe. E na média geral de todos os meses, o seu erro é menor que o dos outros dois.

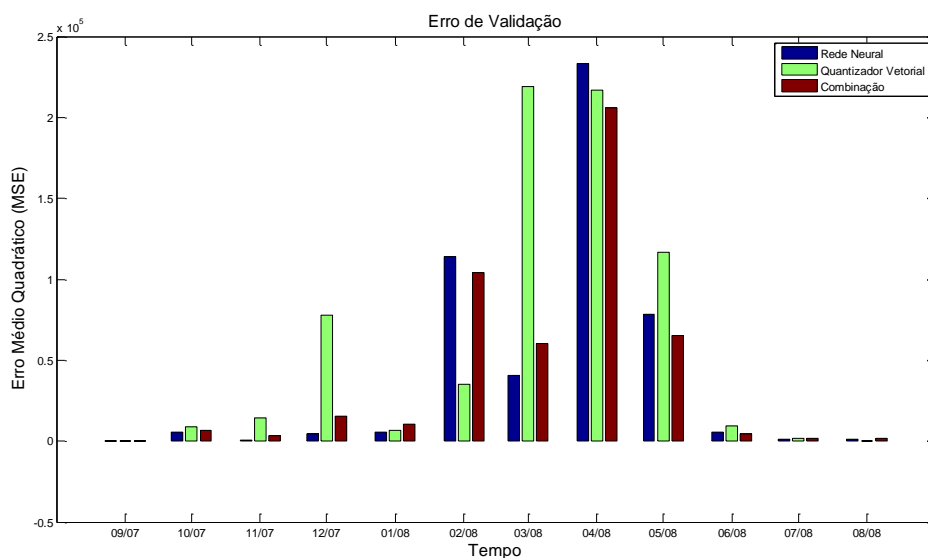


Figura 4.15 – Comparação dos Erros de Validação para todos os métodos

Tabela 4.2 – Comparação dos Erros de Validação para todos os métodos

Tempo	Rede Neural	Quantização Vetorial	Combinação
set/07	26,77	151,26	128,60
out/07	5485,43	9034,48	6566,74
nov/07	719,10	14164,66	3446,57
dez/07	4700,38	77810,79	15394,00
jan/08	5735,82	6814,60	10347,04
fev/08	114165,59	35022,84	104079,68
mar/08	40615,92	219343,01	60122,54
abr/08	233129,50	217076,39	205807,85
mai/08	78142,47	116826,93	65421,11
jun/08	5814,52	9473,03	4391,72
jul/08	1200,17	1557,88	1573,33
ago/08	1217,41	253,14	2001,99
Média	40912,76	58960,75	39940,10

4.6 Conclusões

Esse capítulo apresenta a identificação do modelo de previsão de vazões, até então desconhecido, pelos métodos de Redes Neurais e Quantização Vetorial, que compõe a primeira etapa do processo, e a combinação dos resultados encontrados, que é a segunda etapa.

Como foi visto, o resultado final encontrado apresenta um ganho quando comparado aos resultados encontrados por cada um dos dois métodos da primeira etapa separadamente.

5 Previsão da Vazão

5.1 Introdução

O capítulo anterior tratou da identificação do modelo desconhecido, deixando tudo pronto para que seja possível fazer a previsão da vazão, que será abordada neste capítulo. Assim como para a identificação, o processo é dividido em duas etapas: primeiro será feita a previsão pelos métodos de Redes Neurais e Quantização Vetorial e depois a combinação dos resultados para gerar a saída final.

A previsão da vazão pode ser obtida aqui devido à antecipação existente entre as variáveis explicativas e a vazão, que neste caso ocorrem com 188, 190, 195 e 200 dias de avanço. Em outras palavras, neste caso é possível termos uma previsão para a vazão com até 188 dias à frente.

Nesta altura da pesquisa, gostaríamos de poder analisar quantitativamente quão bom, regular ou ruim é a previsão que geramos. Por essa razão, propositadamente, foi deixado um período de um ano de vazão fora da busca por variáveis explicativas. Ou seja, os dados aqui utilizados estavam atualizados até setembro de 2009 e o período utilizado para a busca foi até o final de agosto de 2008. O período de um ano utilizado nesta análise crítica permite englobar tanto a época úmida como a seca.

5.2 A previsão utilizando a Rede Neural

A previsão de vazões utilizará a Rede Neural que foi treinada e apresentada na Seção 4.3, com pesos e *bias* já otimizados, e o conjunto de dados para teste, que até agora não foram incluídos neste trabalho.

Esse conjunto de dados é formado, em cada instante de tempo, pelas variáveis explicativas e pelos fasores de tempo (que são colocados caso tenham sido utilizados no treinamento). A cada instante de tempo a Rede pega as variáveis explicativas e os fasores, multiplica pelos pesos da primeira camada e adiciona o bias e utiliza a função de ativação, em cada um dos neurônios. As saídas destes neurônios passam pelo mesmo processo na segunda camada e assim sucessivamente até alcançar a última camada, cuja saída será a aproximação da vazão desejada.

Os resultados encontrados, em valores diários, para o conjunto de teste do rio São Francisco estão ilustrados, respectivamente, na Figura 5.1.

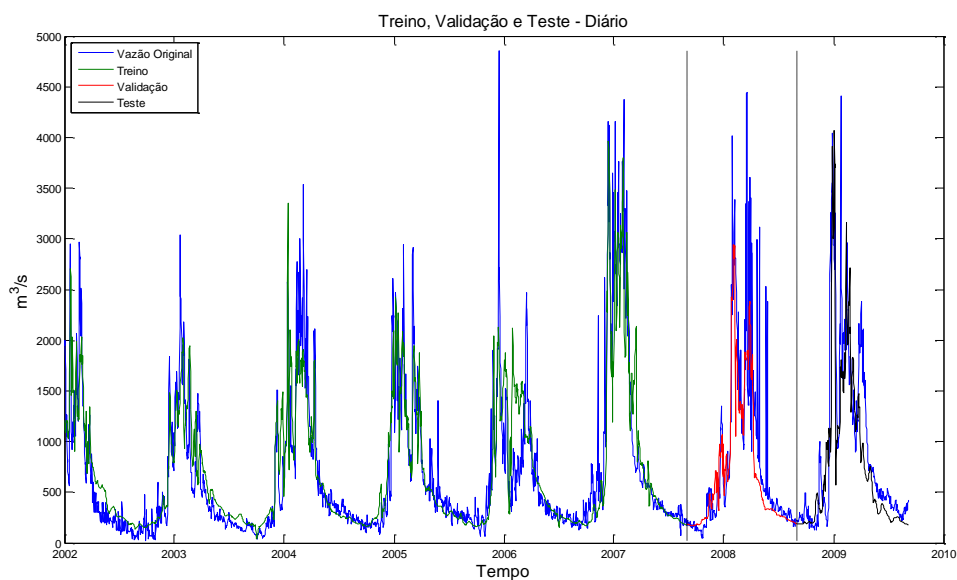


Figura 5.1 – Previsão pela Rede Neural – em escala de tempo diária

Para facilitar a visualização da qualidade da previsão, assim como já ilustrado nas identificações, foi feita uma média mensal dos dados da figura acima, conforme ilustrado na Figura 5.2.

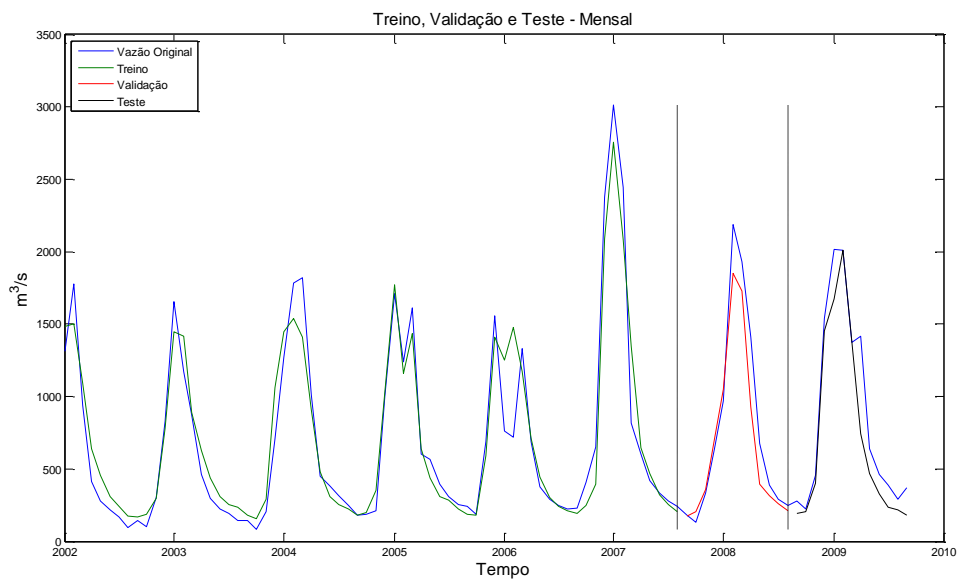


Figura 5.2 – Previsão pela Rede Neural – em escala de tempo mensal

Os resultados encontrados pela Rede Neural acompanham os valores de vazão com uma margem de erro pequena, sendo, portanto uma boa aproximação para valores futuros de previsão.

5.3 A Previsão pelo Quantizador Vetorial

A previsão pelo Quantizador Vetorial utilizará o mesmo conjunto de dados de entrada que foi empregado na Rede Neural da seção anterior e o dicionário que foi otimizado na Seção 4.4.

O conjunto de dados de teste formam, a cada instante de tempo, vetores compostos pelos dados de entrada e pelos fasores (que são colocados caso tenham sido utilizados no treinamento), conforme incluído na equação abaixo.

$$\underbrace{\begin{bmatrix} VE_{1(M+1)} & VE_{1(M+2)} & \cdots & VE_{1(M+M)} \\ VE_{2(M+1)} & VE_{2(M+2)} & \cdots & VE_{2(M+M)} \\ \vdots & \vdots & \ddots & \vdots \\ VE_{I(M+1)} & VE_{I(M+2)} & \cdots & VE_{I(M+M)} \\ F_{1(M+1)} & F_{1(M+2)} & \cdots & F_{1(M+M)} \\ F_{2(M+1)} & F_{2(M+2)} & \cdots & F_{2(M+M)} \end{bmatrix}}_{\text{Teste}} \quad (5.1)$$

Para cada instante de tempo encontra-se qual o centróide, sem utilizar a última linha, que apresente a menor distância ao vetor e atribui-se a este instante o valor desta última dimensão como aproximação da vazão. Repetindo-se este processo para todos os instantes deste conjunto, chegamos ao resultado apresentado na Figura 5.3.

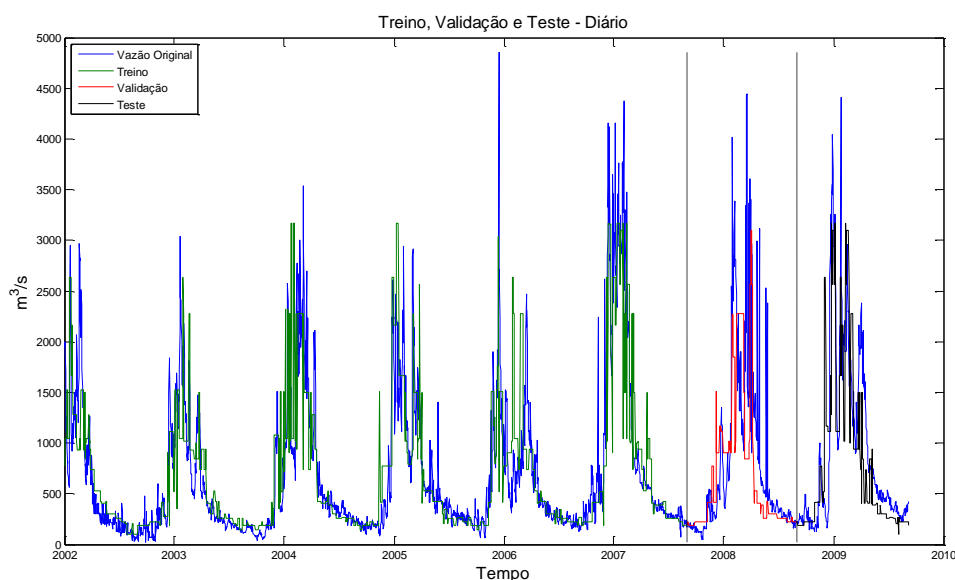


Figura 5.3 – Previsão pelo Quantizador Vetorial – em escala de tempo diária

Assim como para a Rede Neural, foi feita a média mensal destes dados de modo a melhorar a visualização da qualidade dos resultados encontrados, conforme ilustrados na Figura 5.4.

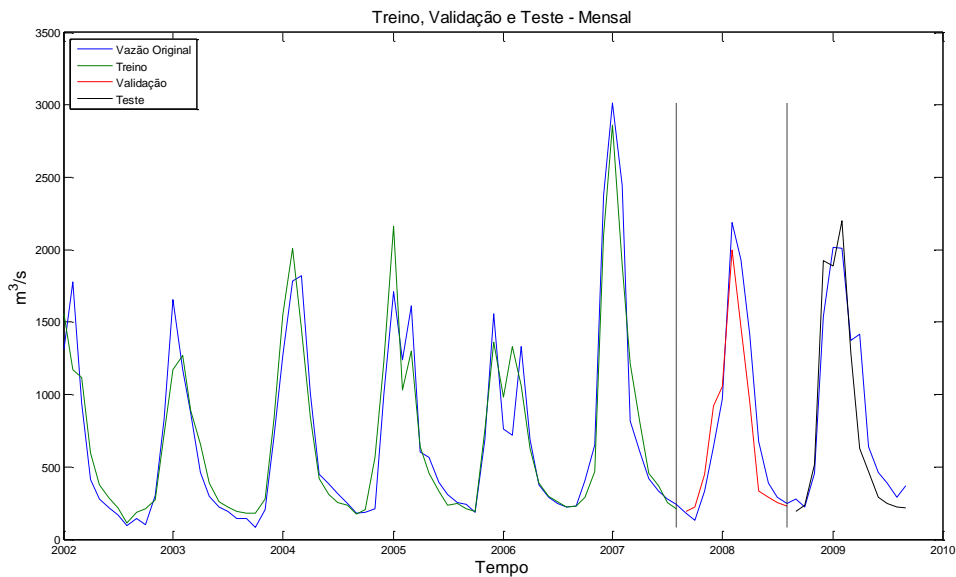


Figura 5.4 – Previsão pelo Quantizador Vetorial – em escala de tempo mensal

Os resultados apresentam uma previsão diferente da obtida na seção anterior, mas também com uma qualidade de aproximação que atende as expectativas.

5.4 Previsão por Combinação

Ao fim da primeira etapa do processo de previsão, temos dois resultados diferentes, ambos com boa qualidade. No intuito de melhorar ainda mais estes resultados, passamos à segunda etapa do processo: fazer a combinação da saída da Rede Neural com a do Quantizador Vetorial.

A combinação agora não mais utilizará as variáveis explicativas como entrada deste sistema, serão utilizados somente os resultados da primeira etapa e os fasores, caso julgados necessários.

A combinação será feita utilizando o Combinador (uma rede neural puramente linear), descrito na Seção 4.5 e os resultados encontrados, em escala de tempo diária,

que encontram-se ilustrados na Figura 5.5. A média mensal deste resultado encontra-se ilustrada na Figura 5.6.

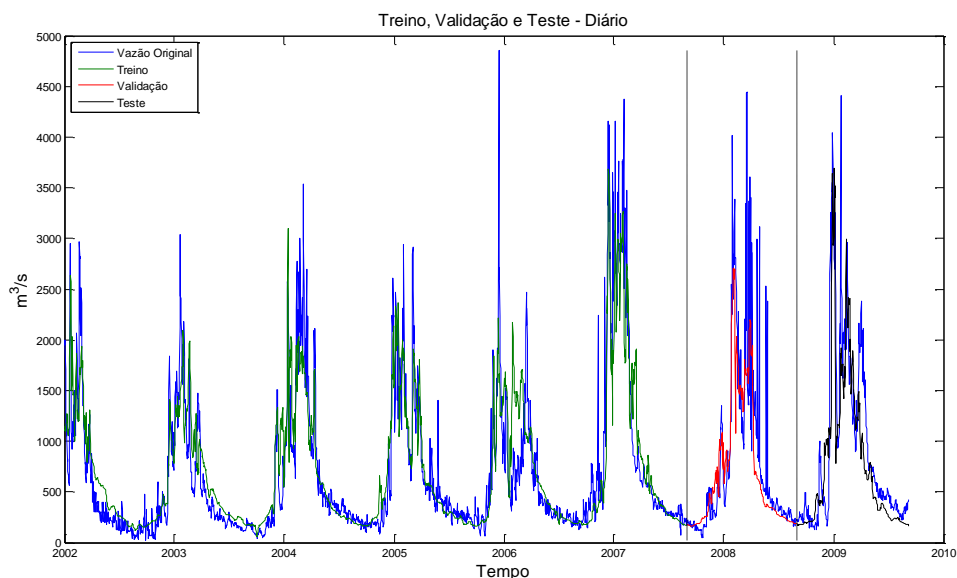


Figura 5.5 – Previsão pelo Combinação – em escala de tempo diária

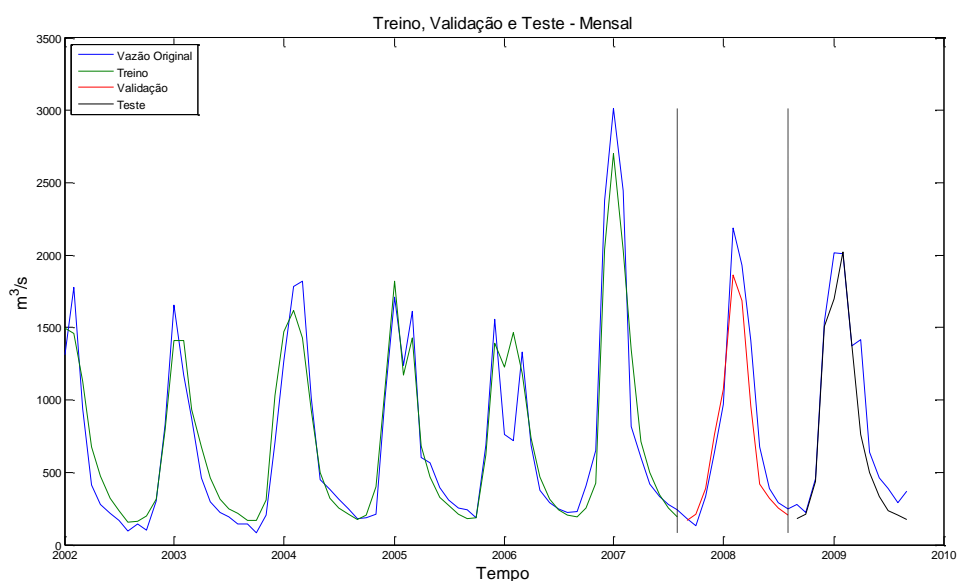


Figura 5.6 – Previsão pelo Combinação – em escala de tempo mensal

Os resultados finais encontrados pelo Combinador tiveram um ganho quando comparado aos resultados parciais encontrados ao final da primeira etapa, isto pode ser observado na Figura 5.7 e na Tabela 5.1.

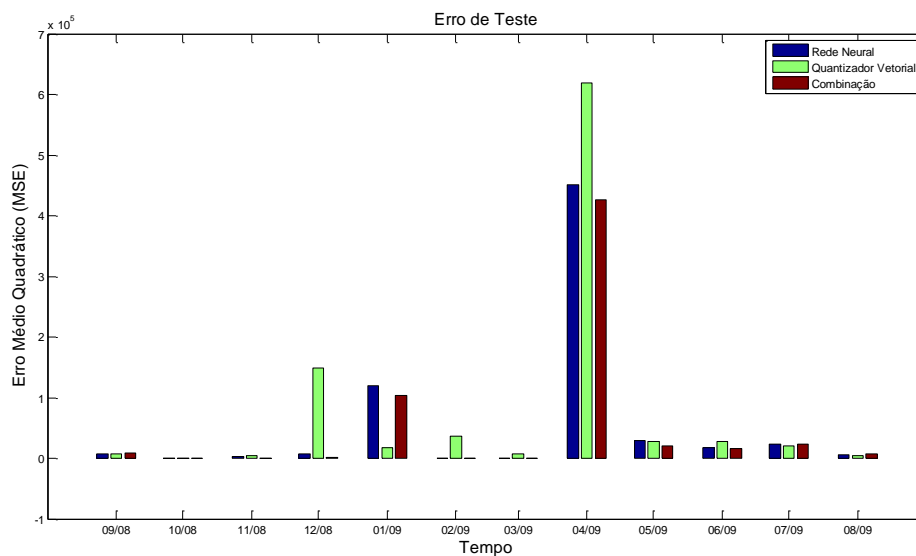


Figura 5.7 – Comparação dos Erros de Teste para todos os métodos

Tabela 5.1 – Comparação dos Erros de Validação para todos os métodos

Tempo	Rede Neural	Quantização Vetorial	Combinação
set/08	7206,44	6703,44	8893,42
out/08	392,04	156,48	135,81
nov/08	2767,38	4045,12	409,42
dez/08	7431,52	149369,31	1009,37
jan/09	119984,61	17065,01	103800,22
fev/09	0,14	36989,35	172,35
mar/09	26,23	7634,20	48,25
abr/09	450721,22	619061,33	426001,63
mai/09	29043,51	28135,19	19866,45
jun/09	17405,82	28375,92	15767,99
jul/09	23218,56	19969,29	23755,09
ago/09	5628,58	4672,11	7583,67
Média	55318,84	76848,06	50620,31

5.5 Conclusões

Este capítulo apresentou os resultados obtidos na previsão da vazão pela Rede Neural, Quantizador vetorial e o Combinador, descritos no capítulo anterior. Os resultados foram analisados para o conjunto de teste em comparação com a vazão referente a este período, que, propositadamente, até então não havia participado deste trabalho.

Ao final foram analisados os erros de cada uma das modelagens e mostrou-se que o resultado obtido pelo Combinador apresenta um ganho quando comparado aos resultados parciais do processo completo.

6 Conclusões

As previsões de vazões no Brasil apresentam uma importância muito grande devido à sua grande participação na geração de energia elétrica por meios das usinas hidrelétricas. Contudo os métodos mais usuais têm sido incapazes de prever eventos que podem trazer efeitos graves sobre todo o país, acarretando situações extremas, como por exemplo, o racionamento de 2001. Isso devido ao fato de que esses modelos, ou fazem uso somente do passado destas vazões, ou de previsão de variáveis físicas, mais especificamente chuva, para fazer a previsão da vazão propriamente dita, e com isso não conseguem se adaptar às mudanças climatológicas que estamos presenciando.

Esse trabalho apresentou uma metodologia de previsão que não faz uso do histórico da série como única fonte para a previsão, ou da previsão de variáveis físicas, e sim um modelo matemático que identifica as relações existentes entre variáveis físicas realizadas (como por exemplo, temperatura do ar, OLR (*Outgoing Longwave Radiation*) e pressão) e a vazão do rio que se deseja prever. A previsão é possível devido à antecipação das variáveis físicas – ditas variáveis explicativas ao longo do texto – com relação à série de vazão.

Inicialmente, no Capítulo 2, foram vistos os conceitos climatológicos, teleconexão e propagação de eventos por todo o globo terrestre, que fundamentam a identificação do modelo desconhecido descrito no mesmo capítulo. Ainda neste capítulo foram mencionados os principais tipos de previsão de vazões utilizados no setor elétrico.

Posteriormente, já no Capítulo 3, foi feita uma revisão dos conceitos matemáticos de Redes Neurais e Quantização Vetorial que foram utilizados na

identificação e previsão do modelo desconhecido já descrito, respectivamente, nos Capítulos 4 e 5.

Esses capítulos descreveram detalhadamente o processo de identificação e previsão da vazão através das variáveis explicativas, que foram divididos em duas etapas. A primeira etapa fez uso das modelagens de Redes Neurais e de Quantização Vetorial que têm, cada uma, como saída uma estimativa da vazão; na segunda etapa foi feita uma combinação desses resultados e a saída final do modelo.

Ao final desses capítulos foi feita uma comparação (Figuras 4.15 e 5.7 e Tabelas 4.1 e 5.1) dos resultados encontrados pelos modelos utilizando redes neurais, quantização vetorial e pela combinação. Utilizando o erro médio quadrático foi constatado que, mesmo com os erros satisfatórios encontrados pelos modelos da primeira etapa, a Combinação reduziu ainda mais estes erros.

6.1 Contribuições

As principais contribuições deste trabalho foram:

- Explorar os conceitos de climatologia como fundamentos para o mapeamento das relações existentes entre as vazões e as variáveis físicas ao longo de todo o globo.
- Utilização destas variáveis como entrada de um modelo de vazão, permitindo previsão da mesma utilizando antecipação temporal, ou seja, eventos específicos, sejam teleconectados ou propagados ao longo do globo, afetam futura vazão.

- Previsão através do conceito de Quantização Vetorial, largamente estudado em compressão de dados, mas pouco utilizado por seu potencial como preditor.

6.2 Sugestões para Trabalhos Futuros

As principais sugestões de trabalhos futuros são:

- Estudar o mapeamento não linear das variáveis explicativas na vazão e metodologias de busca destas variáveis na superfície do globo terrestre.
- Estudar os conceitos climatológicos de teleconexão e propagação que expliquem as mudanças recentes que estamos sofrendo, tanto na vazão como também na própria climatologia.
- Busca de intervalos de validade para os modelos e/ou sua variabilidade no tempo ou como função de outros fenômenos.
- Desenvolver técnicas que separem os efeitos de propagação e teleconexão.

A Apêndice

A.1 Regra da Cadeia

A Regra da Cadeia (Malta, Pesco, & Lopes, 2002) trata da diferenciabilidade de uma função composta a partir da diferenciabilidade das funções componentes, ou seja:

$$y = f(x) \quad \text{A.1}$$

$$x = g(u) \quad \text{A.2}$$

$$\frac{dy}{du} = \frac{dy}{dx} \frac{dx}{du} \quad \text{A.3}$$

A.2 Derivada da função Tangente Hiperbólica

A derivada da função tangente hiperbólica é dada por:

$$\varphi(u) = \tanh(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}} \quad \text{A.4}$$

$$\frac{d\varphi}{du} = \frac{(e^u + e^{-u})^2 - (e^u - e^{-u})^2}{(e^u + e^{-u})^2} = 1 - \tanh^2(u) \quad \text{A.5}$$

$$\frac{d\varphi}{du} = 1 - \varphi^2 \quad \text{A.6}$$

Portanto conforme podemos perceber nas equações (A.4)-(A.6), a derivada da função tangente hiperbólica é dada por 1 menos a própria função ao quadrado.

A.3 O Produto de Kronecker

O produto de Kronecker (Steeb & Shi, 1997) é uma operação entre duas matrizes de quaisquer tamanhos e o resultado é uma matriz com tamanho igual ao produto do tamanhos das matrizes iniciais. Por exemplo, tendo uma matriz **A** m x n e uma matriz **B** p x q, a matriz de saída será (mp) x (nq).

$$\begin{bmatrix} 0 & 2 \\ 3 & 4 \end{bmatrix} \odot \begin{bmatrix} 1 & 2 & 3 \\ 5 & 7 & 13 \end{bmatrix} = \begin{bmatrix} 0 \cdot 1 & 0 \cdot 2 & 0 \cdot 3 & 2 \cdot 1 & 2 \cdot 2 & 2 \cdot 3 \\ 0 \cdot 5 & 0 \cdot 7 & 0 \cdot 13 & 2 \cdot 5 & 2 \cdot 7 & 2 \cdot 13 \\ 3 \cdot 1 & 3 \cdot 2 & 3 \cdot 3 & 4 \cdot 1 & 4 \cdot 2 & 4 \cdot 3 \\ 3 \cdot 5 & 3 \cdot 7 & 3 \cdot 13 & 4 \cdot 5 & 4 \cdot 7 & 4 \cdot 13 \end{bmatrix} \quad \text{A.7}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 2 & 4 & 6 \\ 0 & 0 & 0 & 10 & 14 & 26 \\ 2 & 6 & 9 & 4 & 8 & 12 \\ 15 & 21 & 39 & 20 & 28 & 52 \end{bmatrix}$$

A.4 Equacionamento de um perceptron de duas camadas

O equacionamento de um perceptron de duas camadas é uma extensão do equacionamento de uma camada. A Figura A.6.1 ilustra um perceptron de duas camadas.

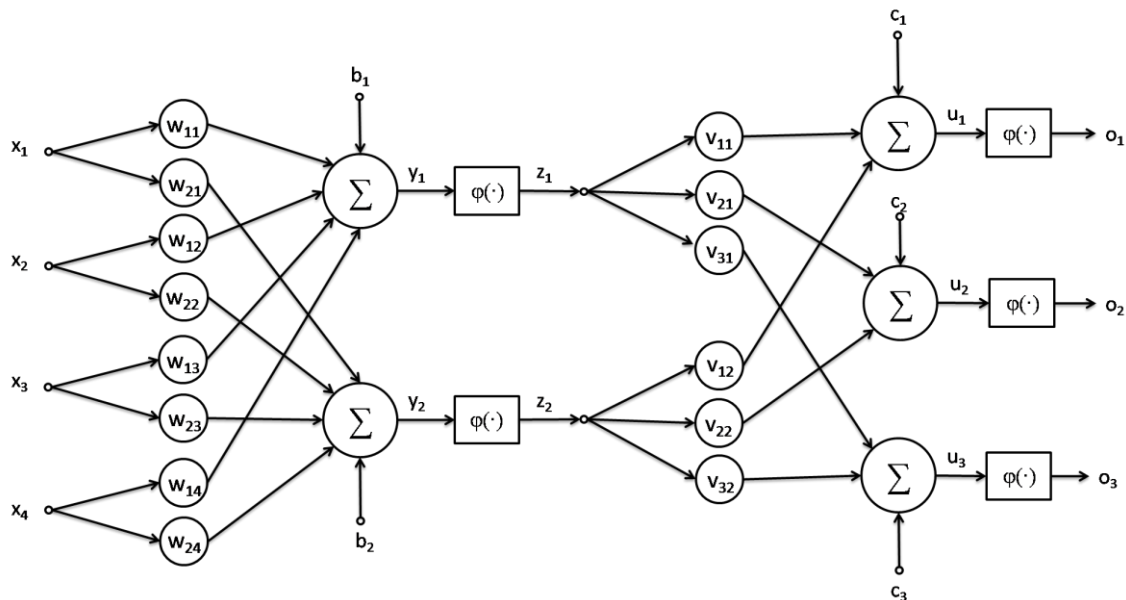


Figura A.6.1 – Perceptron de duas camadas

$$y_i = \left(\sum_j w_{ij} x_j \right) + b_i \quad \text{A.8}$$

$$z_i = \varphi(y_i) \quad \text{A.9}$$

$$u_k = \left(\sum_p v_{kp} z_p \right) + c_k \quad \text{A.10}$$

$$o_k = \varphi(u_k) \quad \text{A.11}$$

$$e_k = t_k - o_k \quad \text{A.12}$$

$$J = \frac{\sum_k e_k^2}{2} \quad \text{A.13}$$

onde:

$j = 1, \dots, Q$ é o número de entradas na primeira camada

$i = 1, \dots, M$ é o número de neurônios na primeira maçada

$p = 1, \dots, M$ é o número de entradas na segunda camada

$k = 1, \dots, N$ é o número de neurônios na segunda camada

As equações (A.8)-(A.13) descrevem matematicamente a Figura A.6.1. Elas podem ser escritas matricialmente da seguinte forma:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1j} & \cdots & w_{1Q} \\ w_{21} & w_{22} & \cdots & w_{2j} & \cdots & w_{2Q} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{i1} & w_{i2} & \cdots & w_{ij} & \cdots & w_{iQ} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{M1} & w_{M2} & \cdots & w_{Mj} & \cdots & w_{MQ} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_Q \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_i \\ \vdots \\ b_M \end{bmatrix} \quad \text{A.14}$$

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b} \quad \text{A.15}$$

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_i \\ \vdots \\ z_M \end{bmatrix} = \begin{bmatrix} \varphi(y_1) \\ \varphi(y_2) \\ \vdots \\ \varphi(y_i) \\ \vdots \\ \varphi(y_M) \end{bmatrix} \quad \text{A.16}$$

$$\mathbf{z} = \varphi(\mathbf{y}) \quad \text{A.17}$$

$$\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \\ \vdots \\ u_N \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1p} & \cdots & v_{1M} \\ v_{21} & v_{22} & \cdots & v_{2p} & \cdots & v_{2M} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_{k1} & v_{k2} & \cdots & v_{kp} & \cdots & v_{kM} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_{N1} & v_{N2} & \cdots & v_{Np} & \cdots & v_{NM} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_p \\ \vdots \\ z_M \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \\ \vdots \\ c_N \end{bmatrix} \quad \text{A.18}$$

$$\mathbf{u} = \mathbf{V}\mathbf{z} + \mathbf{c} \quad \text{A.19}$$

$$\begin{bmatrix} o_1 \\ o_2 \\ \vdots \\ o_k \\ \vdots \\ o_N \end{bmatrix} = \begin{bmatrix} \varphi(u_1) \\ \varphi(u_2) \\ \vdots \\ \varphi(u_k) \\ \vdots \\ \varphi(u_N) \end{bmatrix} \quad \text{A.20}$$

$$\mathbf{o} = \varphi(\mathbf{u}) \quad \text{A.21}$$

$$\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \\ \vdots \\ e_N \end{bmatrix} = \begin{bmatrix} t_1 - o_1 \\ t_2 - o_2 \\ \vdots \\ t_k - o_k \\ \vdots \\ t_N - o_N \end{bmatrix} \quad \text{A.22}$$

$$\mathbf{e} = \mathbf{t} - \mathbf{o} \quad \text{A.23}$$

$$J = \frac{\mathbf{e}^T \mathbf{e}}{2} \quad \text{A.24}$$

As equações abaixo descrevem as derivadas parciais da função custo em relação a cada bias e pesos das duas camadas.

$$\frac{\partial J}{\partial c_k} = \frac{\partial J}{\partial e_k} \frac{\partial e_k}{\partial o_k} \frac{\partial o_k}{\partial u_k} \frac{\partial u_k}{\partial c_k} = -e_k(1 - o_k^2) \quad \text{A.25}$$

$$\frac{\partial J}{\partial v_{kp}} = \frac{\partial J}{\partial e_k} \frac{\partial e_k}{\partial o_k} \frac{\partial o_k}{\partial u_k} \frac{\partial u_k}{\partial v_{kp}} = -e_k(1 - o_k^2)z_p \quad \text{A.26}$$

$$\begin{aligned} \frac{\partial J}{\partial b_i} &= \sum_k \frac{\partial J}{\partial e_k} \frac{\partial e_k}{\partial o_k} \frac{\partial o_k}{\partial u_k} \frac{\partial u_k}{\partial z_i} \frac{\partial z_i}{\partial y_i} \frac{\partial y_i}{\partial b_i} = \sum_k -e_k(1 - o_k^2)v_{ki}(1 - z_i^2) \\ &= -(1 - z_i^2) \sum_k v_{ki}(1 - o_k^2)e_k \end{aligned} \quad \text{A.27}$$

$$\begin{aligned} \frac{\partial J}{\partial w_{ij}} &= \sum_k \frac{\partial J}{\partial e_k} \frac{\partial e_k}{\partial o_k} \frac{\partial o_k}{\partial u_k} \frac{\partial u_k}{\partial z_i} \frac{\partial z_i}{\partial y_i} \frac{\partial y_i}{\partial w_{ij}} = \sum_k -e_k(1 - o_k^2)v_{ki}(1 - z_i^2)x_j \\ &= -x_j(1 - z_i^2) \sum_k v_{ki}(1 - o_k^2)e_k \end{aligned} \quad \text{A.28}$$

Em notação matricial as equações A.25-A.28 se reduzem à:

$$\frac{\partial J}{\partial \mathbf{c}} = \frac{\partial \mathbf{u}}{\partial \mathbf{c}} \left(\frac{\partial \mathbf{o}}{\partial \mathbf{u}} \frac{\partial \mathbf{e}}{\partial \mathbf{o}} \frac{\partial J}{\partial \mathbf{e}} \right) = -(I - \text{diag}^2(\mathbf{o}))\mathbf{e} \quad \text{A.29}$$

$$\frac{\partial J}{\partial \mathbf{V}} = \frac{\partial \mathbf{u}}{\partial \mathbf{V}} \odot \left(\frac{\partial \mathbf{o}}{\partial \mathbf{u}} \frac{\partial \mathbf{e}}{\partial \mathbf{o}} \frac{\partial J}{\partial \mathbf{e}} \right) = -\mathbf{z}^T \odot (I - \text{diag}^2(\mathbf{o}))\mathbf{e} \quad \text{A.30}$$

$$\frac{\partial J}{\partial \mathbf{b}} = \frac{\partial \mathbf{y}}{\partial \mathbf{b}} \left[\frac{\partial \mathbf{z}}{\partial \mathbf{y}} \frac{\partial \mathbf{u}}{\partial \mathbf{z}} \left(\frac{\partial \mathbf{o}}{\partial \mathbf{u}} \frac{\partial \mathbf{e}}{\partial \mathbf{o}} \frac{\partial J}{\partial \mathbf{e}} \right) \right] = -(I - \text{diag}^2(\mathbf{z}))\mathbf{V}^T(I - \text{diag}^2(\mathbf{o}))\mathbf{e} \quad \text{A.31}$$

$$\frac{\partial J}{\partial \mathbf{W}} = \frac{\partial \mathbf{y}}{\partial \mathbf{W}} \odot \left[\frac{\partial \mathbf{z}}{\partial \mathbf{y}} \frac{\partial \mathbf{u}}{\partial \mathbf{z}} \left(\frac{\partial \mathbf{o}}{\partial \mathbf{u}} \frac{\partial \mathbf{e}}{\partial \mathbf{o}} \frac{\partial J}{\partial \mathbf{e}} \right) \right] = -\mathbf{x}^T \odot (I - \text{diag}^2(\mathbf{z}))\mathbf{V}^T(I - \text{diag}^2(\mathbf{o}))\mathbf{e} \quad \text{A.32}$$

7 Referências Bibliográficas

- Ambrizzi, T. (2003). El Niño/Oscilação Sul e Teleconexões atmosféricas no Hemisfério Austral. *Tese de Livre Docência*, 201.
- Costa, F. d., Maceira, M. E., & Damázio, J. M. (2007). Modelos de Previsão Hidrológica Aplicados ao Planejamento da Operação do Sistema Elétrico Brasileiro. *Workshop Previsão de Vazões*. Rio de Janeiro: Operador Nacional do Sistema Elétrico - ONS.
- Drumond, A. R., & Ambrizzi, T. (Julho de 2003). Estudo observacional e numérico de variação da circulação atmosférica nas américas em episódios extremos da oscilação sul. *Revista Brasileira de Meteorologia*, v.18 (n.1), pp. 1-12.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification* (2nd ed.). Wiley.
- Gersho, A., & Gray, R. M. (1992). *Vector Quantization and Signal Compression*. Springer.
- Gomes, J. G. (2007). Fundamentos de Redes Neurais (Mini-Curso). *VIII Congresso Brasileiro de Redes Neurais (CBRN 2007)*. Florianópolis/SC.
- Guilhon, L. G., & Rocha, V. F. (2008). Aplicação de Modelos Chuva-Vazão no Sistema Interligado Nacional - SIN. *II Simpósio de Recursos Hídricos do Sul Sudeste*. Rio de Janeiro: ABRH.
- Hassoun, M. H. (1995). *Fundamentals of Artificial Neural Networks*. Cambridge: The MIT Press.
- Haykin, S. (1996). *Adaptive Filter Theory* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation* (2nd ed.). Prentice Hall.
- Kohonen, T. (2001). *Self-Organizing Maps* (3rd ed.). Berlin: Springer-Verlag.
- LabMet. (s.d.). *El Niño e La Niña*. Acesso em Setembro de 2009, disponível em <http://www.nemrh.uema.br/meteoro/elnino.htm>
- Macêdo, L. H. (2006). O Conceito da Inovação: Uma Ferramenta para a Modelagem, Análise e Compreensão de Fenômenos Climatológicos. D.SC. thesis, PUC-Rio.
- Macêdo, L. H., Szczupak, J., & Pinto, L. (2007). Climatological Analysis and Prediction: a Nonlinear Approach. *8-th International Symposium on Signals, Circuits and Systems, ISSCS2007*. Iasi, Romania.
- Malta, I., Pesco, S., & Lopes, H. (2002). *Cálculo á Uma Variável* (Vol. II). Rio de Janeiro: Editora PUC-Rio.
- Mitra, S. K. (1998). *Digital Signal Processing – A Computer-Based Approach*. McGraw-Hill.
- Müller, G. V., Ferraz, S. T., & Ambrizzi, T. (Março de 2009). Propagação das ondas de Rossby nos invernos de máxima frequência de ocorrência de geadas na Pampa úmida. *Revista Brasileira de Meteorologia*, v.24 (n.1), pp. 56-62.
- Noaa - *Physical Science Division*. (2009, Setembro). (Noaa/Nasa) Retrieved Setembro 2009, from Earth System Research Laboratory: www.cdc.noaa.gov
- Pinto, L., Szczupak, J., Macêdo, L. H., Ambrizzi, T., & Molina, E. (2004). Previsão de Cenários Hidrológicos: Quebrando o Paradigma da Incerteza. *SEPOPE*. Rio de Janeiro.
- Quintão, M. T., Souza, S. A., Costa, F. d., & Damázio, J. M. (2008). Uso dos Modelos PREVIVAZM E PREVIVAZ para previsão de afluências mensais aos

aproveitamentos hidroelétricos Tucuruí e Curuá-Una. *II Simpósio de Recursos Hídricos do Sul-Sudeste*. Rio de Janeiro: ABRH.

Steeb, W.-H., & Shi, T. K. (1997). *Matrix Calculus And Kronecker Product With Applications and C++ Programs* (1st ed.). World Scientific Publishing.

Vapnik, V. N. (1998). *Statistical Learning Theory*. New York: Wiley.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *IRE WESCON Convention Record*, (pp. 96-104).