



PREVISÃO DAS VAZÕES AFLUENTES DIÁRIAS POR REDE NEURAL PARA A OPERAÇÃO DO SISTEMA INTERLIGADO NACIONAL

Natalia Dias Sardinha Mendes

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: Luiz Pereira Calôba

Maurício Tiomno Tolmasquim

Rio de Janeiro

Agosto de 2019

PREVISÃO DAS VAZÕES AFLUENTES DIÁRIAS POR REDE NEURAL PARA A
OPERAÇÃO DO SISTEMA INTERLIGADO NACIONAL

Natalia Dias Sardinha Mendes

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. Maurício Tiomno Tolmasquim, D.Sc.

Prof. Amaro Olímpio Pereira Jr, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

AGOSTO DE 2019

Mendes, Natalia Dias Sardinha

Previsão das Vazões Afluentes Diárias por Rede Neural para Projeção de PLD / Natalia Dias Sardinha Mendes. – Rio de Janeiro: UFRJ/COPPE, 2019.

XV, 80 p.: il.; 29,7 cm.

Orientadores: Luiz Pereira Calôba

Maurício Tiomno Tolmasquim

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia Elétrica, 2019.

Referências Bibliográficas: p. 76-80.

1. Previsão das Vazões Afluentes Diárias. 2. Rede Neural Artificial. 3. Projeção de PLD. I. Calôba, Luiz Pereira *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

A minha família, Fabiano, Julia e Artur.

AGRADECIMENTOS

Ao meu marido e aos filhos pelo amor, apoio e compreensão no decorrer dos quase três anos de dedicação ao mestrado.

Aos meus queridos pais, por terem se empenhado na minha formação educacional, além de terem transparecido a sua grande importância na vida humana.

Ao meu orientador e co-orientador, com os quais tive o imenso privilégio de trabalhar e de tentar adquirir um pouco de seus grandes conhecimentos. A participação deles foi fundamental não apenas na elaboração dessa dissertação, mas na minha formação profissional.

Aos especiais amigos da COPPE, que desde o início me acolheram e dividiram seus conhecimentos para a elaboração dessa dissertação.

Finalizo agradecendo a Deus, pela minha saúde e por tudo que Ele me permitiu construir na minha vida.

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

PREVISÃO DAS VAZÕES AFLUENTES DIÁRIAS POR REDE NEURAL PARA A OPERAÇÃO DO SISTEMA INTERLIGADO NACIONAL

Natalia Dias Sardinha Mendes

Agosto/2019

Orientadores: Luiz Pereira Calôba

Maurício Tiomno Tolmasquim

Programa: Engenharia Elétrica

A previsão de vazões afluentes é um dos principais insumos na formação do Custo Mensal de Operação - CMO do sistema elétrico nacional, bem como, do Preço de Liquidação das Diferenças - PLD, sendo assim, esta área é desafiadora em relação à complexidade na sua previsibilidade, além da possibilidade de gerar altos lucros financeiros para os agentes do mercado energético.

Esse trabalho teve o objetivo de calibrar um modelo de previsão de vazões afluentes diárias na usina hidrelétrica de Três Marias, na bacia do São Francisco, 7 dias à frente, utilizando a técnica de redes neurais, considerando as vazões afluentes naturais do próprio reservatório da usina, e as informações pluviométricas e fluviométricas advindas de postos hidrométricos a montante da região de estudo.

Ressalta-se que o modelo de previsão desta dissertação atingiu seu objetivo, pois produziu resultados superiores aos modelos utilizados pelo sistema elétrico, tais como o tradicional estatístico PREVIVAZH e, até mesmo, ao modelo de rede neural, Neuro3M que foi usado até o início deste ano no mesmo reservatório.

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

DAILY AFLUENT FLOW FORECASTS BY NEURAL NETWORK FOR PLD
PROJECTION

Natalia Dias Sardinha Mendes

August/2019

Advisors: Luiz Pereira Calôba

Maurício Tiomno Tolmasquim

Department: Electrical Engineering

The prediction of affluent flows is one of the main inputs in the definition of the Monthly Operating Cost - CMO of the Brazilian electrical system, as well as, the Market Clearing Price – PLD. Therefore the prediction of affluent flows is a challenging area in relation to the complexity of problems, besides the possibility of generating high financial profits for the energy market agents.

This work aimed to calibrate a daily affluent flow forecasting model at the Três Marias hydroelectric plant in the São Francisco basin, 7 days ahead, using the neural networks technique, considering the natural affluent flows from the plant's own reservoir, and pluviometric and fluviometric information from hydrometric stations upstream of the study region.

The prediction model of this paper using the neural network technique produced better results than the models used by the electrical system, such as the PREVIVAZH traditional statistician and even the Neuro3M neural network model, which was used until the beginning of this year in the same reservoir.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	<i>Aspectos Gerais</i>	<i>1</i>
1.2	<i>Justificativas</i>	<i>1</i>
1.3	<i>Objetivos.....</i>	<i>2</i>
1.4	<i>Estrutura do Trabalho.....</i>	<i>3</i>
2	O SISTEMA ELÉTRICO BRASILEIRO.....	5
2.1	<i>Descrição Física do Sistema</i>	<i>5</i>
2.2	<i>A Cadeia de Modelos de Otimização</i>	<i>6</i>
2.2.1	NEWAVE.....	7
2.2.2	DECOMP	9
2.2.3	Modelos de Previsão de Vazão Diária.....	11
2.3	<i>Planejamento e Programação da Operação.....</i>	<i>12</i>
2.4	<i>Sensibilidade das Vazões Naturais Afluentes sobre o Preço de Liquidação das Diferenças.....</i>	<i>13</i>
2.4.1	Custo Marginal de Operação	13
2.4.2	Preço de Liquidação das Diferenças.....	16
3	REVISÃO DA LITERATURA	18
3.1	<i>Escolha do melhor algoritmo de previsão.....</i>	<i>18</i>
3.2	<i>Séries Temporais</i>	<i>20</i>
3.2.1.	Decomposição de Séries Temporais.....	22
3.2.2.	Tendência.....	23
3.2.3.	Sazonalização e Ciclos senoidais	24
3.2.4.	Preditor Não Linear	25
3.3	<i>Redes Neurais Artificiais.....</i>	<i>26</i>
3.3.1.	Função de Ativação	28
3.3.2.	Normalização dos Dados	29
3.3.3.	Funções Custo	30

3.3.4.	Configuração	32
3.3.5.	Rede Multi Layer Perceptron com Backpropagation	33
3.3.6.	Rede Multi Layer Perceptron com Levenberg-Marquardt	34
3.3.7.	Rede Non-linear Sigmoidal Regression Blocks Networks.....	35
4	DESCRIÇÃO DO ESTUDO DE CASO	37
4.1	<i>Bacia do São Francisco</i>	37
4.2	<i>Previsão de Vazões Afluentes Naturais em Três Marias</i>	39
4.3	<i>Postos Fluviométricos e Pluviométricos</i>	41
5	METODOLOGIAS E APLICAÇÕES	44
5.1	<i>Base de Dados</i>	44
5.1.1	Informações Pluviométricas e Fluviométricas	48
5.1.2	Decomposição da Série Temporal.....	50
5.1.3	Análise Estatística	53
5.1.4	Normalização.....	55
5.2	<i>Treinamento da Rede Neural Artificial</i>	56
6	ANÁLISES E RESULTADOS	59
6.1	<i>Análises do Pré-processamento</i>	59
6.2	<i>Análises das Simulações</i>	66
7	CONCLUSÕES E RECOMENDAÇÕES	74
8	BIBLIOGRAFIA	76

ÍNDICE DE FIGURAS

Figura 1 – Acoplamento de Médio Prazo e de Curto Prazo	9
Figura 2 – Participação dos modelos hidrológicos na cadeia do planejamento da operação do SIN	11
Figura 3 – Esquemático do PMO.....	13
Figura 4 – Dilema da Operação no Despacho Hidrotérmico.....	14
Figura 5 – Curvas de Função de Custo Futuro e Imediato	15
Figura 6 – Decomposição do PLD.....	17
Figura 7 – Modelo básico de neurônio	27
Figura 8 – Rede MLP.....	28
Figura 9 – Rede NSRBN.....	35
Figura 10 – Localização das sub-bacias da região do Alto São Francisco	37
Figura 11 – Disposição dos aproveitamentos da Bacia do São Francisco.....	40
Figura 12 – Sub-bacias da área de drenagem de Três Marias.....	43
Figura 13 – Vazões históricas do reservatório de Três Marias do ano 2005	45
Figura 14 – Vazões históricas dos postos fluviométricos à montante do ano 2005	45
.....	45
Figura 15 – Vazões históricas do reservatório de Três Marias do Período Seco 2006	47
Figura 16 – Vazões históricas do reservatório de Três Marias do Período Úmido 2005-2006.....	47
Figura 17 – Vazões históricas do reservatório de Três Marias do Período Úmido 2006-2007.....	48
Figura 18 – Logaritmo neperiano das vazões históricas do reservatório de Três Marias do Período Seco 2006.....	49
Figura 19 – Logaritmo neperiano das vazões históricas do reservatório de Três Marias do Período Úmido 2005-2006	49
Figura 20 – Valores médios dos postos pluviais do Período Seco 2006	50
Figura 21 – Valores médios dos postos pluviais do Período Úmido 2005-2006..	50
Figura 22 – Vazões históricas sem Tendência do reservatório de Três Marias do Período Seco 2006.....	52

Figura 23 – Vazões históricas sem Média do reservatório de Três Marias do Período Úmido 2005-2006	52
Figura 24 – Valores médios dos postos pluviais sem Média do Período Úmido 2005-2006.....	53
Figura 25 – Vazões históricas normalizadas do reservatório de Três Marias do Período Seco 2006.....	56
Figura 26 – FAC da Vazão do Reservatório de Três Marias do Período Seco 2006	59
Figura 27 – FAC da Vazão do Reservatório de Três Marias do Período Úmido 2005-2006.....	60
Figura 28 – FAC dos Valores médios dos postos pluviais do Período Úmido 2005-2006	60
Figura 29 – FACP da Vazão do Reservatório de Três Marias do Período Seco 2006	61
Figura 30 – FCC da Vazão de Indaiá com a vazão do Reservatório de Três Marias do Período Úmido 2005-2006	62
Figura 31 – FCC da Vazão de Porto das Andorinhas com a vazão do Reservatório de Três Marias do Período Úmido 2005-2006	62
Figura 32 – FCC dos valores médios dos Postos Pluviais com a vazão do Reservatório de Três Marias do Período Úmido 2005-2006.....	63
Figura 33 – Ciclos senoidais da Vazão do Reservatório de Três Marias dos Períodos Secos.....	64
Figura 34 – Espectro médio de frequências da Vazão do Reservatório de Três Marias do Período Seco 2006.....	65
Figura 35 – Espectograma da Vazão do Reservatório de Três Marias do Período Seco 2006	65
Figura 36 – Previsão de 7 dias da Vazão do Reservatório de Três Marias do Período Seco 2006.....	69
Figura 37 – Previsão de 7 dias da Vazão do Reservatório de Três Marias do Período Úmido 2005-2006	70
Figura 38 – Histograma do erro de Previsão de 7 dias da Vazão do Reservatório de Três Marias do Período Seco 2006.....	70
Figura 39 – Histograma do erro de Previsão de 7 dias da Vazão do Reservatório de Três Marias do Período Úmido 2005-2006	71

Figura 40 – Dispersão entre Séries Original e Preditada da Vazão do Reservatório de Três Marias do Período Seco 2006.....	72
Figura 41 – Dispersão entre Séries Original e Preditada da Vazão do Reservatório de Três Marias do Período Úmido 2005-2006	72

ÍNDICE DE TABELAS

Tabela 1 – Postos Pluviométricos e Fluviométricos selecionados.....	42
Tabela 2 – Divisão dos períodos hidrológicos anuais.....	46
Tabela 3 – Relação de postos com suas defasagens temporais.....	54
Tabela 4 – Resultados RMSE dos testes da rede RNA do Período Seco	66
Tabela 5 – Resultados MAPE dos testes da rede RNA do Período Seco	66
Tabela 6 – Resultados RMSE dos testes da rede RNA do Período Úmido	67
Tabela 7 – Resultados MAPE dos testes da rede RNA do Período Úmido.....	67
Tabela 8 – Resultados RMSE dos Períodos Secos e Úmidos.....	68
Tabela 9 – Resultados MAPE dos Períodos Secos e Úmidos.....	68

ÍNDICE DE ABREVIATURAS

ACL - Ambiente de Contratação Livre

ACR - Ambiente de Contratação Regulada

ANA – Agência Nacional de Águas

ANEEL - Agência Nacional de Energia Elétrica

BDT - Banco de Dados Técnicos

CCEE - Câmara de Comercialização de Energia Elétrica

CEPEL - Centro de Pesquisa de Energia Elétrica

CMO - Custo Marginal de Operação

CVU – Custo Variável Unitário

DECOMP - Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Curto Prazo

EAR - Energia Armazenada

ENA - Energia Natural Afluyente

FAC - Função de Autocorrelação

FACP - Função de Autocorrelação Parcial

FCC - Função de Correlação Cruzada

FCF – Função de Custo Futuro

FCI - Função de Custo Imediato

FDP - Função de Densidade de Probabilidade

FFT - Fast Fourier Transform

LM - Levenberg-Marquardt

MAPE - Erro Médio Percentual Absoluto

MCP - Mercado de Curto Prazo

MLP – Multi Layer Perceptron

MLT – Média de Longo Termo

MSE - Erro Médio Quadrático

NEWAVE - Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Longo e Médio Prazo

NSRBN – Non-linear Sigmoidal Regression Blocks Networks

ONS – Operador Nacional do Sistema Elétrico

PDE - Plano Decenal de Expansão de Energia

PDO - Programação Diária da Operação

PMO – Planejamento Mensal da Operação

PLD - Preço de Liquidação das Diferenças

REE - Reservatório de Energia Equivalente

RMSE - Raiz do Erro Médio Quadrático

RNA - Redes Neurais Artificiais

SEB - Sistema Elétrico Brasileiro

SIN – Sistema Interligado Nacional

ZCAS - Zona de Convergência do Atlântico Sul

1 INTRODUÇÃO

1.1 Aspectos Gerais

As vazões fluviais naturais brasileiras são oriundas das precipitações pluviométricas. O comportamento inconstante das vazões nacionais dificulta a sua previsibilidade, sendo usualmente modelada sob o efeito de componentes aleatórios.

O sistema brasileiro de geração elétrica, por ser composto predominantemente por usinas hidrelétricas de grandes reservatórios de acumulação de água conectados em cascatas, executa a regularização dos regimes fluviais, através do armazenamento de água nos períodos de maiores afluências de vazões fluviais, para ser utilizada durante os períodos mais secos, quando são escassas essas vazões. Sendo esta regularização a base do planejamento energético elétrico brasileiro, e fundamental para manter a continuidade do fornecimento da geração hidráulica, devido à irregularidade das vazões fluviais.

As coordenações centralizadas do sistema de reservatórios de água das usinas hidrelétricas, bem como das gerações de usinas termoeletricas, sendo ambas executadas pelo Operador Nacional do Sistema Elétrico - ONS no âmbito do chamado Planejamento Mensal da Operação - PMO do Sistema Interligado Nacional - SIN, permitem o melhor aproveitamento das vazões naturais, evitando o desperdício de água e gastos excessivos com combustíveis para as usinas térmicas, alcançando a melhor relação entre o menor custo e a máxima segurança operativa no atendimento total da demanda.

1.2 Justificativas

Historicamente, o modelo estocástico univariado PREVIVAZH de previsão de vazão diária, que não considera informações de precipitação, apresenta resultados com erros elevados. Dessa forma, a partir de 2008, o ONS passou a substituir seu uso gradativamente em algumas bacias do SIN por outros da área de Inteligência Artificial -

IA, tal como Redes Neurais Artificiais - RNA, Máquinas de Vetor Suporte (em inglês, *Support Vector Machine* - SVM), Algoritmos Genéticos - AG e Lógica Difusa [1].

A opção do ONS em usar o modelo RNA foi na usina hidrelétrica de Três Marias, localizada no alto da Bacia do São Francisco, através de uma plataforma privada chamada Neuro3M, oriunda do software Neurohidro e baseada no método NSRBN. A partir desta implementação, a aceitação da RNA pelo mercado energético tem sido crescente, pois ao contrário de outros métodos baseados em modelos tradicionais, a RNA é um método auto-adaptativo (aproximador universal) que busca capturar relações lineares ou não-lineares nos dados históricos reais. Além disso, apresenta capacidade em lidar com dados ruidosos, incompletos ou imprecisos, revelando-se uma poderosa metodologia para previsão com desvios menores do que os das redes tradicionais.

O ONS passou também a utilizar informações de precipitação como dado de entrada nos novos modelos, acarretando ganho significativo na qualidade das previsões de vazões, pois eliminou o “efeito sombra”, ou seja, deixou de ocorrer defasagem temporal entre as previsões e as vazões verificadas. Além disso, o uso de observações das estações fluviométricas e pluviométricas na própria região de interesse proporcionou um ganho na qualidade do acompanhamento das condições hidrológicas da bacia [1].

1.3 Objetivos

O planejamento da operação é totalmente dependente do tempo, uma vez que qualquer decisão de deplecionamento dos reservatórios de água, no presente, influencia a capacidade de geração do sistema no futuro, já que esta capacidade depende principalmente das afluições hidrológicas futuras. Portanto, a qualidade das previsões hidrológicas afeta o desempenho da operação do sistema, estando diretamente correlacionada com a confiabilidade e os custos da operação.

O planejamento da operação do sistema hidrotérmico brasileiro, bem como todo o mercado de energia, podem se beneficiar largamente do contínuo aprimoramento dos modelos de previsão hidrológica que, devido à complexidade da previsibilidade, faz-se

uso de uma cadeia de modelos desenvolvidos em sua maioria no Centro de Pesquisas de Energia Elétrica - CEPEL no Brasil [2].

No presente trabalho, a série alvo estudada é de vazão natural afluyente diária do Reservatório de Três Marias e as séries auxiliares são de vazão natural afluyente e de precipitação diários dos postos hidrométricos à montante do reservatório, tal como a rede RNA (Neuro3M), usada pelo ONS, está configurada.

Todas as séries temporais históricas estudadas são oriundas do Banco de Dados Técnicos - BDT do ONS e selecionadas entre os anos de 2000 e 2016. Sendo assim, abaixo estão detalhados os três anos mais afluentes e os três anos menos afluentes escolhidos para representar o período úmido e o período seco, e tentar dimensionar os resultados limítrofes:

- Período úmido, os anos de 2005-2006, 2006-2007 e 2011-2012 para as maiores vazões, e 2012-2013, 2013-2014 e 2014-2015 para as menores vazões; e
- Período seco, os anos de 2005, 2006 e 2011 para as maiores vazões, e 2013, 2014 e 2015 para as menores vazões.

Essas séries irão passar por um pré-processamento, a fim de extrair as suas componentes lineares e, posteriormente, um modelo de Redes Neurais Artificiais - RNA será aplicado para a predição, em até 7 dias, do sinal residual de vazões diárias do Reservatório de Três Marias.

Para avaliar e comprovar a capacidade do modelo desenvolvido de prever vazões com melhor precisão do que as previsões da RNA de uso atual do ONS, recursos gráficos serão utilizados, bem como métricas de distância entre o perfil real e o perfil predito pelo modelo. Finalmente, os desvios encontrados pelo modelo proposto serão comparados com o desempenho dos modelos usados atualmente pelo ONS.

1.4 Estrutura do Trabalho

No capítulo 2 são descritos os fundamentos do Sistema Elétrico Brasileiro - SEB, incluindo a apresentação da cadeia de modelos de otimização, o processo de previsão de vazões no ONS, além da importância das previsões na formação do preço da energia.

O capítulo 3 inicia-se com a apresentação de algumas publicações acadêmicas que detalham metodologias similares ao deste trabalho, descrevendo algumas formas já publicadas de dirimir o problema. Em seguida, a teoria que fundamenta este projeto é descrita, através do conceito de séries temporais, das formas de realizar sua decomposição, tendência, e sazonalidade, bem como o detalhamento do modelo de RNA. Já no capítulo 4, é detalhado todo o estudo de caso.

O método seguido para a realização do projeto será descrito no capítulo 5, que abordará assuntos como a leitura dos dados, pré-processamento, treinamento do modelo e métricas. Os resultados serão apresentados no capítulo 6, com gráficos temporais, histogramas e *scatter plot* das previsões comparados ao valor esperado. Por fim, o capítulo 7 será destinado às considerações finais e sugestões de trabalhos futuros.

2 O SISTEMA ELÉTRICO BRASILEIRO

2.1 Descrição Física do Sistema

O Sistema Elétrico Brasileiro - SEB possui dimensão intercontinental e características únicas em âmbito mundial. A sua matriz energética é diversificada, mas pode-se afirmar que é hidrotérmica, com forte predominância de usinas hidrelétricas, além de possuir múltiplos proprietários.

O SIN é dividido em quatro subsistemas (Sul, Sudeste / Centro-Oeste, Norte e Nordeste), onde as grandes usinas hidrelétricas estão situadas a distâncias consideráveis dos centros de carga, e, portanto, conectadas por uma extensa malha de transmissão. Os subsistemas Sul e Sudeste / Centro-Oeste do país são responsáveis por aproximadamente 80% do consumo de energia elétrica. E os subsistemas Nordeste e Norte, correspondem a outros 20% do consumo de eletricidade.

O Brasil possui mais de 160 GW de potência instalada, interconectada através do SIN [3]. Em 2018, a geração hidrelétrica respondeu por 69,94% do total gerado, seguida pela térmica, eólica e solar com respectivamente 19,83%, 7,37% e 2,7% de participação [4]. Ou seja, a geração do SIN é de predominância hidráulica, cujo regime hidrológico é totalmente dependente das chuvas.

Contudo este cenário vem se modificando, analisando a capacidade a instalar já contratada nos leilões realizados, nota-se para 2022 uma forte participação das fontes térmicas (23,5%), eólica (8,8%) e solar (2,1%), e uma redução da participação das usinas hidrelétricas (65,6%) [3], caminhando para uma matriz energética mais equilibrada entre as diversas fontes que a compõe.

A operação do SIN é realizada de forma coordenada e centralizada pelo ONS, objetivando alcançar ganhos sinérgicos com o menor custo financeiro. Tais ganhos são possíveis devido à diversidade hidrológica existente nas bacias hidrográficas de cada subsistema, bem como à significativa economia de escala de grandes usinas com reservatórios de regularização.

A coordenação centralizada se faz necessária, porque a operação de uma determinada usina depende das vazões liberadas a montante por outras usinas

pertencentes na mesma cascata, ao mesmo tempo em que sua operação afeta as usinas a jusante, de forma análoga. Podendo, inclusive, serem de domínio de diferentes empresas (agentes de geração).

Outro importante objetivo do ONS é sempre basear suas decisões operativas em busca da geração energética mais barata, sendo que os principais parâmetros responsáveis pela determinação do custo da geração são a energia armazenada nos reservatórios (% de volume útil preenchido) e a afluência aos mesmos. Porém, a mais importante das duas é a afluência nos reservatórios, pois é esta que irá determinar se mais energia será armazenada ou se os reservatórios serão esvaziados no futuro.

Sendo assim, a partir do conhecimento das previsões de vazões naturais afluentes nas bacias e sub-bacias, uma região do país pode tornar-se estratégica para o setor elétrico nacional. Os modelos matemáticos de previsões baseados em técnicas estocásticas já são conceituados no mercado, mas o ONS tem investido cada vez mais no uso dos modelos hidrológicos determinísticos do tipo chuva-vazão, que utilizam os dados de chuva para prever eventos de vazões, pois eles têm apresentado melhores resultados, apesar de ter iniciado o uso de previsões calculadas por modelos estocásticos.

Como o nível dos reservatórios é definido pela água oriunda das chuvas e estas são de difícil previsibilidade, o problema de planejamento da operação torna-se essencialmente estocástico. O ONS para ser capaz de operar um sistema de grandes incertezas necessita de modelos matemáticos capazes de auxiliar na otimização do planejamento da operação, de forma a encontrar a melhor decisão operativa do presente até um determinado período futuro.

2.2 A Cadeia de Modelos de Otimização

Os modelos econômicos de otimização energética utilizados pelo setor elétrico brasileiro, desenvolvidos pelo CEPEL, objetivam otimizar o planejamento da operação eletro-energética, através da minimização do custo total por intermédio de decisões de geração energética, de intercâmbios entre regiões e do corte de carga (déficit energético). Baseados no princípio da otimalidade [5], o qual afirma que as decisões

futuras devem constituir uma política ótima no que se refere ao estado resultante da decisão atual, os modelos de otimização possuem diferentes horizontes de estudo, além de possibilitarem o acoplamento entre si, ou seja, os resultados de um modelo são utilizados como dados de entrada na simulação de outro modelo.

O custo total da operação é composto pelos gastos com combustíveis de usinas térmicas (Custo Variável Unitário - CVU), pelo valor da água (fornecido pelo modelo NEWAVE), bem como, pelo custo de déficit de energia, que compreende as penalidades associadas ao não atendimento da demanda energética que pode vir a ocorrer.

A cadeia principal de modelos, utilizada atualmente para o planejamento do SEB, é composta pelos programas NEWAVE (médio prazo) e DECOMP (curto prazo). Já o modelo DESSEM de curtíssimo prazo (horizonte de até 1 semana, com discretização semi-horária), até a data deste trabalho, encontra-se em fase de testes para ser colocado em operação em 2020 juntamente com os outros modelos.

2.2.1 NEWAVE

O programa NEWAVE (Modelo de Despacho Hidrotérmico a Reservatórios Equivalentes) é largamente utilizado no setor elétrico nacional como balizador de importantes assuntos, tais como, o Plano Decenal de Expansão de Energia - PDE, o planejamento do PMO, a definição e o cálculo da garantia física (energia assegurada) de empreendimentos de geração, elaboração de diretrizes para os leilões de energia, definição de estratégias corporativas de empresas e agentes, além da comercialização de energia através do cálculo do Preço de Liquidação das Diferenças - PLD. Sendo assim, o grande objetivo do uso deste modelo de otimização no planejamento da operação é minimizar o valor esperado do custo total de operação, para um horizonte de cinco anos de estudo, discretizado mensalmente, através de estratégias de combinação da geração hidráulica e térmica. Adicionalmente, o programa também é responsável por fornecer os impactos oriundos da utilização da água armazenada (Energia Armazenada - EAR) nos reservatórios, através da construção da Função de Custo Futuro - FCF, a qual será acoplada no modelo de curto prazo, DECOMP.

O NEWAVE é composto por quatro módulos computacionais que serão aplicados na criação da sua política de operação. E utiliza para a resolução do problema a metodologia conhecida como Programação Dinâmica Dual Estocástica - PDDE, possibilitando que o problema principal seja dividido em um conjunto de problemas menores, tornando mais simples a resolução do mesmo.

Dentre os módulos que o compõe, há o de “Geração dos Cenários de Afluências”, onde são gerados os cenários probabilísticos de Energia Natural Afluente - ENA, através das informações de previsões mensais recebidas dos modelos PREVIVAZM [6] e GEVAZP [7] [8].

O modelo estocástico univariado PREVIVAZM se baseia em combinações de uma grande quantidade de modelagens estocásticas, bem como em métodos de estimação de parâmetros para conseguir prever vazões em base mensal, num horizonte de até 12 (doze) meses à frente, a partir de um histórico de vazões naturais médias mensais.

O modelo estocástico multivariado GEVAZP realiza consultas no registro histórico de ENA, observadas de 1931 até o ano atual, e fazendo uso da metodologia linear $PAR(p)$, Auto Regressivo Periódico de ordem p , gera as séries sintéticas de energia. Essas séries são estatisticamente similares às séries históricas e têm a mesma probabilidade de ocorrer, ou seja, são equiprováveis.

Sendo assim, a metodologia $PAR(p)$ tem a função de verificar no histórico as vazões e as ENAs que ocorreram, e calcular correlações que servirão de base para a construção dos cenários de energia. Para cada mês e para cada Reservatório de Energia Equivalente - REE há um valor diferente para a ordem de regressão p , sendo que, para a geração dos cenários de energia utilizados pelo NEWAVE, a ordem de regressão máxima é de 6 (seis) meses, ou seja, são consultados no máximo seis meses do histórico para a criação das correlações.

Essas correlações são importantes para que as incertezas com relação às afluências sejam minimizadas, através da reprodução do maior número possível de cenários no cálculo, além de serem essenciais para a construção da FCF.

2.2.2 DECOMP

O programa DECOMP (Determinação da Coordenação da Operação de Curto Prazo) também foi desenvolvido com a metodologia de Programação Dinâmica Dual - PDD e é aplicado, em conjunto com o NEWAVE, no planejamento da operação de sistemas hidrotérmicos.

Sendo considerado um modelo de curto prazo, então possui um horizonte de estudo de até um ano, discretizando o primeiro mês em semanas e o segundo mês em diante, mensalmente.

A previsão de vazões utilizada como entrada para o DECOMP é considerada de médio prazo, porque abrange um horizonte de dois meses à frente, sendo o primeiro mês determinístico e o segundo estocástico. Por conter um horizonte de estudo menor do que o NEWAVE, ele possui, além de menores incertezas, um maior detalhamento no que tange a representação das usinas, seus armazenamentos e restrições operativas de modo individualizado. Na Figura 1 é possível observar o acoplamento do DECOMP com o NEWAVE.

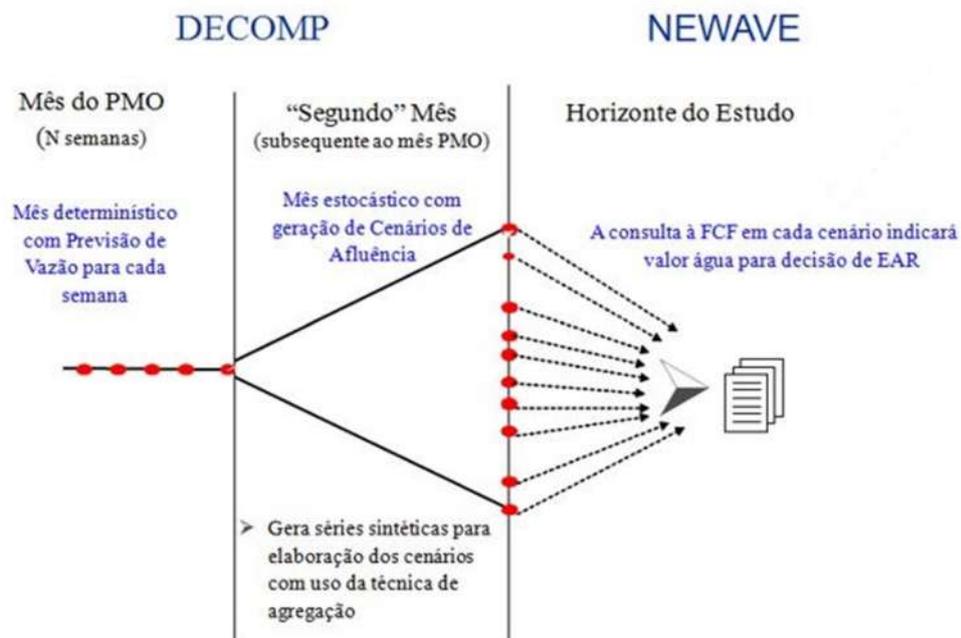


Figura 1 – Acoplamento de Médio Prazo e de Curto Prazo

Para o DECOMP realizar o cálculo dos cenários de ENAs de médio prazo, o ONS utiliza as previsões calculadas pelos modelos PREVIVAZ [9] e GEVAZP, sendo este último já mencionado. O primeiro modelo gera as informações de vazões médias semanais que chegam às usinas no decorrer do primeiro mês. Enquanto o segundo modelo, gera séries sintéticas de vazões médias mensais para o segundo mês, realizando dessa forma a simulação estocástica para o período desejado.

O modelo estocástico univariado PREVIVAZ prevê afluências em base semanal para um horizonte de até 6 (seis) semanas à frente, baseando-se em uma grande quantidade de combinações de estruturas de correlação e métodos de estimação de parâmetros, utilizando informações de até 4 (quatro) semanas anteriores.

No planejamento da operação do SIN, o primeiro mês é dividido em semanas, denominadas semanas operativas, e as afluências da primeira semana já são conhecidas, ou seja, diz-se para o modelo PREVIVAZ a quantidade de água diária (vazão natural afluente) que vai chegar a cada usina dos subsistemas, realizando uma otimização determinística.

O modelo estocástico PREVIVAZH [10] é um dos modelos utilizados para prever essas vazões em base diária para um horizonte de até 14 (quatorze) dias à frente, baseando-se na desagregação de previsão semanal a partir da tendência inferida das últimas vazões passadas e de séries sintéticas diárias de vazões naturais de forma ponderada. É importante ressaltar que este modelo deve passar por uma calibração anual e foi utilizado pelo ONS predominantemente em todas as bacias até 2008.

Na Figura 2 consta um esquemático para melhor explicar a funcionalidade de cada modelo hidrológico na cadeia de modelos do planejamento da operação do SIN.

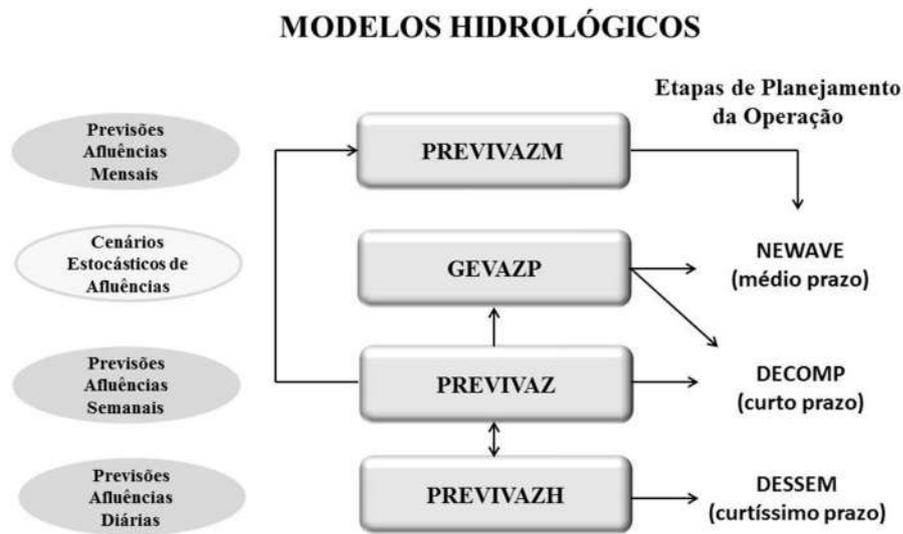


Figura 2 – Participação dos modelos hidrológicos na cadeia do planejamento da operação do SIN

2.2.3 Modelos de Previsão de Vazão Diária

O ONS faz uso atualmente de uma diversidade de modelos para prever essas vazões afluentes em base diária, denominadas previsões de curto prazo, tais como: PREVIVAZH, SMAP (modelo determinístico chuva-vazão), MEL (modelo estocástico linear e multivariado), MGB (modelo conceitual distribuído), Lógica Difusa (Fuzzy), Redes Neurais Artificiais (RNA), dentre outras.

O ONS vem buscando, continuamente, a melhoria da modelagem hidrológica, bem como de seus principais insumos. Nesse contínuo aprimoramento, pode-se destacar o processo de concorrência de modelos hidrológicos os quais passaram a considerar a precipitação observada e prevista na bacia para a modelagem chuva-vazão. Tal processo culminou, em 2009, com a substituição do modelo PREVIVAZH em algumas bacias do SIN pelos modelos MGB (bacia do rio Paranaíba e bacia incremental à UHE São Simão), SMAP-MEL (bacia do rio Paraná no trecho incremental à UHE Itaipu), e Fuzzy (bacia do rio Iguaçu) para a previsão de vazões diárias da primeira semana operativa [11].

Após a entrada desses modelos houve, ainda, a incorporação do modelo Neuro3M (método NSRBN) de redes neurais no alto da bacia do rio São Francisco e, principalmente, do modelo SMAP/ONS que atualmente é utilizado nas bacias dos rios Grande, Paranaíba (a montante da UHE Itumbiara), Paranapanema.

2.3 Planejamento e Programação da Operação

O planejamento da operação do SIN é realizado pelo ONS e divulgado mensalmente no PMO, com revisões semanais. Neste programa são declaradas todas as informações utilizadas pelos modelos de otimização, bem como seus resultados, sendo, portanto, de grande interesse por todos os agentes do SEB.

Como exemplo de entradas de informações tem-se as energias de armazenamentos dos reservatórios de água das usinas hidrelétricas, as previsões de aflúências semanais, as restrições hidráulicas e elétricas do SIN, os limites de intercâmbio elétrico entre os subsistemas, a inflexibilidade energética e os CVU das usinas térmicas. Já com relação aos resultados obtidos, podemos destacar o Custo Marginal de Operação - CMO, bem como outras diretrizes operativas para o horizonte mensal.

Para obter a previsão de vazões médias semanais para todos os locais de aproveitamentos hidrelétricos, desde a primeira até a última semana operativa de cada mês, o ONS necessita de informações das vazões médias diárias da semana em curso de elaboração do PMO, sendo estas solicitadas aos Agentes de Geração responsáveis pelos aproveitamentos das bacias, ou obtidas através dos modelos de previsão de curto prazo.

O PMO é divulgado sempre nas últimas quintas e sextas de cada mês. Sendo assim, a cadeia de modelos de otimização inicia seu processamento na última terça-feira e finaliza somente na última sexta-feira. Em seus resultados constam informações das futuras 6 (seis) semanas operativas, sendo que toda semana corresponde ao período de sábado à sexta. Já as suas revisões semanais ocorrem às quartas-feiras, quando as previsões de vazões médias semanais para todas as semanas remanescentes são sempre corrigidas.

As vazões médias diárias necessárias para a elaboração do PMO são as observadas de sábado até terça-feira, bem como as previstas de quarta-feira até sexta-feira, completando assim uma semana operativa. Ou seja, os Agentes de Geração dos aproveitamentos enviam as vazões naturais diárias verificadas, além das previsões para a semana em curso, possibilitando ao ONS usar essas informações para obter as vazões médias semanais estimadas das semanas operativas futuras. Em paralelo, o ONS executa os seus modelos de curto prazo, cujas previsões médias diárias poderão substituir as dos Agentes de Geração em caso de falta ou inconsistência das mesmas.

Os Agentes de Geração enviam as vazões naturais diárias verificadas em base horária, restando para o ONS a responsabilidade de integralizar esses dados recebidos, tratar e torná-los consistidos em média diária, e conseqüentemente, serem usados como base de dados de entrada de seus modelos de previsão de curto prazo.

A Figura 3 ilustra o esquema do PMO com as informações necessárias para cada dia da semana.

							SEMANAS OPERATIVAS								
							1ª SEMANA				2ª SEMANA ... 6ª SEMANA				
sáb	dom	seg	ter	qua	qui	sex	sáb	dom	seg	ter	qua	qui	sex	sáb ...sex	
					PMO		previsão de vazão média semanal (PREVIVAZ)						previsão de vazão média semanal (PREVIVAZ)		
vazão diária verificada				previsão de vazão diária e mensal											

PMO - 4 dias de vazão diária verificada (Agentes) + 3 dias de previsão de vazão diária (Agentes ou ONS)

Figura 3 – Esquemático do PMO

2.4 Sensibilidade das Vazões Naturais Afluentes sobre o Preço de Liquidação das Diferenças

2.4.1 Custo Marginal de Operação

O planejamento da operação energética possui uma política de operação centralizada (modelo inglês *Tight Pool*), que objetiva atender a demanda energética nacional através da minimização do custo de operação do sistema, sofrendo este uma

maior influência do custo do combustível utilizado por cada usina para a produção de energia.

As usinas térmicas possuem um CVU de operação proporcional ao custo do combustível utilizado (gás natural, carvão, óleo etc.). Já as usinas hidrelétricas, dependem da água armazenada para a produção de energia, que apesar de não existir um gasto financeiro na compra deste recurso natural, há um “custo” envolvido inerente ao seu uso. A base da lógica do cálculo do “custo” da água não é tão simples, pois existe um acoplamento temporal entre a decisão tomada no presente e sua consequência futura.

Se no presente for utilizada demasiadamente a água dos reservatórios e as afluências naturais não forem suficientes para repor o nível dos reservatórios, no futuro, o custo de operação do sistema tende a ser muito alto, pois o atendimento à demanda energética será através da geração térmica, ou até mesmo, será necessário cortar carga elétrica (déficit de energia) se a geração total do sistema não for suficiente. Por outro lado, se decidir economizar a água dos reservatórios, utilizando a geração térmica no presente, e ocorrer grandes vazões afluentes naturais no futuro, acarretará o vertimento turbinável da água dos reservatórios, ou seja, esse recurso natural energético será desperdiçado. De forma simplificada, a Figura 4 abaixo caracteriza esse dilema do planejamento da operação.

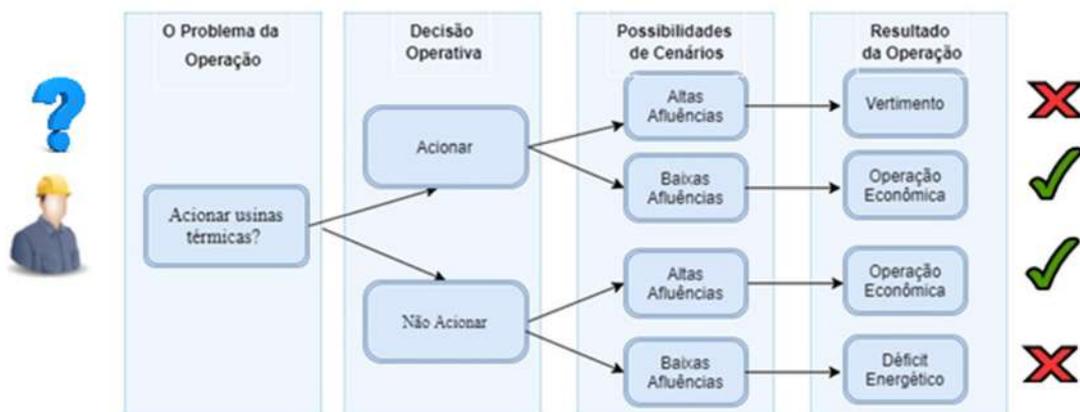


Figura 4 – Dilema da Operação no Despacho Hidrotérmico

A otimização do problema consiste em comparar o benefício imediato da utilização dos recursos hídricos e o benefício futuro de seu armazenamento, estando ambos os cenários intrinsecamente relacionados com as afluências futuras.

A Figura 5 abaixo representa a análise realizada dos custos imediatos e futuros em função do volume final de energia hidráulica armazenada nos reservatórios, por meio das suas respectivas funções de custo.

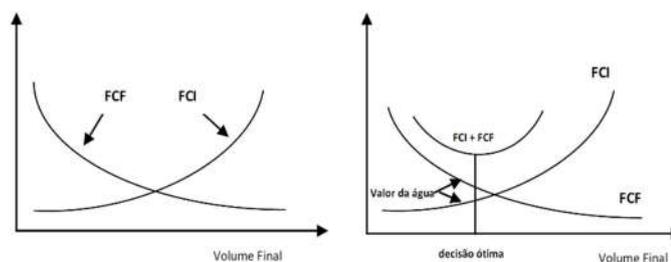


Figura 5 – Curvas de Função de Custo Futuro e Imediato

A Função de Custo Imediato - FCI representa os custos de geração térmica no presente. Pode-se observar que o custo imediato irá diminuir à medida que for diminuindo o volume final disponível no futuro, pois significa que foi decidido despachar poucas usinas térmicas no presente, em detrimento de maior geração hidráulica. Por sua vez, seguindo o raciocínio desta decisão presente, a FCF irá aumentar, pois o volume armazenado final irá diminuir, acarretando menos energia hidráulica disponível no futuro. Sendo assim, o uso ótimo da água armazenada corresponde ao ponto que minimiza a soma dos custos imediatos e futuros.

A partir da convergência do modelo DECOMP em encontrar o ponto mínimo da soma das FCF, considerando para isso informações como o armazenamento dos reservatórios, os CVUs das térmicas, as cargas do SEB, as afluências previstas, dentre outros dados, o ONS divulga o CMO em todas as sextas-feiras para cada submercado (Sudeste, Sul, Nordeste e Norte) e para os três patamares de carga do sistema (pesada, média e leve). Sendo um custo válido para a próxima semana operativa, que se inicia no sábado e termina na sexta-feira seguinte, ou seja, é calculado ex-ante. Resumindo-se, o CMO é o valor que corresponde ao custo para se produzir o próximo consumo de energia elétrica que o SEB necessita, sempre refletindo o menor custo para atender a uma demanda adicional.

2.4.2 Preço de Liquidação das Diferenças

Paralelamente ao cálculo do CMO, a Câmara de Comercialização de Energia Elétrica - CCE divulga o PLD. Para calcular o PLD, faz-se uso dos mesmos modelos matemáticos de otimização e dados de entrada que o ONS, sendo que para este cálculo são retiradas as restrições conjunturais de transmissão internas de cada submercado e as restrições das usinas em teste. Além disso, o PLD é limitado por um piso e um teto, que é basicamente o preço mínimo e máximo que ele pode alcançar, sendo estabelecidos e atualizados pela Agência Nacional de Energia Elétrica - ANEEL anualmente.

O PLD é o preço ex-ante da energia, ou seja, o preço previsto para a próxima semana operativa e divulgado no presente (última sexta-feira do mês), simbolizando o sinal econômico determinante para as diretrizes da Programação Diária da Operação - PDO, bem como do mercado financeiro energético.

No modelo vigente do setor elétrico brasileiro, a comercialização de energia pode ser realizada basicamente através de dois ambientes, o Ambiente de Contratação Regulada - ACR e o Ambiente de Contratação Livre - ACL. Todos os contratos, tanto no ACR quanto no ACL, são registrados na CCEE e servem de base para a contabilização e liquidação das diferenças de energia no Mercado de Curto Prazo - MCP.

A contabilização para efeito de liquidação financeira envolve o cálculo da diferença entre a energia gerada e a contratada pelos Agentes do mercado, sendo valorada pelo PLD, representando grande importância econômica para o setor.

Conforme mostra a Figura 6, o peso das vazões naturais afluentes na composição do preço da energia é bem elevado, e adicionalmente podemos afirmar que é um ponto bastante frágil, por apresentar historicamente no setor um grande volume de erros nos resultados dos modelos de previsão em uso atualmente. No entanto, a implementação de modelos que consideram informações de chuva em conjunto com os de vazão vem apresentando uma assertividade bastante significativa, corrigindo assim a principal distorção entre os insumos previstos e verificados.



Figura 6 – Decomposição do PLD [12]

Sendo assim, a aplicação de novas técnicas de predição de vazões afluentes é um assunto de elevado interesse no setor elétrico energético, pois com o aumento da precisão dos modelos de previsão de vazões, além de minimizar as incertezas futuras sobre as informações das ENAs, teremos resultados de CMO e de PLD mais aderentes ao cenário verificado.

O PLD é a base de cálculo para toda a contabilização financeira do setor energético, tais como: tratamentos de exposições financeiras ao mercado de curto prazos; encargos devido a restrições de operação e segurança energética; precificação das penalidades de energia e por falta de combustível; contratação de energia de reserva e abatimento do encargo a pagar pelos consumidores; etc. Dessa forma, é de vital importância para todos os agentes de distribuição, geração e comercialização de energia prever vazões afluentes com menores desvios para poder enfim realizar estudos de projeção do PLD com maior confiabilidade.

3 REVISÃO DA LITERATURA

3.1 Escolha do melhor algoritmo de previsão

Os modelos clássicos estocásticos de previsão, tais como os modelos hidrológicos desenvolvidos pelo CEPEL, fazem uso de equações matemáticas com coeficientes estimados, e estruturas e número de termos estatisticamente identificados [13]. As expressões matemáticas que descrevem os fenômenos da natureza são sempre tentativas de aproximações da realidade.

Sendo assim, o uso dos métodos clássicos de previsão baliza-se na constante busca por métodos resistentes às deficiências de identificação do modelo e da estimação dos parâmetros.

Historicamente, os resultados da aplicabilidade desses modelos para prever vazões afluentes naturais não têm sido satisfatórios pelo mercado energético, acarretando uma grande busca por outros modelos mais eficientes.

O comportamento das vazões dos rios tem sido bastante estudado por vários autores da área de inteligência artificial, e especificamente, de RNA. Esta rede faz uma analogia com os neurônios humanos baseada na transmissão de sinais elétricos. De maneira simplificada, a rede pode aceitar uma série de dados de entrada e gerar, a partir destes, uma ou mais saídas, através da busca de uma função que modele o sistema.

Assim, visando conhecer um pouco mais sobre a aplicação da técnica de redes neurais, são descritos a seguir alguns trabalhos.

Em 1997, Shamseldin [14] usou o modelo de redes neurais com a técnica chuva-vazão, ou seja, fez uso de informações de chuvas e vazões históricas de postos da área em estudo. Ao comparar os resultados com outros modelos, tal como, modelo simples linear - SLM, modelo sazonal baseado na perturbação linear - LMP e o modelo linear de perturbação - NNLP, as redes neurais revelaram um bom desempenho, mas com resultados variáveis.

No mesmo ano, Valença [15] implementou o modelo de redes neurais do tipo *Multi Layer Perceptron* – MLP pelo método *backpropagation* de duas formas distintas

em paralelo, ou seja, previu vazões médias mensais usando informações de chuva e vazão, bem como, uma segunda previsão com informações apenas de vazão. Segundo o autor, os dois estudos com o uso de redes neurais alcançaram resultados de ótima qualidade, quando comparados com as metodologias tradicionais. O primeiro modelo foi comparado com os tradicionais MOHTSAR e SMAP [16], enquanto o segundo, foi comparado com a metodologia estatística tradicional de Box e Jenkins [13].

Sperb et al. [17] demonstraram a aplicabilidade de um modelo de redes neurais artificiais na previsão de enchentes (vazão em tempo real) para a Bacia do Rio Itajaí-Açu até a cidade de Blumenau, utilizando dados de precipitação de estações pluviométricas distribuídas na bacia. O algoritmo de treinamento escolhido foi do tipo MLP pelo método de *Levenberg-Marquardt* - LM, justificado pela não linearidade dos dados. Como os resultados atingiram 93% de acerto nas previsões das vazões em tempo real, a RNA ganhou grande notoriedade na área de recursos hídricos.

No ano de 2000, Kyrillos [18] criou um modelo de redes neurais (Qnet) para previsão de vazões em tempo real do posto de Morpará, na bacia do rio São Francisco, objetivando auxiliar no gerenciamento dos recursos hídricos, através de um Sistema de Suporte à Decisão - SSD. Foram utilizados dados diários de vazão e de chuva média dos postos da bacia. Os melhores resultados alcançados foram das redes com previsão de curtíssimo prazo (6 dias) e as que fizeram uso de maior quantidade de neurônios na entrada.

Lima e Ferreira [19] simularam em 2003 alguns modelos de RNA para previsão média mensal de vazão na bacia da Fazenda Cajazeiras, no Estado do Ceará, fazendo uso de dados históricos de chuva e de vazão (tipo chuva-vazão) compreendidos entre os anos 1963 e 2000. Os modelos possuíam configurações distintas para o número de neurônios nas camadas de entrada e intermediária, além da quantidade de anos separados para o treinamento e para a validação da rede. A partir dos resultados obtidos, puderam observar que o melhor desempenho atingido foi quando o conjunto de treinamento assumiu a metade da série histórica, e também quando adotou um número maior de nós na entrada da rede. Inclusive, fazendo uso do dado de precipitação (t , $t-1$ e $t-2$) e de vazão ($t-1$) em diferentes atrasos no tempo. Já em relação ao número de neurônios na camada intermediária, constatou-se que a variação deste número não impactou significativamente nos resultados. Além de ser uma quantidade definida

somente através das simulações, pois não existe nenhum método que determine um valor ideal.

Méndez et al. [20] analisaram as performances dos modelos do tipo Box e Jenkins e do tipo RNA, que fizeram uso de informações de chuva e de vazão, tanto para previsões de vazões mensais, quanto para vazões diárias no rio Xallas, localizado no Nordeste da Espanha, com os dados de 1992 até 1999. Os resultados revelaram que o modelo Box-Jenkins possui melhor desempenho para previsões mensais e as redes RNA, para previsões diárias.

No ano de 2005, Valença e Valença [21] simularam dois diferentes modelos de RNA, utilizando o programa Neurohidro, de código fechado e baseado num método chamado NSRBN, para a previsão de vazões naturais médias mensais afluentes ao reservatório de Boa Esperança. Um modelo utilizando informações de chuva e de vazão, e um outro, apenas com informações de vazão. No primeiro modelo, os dados utilizados foram entre 1978 e 1986, e o erro percentual médio absoluto alcançado foi de 53,9% com desvio padrão de 4,1 m³/s. Já no segundo modelo, com o período de estudo entre 1931 e 2003, obtendo um erro percentual médio absoluto de 22,1% com desvio padrão de 111,5 m³/s.

Ao longo das últimas décadas, muitos trabalhos têm comprovado a eficiência da aplicação das redes RNA em diversas áreas, como, por exemplo, na previsão de vazões diárias e mensais. Além de enfatizarem a superioridade dos resultados provenientes das redes RNA, em detrimento dos resultados obtidos dos tradicionais modelos estatísticos [22] [23] [24].

Contudo, poucos trabalhos têm mostrado o pré-processamento dos dados de entrada da rede neural, tal como o processo de seleção, decomposição e de análises. Essa preparação dos dados pode melhorar ainda mais os resultados, visto que a aplicabilidade de diversas metodologias de redes neurais já está consolidada.

3.2 Séries Temporais

Uma série temporal (ou histórica) é uma sequência de valores que representa o comportamento de um conjunto de observações feitas ao longo do tempo [25]. Uma série temporal, com N observações, pode ser representada como (1).

$$S = S1; S2; \dots ; SN \quad (1)$$

Onde S é a série temporal e as variáveis $S1$ até SN são as amostras da série no período de tempo $t = 1; 2; \dots ; N$.

Segundo Chris Chateld, o estudo de uma série temporal tem por objetivo alcançar uma descrição, explicação, predição, ou um controle do comportamento. A descrição da série consiste em obter as medidas descritivas simples das suas principais propriedades, e isto é realizado primeiramente através de um gráfico das observações ao longo do tempo. É possível observar no gráfico a periodicidade, os crescimentos ou decréscimos médios, e até os pontos discrepantes (*outliers*¹) da série. Já com relação a obtenção da explicação da série, muitas das vezes, faz-se uso da observação da influência sofrida por uma segunda série. Essa abordagem também pode ser utilizada para o controle de um sistema físico ou econômico. Por fim, tem-se a predição de valores futuros de uma série temporal, sendo este o objetivo desta dissertação.

Uma série temporal é chamada contínua quando suas observações são feitas continuamente ao longo do tempo, independentemente se as variáveis medidas só podem assumir valores discretos. Por outro lado, a série temporal é chamada discreta quando as observações ocorrem em intervalos de tempo específicos. A série contínua no tempo, $S(t_{contínuo})$, é representada pela série discreta $S(t\Delta T)$, onde $\Delta T = 1, 2, \dots, N$. Após a amostragem, é realizada a normalização dos dados do intervalo para $\Delta t = 1$ e, conseqüentemente, a frequência de amostragem para $f_s = \frac{1}{\Delta T} = 1$. E com isso, a série discreta passa a ser $S(t)$, onde $t = 1, \dots, N$ [26].

A predição de valores futuros da série, estando no instante t , é representada como $S(t+k)$, onde $k > 0$. Se uma série temporal pode ser predita sem erros associados, ela é dita como determinística [25]. Porém, na prática a maioria das séries é estocástica, o que significa que as amostras futuras serão estimadas com um erro associado, ou seja, serão apenas parcialmente determinadas pelas amostras passadas.

¹ São pontos de observação dentro de um conjunto de dados que se encontram consideravelmente distantes das outras observações deste mesmo conjunto [27].

3.2.1. Decomposição de Séries Temporais

A forma mais prática de analisar uma série temporal é decompô-la em séries mais simples [26]. Esta ação de decomposição significa separar da série original os componentes determinísticos no tempo, e a diferença entre a série completa e esses componentes mais simples é chamada de série residual. Sobrando na série residual dois componentes: um primeiro, cujos valores possuem uma dependência complexa com o tempo (normalmente, não linear), e um segundo, que apresenta ruídos não correlatos (não previsíveis). Por outro lado, os componentes da série determinística são geralmente tendência, sazonalidade, e ciclos senoidais.

No processo de decomposição, que pode ser aditivo, multiplicativo ou misto, é muito importante observar a energia de cada um desses componentes. Principalmente no acompanhamento da perda de energia ao se retirar cada um dos componentes por vez. Em estatística, uma forma de estimar a energia de uma série temporal é calculando a sua variância ao longo do tempo. Na área de engenharia é comum realizar uma analogia com sinais elétricos, referindo-se a esta variância como espectro de potência, mesmo quando não há potência física envolvida [28].

Os componentes da série determinística possuem uma energia consideravelmente maior do que a energia da série residual. Este fato pode dificultar o processo de aprendizado de um modelo que visa prever a série residual, já que esta última possui dependência não linear com o tempo. Sendo assim, torna-se importante reforçar não somente a necessidade do processo de decomposição de séries temporais, como também a análise da energia remanescente a cada etapa de retirada de um componente.

O principal objetivo da decomposição é tentar obter uma série residual que seja estritamente estacionária no tempo, ou seja, com todos os momentos invariantes no tempo, para que os valores passados da série possam ser usados para caracterizar estatisticamente a série em qualquer tempo passado e futuro.

Não obstante a esta decomposição ideal, na prática, limitamos a garantir a obtenção de uma série, pelo menos, fracamente estacionária no tempo, onde seus dois primeiros momentos, a média e a variância, são estáticos.

3.2.2. Tendência

Como já foi dito, a tendência é uma componente determinística da série temporal e representa-a graficamente de forma mais simplificada, identificando as mudanças de longo prazo do comportamento da série e excluindo as mudanças periódicas ou quase periódicas.

Para identificar a tendência é necessário ajustar os dados em uma das funções paramétricas descritas pelas equações (2), (3), (4) e (5), onde μ_t é o nível médio local no tempo t [29].

- Tendência linear global (determinística)

$$\mu_t = \alpha + \beta t \quad (2)$$

Onde α e β são constantes que representam a interceptação da reta com o eixo vertical e coeficiente angular da reta, respectivamente.

- Tendência linear (estocástica)

$$\mu_t = \alpha_t + \beta_t t \quad (3)$$

Onde os parâmetros α_t e β_t evoluem com o tempo e representam, respectivamente, a interceptação local da reta e coeficiente angular local da reta.

- Curva de Gompertz

$$\log(\mu_t) = a + br^t \quad (4)$$

Onde $0 < r < 1$, e os parâmetros a e b definem a curva.

- Curva logística

$$\mu_t = \frac{a}{(1 + be^{-ct})} \quad (5)$$

Onde a , b e c são parâmetros que definem a curva.

Dentre os tipos de tendência que podem ser extraídos, o mais utilizado é a tendência linear. Mas, independentemente do tipo de tendência escolhida, seu comportamento deve variar ao longo do tempo. Por este motivo, pode ser necessário

fazer a extração da tendência local, em uma janela de tempo menor do que a janela de aquisição total.

Uma vez determinados os parâmetros da curva escolhida para a representação da tendência, é possível realizar a sua extração da série original, conforme a equação (6), onde $S(t)$ é a série original e $S_I(t)$ é a série sem tendência.

$$S_I(t) = S(t) - \mu_t \quad (6)$$

Na prática, através das simulações, é necessário se atentar para o fato de que a escolha de uma tendência que represente a série de forma muito detalhada pode indicar uma boa interpolação para os pontos do passado, mas talvez não apresente uma capacidade de generalização para os pontos futuros

3.2.3. Sazonalização e Ciclos senoidais

A sazonalidade é a componente que representa padrões que se repetem em intervalos de tempo regulares, ou períodos de tempo P . Já os ciclos senoidais são um caso particular da sazonalidade, pois a identificação da sua presença se revela na existência da forma de uma senoide, com período P , na série [26], podendo ser modelada como um comportamento amortecido de uma senoide [30].

Geralmente, a estimativa da sazonalidade, $seas(i)$, para cada instante $i = 1, 2, \dots, P$ do período, pode ser obtida calculando o valor médio da série neste instante a cada período, conforme representado em (7).

$$seas(i) = \frac{1}{round(\frac{N}{P})} \sum_{k=0}^{round(\frac{N}{P})-1} S_I(i + kP), i = 1, \dots, P \quad (7)$$

Onde $S_I(t)$ é a série já sem tendência, N é o número de amostras, e $round(\frac{N}{P})$ é o resultado inteiro da divisão $\frac{N}{P}$, aproximando o resultado para baixo, que representa o número de períodos completos de sazonalidade que existem na série.

Adicionalmente, existem alguns tratamentos clássicos que podem ser aplicados às séries temporais para detectar as sazonalidades e os ciclos senoidais [31]. Um deles é a

aplicação da transformada de Fourier, pelo algoritmo *Fast Fourier Transform - FFT*, que tem como saída os parâmetros a e b da expressão senoidal, representada em (8).

$$cs(t) = a * \cos(2\pi ft) + b * \sin(2\pi ft) \quad (8)$$

O outro método é calcular a autocorrelação de uma série temporal que, por sua vez, é definida como a correlação entre a série no tempo local, $S(t)$, com a mesma série atrasada de k unidades de tempo, $S(t - k)$ [26].

Deste modo, a sazonalidade é identificada quando há uma autocorrelação significativa em atrasos múltiplos de P (nP , onde n é inteiro), ou quando o espectrograma tem raias em $n = \frac{1}{P}$, onde n é inteiro. Já para o caso de ciclos senoidais, a autocorrelação tem que apresentar um aspecto de senoide, e seu espectrograma constar uma raia de frequência $f = \frac{1}{P}$ e amplitude significativamente acima da média.

Finalmente, para realizar a extração da sazonalidade e dos ciclos senoidais, basta subtrair os mesmos da série temporal já sem tendência [26], no caso de uma decomposição aditiva, conforme representado em (9).

$$S_2(t) = S_1(t) \pm seas \left[R \left(\frac{t}{P} \right) \right] - cs(t) \quad (9)$$

Onde $S_2(t)$ é a série residual, $cs(t)$ é um ciclo senoidal e $R \left(\frac{t}{P} \right)$ é o resto da divisão $\frac{t}{P}$, que representa o instante i do período da sazonalidade no tempo t da série.

Assim como na extração da tendência, é importante observar e certificar se o comportamento da sazonalidade varia ao longo do tempo. Deste modo, pode ser necessário fazer a extração da sazonalidade localmente, em uma janela de tempo menor do que a janela de aquisição total.

3.2.4. *Preditor Não Linear*

À princípio, a série residual $S_2(t)$ não tem mais componentes que sejam facilmente determinadas como função de tempo, mas seus valores podem ainda ter dependência de valores passados.

Usualmente esta série residual apresentará uma autocorrelação decrescente com o atraso. Se forem encontradas correlações significativas de valores atrasados com os atuais, então os valores futuros podem ser expressos em função dos valores passados. É conveniente permitir que os valores futuros sejam previstos como uma função não linear dos valores passados, implementada por uma rede neural, atuando como um preditor.

3.3 Redes Neurais Artificiais

Em 1943, McCulloch e Pitts criaram o conceito de RNA, conforme a Figura 7, inspirados no funcionamento do cérebro e do neurônio biológico [32].

Segundo Haykin [33], um modelo neural é composto por três elementos básicos: sinapses, um somador e uma função de ativação. As sinapses são caracterizadas por um peso, ou força, atribuído a elas, isto é, um sinal x_i aplicado à entrada da sinapse i , conectada ao neurônio j , é multiplicado pelo peso sináptico w_{ij} , modulando os sinais que passam pelas sinapses. O somador é o elemento que soma os sinais de entrada, ponderados pelas respectivas sinapses do neurônio. A função de ativação pode limitar a amplitude da saída de um neurônio e inserir uma não linearidade.

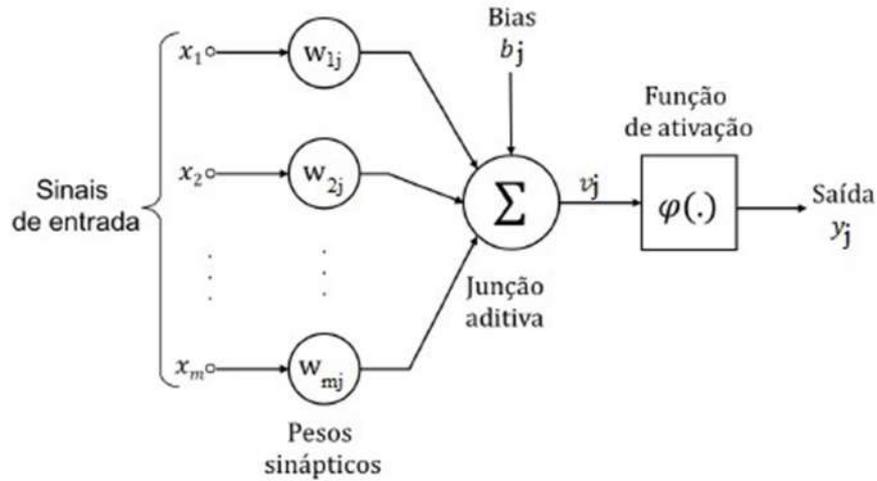


Figura 7 – Modelo básico de neurônio

A topologia da rede define a maneira com que os vários neurônios são interconectados. Como os neurônios possuem grande número de dendritos, eles podem receber sinais de vários neurônios adjacentes simultaneamente.

A entrada líquida é dada pela seguinte equação:

$$net_j = \sum_{i=1}^N w_{ij} x_i - b_j \quad (10)$$

Onde w_{ij} é o peso; x_i é o valor da entrada; e b_j é o limiar.

Em 1958, Rosenblatt [34] criou o algoritmo *perceptron* para reconhecimento de padrões. Neste modelo, os neurônios são organizados em camadas, uma de entrada e outra de saída e os pesos das conexões são adaptados a fim de se atingir a eficiência sináptica. Foi comprovado que este algoritmo é capaz de aprender apenas os padrões linearmente separáveis [35].

Posteriormente, o algoritmo *perceptron* foi aprimorado para o modelo *Multi Layer Perceptron - MLP*, que consiste tipicamente numa camada de entrada, uma ou mais camadas escondidas, e uma camada de saída. O Teorema da Aproximação Universal afirma que basta uma única camada escondida na rede MLP, contendo um número finito de neurônios e uma função de ativação não linear para obter um aproximador universal [36].

A Figura 8 representa um exemplo de rede MLP com duas camadas escondidas. O sinal de entrada se propaga através da rede, camada por camada, até alcançar a saída (sentido *forward*) [33]. Adicionalmente, o algoritmo *backpropagation*, que será detalhado na subseção 3.3.5, é utilizado para o treinamento da rede neural em prol da implementação da adaptação dos pesos.

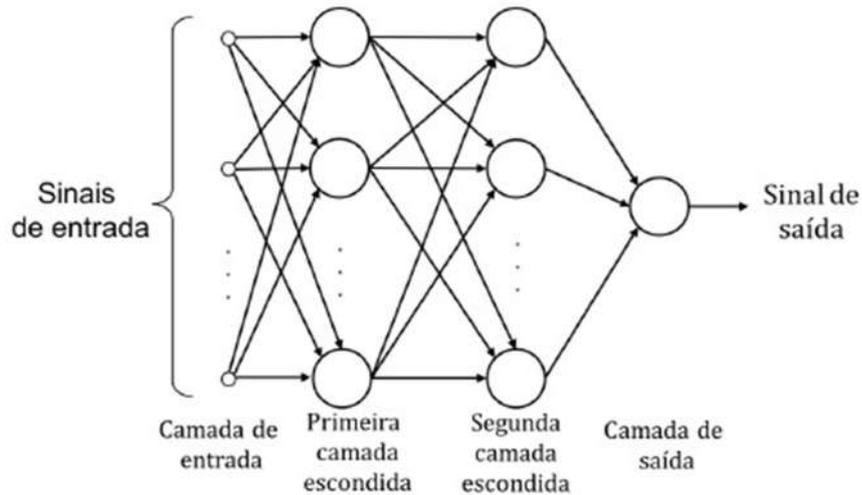


Figura 8 – Rede MLP

3.3.1. Função de Ativação

A função de ativação, ou função de transferência, $\varphi(.)$ de um neurônio artificial (Figura 7) tem como objetivo calcular o valor da saída deste neurônio, a partir do valor do potencial de ativação (v_j). Este potencial é responsável pela definição da amplitude do sinal de saída.

A seguir, estão as funções de ativação mais utilizadas: degrau (11), linear (12), tangente hiperbólica (13) e sigmoide (14).

- Função degrau

$$\varphi(v) = \begin{cases} 1, & \text{se } v \geq 0 \\ 0, & \text{se } v < 0 \end{cases} \quad (11)$$

- Função linear

$$\varphi(v) = \begin{cases} 1, & \text{se } v \geq +\frac{1}{2} \\ v, & \text{se } +\frac{1}{2} > v > -\frac{1}{2} \\ 0, & \text{se } v \leq -\frac{1}{2} \end{cases} \quad (12)$$

- Função tangente hiperbólica

$$\varphi(v) = \tanh(v) \quad (13)$$

- Função sigmoide

$$\varphi(v) = \frac{1}{1 + e^{-av}} \quad (14)$$

Onde a é o parâmetro de inclinação da sigmoide.

A função de ativação mais utilizada na construção de RNA é a função sigmoide quando se deseja um intervalo entre $[0,1]$ na saída. Ou, a função tangente hiperbólica, para saídas entre $[-1,1]$. Uma característica importante destas funções é que elas são deriváveis, diferentemente da função degrau [36].

Caso o intuito do modelo seja obter saídas contínuas, como é o caso de aplicações em regressão, geralmente faz-se uso de neurônios lineares na camada de saída. Por outro lado, para obter saídas discretas (classificação), utiliza-se neurônios não lineares.

O uso de pelo menos um neurônio linear na camada intermediária proporciona a ocorrência de um mapeamento de possíveis relações lineares que possam existir entre os dados de entrada e os da saída [26].

3.3.2. Normalização dos Dados

Geralmente, os dados de entrada passam pelo processo de normalização antes do processo de treinamento para evitar a saturação de seus valores [37], pois as funções de ativação não lineares podem restringir a amplitude dos valores de saída de um neurônio.

O processo de normalização consiste em garantir que os dados de entrada estejam dentro de um intervalo previamente estabelecido. Alguns exemplos de funções de normalização são:

- Transformação linear para o intervalo $[0,1]$

$$x_n = \frac{x_0 - x_{min}}{x_{max} - x_{min}} \quad (15)$$

- Transformação linear para o intervalo $[a,b]$

$$x_n = \frac{(b - a)(x_0 - x_{min})}{x_{max} - x_{min} + a} \quad (16)$$

- Normalização estatística

$$x_n = \frac{(x_0 - \bar{x})}{\sigma} \quad (17)$$

- Normalização simples

$$x_n = \frac{x_0}{x_{max}} \quad (18)$$

- Normalização pela tangente hiperbólica

$$x_n = C \tanh(Dx_0) \quad (19)$$

Onde x_n corresponde aos dados normalizados; x_0 aos dados originais; x_{min} , x_{max} , \bar{x} e σ , ao valor mínimo, valor máximo, à média e ao desvio padrão dos dados originais, respectivamente. As variáveis C e D são parâmetros que ajustam a intensidade da normalização pela tangente hiperbólica.

A normalização mais comum de ser utilizada é a estatística, onde as variáveis originais de entrada da rede são escaladas para a média nula e desvio padrão unitário. Entretanto, a normalização pode ser feita com uma transformação linear para o caso de dados com comportamento linear e, para o caso de dados com comportamento não linear, pode usar transformação logarítmica.

3.3.3. Funções Custo

Em 1950, surgiu o conceito de função custo na área de estatística [38] para representar uma medida de dissimilaridade entre a predição e o valor esperado (alvo). Sendo assim, a função custo é sempre utilizada para avaliar o desempenho de modelos

de previsão, já que o processo de treinamento supervisionado de uma RNA, por exemplo, consiste em minimizar o erro entre a saída predita e o valor desejado.

A função de erro mais comumente utilizada é o Erro Médio Quadrático (do inglês, *Mean Squared Error - MSE*). A seguir, estão todas as funções utilizadas na literatura [31].

- Erro Médio Quadrático - *MSE*

$$MSE = \frac{1}{N} \sum_{i=1}^N (e_i)^2 \quad (20)$$

- Raiz do Erro Médio Quadrático - *RMSE*

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (e_i)^2} \quad (21)$$

- Erro Médio Percentual Absoluto - *MAPE*

$$MAPE = \frac{100}{N} \sum_{i=1}^N \frac{|e_i|}{|x_i|} \quad (22)$$

- Correlação - *r*

$$r = \frac{cov(X, Y)}{\sqrt{var(X)}\sqrt{var(Y)}} \quad (23)$$

Onde $cov(X, Y)$ é a covariância entre X e Y e $var(X)$ e $var(Y)$ são as variâncias de X e Y .

O coeficiente de correlação de Pearson $r(X, Y)$ é uma medida da dependência linear entre os valores estimados (X) e os valores reais (Y), que varia no intervalo $[-1, 1]$. Os valores de correlação próximos dos limites indicam que existe significativa dependência linear entre as variáveis, por outro lado, os valores no entorno de zero indicam que não existe dependência.

Se calcularmos a estatística de r a partir de N pares (X_i, Y_i) , $i = 1, \dots, N$, verificaremos que a distribuição é normal, a média é efetivamente nula, e o desvio padrão é $\sigma_r = \frac{1}{\sqrt{N}}$. Isto significa que, num nível de confiança de 95%, os valores de

correlação de duas variáveis randômicas estarão entre $\pm 2\sigma_r = \pm \frac{2}{\sqrt{N}}$ e que praticamente nenhum valor excederá $\pm 3\sigma_r$. Sendo assim, conclui-se que existe dependência entre duas variáveis quando $|r| > \frac{2}{\sqrt{N}}$.

3.3.4. Configuração

Na literatura não existe uma regra específica que defina o número de camadas escondidas, nem o seu número de neurônios. Contudo, não é utilizado um grande número de camadas escondidas, porque o resultado se torna menos preciso com o aumento do erro médio na atualização dos pesos das sinapses. Com relação ao número de neurônios nas camadas escondidas, geralmente é definido empiricamente através das simulações.

A rede neural não pode ser tão pequena ao ponto de não armazenar todos os padrões do conjunto de treinamento e correr o risco de não convergir. Entretanto, se a rede for muito grande, ou seja, com muitos parâmetros (pesos), poderá conseguir modelar até o ruído.

É muito importante evitar o supertreinamento (*over-training* ou *over-fitting*), que se caracteriza pela especialização da rede em relação aos exemplos contidos nos dados de entrada, ou seja, evitar que a rede se torne sobredimensionada [39], através da implementação de um teste conhecido como validação cruzada (*cross-validation*) [40].

Basicamente, para alcançar o ponto de parada ótimo de treinamento, a validação cruzada interrompe o treinamento quando o erro no processo de validação começa a crescer de forma consistente [41] [42].

Para isso, os dados de entrada são divididos em três conjuntos independentes, o de treinamento para modificar os pesos, o conjunto de validação para verificar o problema do supertreinamento, e por último o conjunto de teste para dimensionar o desempenho do modelo.

Adicionalmente, para que a rede neural realize um bom aprendizado são necessários alguns processamentos nos dados de entrada antes de iniciar o treinamento, tais como a seleção baseada na correlação dos dados, já vista na subseção 3.3.3, e a

normalização deles, explicada na subseção 3.3.2. Estes procedimentos conseguem diminuir o tamanho da rede, acelerar o aprendizado e ainda melhorar a sua generalização.

3.3.5. Rede Multi Layer Perceptron com Backpropagation

O modelo *MLP* com algoritmo *backpropagation*, conhecida como rede neural supervisionada, é geralmente aplicada em método de otimização, conhecido em estatística como aproximação estocástica.

O processo de treinamento se baseia no método de gradiente descendente [43], e se divide em duas etapas. Na primeira etapa, uma entrada é aplicada na rede neural e seus efeitos são propagados no sentido *forward*, camada por camada, até chegar na última camada, e produzir um vetor de saída. Na segunda etapa, *backward*, os pesos sinápticos são ajustados de acordo com uma regra para correção do erro, e sempre buscar minimizar o erro global da camada de saída.

Resumidamente, a saída atual da rede neural é subtraída do valor esperado (alvo) gerando um erro, que é propagado no sentido oposto das conexões dos pesos sinápticos (retropropagação do erro, do inglês, *error backpropagation*). Por último, os pesos sinápticos são somados ao fator Δw_{ij} , representado em (24), para reduzir o erro, isto é, aproximar a saída atual do valor esperado. Ou seja, cada peso sináptico é atualizado proporcionalmente ao negativo da derivada parcial do erro com relação ao peso.

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \quad (24)$$

Onde w_{ij} é o peso sináptico de um neurônio, E é uma função custo que mede a dissimilaridade entre a saída atual e o valor esperado, e η é a velocidade de treinamento (*learning rate*), cujo parâmetro define o tamanho do passo de treinamento e, conseqüentemente, da atualização do peso w_{ij} .

Em prol de evitar a convergência do treinamento em algum mínimo local, deve-se introduzir na equação uma constante chamada momento, α , que multiplica a variação de

pesos da iteração anterior, permitindo o aumento da taxa de aprendizagem sem que ocorram oscilações nos pesos ajustados, além de atuar no aumento da velocidade de convergência.

$$w \leftarrow w - \eta \cdot \nabla E(w) + \alpha \cdot \delta w_{old} \quad (25)$$

O erro máximo e o número de iterações são parâmetros que são definidos pelo usuário antes de iniciar a execução do algoritmo de treinamento da rede. Dessa forma, o procedimento de aprendizado é repetido diversas vezes até alcançar o limite de iterações, enquanto o erro permanecer menor do que o máximo especificado.

O algoritmo de *backpropagation* apresenta ótimos resultados na maioria dos casos, mas deve-se atentar em evitar problemas, tais como a possível convergência em mínimos locais, o processo longo de aprendizado (minimizado com novas escolhas para η e α), e fenômenos como paralisia do treinamento (*network paralysis*), que é contornada diminuindo-se o valor de η .

3.3.6. Rede Multi Layer Perceptron com Levenberg-Marquardt

O algoritmo LM é uma variação do *backpropagation* padrão por utilizar uma aproximação pelo método de Newton, em vez da descida de gradiente como método de aproximação do mínimo da função erro.

Esta aproximação é obtida a partir da modificação do método de *Gauss-Newton* introduzindo-se o parâmetro μ , conforme a equação a seguir, proporcionando maior velocidade de convergência e tornando-se num dos métodos mais rápidos para o treinamento de redes neurais *feedforward* de tamanho moderado:

$$\Delta w = (J^T J + \mu I)^{-1} J^T \varepsilon \quad (26)$$

Onde:

- w representa a diferença entre os pesos inicial e final;

- J é o jacobiano dos erros da camada de saída. Cada elemento desta matriz J representa uma derivada parcial de um elemento da matriz de erros com o seu correspondente peso;
- I é a matriz identidade. Como ela é multiplicada pela constante α , gera-se uma nova matriz contendo como elementos apenas valores de α ;
- μ é um escalar que controla a derivação dos erros, permitindo que o termo (JTJ) possa ser invertido;
- ε é o vetor de erros calculados.

3.3.7. Rede Non-linear Sigmoidal Regression Blocks Networks

A rede *Non-linear Sigmoidal Regression Blocks Networks* - NSRBN, proposta por [44] e usada pelo modelo Neuro3M, atualmente implementado pelo ONS, é composta por blocos de polinômios homogêneos, na qual as unidades escondidas desses blocos realizam uma regressão sigmoide linear das entradas (tal qual uma rede MLP), porém, na camada de saída, há uma ativação de regressão logística não linear com relação aos neurônios escondidos. A Figura 9 abaixo mostra a arquitetura de uma rede NSRBN.

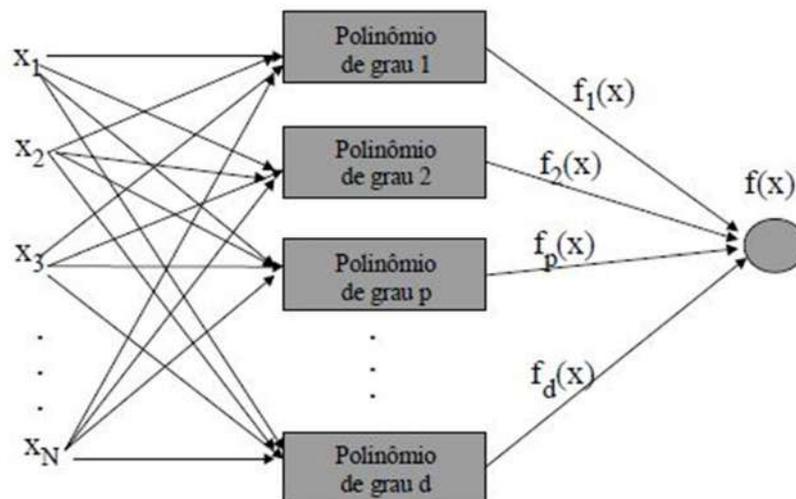


Figura 9 – Rede NSRBN [1]

A rede NSRBN é formada por um algoritmo do tipo construtivo (arquitetura dinâmica), onde o número de camadas e unidades de processamento varia gradualmente

durante o treinamento até que se atinja o valor mínimo global com uma estrutura de modelo de tamanho ótimo.

O modelo tem como princípio manter inalterados os valores dos pesos de todos os blocos anteriores ao atual, de tal forma que o número de pesos a ser otimizado, a cada iteração, não aumente explosivamente, como no algoritmo de análise combinatória tradicional.

Qualquer função contínua, $f(x)$, pode ser uniformemente aproximada quando dividida em blocos homogêneos de ordem 1 até o bloco de ordem d . Dessa forma, a função total (polinômio completo) pode ser aproximada, de maneira incremental, pela soma das funções parciais $\hat{f}_p(x), p = 1, \dots, d$ (polinômios homogêneos), tal como a equação abaixo:

$$f(x) = g(\hat{f}_1(x) + \hat{f}_2(x) + \dots + \hat{f}_d(x)) \quad (27)$$

A variável p representa um passo do algoritmo para o qual o bloco $\hat{f}_p(x)$ é adicionado à rede, e a variável g representa uma função de ativação adequada para a unidade de saída.

4 DESCRIÇÃO DO ESTUDO DE CASO

4.1 Bacia do São Francisco

A ANEEL divide a bacia do rio São Francisco, localizado no estado de Minas Gerais, em nove sub-bacias. A região estudada corresponde a área de drenagem do Alto São Francisco, especificamente, a área de drenagem do reservatório de Três Marias, conforme a Figura 10. Segundo a Agência Nacional de Águas - ANA, o Alto São Francisco se localiza numa zona climática tropical e estende-se desde a cidade de São Roque de Minas, nas cabeceiras da Serra da Canastra, até Pirapora (MG).



Figura 10 – Localização das sub-bacias da região do Alto São Francisco [45]

A usina hidrelétrica de Três Marias, com potência instalada de 396 MW, é de propriedade da empresa CEMIG. O seu reservatório de armazenamento de água tem uma área de drenagem de 50.732 km², com níveis operativos, máximo e mínimo,

respectivamente, de 572,50 m e 549,20 m, e Vazão Média de Longo Termo - MLT² de 689 m³/s [1].

O Alto São Francisco apresenta topografia ligeiramente acidentada, com serras e terrenos ondulados e altitudes começando em 1.600 m e diminuindo até alcançar 600 m. O clima predominante na região é o tropical úmido e o temperado de altitude, e a região contribui com vazão média anual de 42%.

As características climáticas mais importantes na região em estudo são as temperaturas de mediana à elevada durante todo o ano, e a existência de duas estações bem marcadas: uma estação chuvosa no verão e outra seca no inverno.

Com relação às precipitações, os totais pluviométricos anuais variam, de forma geral, entre 1.700 mm nas cabeceiras da bacia à 1.000 mm na confluência do rio São Francisco com o rio das Velhas. O trimestre mais chuvoso contribui em média com cerca de 55% a 60% do total anual precipitado, correspondendo, para quase a totalidade da bacia, aos meses de novembro, dezembro e janeiro. A exceção ocorre na extremidade ocidental da bacia, onde os meses mais chuvosos são dezembro, janeiro e fevereiro. Já o trimestre mais seco, correspondente aos meses de junho, julho e agosto, contribuindo em média com menos de 5% da precipitação anual.

Por se situar na região Sudeste, o Alto São Francisco é afetado pela passagem de sistemas climáticos frontais, e pela interação desses com a convecção tropical, denominada de Zona de Convergência do Atlântico Sul - ZCAS. Um percentual significativo da precipitação nesse trecho da bacia deve-se também à atuação de sistemas de escala subsinótica [46], tais como linhas de instabilidade, que ocasionam precipitação. Estas são de curta duração, em oposição às associadas às frentes e à ZCAS que duram alguns dias e abrangem áreas maiores.

A floresta é predominante na região úmida, ao longo dos rios e riachos, onde ocorre maior umidade do solo, formando floresta de galerias ou mata ciliar. O percurso dos rios da bacia também passa por regiões subúmidas e transicionais para semi-árido, finalizando nas regiões de solos de baixa fertilidade, caracterizado pelo Cerrado.

² MLT é a média de ENA calculada com base em uma série histórica desde 1931. Já a ENA, é a energia que pode ser produzida com a vazão de água de um determinado rio a um reservatório de uma usina hidrelétrica.

4.2 Previsão de Vazões Afluentes Naturais em Três Marias

A vazão natural é considerada como as vazões obtidas através do balanço hídrico do reservatório, adicionando evaporação no reservatório, bem como, vazões consuntivas da área de drenagem do reservatório. Sempre procurando retirar a interferência do homem na natureza.

Para o subsistema Nordeste, o ONS divulga no PMO as previsões de vazões naturais dos seguintes aproveitamentos principais da Bacia do São Francisco: Sobradinho, Itaparica, Complexo Paulo Afonso, Xingó, Boa Esperança, Pedra do Cavalo e Itapebi.

Especificamente para os aproveitamentos da bacia do São Francisco, indicados na Figura 11, a vazão natural para o reservatório de Sobradinho é formada pela vazão natural de Três Marias e de Queimado, adicionada à vazão incremental do trecho Três Marias/Queimado até Sobradinho.



Figura 11 – Disposição dos aproveitamentos da Bacia do São Francisco [47]

Já as vazões naturais dos aproveitamentos de Itaparica (Luiz Gonzaga), Complexo Paulo Afonso e Xingó são calculadas levando em consideração a vazão natural de Sobradinho, adicionando-se as vazões incrementais em cada trecho.

Dessa forma, as previsões de vazões afluentes naturais de Sobradinho, e conseqüentemente para toda a cascata, são afetadas pela precisão das previsões realizadas para Três Marias, Queimado e o trecho incremental de Três Marias/Queimado até Sobradinho.

Como o reservatório de Três Marias, localizado na cabeceira do rio São Francisco, possui maior representatividade em relação ao reservatório de Queimado, com representação de 26% das aflúncias naturais do rio São Francisco [1], a previsão

de vazões afluentes naturais e a operação do aproveitamento de Três Marias são de extrema relevância no processo de operação dos demais reservatórios a jusante.

Conforme [1], para o reservatório de Três Marias, o ONS faz uso do modelo de rede neural do tipo NSRBN, denominado Neuro3M, descrita na subseção 3.3.7, para obter as previsões das vazões naturais médias diárias (séries temporais não estacionárias). A calibração desse modelo foi determinada para ocorrer a cada 10 anos, entretanto a última calibração foi realizada em 2005 (com dados entre 1995 e 2005), quando na verdade, deveria ter sido realizada uma recalibração em 2015 [11].

Em 2018, o ONS emitiu a [11] para explicar os motivos para o qual decidiu retornar com a padronização para todas as bacias do SIN do uso de apenas um único modelo de previsão de vazões, cuja escolha foi pelo modelo SMAP/ONS e sua implantação será gradativa com término previsto para 2020.

Nesta nota técnica foi revelado que a generalização visa trazer maior qualidade e agilidade para a previsão de vazões dos aproveitamentos do SIN, além de tornar o processo mais reprodutível e de mais fácil compreensão pelos agentes do SEB. A escolha pelo SMAP/ONS foi principalmente devido ao fato de que este modelo já estava em uso em 33% das bacias e apresentando bons resultados, além de ser de fácil compreensão metodológica e de código aberto e livre.

4.3 Postos Fluviométricos e Pluviométricos

As previsões de vazões naturais no ONS são elaboradas de forma direta por um modelo matemático para os locais denominados como postos base de telemedição, ou, de forma indireta para os postos não base, a partir de regressões lineares obtidas a partir das previsões realizadas para um posto base próximo.

A definição dos postos base e a obtenção das equações de regressão para o cálculo da vazão prevista dos postos não base são oriundos do estudo da série histórica de vazões de cada aproveitamento e das relações estatísticas entre postos de uma determinada bacia.

As informações pluviométricas e fluviométricas médias diárias de parte da bacia do Alto São Francisco fazem parte da composição do Banco de Dados Técnicos - BDT do ONS e são usadas para a realização das previsões de vazões naturais do reservatório de Três Marias.

Os postos pluviométricos e fluviométricos que foram utilizados para o treinamento da rede em uso atualmente pelo ONS, com dados de 1987 até 2007, foram os da CEMIG, com complementações de informações da ANA, tal como são revelados na Tabela 1.

Tabela 1 – Postos Pluviométricos e Fluviométricos selecionados

POSTOS DE MEDIÇÃO	
PLUVIAIS	FLUVIAIS
Três Marias Barramento	Reservatório Três Marias -TM
Porto das Andorinhas	Porto das Andorinhas - PA
Iguatama	Iguatama - IG
Porto Pará	Porto Pará - PR
Ponte dos Vilelas	Ponte dos Vilelas Jusante - PV
Porto Mesquita	Porto Mesquita - ME
Porto Indaiá	Porto Indaia - ID
Dores do Indaiá	
Cajuru Barramento	
Vargem Bonita	
Fazenda Campo Grande	
Pitangui	

O trabalho realizado pelo ONS de seleção dos postos e de análise das informações recebidas se baseou num banco de dados que não poderia conter muitas falhas. Sendo assim, as informações foram submetidas a uma análise mensal de consistência pelo método estatístico do vetor regional. E com relação ao preenchimento diário de falhas, no caso de dados pluviométricos, ocorreu pelo método estatístico da ponderação

regional com os três postos de maior correlação. E para os dados fluviométricos, o preenchimento foi realizado através da regressão linear múltipla a partir do posto mais correlacionado.

A Figura 12 ilustra a distribuição espacial desses postos, bem como as áreas de drenagem existentes a montante do reservatório de Três Marias.



Figura 12 – Sub-bacias da área de drenagem de Três Marias [1]

Conforme a [11], os dados de entrada do modelo de previsão são os totais diários de precipitação previstos, a média dos dados diários de precipitação observados a montante do Reservatório de Três Marias, e as vazões diárias observadas dos postos fluviais a montante, bem como do próprio reservatório.

O cálculo da média dos dados diários de precipitação observados na área da bacia pondera o peso de cada estação pluviométrica, ou seja, os coeficientes são proporcionais à área de influência de cada posto. Entretanto, conforme divulgação nos Boletins Diários do ONS e seus PMO, esses coeficientes ainda não são usados, tornando a média ponderada em média aritmética.

5 METODOLOGIAS E APLICAÇÕES

5.1 Base de Dados

Neste trabalho foram usadas informações de estações pluviométricas e fluviométricas presentes à montante do reservatório de Três Marias, denominadas séries auxiliares, além das vazões naturais afluentes do próprio reservatório, representando a série alvo. Todas essas informações utilizadas correspondem às já existentes na base de dados históricos consistidos do ONS.

O banco BDT do ONS possui informações de vazões atualizadas desde 1931 e de precipitações desde 1987. O presente trabalho de pesquisa obteve o referido histórico até o fim do ano 2016, através da Gerência de Hidrologia da empresa. Como os dados são diários e o excesso de informações acarretaria um alto e lento processamento, foram selecionados alguns anos, dentre os últimos 12 anos, para treinar, validar e testar a rede. Sendo assim, foram escolhidos os três anos mais afluentes e os três anos menos afluentes, totalizando assim 2.080 dias para cada posto de medição.

Dessa forma, os anos estudados foram definidos tal como descrito abaixo:

- Período seco: 2005, 2006 e 2011 representando os anos com as maiores vazões, e 2013, 2014 e 2015 com as menores vazões; e
- Período úmido: 2005-2006, 2006-2007 e 2011-2012 com as maiores vazões, e 2012-2013, 2013-2014 e 2014-2015 com as menores vazões.

Foram criadas duas redes distintas, de forma a separar os períodos hidrológicos secos e úmidos, devido às grandes diferenças comportamentais entre eles, tal como consta na Figura 13 o exemplo das vazões históricas do reservatório de Três Marias (série alvo) durante todo o ano de 2005, cujos períodos úmidos se localizam nas extremidades do ano. Da mesma forma, a Figura 14 revela o comportamento de todas as vazões dos postos fluviométricos selecionados à montante (séries auxiliares) do mesmo ano.

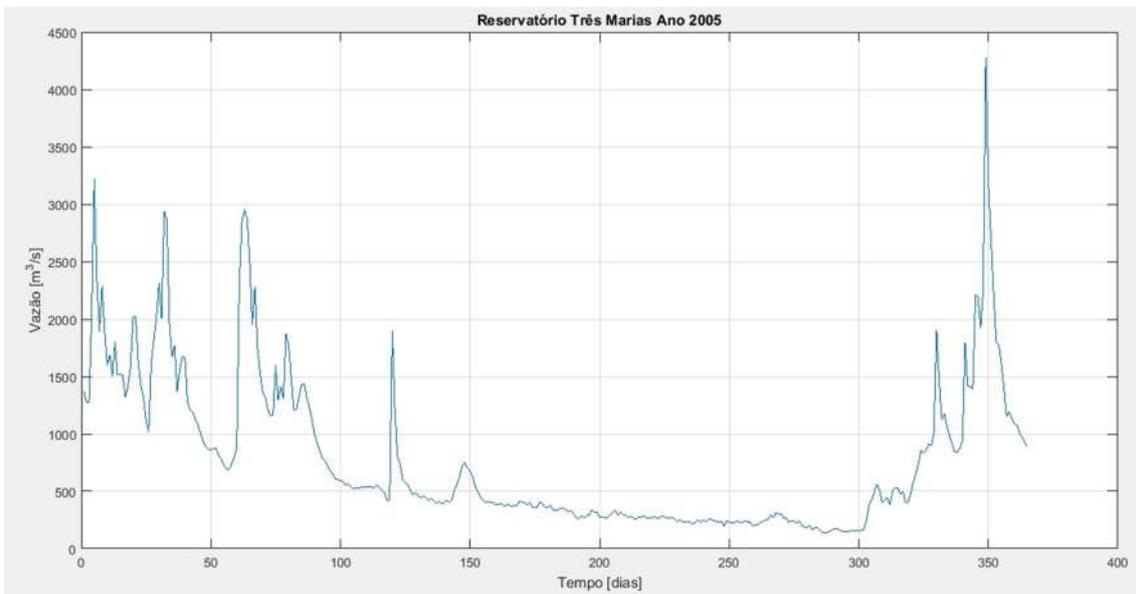


Figura 13 – Vazões históricas do reservatório de Três Marias do ano 2005

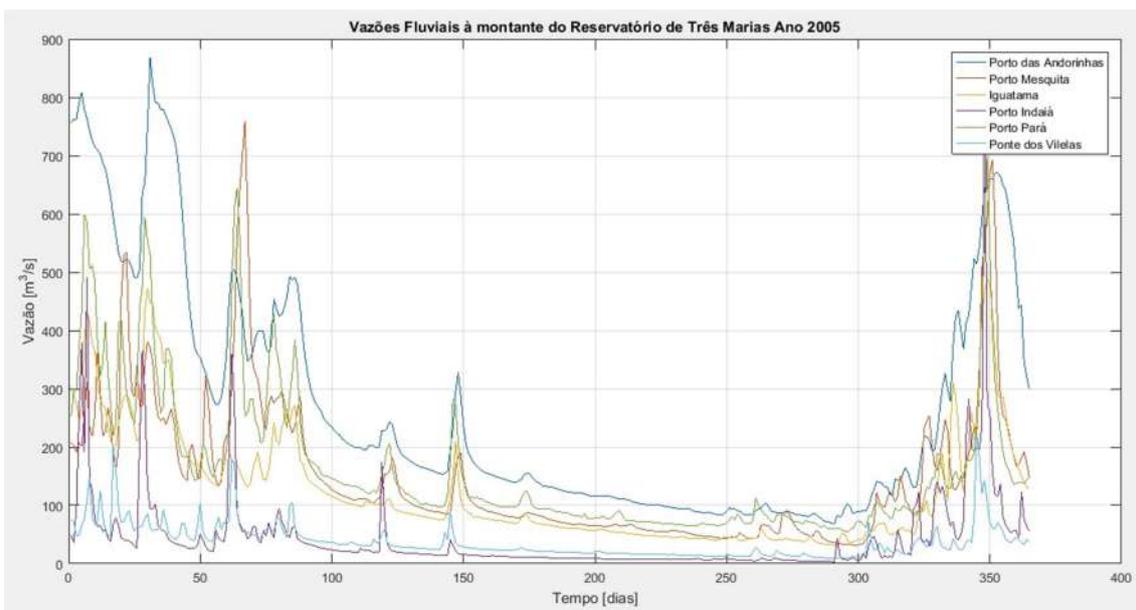


Figura 14 – Vazões históricas dos postos fluviométricos à montante do ano 2005

Para cada ano foi realizada uma análise das séries, através de seus gráficos, para determinar quando os períodos se iniciam e terminam, e a definição foi resumida na Tabela 2.

Tabela 2 – Divisão dos períodos hidrológicos anuais

PERÍODO	ANO	DATA	DIAS
Seco	2005	07/06 - 16/09	102
	2006	01/05 - 13/09	136
	2011	01/05 - 30/09	153
	2013	10/06 - 21/09	104
	2014	08/08 - 14/10	68
	2015	21/05 - 24/08	96
Úmido	2005-2006	17/09 - 30/04	226
	2006-2007	14/09 - 08/09	360
	2011-2012	01/10 - 30/04	212
	2012-2013	25/10 - 30/04	188
	2013-2014	06/09 - 30/04	237
	2014-2015	15/10 - 30/04	198

A prerrogativa adotada foi primeiramente identificar o período seco, cuja definição se restringiu ao simples comportamento médio de declínio linear das vazões naturais, e, conseqüentemente por exclusão, estabelecer o período úmido. As Figuras 15, 16 e 17 exemplificam o detalhamento de uma análise realizada, no qual foi definido o início e o fim do período seco de 2006 para somente depois estabelecer os períodos úmidos adjacentes.

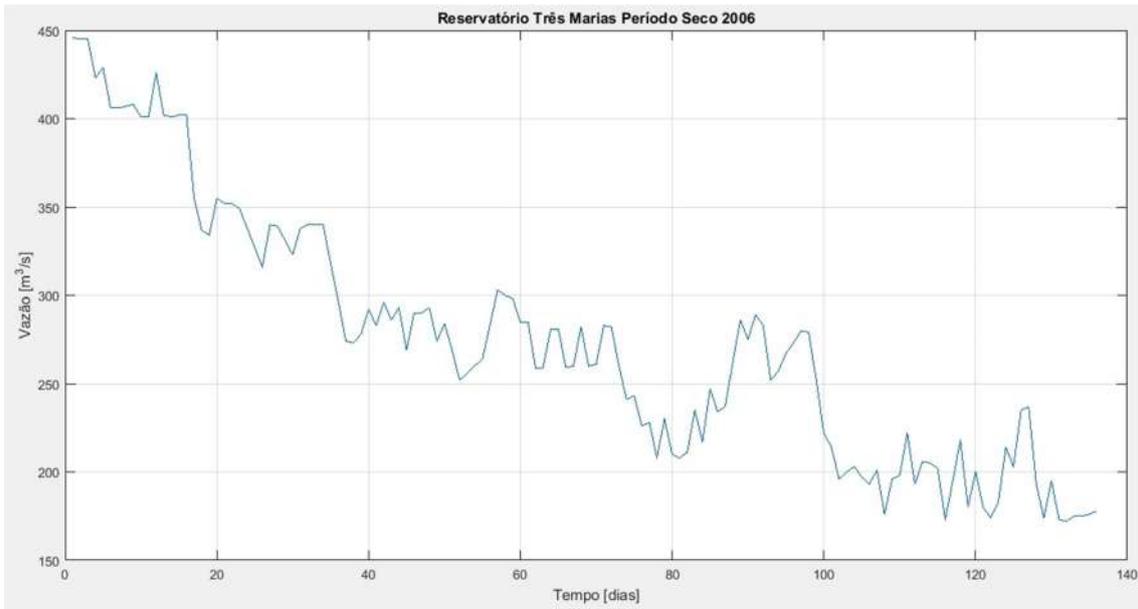


Figura 15 – Vazões históricas do reservatório de Três Marias do Período Seco 2006

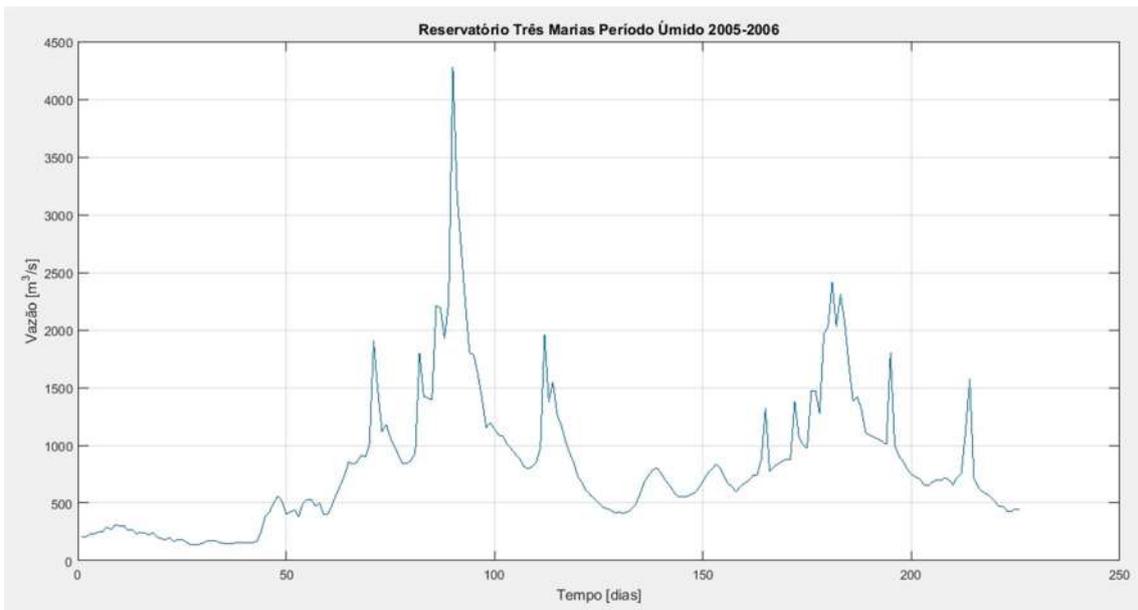


Figura 16 – Vazões históricas do reservatório de Três Marias do Período Úmido 2005-2006

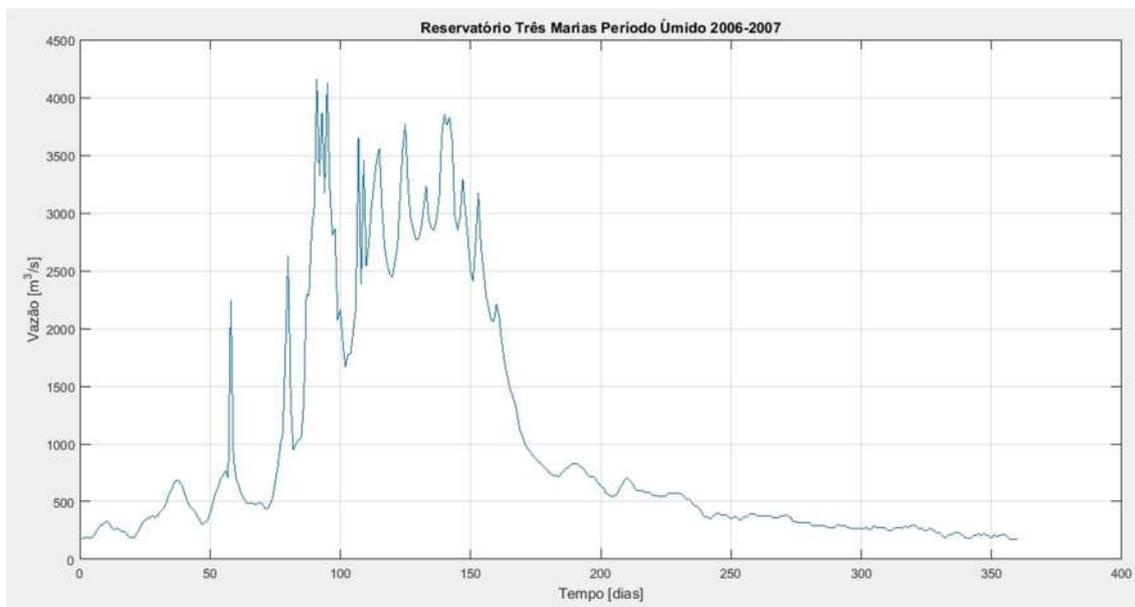


Figura 17 – Vazões históricas do reservatório de Três Marias do Período Úmido 2006-2007

5.1.1 Informações Pluviométricas e Fluviométricas

Os postos fluviais utilizados pelo ONS, conforme já explicitado na subseção 4.3, bem como os adotados neste trabalho, são os que seguem: Reservatório de Três Marias - TM, Porto das Andorinhas - PA, Porto Mesquita - ME, Iguatama - IG, Ponte dos Vilelas - PV, Porto Pará - PR, Porto Indaiá - ID.

O cálculo da média aritmética simples pluvial diária, representado no presente trabalho pela sigla MdPI, realizado pelo ONS, é composto pelas precipitações diárias dos seguintes postos de precipitação localizados à montante do reservatório de Três Marias: Cajuru Barramento; Iguatama montante; Vargem Bonita; Dores do Indaiá; Fazenda Campo Grande; Ponte dos Vilelas; Porto das Andorinhas; Porto do Indaiá; Porto Mesquita; Porto Pará; Três Marias Barramento; Pitangui.

Inicialmente, para reduzir a grande variabilidade das informações e dar a mesma importância a todas, tornando os valores mais relativos entre si, foi realizado o logaritmo neperiano (ou logaritmo natural) das séries temporais de vazões para cada posto de medição fluvial, tal como revelam as Figuras 18 e 19.

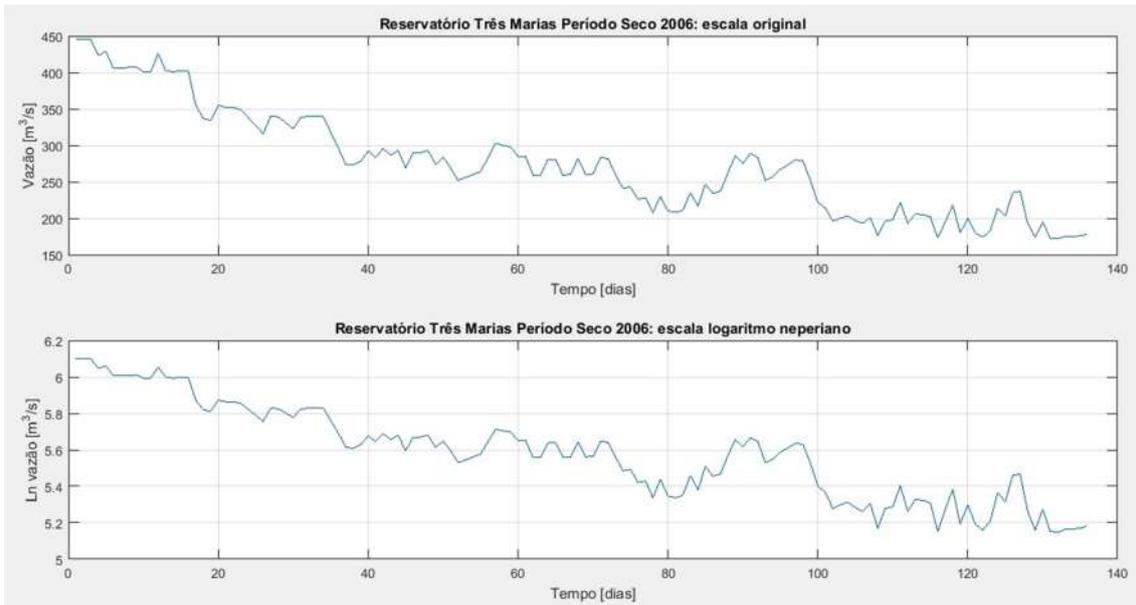


Figura 18 – Logaritmo neperiano das vazões históricas do reservatório de Três Marias do Período Seco 2006

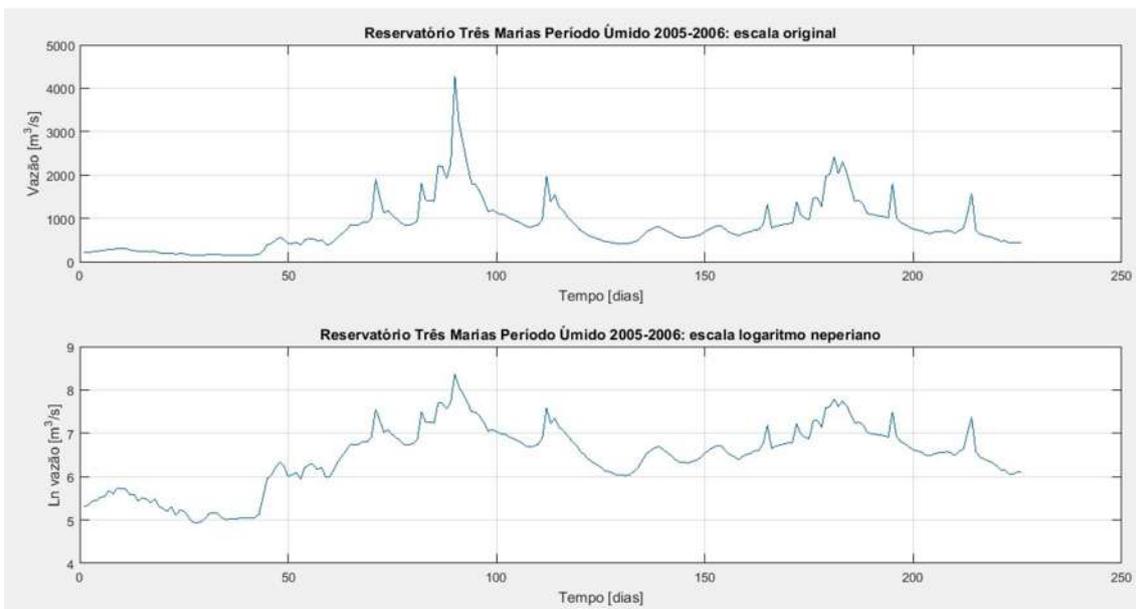


Figura 19 – Logaritmo neperiano das vazões históricas do reservatório de Três Marias do Período Úmido 2005-2006

As séries temporais de precipitação não precisaram ter seus valores originais transformados, pois já não apresentavam diferenças tão significantes, tal como exemplificam as Figuras 20 e 21.

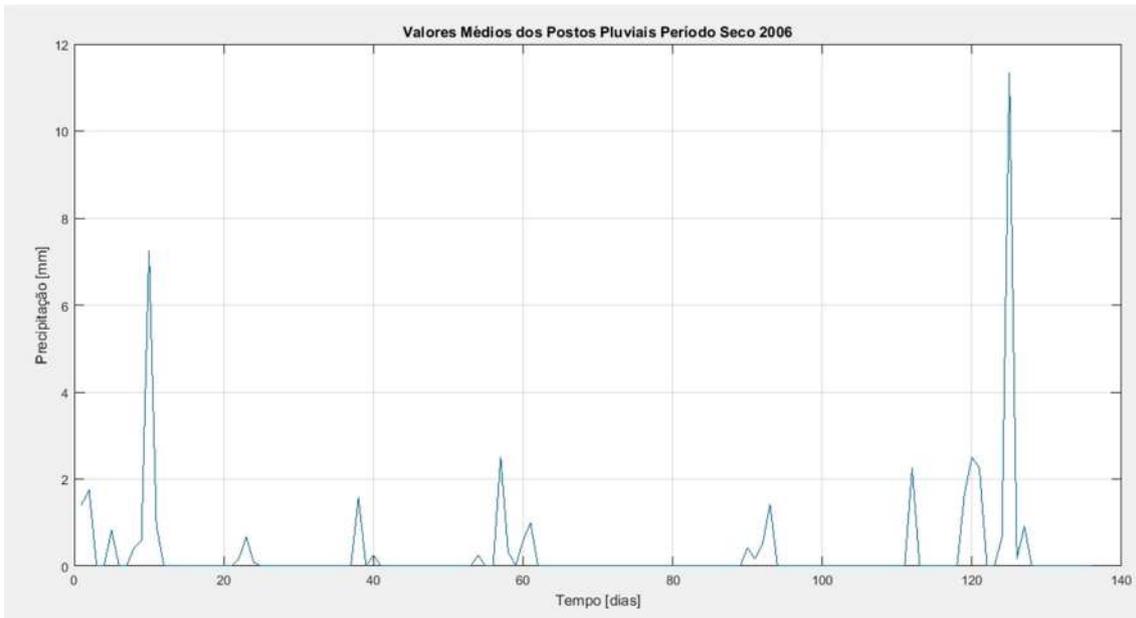


Figura 20 – Valores médios dos postos pluviais do Período Seco 2006

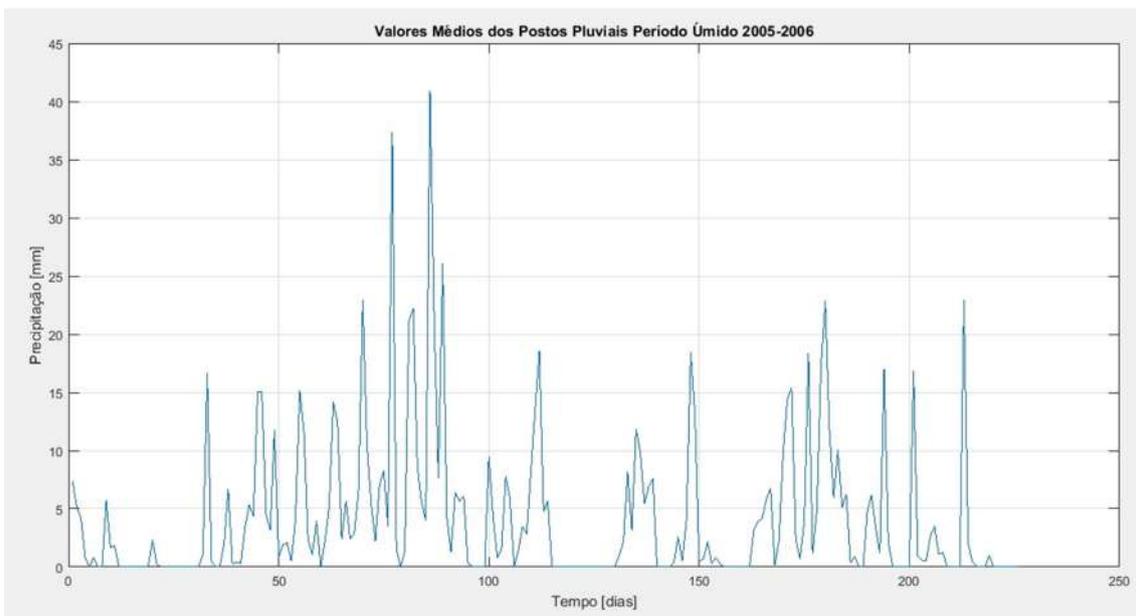


Figura 21 – Valores médios dos postos pluviais do Período Úmido 2005-2006

5.1.2 Decomposição da Série Temporal

Como citado na subseção 3.2.1, uma série temporal pode ser composta por:

- Componentes determinísticas (tendência e sazonalidade);

- Componentes não lineares; e
- Ruídos não correlatos (imprevisíveis).

O objetivo da decomposição é extrair as componentes determinísticas (tendência e sazonalidade), gerando uma série residual, formada pelas componentes não lineares e pelos ruídos não correlatos. Esta série residual é utilizada para treinar o modelo de modo que as componentes não lineares sejam previstas. E ao término da predição, realiza-se uma combinação do resultado da previsão não linear com as componentes determinísticas para compor a predição final.

Conforme análises mostradas no capítulo 6, através do auxílio de um conjunto de ferramentas analíticas, constatou-se que as séries não apresentam sazonalidade, mas sim, ciclos senoidais. Esses, por sua vez, não foram extraídos da série original, pois as frequências dos ciclos senoidais com maior densidade de probabilidade não apresentam um padrão de repetição de frequência entre os anos, revelando um comportamento não estacionário.

O processo adotado de decomposição foi diferente para cada período hidrológico. Nos períodos secos, foi realizada a decomposição da série temporal de vazões para cada ano, iniciada pela estimação da sua tendência, e conseqüentemente com a sua extração da série original. Resultando, portanto, numa série temporal de vazões sem tendência.

Foi implementado um algoritmo para estimar a tendência linear de cada um dos anos selecionados para cada posto de medição, utilizando a função de regressão linear polinomial de 1º grau da biblioteca *Curve Fitting* (cftool) do software *MathWorks MatLab*, pois esta apresentou uma boa capacidade de generalização para os pontos futuros. A Figura 22 exemplifica o processo de extração da tendência das vazões do Reservatório de Três Marias no período seco de 2006.

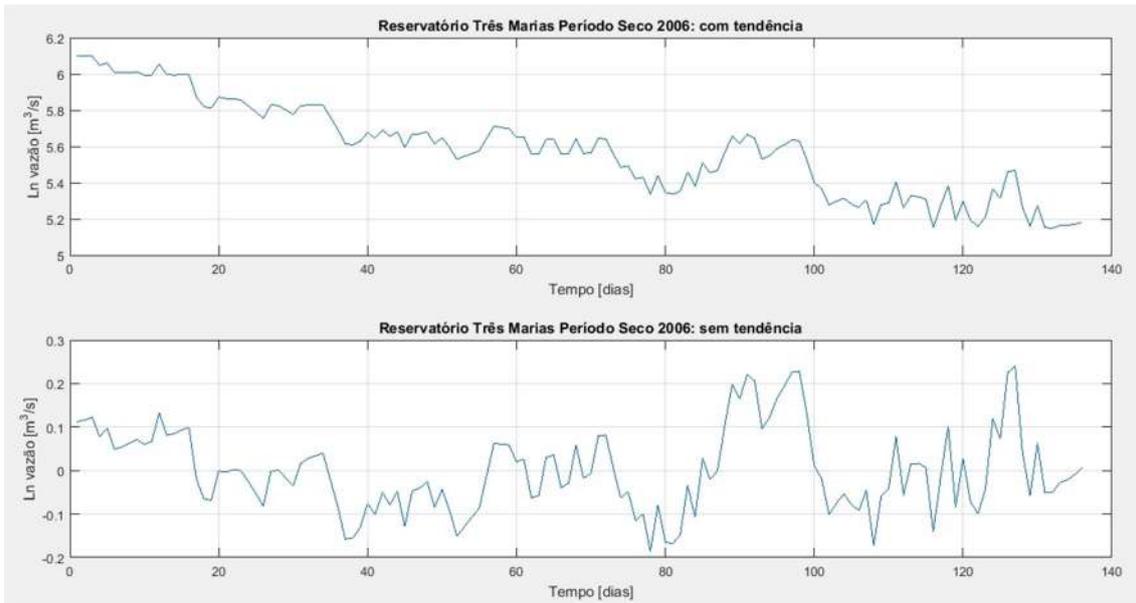


Figura 22 – Vazões históricas sem Tendência do reservatório de Três Marias do Período Seco 2006

Com relação ao período hidrológico úmido, foram obtidos bons resultados através da simples extração do valor médio (eliminando a componente DC, do inglês, *Direct Current*) das séries de vazões para cada ano selecionado de cada posto, em vez de tentar estimar uma tendência. Como efeito de comparação, a Figura 23 exemplifica essa extração realizada no período úmido 2005-2006.

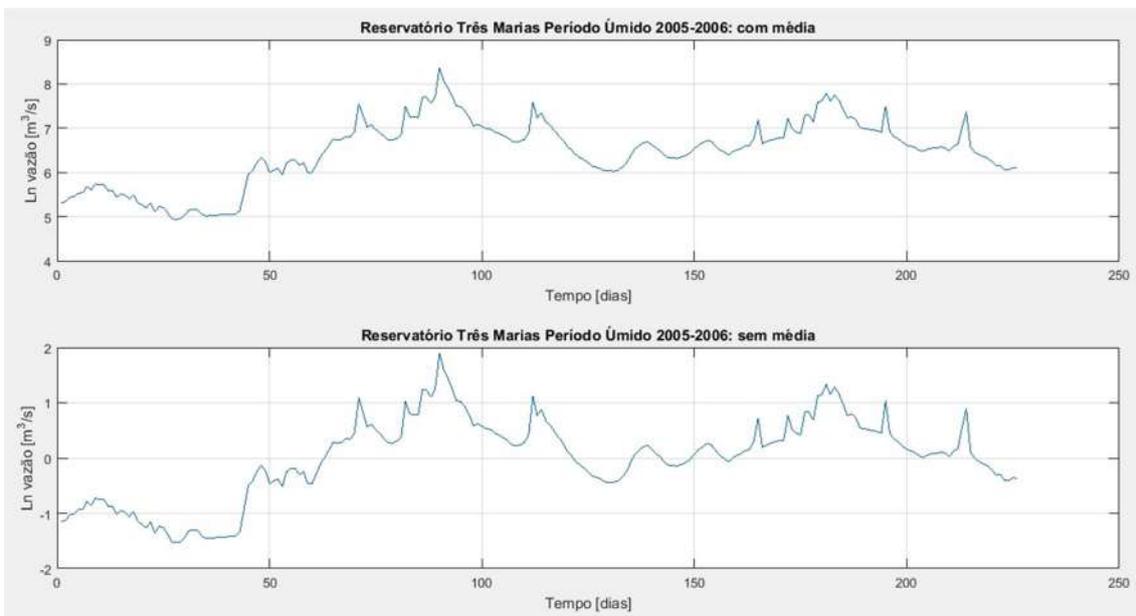


Figura 23 – Vazões históricas sem Média do reservatório de Três Marias do Período Úmido 2005-2006

Tratando-se dos dados de precipitações dos postos de medição à montante do reservatório de Três Marias, a tendência também não foi extraída e ambos os períodos tiveram suas médias retiradas da informação original, conforme exemplifica a Figura 24.

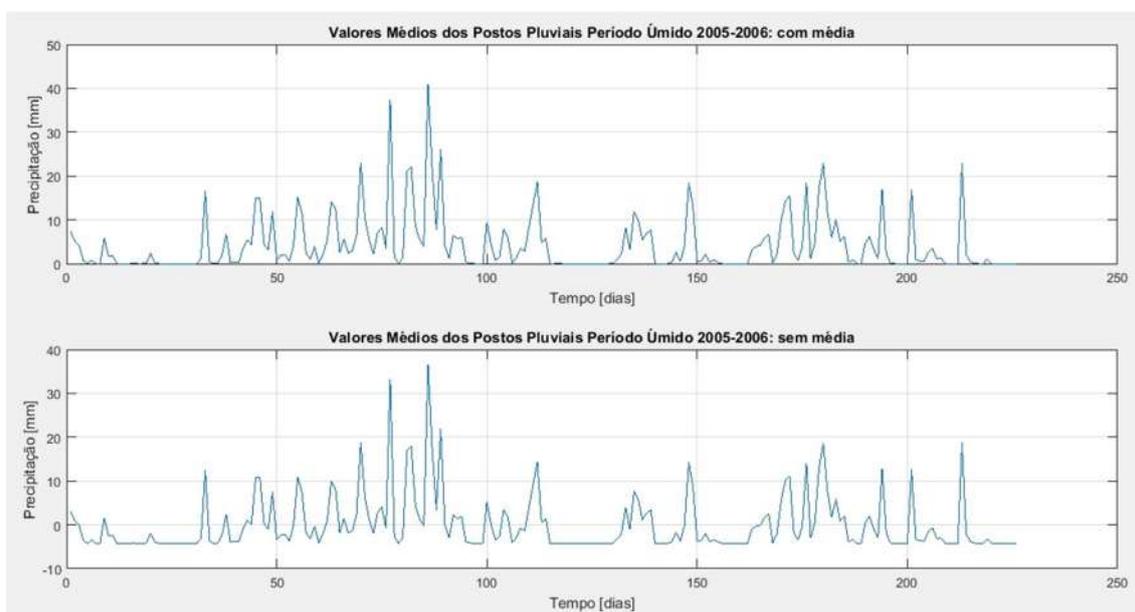


Figura 24 – Valores médios dos postos pluviais sem Média do Período Úmido 2005-2006

5.1.3 Análise Estatística

Após a obtenção das séries residuais de todos os postos de medição para cada ano selecionado, foi necessário realizar a escolha da defasagem temporal (*lag*) de cada entrada da rede RNA. Para fazer esta escolha, foi utilizada a Função de Autocorrelação Parcial - FACP da vazão do próprio reservatório de Três Marias, bem como a Função de Correlação Cruzada - FCC da vazão natural de Três Marias com os postos fluviométricos e com a precipitação média da área, obtendo uma janela de defasagem que varia entre 7 e 10 dias.

A participação de postos que conseguiram preservar a dependência da série com no mínimo 7 dias de defasagem foi essencial para possibilitar a obtenção da previsão de vazão natural de até 7 dias. Na Tabela 3, consta a relação dos postos de medição com as respectivas defasagens temporais definidas, cujas seleções basearam-se nos melhores

resultados alcançados, onde t corresponde à variável no dia da previsão, $t-1$ com um dia de defasagem, $t-2$ com dois dias de defasagem, e assim por diante.

Tabela 3 – Relação de postos com suas defasagens temporais

	PERÍODO SECO					
POSTOS DE MEDIÇÃO	2005	2006	2011	2013	2014	2015
Reservatório Três Marias - TM	t-7	t-7	t-7	t-7	t-7	t-7
Porto das Andorinhas - PA	t-7	t-7,t-8,t-9	t-7,t-8,t-9	t-7	t-7	t-7,t-8,t-9
Iguatama - IG	t-7	t-8,t-9,t-10	t-7,t-8,t-9	t-7,t-8,t-9	t-7,t-8,t-9	t-7,t-8,t-9
Porto Pará - PR	t-8	t-8,t-9,t-10	t-7,t-8,t-9	t-7,t-8,t-9	t-7,t-8,t-9	t-7,t-8,t-9
Ponte dos Vilelas Jusante - PV	t-8,t-9,t-10	t-8,t-9,t-10	t-7,t-8,t-9	t-7,t-8,t-9	t-8	t-7,t-8,t-9
Porto Mesquita - ME	t-7,t-8,t-9	t-7,t-8,t-9	t-7,t-8,t-9	t-7	t-7,t-8,t-9	t-7,t-8,t-9
Porto Indaia - ID	t-7	t-7,t-8,t-9	t-7	t-7,t-8,t-9	t-8	t-8
Média Postos Pluviais - MdPl	t-7	t-7	t-7	t-7	t-7	t-7
	PERÍODO ÚMIDO					
POSTOS DE MEDIÇÃO	2005-2006	2006-2007	2011-2012	2012-2013	2013-2014	2014-2015
Reservatório Três Marias - TM	t-7	t-7	t-7	t-7	t-7	t-7
Porto das Andorinhas - PA	t-7,t-8,t-9	t-7,t-8,t-9	t-7,t-8,t-9	t-7,t-8,t-9	t-7,t-8,t-9	t-7,t-8,t-9
Iguatama - IG	t-7,t-8	t-7,t-8	t-7,t-8	t-7,t-8	t-7,t-8	t-9,t-10
Porto Pará - PR	t-7	t-7	t-7	t-7	t-11	t-7
Ponte dos Vilelas Jusante - PV	t-7	t-12	t-10	t-7	t-7	t-7
Porto Mesquita - ME	t-7,t-8	t-7,t-8	t-7,t-8	t-7,t-8	t-9,t-10	t-7,t-8

Porto Indaia - ID	t-7	t-10	t-7	t-7	t-10	t-9
Média Postos Pluviais - MdPI	t-7	t-7	t-7	t-7	t-7	t-7

Dependendo do ano, pode variar entre 12 e 20 a quantidade de séries de leitura na camada de entrada da rede. Vale ressaltar que a rede RNA supõe que todas as entradas são independentes entre si no tempo, isto é, não há uma noção de temporalidade entre elas.

5.1.4 Normalização

O processo de normalização faz parte da fase de pré-processamento dos dados, e como foi escolhida a normalização estatística, as séries residuais foram transformadas para apresentar uma média zero e um desvio padrão unitário, antes delas serem lidas pela camada de entrada do modelo.

As séries residuais normalizadas para toda a janela de aquisição servirão de entrada para o treinamento, teste e validação do modelo. Cabe salientar que os dados de saída do modelo também estarão normalizados. A Figura 25 exemplifica uma série residual antes e após o processo de normalização.

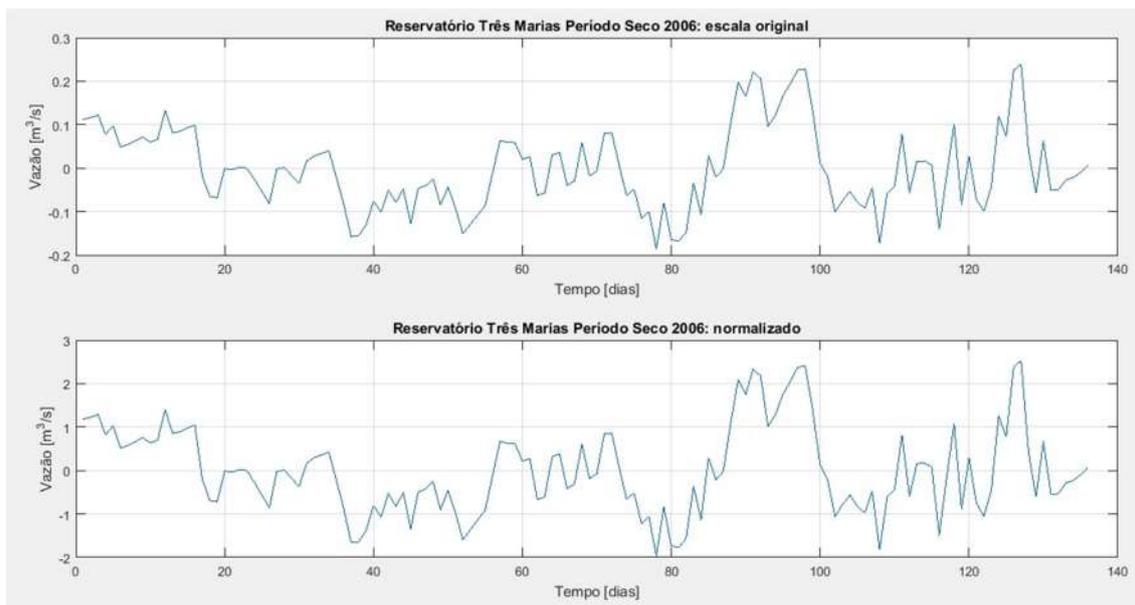


Figura 25 – Vazões históricas normalizadas do reservatório de Três Marias do Período Seco 2006

5.2 Treinamento da Rede Neural Artificial

Após o processo de decomposição das séries temporais e da normalização, foram obtidas séries residuais normalizadas que foram utilizadas para o treinamento das redes RNA dos períodos seco e úmido, com previsão de 7 dias de vazões médias diárias para o reservatório de Três Marias.

Foram considerados como dados de entrada para as redes neurais as informações dos postos pluviométricos, fluviométricos e da vazão natural do reservatório de Três Marias (série alvo), com diversas defasagens, conforme explicações da subseção 5.1.3. Dessa forma, cada ano foi configurado com uma quantidade específica de dados para serem lidos na camada de entrada, tal como o detalhamento da Tabela 3.

As redes RNA foram configuradas com 1 camada de entrada, 1 camada intermediária e 1 camada de saída, e treinadas com o algoritmo de aprendizado supervisionado LM, através do software *MathWorks MatLab* (versão 2016) e utilização da biblioteca (Toolbox) de Redes Neurais (nntool).

As funções de transferência tangente hiperbólica (*tansig*) e linear (*purelin*) foram implementadas na camada intermediária e na camada de saída, respectivamente.

Os seguintes parâmetros foram definidos no treinamento da rede:

- número máximo de épocas, repetições, para o treinamento (`net.trainParam.epochs = 100`);
- número de épocas entre os progressos mostrados (`net.trainParam.show = 25`);
- tempo máximo para treinar em segundos (`net.trainParam.time = 120`);
- meta de desempenho (`net.trainParam.goal = 0`);
- número máxima de falhas para validação (`net.trainParam.max_fail = 6`);
- mínimo valor do gradiente de performance (`net.trainParam.min_grad = 1e-7`);
- parâmetro Marquardt inicial ajustado (`net.trainParam.mu = 0.001`);
- fator de acréscimo (`net.trainParam.mu_inc = 10`);
- fator de decréscimo (`net.trainParam.mu_dec = 0.1`); e
- máximo valor (`net.trainParam.mu_max = 1e10`).

Levando em consideração os parâmetros mostrados, o modelo foi configurado para interromper o treinamento, e conseqüentemente apresentar os resultados, quando apenas uma das condições abaixo já fosse satisfeita:

- número máximo de épocas fosse alcançado;
- performance fosse minimizada até a meta;
- número de falhas da validação ultrapassasse o valor máximo;
- gradiente de performance ficasse abaixo do valor mínimo;
- parâmetro *mu* excedesse o valor máximo; e
- tempo máximo de simulação fosse excedido.

Na inicialização dos pesos e bias foi usado o método de *Nguyen-Widrow*, que se caracteriza basicamente por gerar valores randômicos para esses parâmetros. Com isso, cada vez que se executou o modelo, obteve uma diferente resposta da rede neural.

Dentre as diversas técnicas existentes na literatura para a implementação da validação cruzada [48], neste trabalho optou-se por escolher aleatoriamente as amostras para compor o conjunto de treinamento e o conjunto de teste, sendo ambos mutualmente exclusivos, com dados escolhidos randomicamente.

A cada simulação os dados são divididos aleatoriamente (função *dividerand*) da seguinte forma: 60% dos dados para o conjunto de treinamento, 20% para validação e 20% para testes. Devido o conjunto total de dados ser bem grande, é muito provável que

os resultados das divisões tiveram sido sempre diferentes, apesar de não existir garantia desta ocorrência.

Antes de calcular as funções custo RMSE e MAPE (%) adotadas, os dados de saída normalizados da rede tiveram que passar pelo processo contrário da normalização estatística, desfeita a transformação logarítmica, além de serem recompostos com a parcela extraída na etapa inicial de decomposição das séries, ou seja, os dados da rede de vazão do período seco foram adicionados com as respectivas tendências lineares, e os dados da rede do período úmido, bem como os dados da média de precipitações, foram adicionados com os respectivos valores médios.

Tanto para o período úmido, quanto para o seco, foram feitas diversas simulações para definir o número de neurônios da camada intermediária, com variações entre 1 e 11 neurônios, e foi escolhida a topologia com 9 neurônios, pois apresentou o menor RMSE para o conjunto de teste.

Assim, após todas as definições alinhadas e as redes configuradas, o modelo foi executado 10 vezes, e conseqüentemente, calculado o valor médio (erro *batch*) do RMSE e do MAPE das 10 execuções, embora as variações das funções custo tenham sido pequenas, apresentando desvio médio padrão entre 0,2 e 0,9. O capítulo 6 mostra em detalhes os resultados de cada execução realizada.

6 ANÁLISES E RESULTADOS

Realizar análises dos dados disponíveis, através de um pré-processamento das informações, tem um efeito significativo no desempenho do modelo. Sendo assim, esta seção se reserva em detalhar os resultados das análises dos dados de entrada da rede neural, bem como os resultados obtidos com as simulações.

6.1 Análises do Pré-processamento

A Função de Autocorrelação - FAC pode ser analisada para verificar se a série pode ser adequadamente modelada por um modelo autoregressivo linear, ou seja, se a previsão de um registro pode ser realizada por combinação linear dos registros anteriores. E para isto, o gráfico de autocorrelação tem que ser semelhante a uma reta. Opostamente ao que foi obtido neste trabalho, através das Figuras 26 e 27, percebemos que os dados das séries de vazão do reservatório de Três Marias - TM possuem correlação altamente não linear com a defasagem temporal.

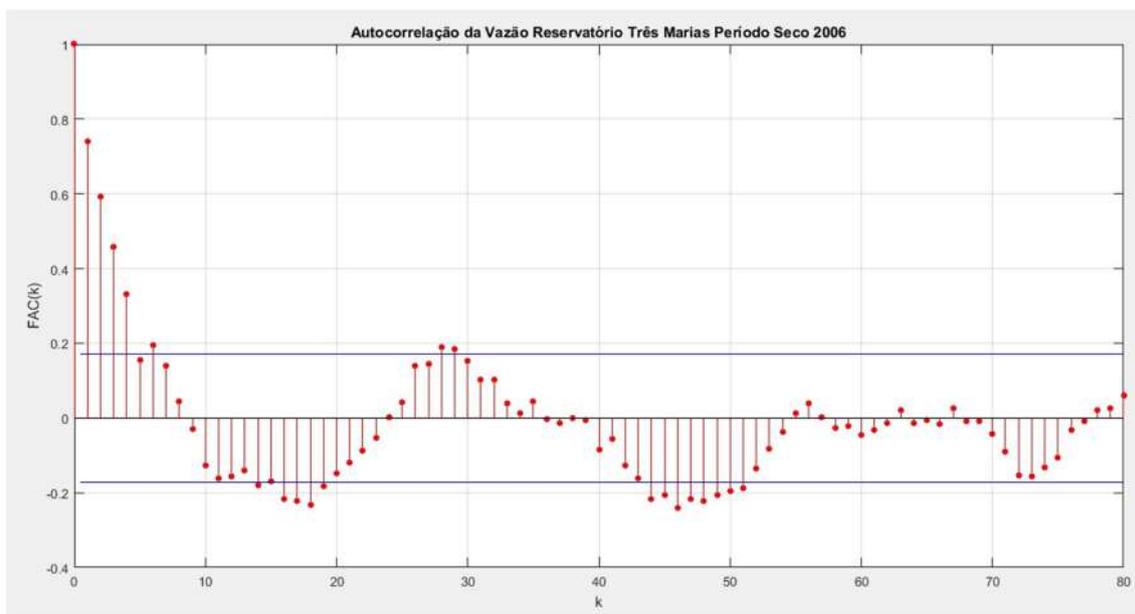


Figura 26 – FAC da Vazão do Reservatório de Três Marias do Período Seco 2006

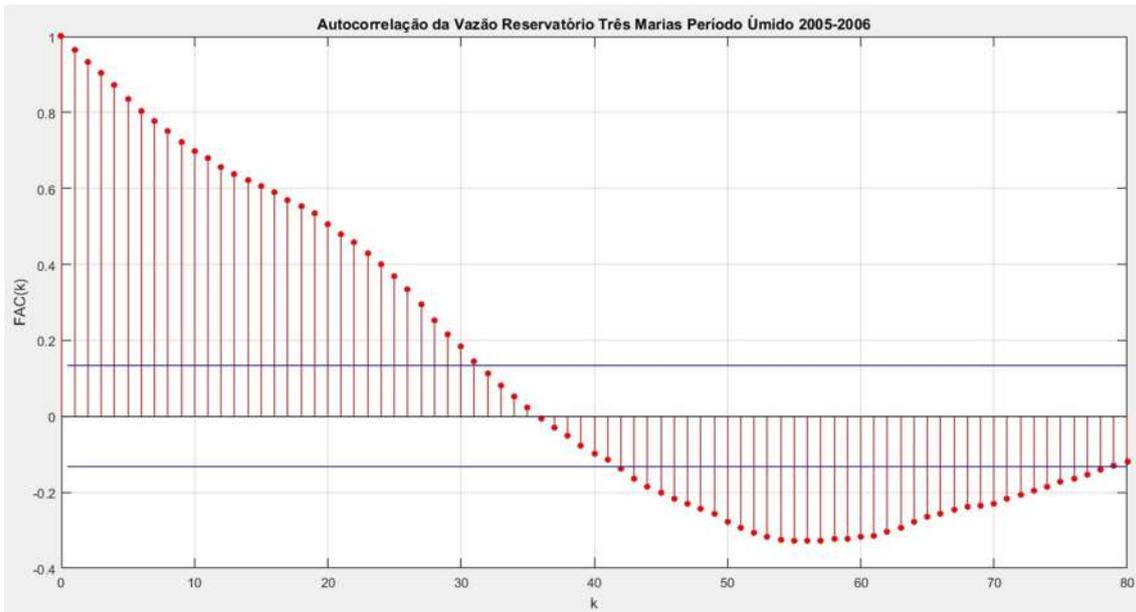


Figura 27 – FAC da Vazão do Reservatório de Três Marias do Período Úmido 2005-2006

Na Figura 28, observa-se que a série de precipitações é menos correlacionada do que a série de vazões, mas também não pode ser considerada como uma sequência aleatória. Embora a correlação seja fraca, a chuva é importante para sinalizar para a rede se o período é de alta ou baixa vazão. Sendo assim, a modelagem deve considerar informações adicionais (auxiliares) oriundas das medidas de chuva na área da bacia.

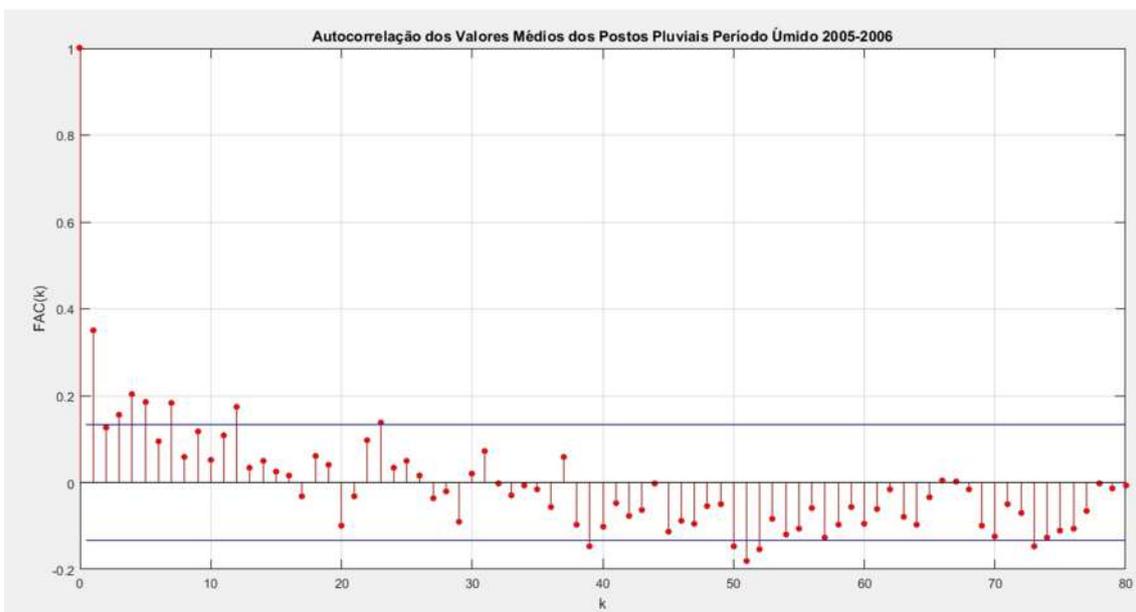


Figura 28 – FAC dos Valores médios dos postos pluviais do Período Úmido 2005-2006

Para a vazão natural de Três Marias, a Figura 29 revela que a Função de Autocorrelação Parcial - FACP diminui consideravelmente após 7 dias de atraso, sendo assim, foi escolhida a série com exatamente os 7 dias de atraso, $x(t-7)$, para possibilitar que a rede do trabalho em questão possa informar a vazão com antecedência de 7 dias.

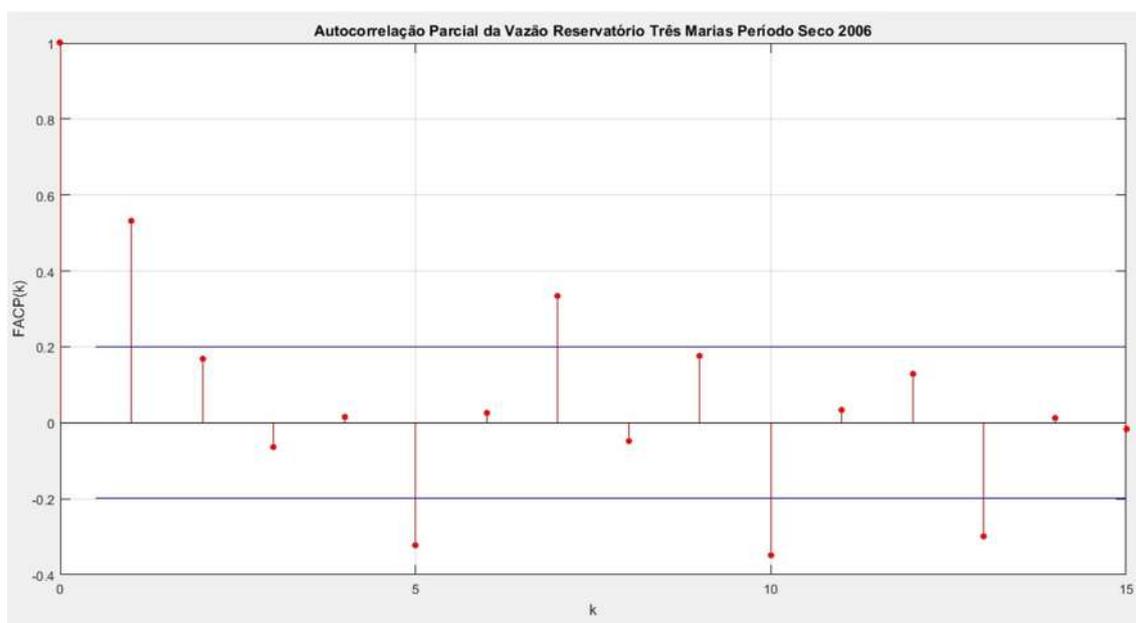


Figura 29 – FACP da Vazão do Reservatório de Três Marias do Período Seco 2006

A plotagem da Função de Correlação Cruzada - FCC de duas variáveis com o *lag* apresenta como varia a correlação entre duas séries de acordo com a defasagem de uma delas. O resultado está relacionado com a representatividade e a área de drenagem do posto fluviométrico. Assim, analisando a Figura 30, observa-se que o posto de Porto Indaiá, que é o posto fluviométrico mais próximo do reservatório, mas tem menor área de drenagem que os outros postos de mesma métrica, portanto, revela ter a menor correlação com a vazão natural de Três Marias (coeficiente de 0,69 para *lag* 7). Por outro lado, na Figura 31, os resultados obtidos para o posto de Porto das Andorinhas foram os melhores (coeficiente de 0,85 para *lag* 7), já que este posto, localizado no curso principal do rio São Francisco, possui a maior área de drenagem, apresentando maior representatividade e defasagem temporal com a vazão natural de Três Marias.

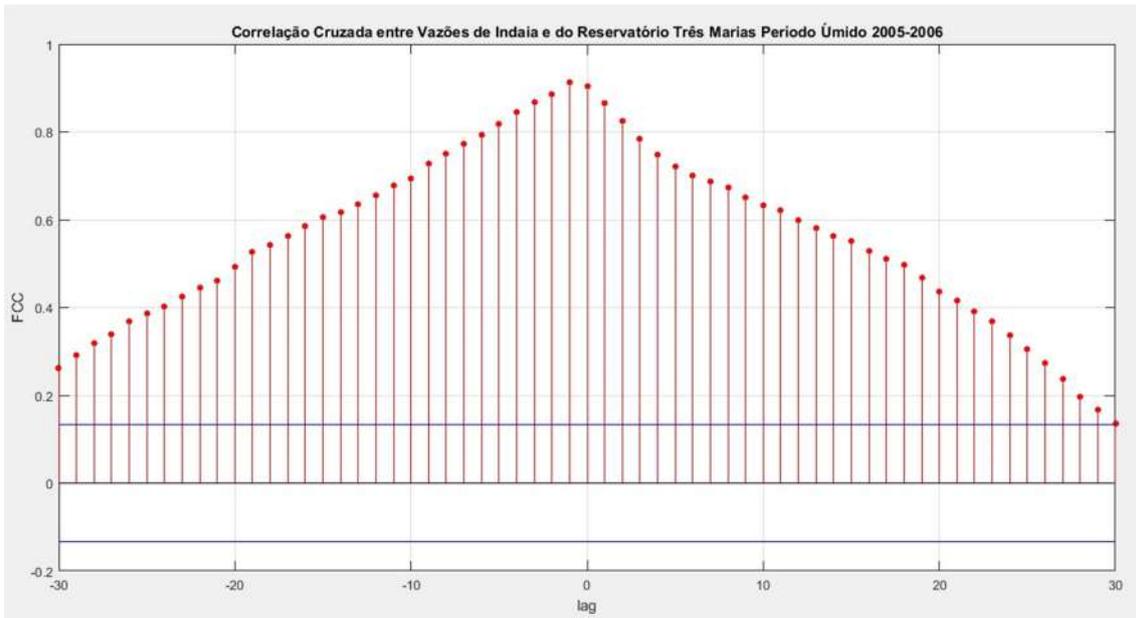


Figura 30 – FCC da Vazão de Indaiá com a vazão do Reservatório de Três Marias do Período Úmido 2005-2006

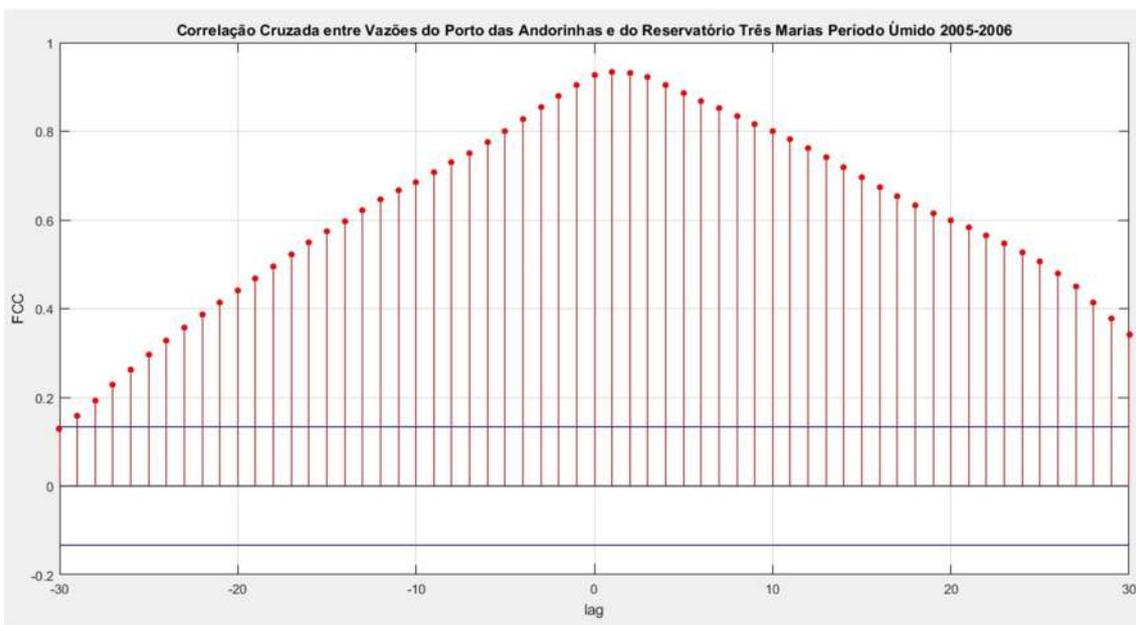


Figura 31 – FCC da Vazão de Porto das Andorinhas com a vazão do Reservatório de Três Marias do Período Úmido 2005-2006

Nessas figuras observa-se que, excluindo-se a vazão natural de Três Marias da análise, o posto de Porto das Andorinhas passa a ser o posto mais representativo da variabilidade comportamental das séries.

A Figura 32 mostra que a vazão natural de Três Marias tem uma relação temporal menor com a precipitação média da área do que as séries de vazões, mas ainda assim, preserva uma boa correlação (coeficiente de 0,12 para *lag* 7).

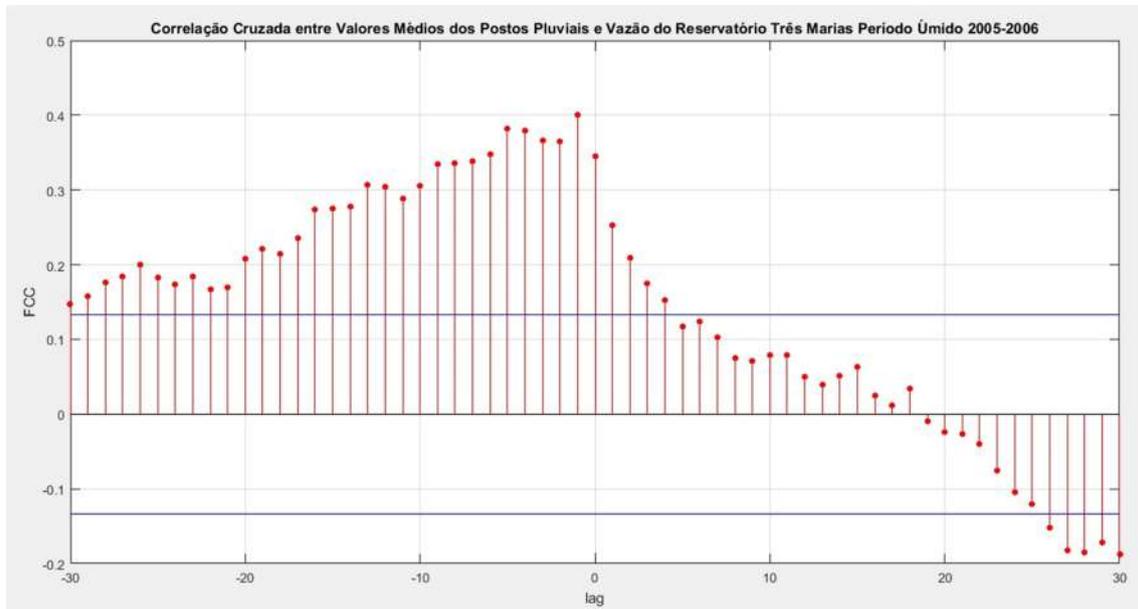


Figura 32 – FCC dos valores médios dos Postos Pluviais com a vazão do Reservatório de Três Marias do Período Úmido 2005-2006

Observa-se na Tabela 3, já mostrada na subseção 5.1.3, que o melhor modelo de previsão diária dos 7 dias, em termos de RMSE e MAPE (%) médios, considera informações dos postos fluviométricos ocorridas para diversas defasagens. Enquanto que para os postos pluviométricos e as vazões naturais afluentes do Reservatório de Três Marias, foram usadas as medições com apenas 7 dias de atraso.

A FAC serve também para verificar a existência de sazonalidade na série, que neste caso a sua presença não é marcante, tal como revelam as Figura 10, e aparece apenas na forma de senoide, pois entre um *lag* e outro a função varia muito. Cabe ressaltar que a mesma análise foi realizada para uma janela menor de tempo, ou seja, o período completo foi dividido em 3 partes, de forma que pudesse certificar a ausência da sazonalidade local.

A Transformada de Fourier, usando o algoritmo *Fast Fourier Transform - FFT*, foi calculada para auxiliar também na análise da existência de sazonalidade e ciclos senoidais nas séries temporais. Sendo realizada também paralelamente numa janela de tempo completa e em janelas de tempos menores.

Dando sequência ao processo de decomposição do sinal, foi usada a Função de Densidade de Probabilidade - FDP de distribuição qui-quadrática (X^2) com dois graus de liberdade e nível de significância de 5% para identificar em cada série quais os ciclos senoidais que apresentam densidade de probabilidade maior do que o ponto crítico (X^2_c), pois neste caso, os valores das frequências observadas não são considerados independentes.

Esses ciclos identificados são extraídos da série, em ordem decrescente de densidade de probabilidade, com acompanhamento e certificação da ocorrência da queda de potência do sinal residual após cada eliminação. Na Figura 33 consta um gráfico que apresenta os ciclos senoidais identificados de cada ano do período seco. E, a partir desta última figura, podemos notar que não existe um padrão de repetição de frequências da série entre os anos. Sendo assim, esse comportamento não estacionário direciona para a decisão de não mais extrair os ciclos senoidais das séries originais.

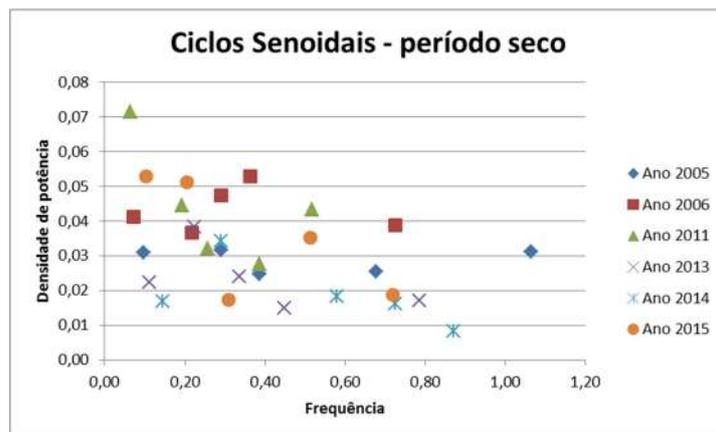


Figura 33 – Ciclos senoidais da Vazão do Reservatório de Três Marias dos Períodos Secos

Com o auxílio do espectro médio de frequências em conjunto com o espectrograma, revelados nas Figuras 34 e 35, analisamos as séries residuais sem os ciclos e percebemos que a densidade da potência do sinal permanece alta em poucas quantidades de frequências, e ainda sim, ocorre uma rápida diminuição dessa energia com o passar dos primeiros dias.

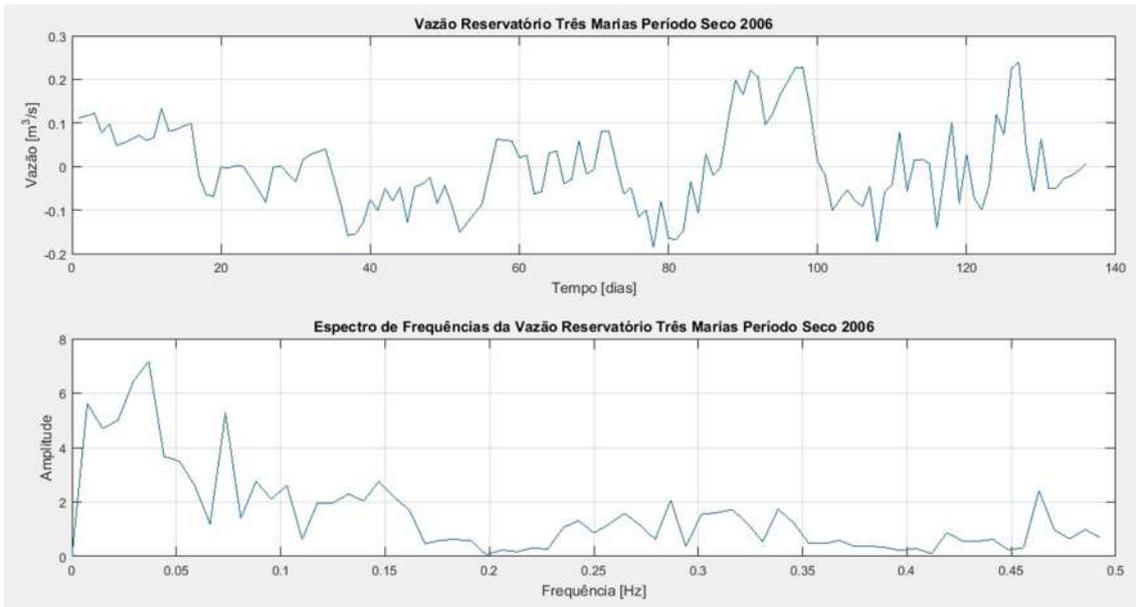


Figura 34 – Espectro médio de frequências da Vazão do Reservatório de Três Marias do Período Seco 2006

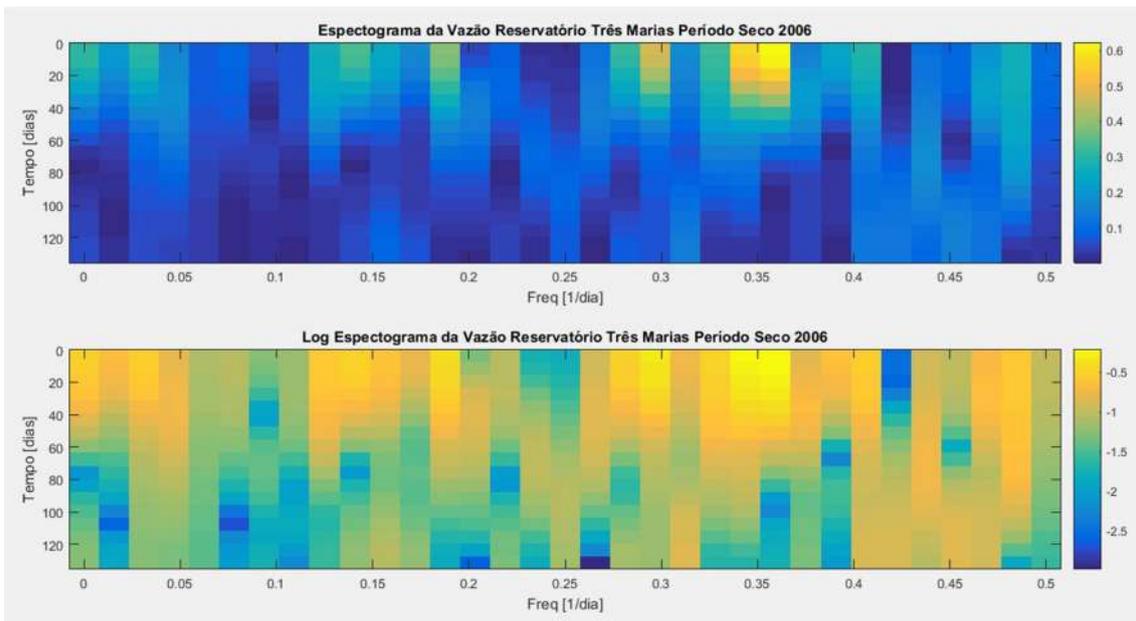


Figura 35 – Espectrograma da Vazão do Reservatório de Três Marias do Período Seco 2006

O pré-processamento dos dados de entrada que foi detalhado nesta subseção acarretou bons resultados nas simulações das redes neurais a serem mostrados em detalhes a seguir.

6.2 Análises das Simulações

As Tabelas 4 e 5 mostram, respectivamente, os resultados RMSE e MAPE (%) dos testes com a melhor rede neural MLP tipo LM do período seco, já treinada e validada.

Sendo denominado erro ótimo como o menor erro obtido dentre as 10 simulações e o erro *batch*, como a média dos erros.

Tabela 4 – Resultados RMSE dos testes da rede RNA do Período Seco

RMSE													
Seco	1	2	3	4	5	6	7	8	9	10	Erro ótimo	Erro <i>batch</i>	Desvio Padrão
2005	4,2	3,1	2,6	2,6	2,6	2,6	2,6	2,6	2,6	2,6	2,6	2,8	0,5
2006	11,4	11,4	11,4	10,9	10,9	10,9	10,9	10,9	10,9	10,9	10,9	11,1	0,2
2011	7,5	7,4	7,5	7,5	7,5	7,5	6,6	6,6	6,6	6,6	6,6	7,2	0,4
2013	4,5	4,5	4,5	4,5	4,5	4,5	4,5	4,5	4,1	4,1	4,1	4,4	0,1
2014	1,7	1,7	1,7	1,7	1,7	1,6	1,6	1,5	1,5	1,5	1,5	1,6	0,1
2015	4,6	4,6	4,6	4,4	4,4	4,5	4,5	4,5	4,4	4,4	4,4	4,5	0,1
Média											5,0	5,3	0,2

Tabela 5 – Resultados MAPE dos testes da rede RNA do Período Seco

MAPE (%)													
Seco	1	2	3	4	5	6	7	8	9	10	Erro ótimo	Erro <i>batch</i>	Desvio Padrão
2005	7,3	6,9	6,9	6,9	6,9	6,9	6,9	6,9	6,9	6,9	6,9	6,9	0,2
2006	29,4	29,4	29,4	29,4	29,4	29,4	29,4	23,2	23,2	23,2	23,2	27,5	3,0
2011	16,7	16,7	16,7	16,7	16,7	16,7	16,4	16,4	14,6	14,6	14,6	16,2	0,7
2013	9,9	9,9	9,9	9,9	9,9	9,9	9,9	9,9	9,9	6,6	6,6	9,5	1,0
2014	2,8	2,8	2,8	2,8	2,8	2,8	2,5	2,5	2,2	2,2	2,2	2,6	0,2
2015	11,0	11,0	9,5	9,5	9,5	9,9	9,9	9,7	9,7	9,4	9,4	9,9	0,6
Média											10,5	12,1	0,9

As Tabelas 6 e 7 mostram, respectivamente, os resultados RMSE e MAPE (%) dos testes com a melhor rede neural MLP tipo LM do período úmido, já treinada e validada.

Tabela 6 – Resultados RMSE dos testes da rede RNA do Período Úmido

RMSE													
Úmido	1	2	3	4	5	6	7	8	9	10	Erro ótimo	Erro batch	Desvio Padrão
2005- 2006	2,2	2,2	2,2	2,2	2,2	2,2	1,6	1,6	1,6	1,6	1,6	2,0	0,3
2006- 2007	3,3	3,3	3,3	3,3	3,3	2,7	2,7	2,7	2,7	2,7	2,7	3,0	0,3
2011- 2012	2,5	2,5	2,5	2,5	2,5	2,5	2,5	1,0	1,0	1,0	1,0	2,0	0,7
2012- 2013	1,3	1,3	1,3	1,3	1,3	1,3	1,1	1,1	1,1	1,1	1,1	1,2	0,1
2013- 2014	3,2	3,2	3,2	3,2	3,2	3,2	2,1	2,1	2,1	2,1	2,1	2,8	0,6
2014- 2015	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,2	1,2	1,2	1,2	1,3	0,1
Média											1,6	2,1	0,4

Tabela 7 – Resultados MAPE dos testes da rede RNA do Período Úmido

MAPE (%)													
Úmido	1	2	3	4	5	6	7	8	9	10	Erro ótimo	Erro batch	Desv. Pad.
2005- 2006	9,2	9,2	9,2	9,2	9,2	9,2	7,4	7,4	7,4	7,4	7,4	8,5	0,9
2006- 2007	12,6	12,6	12,6	12,6	12,6	11,1	11,1	11,1	11,1	11,1	11,1	11,9	0,8
2011-	10,4	10,4	10,4	10,4	10,4	10,4	10,4	8,7	8,7	8,7	8,7	9,9	0,8

2012													
2012-2013	3,1	3,1	3,1	3,1	3,1	3,1	2,2	2,2	2,2	2,2	2,2	2,7	0,4
2013-2014	11,3	11,3	11,3	11,3	11,3	11,3	9,3	9,3	9,3	9,3	9,3	10,5	1,0
2014-2015	3,3	3,3	3,3	3,3	3,3	3,3	3,3	2,6	2,6	2,6	2,6	3,1	0,3
Média											6,9	7,8	0,7

Resumidamente, as Tabelas 8 e 9 concatenam os resultados RMSE e MAPE (%) alcançados para os períodos secos e úmidos.

Tabela 8 – Resultados RMSE dos Períodos Secos e Úmidos

RMSE					
Período Seco	Erro ótimo	Erro batch	Período Úmido	Erro ótimo	Erro batch
2005	2,63	2,83	2005-2006	1,58	1,97
2006	10,94	11,08	2006-2007	2,75	3,05
2011	6,64	7,16	2011-2012	1,04	2,04
2013	4,07	4,38	2012-2013	1,05	1,23
2014	1,48	1,59	2013-2014	2,08	2,77
2015	4,43	4,48	2014-2015	1,16	1,31
Média	5,03	5,25	Média	1,61	2,06

Tabela 9 – Resultados MAPE dos Períodos Secos e Úmidos

MAPE (%)					
Período Seco	Erro ótimo	Erro batch	Período Úmido	Erro ótimo	Erro batch
2005	6,89	6,91	2005-2006	7,42	8,48
2006	23,18	27,55	2006-2007	11,11	11,88
2011	14,62	16,25	2011-2012	8,70	9,89
2013	6,57	9,54	2012-2013	2,23	2,74
2014	2,19	2,60	2013-2014	9,26	10,47
2015	9,35	9,90	2014-2015	2,65	3,13
Média	10,47	12,12	Média	6,90	7,76

No gráfico superior da Figura 36, está representado em azul o valor esperado da série original de vazões diárias do Reservatório de Três Marias para o período seco de 2006 e em vermelho, a previsão da RNA para o conjunto de teste.

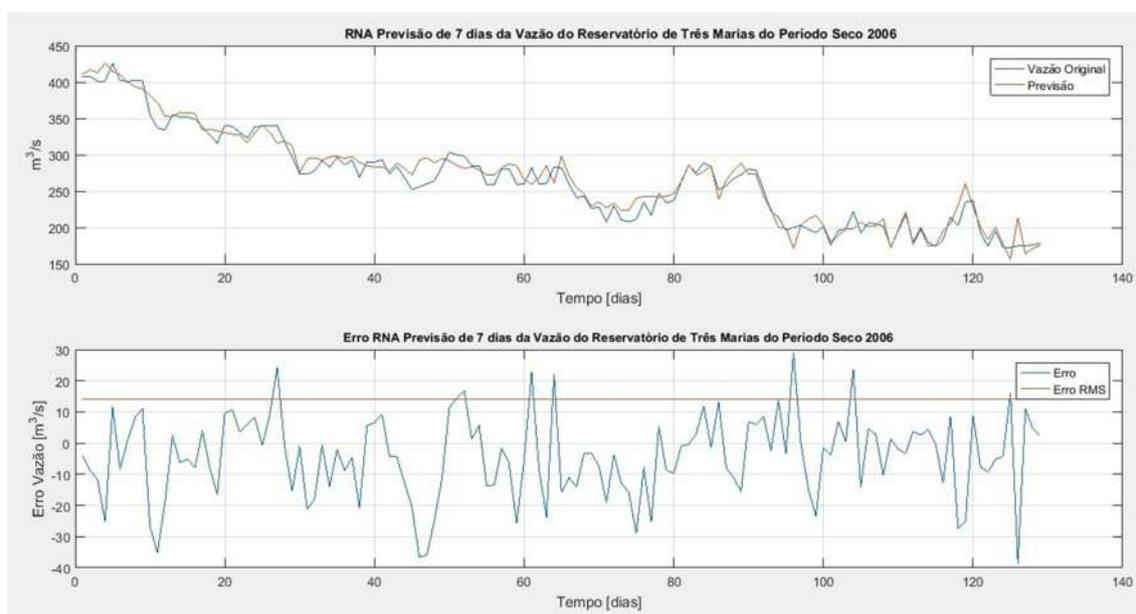


Figura 36 – Previsão de 7 dias da Vazão do Reservatório de Três Marias do Período Seco 2006

Neste gráfico é possível observar que a RNA foi treinada adequadamente e consegue acompanhar razoavelmente o comportamento complexo da série original, sendo capaz de generalizar para o conjunto de teste. É interessante ressaltar que, algumas vezes, a RNA não chega na amplitude esperada, embora seja capaz de indicar a existência de um pico. Além disso, há ocasiões em que a rede também não acompanha as mudanças bruscas de comportamentos, tal como no início dos dias (tempo < 20 dias) com vazão variando entre 400 e 450 m³/s, bem como próximo dos 45 dias com vazão variando entre 250 e 300 m³/s.

Já na Figura 37, encontra a previsão de vazões diárias do Reservatório de Três Marias da RNA para o período úmido 2005-2006, onde podemos notar que a RNA consegue acompanhar o comportamento da série original com bem menos erros do que no período seco. Não obstante, entre os dias 50 e 100, com vazões próximas de 2.000 e 2.500 m³/s, o erro da previsão aumentou nos momentos de picos e nas mudanças acentuadas de comportamento.

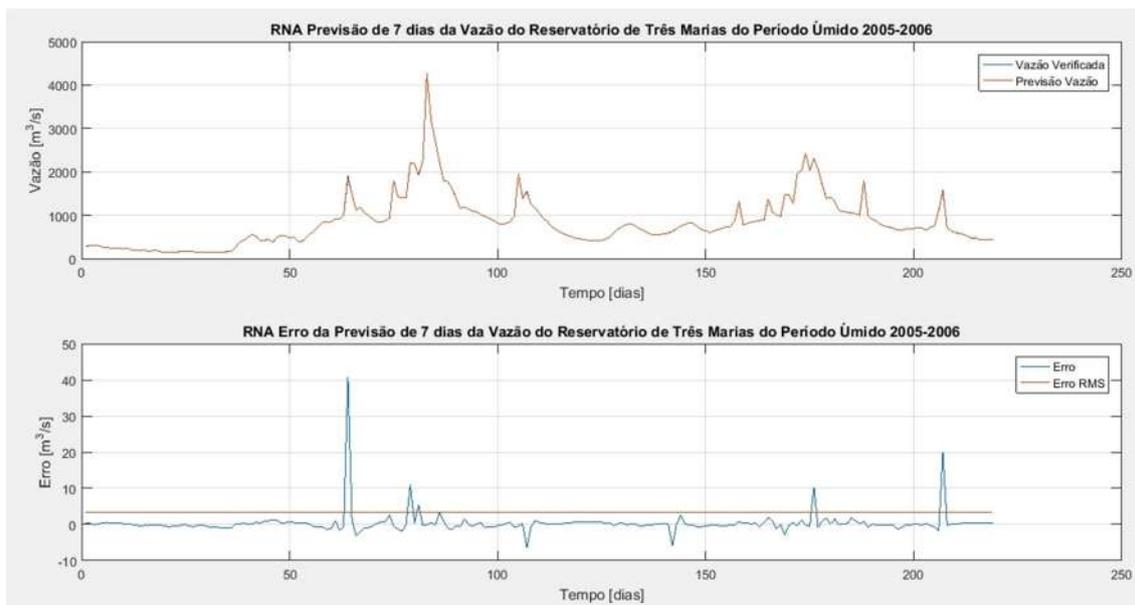


Figura 37 – Previsão de 7 dias da Vazão do Reservatório de Três Marias do Período Úmido 2005-2006

Um ponto importante a ser avaliado é a estimativa do comportamento estático do erro. Nas Figuras 38 e 39, estão representados os histogramas do erro entre a saída da RNA e o valor esperado. Observa-se que em ambos os histogramas o erro apresenta uma distribuição aproximadamente centralizada em zero e simétrica, o que indica que os modelos de previsão não possuem tendência de estimativa intrínseca, ou seja, não tem viés positivo, nem negativo.

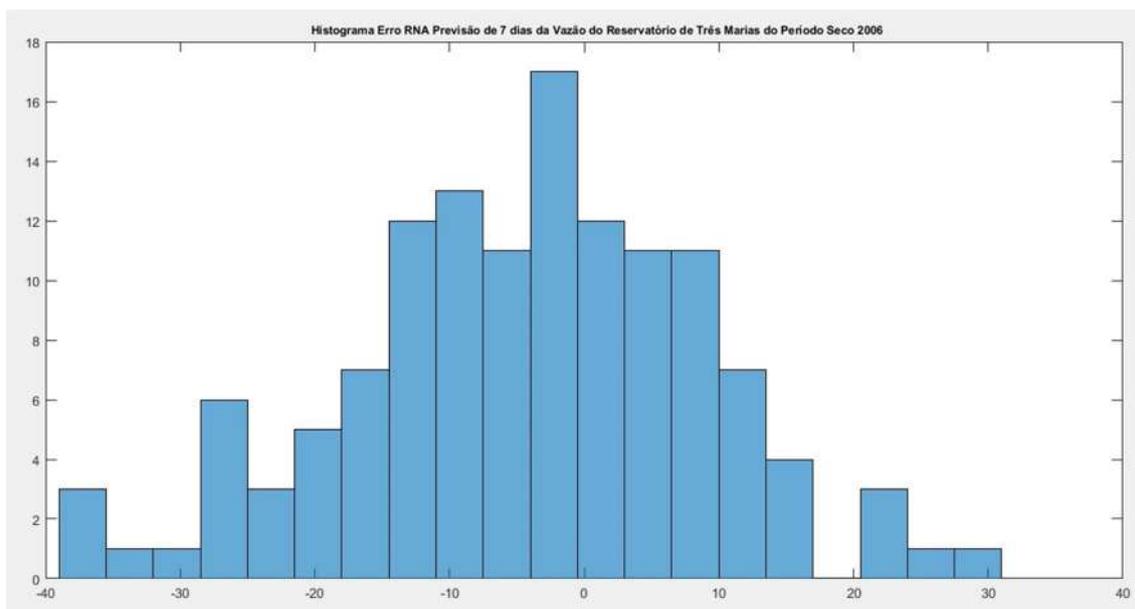


Figura 38 – Histograma do erro de Previsão de 7 dias da Vazão do Reservatório de Três Marias do Período Seco 2006

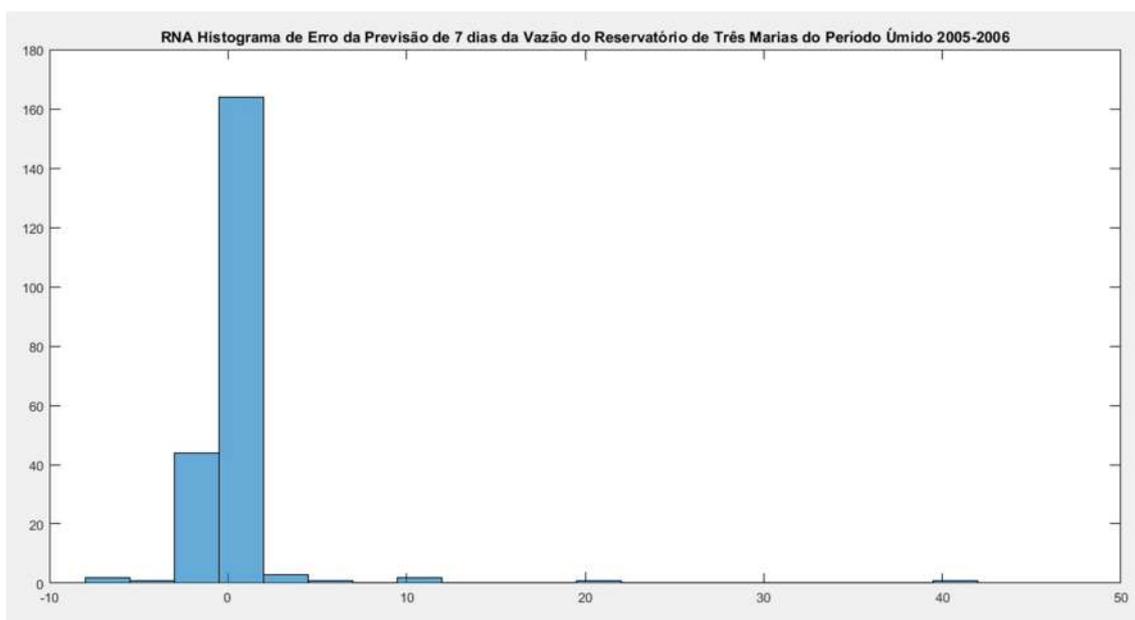


Figura 39 – Histograma do erro de Previsão de 7 dias da Vazão do Reservatório de Três Marias do Período Úmido 2005-2006

Os gráficos de dispersão (*scatter plot*) entre a série original e a série predita pelo modelo RNA estão ilustrados nas Figuras 40 e 41. Esses gráficos confirmam a dificuldade que o modelo tem de prever os picos da série original, além das acentuadas variações de vazões, que ocorreram, por exemplo, com vazões entre 400 e 450 m³/s e entre 250 e 300 m³/s do período seco de 2006.

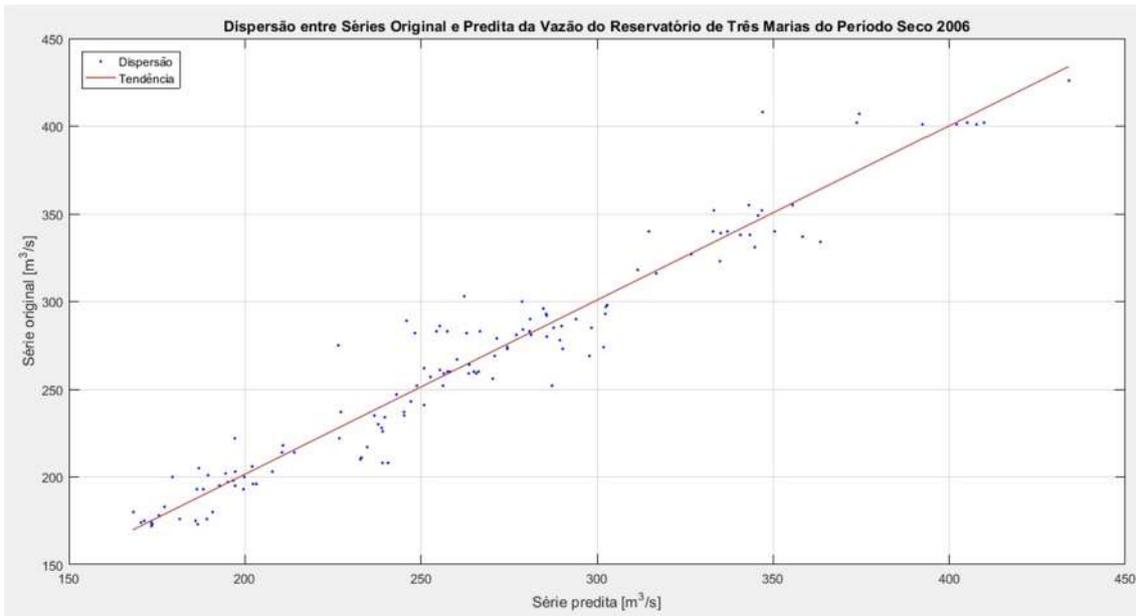


Figura 40 – Dispersão entre Séries Original e Predita da Vazão do Reservatório de Três Marias do Período Seco 2006

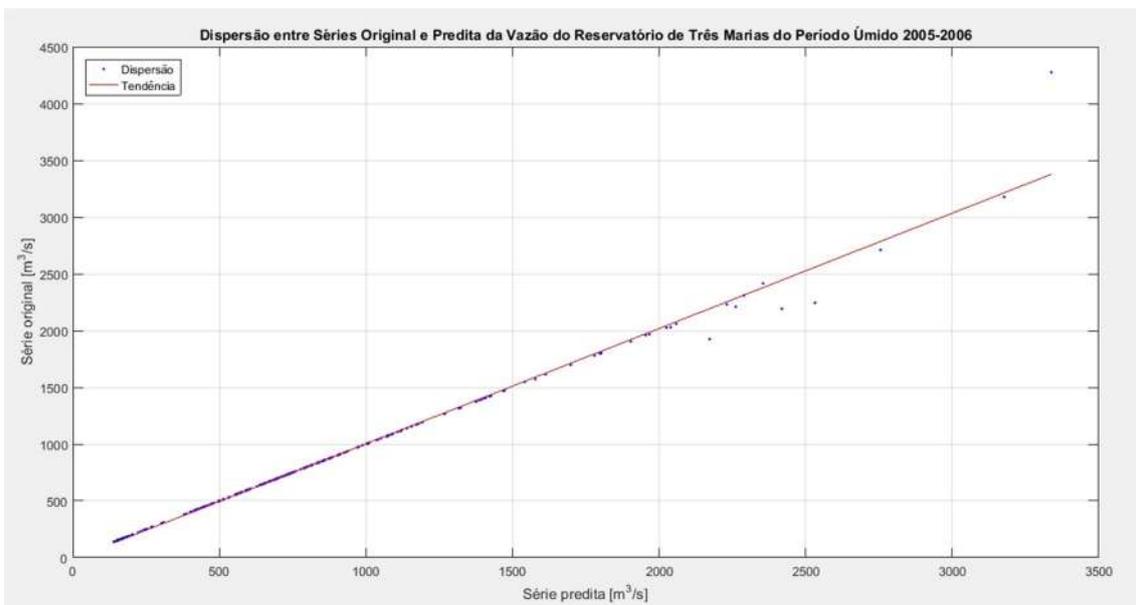


Figura 41 – Dispersão entre Séries Original e Predita da Vazão do Reservatório de Três Marias do Período Úmido 2005-2006

Através dos gráficos de dispersão podemos detectar se os modelos de previsão fornecem uma superestimação, quando o valor previsto tende a ficar cada vez menor do que o valor real com o decorrer dos dias, ou subestimação, que apresenta um efeito reverso.

Adicionalmente, os gráficos de dispersão também revelam se há no modelo uma tendência aditiva ou subtrativa.

E seus resultados vão ao encontro dos resultados dos histogramas, ou seja, os modelos apresentaram resultados aproximados sem tendências de erro.

Para comparar os resultados da rede deste trabalho, já expostos nas Tabelas 8 e 9, e comprovar sua eficácia, foram obtidos os erros históricos médios da rede neural Neuro3M (método NSRBN) utilizada pelo ONS [1], cujo erro MAPE médio (erro *batch*) para previsão de até 3 dias de antecedência é de aproximadamente 16 % no ano, 14 % no período seco e 18 % no período úmido.

Também constavam resultados oriundos da rede PREVIVAZH, que era usada antes da sua substituição pela rede RNA atual, e os erros eram de 29 % no ano, 18 % no período seco e 39 % no período úmido.

7 CONCLUSÕES E RECOMENDAÇÕES

Esforços se fazem necessários na obtenção de uma melhor previsão de vazões aos reservatórios do SIN, visto que esses resultados interferem diretamente na otimização do despacho das gerações das usinas hidrelétricas, e conseqüentemente, no custo operativo do sistema e no preço da energia.

Para isso, foi proposta a calibração de um modelo de previsão de vazões afluentes diárias na usina hidrelétrica de Três Marias, na bacia do São Francisco, 7 dias à frente, utilizando a técnica de redes neurais MLP tipo LM, considerando as vazões afluentes naturais do próprio reservatório da usina, e as informações pluviométricas e fluviométricas advindas de postos hidrométricos a montante da região de estudo. E em prol de uma correta comparação de resultados com a rede neural Neuro3M (método NSRBN) em uso na mesma região pelo ONS, as informações usadas no modelo proposto foram equivalentes.

Foram realizadas diversas análises estatísticas no pré-processamento das informações recebidas do BDT do ONS objetivando consolidar dados de entrada no modelo com boa representatividade e qualidade, já que esses dados são fundamentais para alcançar bons resultados.

Indo ao encontro da opinião do ONS [1], a partir da comparação entre os desempenhos da rede neural e do antigo modelo PREVIVAZH, concluiu-se na prática que a aplicação da técnica de redes neurais para a previsão de vazões do reservatório de Três Marias proporcionou as seguintes vantagens:

- redução significativa dos erros médios da previsão de vazões;
- previsão sem tendência de estimativa intrínseca, ou seja, não tem viés positivo, nem negativo;
- redução da ocorrência de grandes erros da previsão, geralmente associadas com a ausência da variável precipitação em sua modelagem; e
- ausência do efeito “sombra”, ou seja, devido à utilização de precipitação como dado de entrada, não é observada defasagem temporal na previsão de vazões.

Corroborando com os resultados deste trabalho, na revisão da literatura há diversos autores que revelaram resultados superiores ao compararem as aplicações dos tradicionais modelos estatísticos com os modelos que usam a técnica de redes neurais.

Observou-se ainda que os resultados aqui obtidos da rede MLP do tipo LM foram melhores do que os da rede Neuro3M (método NSRBN).

Existe a necessidade de continuar as pesquisas para desenvolver modelos de previsão de vazões cada vez mais precisos. Como sugestão para futuras pesquisas pode-se testar:

- elaborar análises geoestatísticas para aplicar pesos para cada posto pluviométrico de forma a obter uma média ponderada para os dados de precipitação, em vez de média aritmética simples, permitindo que o modelo incorpore a representatividade de cada posto;
- testar outras metodologias de redes neurais com noção de temporalidade entre as amostras apresentadas, tal como modelo RNN (do inglês, *Recurrent Neural Network*), construído com LSTM (do inglês, *Long Short-Term Memory*), conhecido como *Deep Learning*; e
- combinar diferentes modelos para tentar lidar com a dificuldade de previsão nos picos muito grandes e pouco frequentes do sinal.

Recomenda-se ainda a realização de estudos em outras bacias do SIN, com a implementação da mesma metodologia proposta neste trabalho, para que possa ratificar os bons resultados da rede neural na previsão de vazões diárias, principalmente em bacias com comportamento hidrológico distinto do alto da bacia do São Francisco.

8 BIBLIOGRAFIA

- [1] Operador Nacional do Sistema Elétrico - ONS, “Previsão de vazões Diárias ao Reservatório de Três Marias Usando a Técnica de Redes Neurais”. **Nota Técnica ONS 051/09**, revisão 1, Dezembro 2009.
- [2] MACEIRA, M.E.P.; TERRY, L.A.; COSTA, F.S.; DAMÁZIO, J.M.; MELO, A.C.G., “Chain of Optimization Models for Setting the Energy Dispatch and Spot Price in the Brazilian System”. **XIV Power Systems Computation Conference**, session 43, paper 1, March 2002.
- [3] Agência Nacional de Energia Elétrica - ANEEL, **Banco de Informações de Geração – BIG**. Disponível em: <<http://www2.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm>>. Acesso em: Outubro 2018.
- [4] Operador Nacional do Sistema Elétrico - ONS, **Resultados da Operação – Histórico da Operação**. Disponível em: <http://ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/geracao_energia.aspx>. Acesso em: Outubro 2018.
- [5] BELLMAN, R., **Dynamic Programming**. 2 ed. NJ, Princeton University Press, 1957.
- [6] MACEIRA, M. E. P., COSTA, F. S., DAMÁZIO, J. M., DENÍCIO, M., GUILHON, L. G., "Modelo Estocástico de Previsão de Vazões Mensais PREVIVAZM". **XV Simpósio Brasileiro de Recursos Hídricos**, Brasil, Setembro 2003.
- [7] MACEIRA, M.E.P., MERCIO, C.M.V.B., “Stochastic Streamflow Model for Hydroelectric Systems”. **V International Conference PMAPS – Probabilistic Methods Applied to Power Systems**, Canada, August 1997.
- [8] JARDIM, D.L.D.D., MACEIRA, M.E.P., FALCÃO, D.M., “Stochastic Streamflow Model for Hydroelectric Systems Using Clustering Techniques”. **IEEE Porto Power Tech Conference**, v. 3, Portugal, August 2001.

- [9] Centro de Pesquisa de Energia Elétrica - CEPEL, “Modelo de Previsão de Vazões Semanais Aplicado ao Sistema Hidroelétrico Brasileiro – PREVIVAZ”. **Relatório Técnico DPP/PEL 125/97**, Rio de Janeiro, Brasil, 1997.
- [10] CARVALHO, A. R. L. de, **Previsão de Afluências Diárias para Horizontes de 1 a 7 dias Através de uma Abordagem Estocástica – Modelo PREVIVAZH**. Dissertação de M. Sc., COPPE/ UFRJ, RJ, Brasil, 2001.
- [11] Operador Nacional do Sistema Elétrico - ONS, “Generalização do Modelo SMAP/ONS para Previsão de Vazões no Âmbito do SIN”. **Nota Técnica ONS 0056/2018**, revisão 2, Julho 2018.
- [12] Câmara de Comercialização de Energia Elétrica - CCEE, "Relatório InfoPLD Janeiro de 2019". **Relatório InfoPLD**, Janeiro 2019.
- [13] BOX, G. E. P., JENKINS, G. M., **Time Series Analysis-Forecasting and Control**. 2 ed. California, Holden-Day, 1970.
- [14] SHAMSELDIN, A. Y., “Application of a neural network technique to rainfall-runoff modeling”, **Journal of Hydrology** v. 199, p. 272-294, 1997.
- [15] VALENÇA, M. J. S., “Aplicação de redes neurais na área de recursos hídricos”. **XII Simpósio Brasileiro de Recursos Hídricos**, Brasil, Setembro 1997.
- [16] LOPES, J.E.G., BRAGA JR., B.P.F., CONEJO, J.G.L., “Simulação hidrológica: Aplicações de um modelo simplificado”. **III Simpósio Brasileiro de Recursos Hídricos**, v.2, 42-62, Brasil, Setembro 1981.
- [17] SPERB, R.M., MODRO, N. R., MARCONDES, P, FOES, P.A., BARCIA, R. M., PACHECO, R. C. S, TAVARES, G., “Prevent: protótipo de um sistema de previsão de enchentes baseado em redes neurais”. **Anais do XIII Simpósio Brasileiro de Recursos Hídricos**, Brasil, Setembro 1999.
- [18] KYRILLOS, D. de S., **Sistema de suporte ao planejamento dos recursos hídricos. Estudo de caso: bacia do rio São Francisco**. Dissertação de M. Sc., UFPE, PE, Brasil, 2000.

- [19] LIMA, C. H. R., FERREIRA, W. M. F., “Análise de modelos de redes neurais aplicados ao processo chuva-deflúvio no Semi-árido”. **Anais do XV Simpósio Brasileiro de Recursos hídricos**, Brasil, Setembro 2003.
- [20] MÉNDEZ, M. C., MANTEIGA, W. G., BANDE, M. F., SÁNCHEZ, J. M. P., CALDERÓN, R. L. “Modelling of the monthly and daily behaviour of the runoff of the Xallas river using Box–Jenkins and neural networks methods”, **Journal of Hydrology** v. 296, n. 1-4, 20, p. 38-58, 2004.
- [21] VALENÇA, M., VALENÇA, A., “Modelagem do processo chuva x vazão e vazão x vazão utilizando redes neurais”. **Anais do XVI Simpósio Brasileiro de Recursos Hídricos**, Brasil, Setembro 2005.
- [22] FAYAL, M. A. de A., **Previsão de Vazão por Redes Neurais Artificiais e Transformada Wavelet**. Dissertação de M.Sc., PUC-Rio, RJ, Brasil, 2008.
- [23] SOUSA, W. dos S., SOUSA, F. de A. S. "Rede neural artificial aplicada à previsão de vazão da Bacia Hidrográfica do Rio Piancó", **Revista Brasileira de Engenharia Agrícola e Ambiental** v.14, n.2, p.173–180, 2010.
- [24] FLORES, G. F., FERREIRA, V. H., ZAMBOTI, M., "Modelo Chuva-vazão para Previsão de Vazão Afluente Diária Utilizando Redes Neurais Artificiais". **XIII Simpósio Brasileiro de Automação Inteligente**, Brasil, Outubro 2017.
- [25] CHATFIELD, C., **The Analysis of Time Series: An Introduction**. 6 ed. London, Chapman and Hall/CRC, 2003.
- [26] CALÔBA, L. P., “Introdução ao Uso de Redes Neurais na Modelagem de Sistemas Dinâmicos e Séries Temporais”. **XIV Congresso Brasileiro de Automática**, Brasil, 2002.
- [27] GRUBBS, F. E. "Procedures for Detecting Outlying Observations in Samples", **Technometrics** v. 11, n. 1, pp. 1-21, 1969.
- [28] OPPENHEIM, A. V., SCHAFER, R. W., BUCK, J. R., **Discrete-Time Signal Processing**. 2 ed. New Jersey, Prentice Hall, 1999.

- [29] CHATFIELD, C., **Time-Series Forecasting**. 5 ed. London, Chapman and Hall/CRC, 2000.
- [30] MORETTIN, P. A., TOLOI, C. M. C., **Análise de Séries Temporais**. 2 ed. São Paulo, Edgard Blucher, 2004.
- [31] POMMERANZENBAUM, I. R., **Redes Neurais Artificiais na Predição das Principais Séries do Índice Ibovespa e suas Aplicações em Sistemas Automatizados de Negociação**. Dissertação de M.Sc., COPPE/UFRJ, RJ, Brasil, 2014.
- [32] MCCULLOCH, W. S., PITTS, W. "A Logical Calculus of Ideas Immanent in Nervous Activity", **Bulletin of Mathematical Biophysics** p. 115-133, 1943.
- [33] HAYKIN, S., **Neural Networks, A Comprehensive Foundation**. 2 ed. New York, Prentice Hall, 1999.
- [34] ROSENBLATT, F. "The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain", **Psychological Review** v. 65, n. 6, pp. 386-408, 1958.
- [35] MINSKY, M., PAPERT, S., **Perceptrons: An Introduction to Computational Geometry**. 2 ed. Massachusetts, M.I.T. Press, 1969.
- [36] CSÁJI, B. C., **Approximation with Artificial Neural Networks**. Dissertação de M.Sc., Faculty of Sciences, Eotvos Loránd University, Hungria, 2001.
- [37] ZHANG, G., PATUWO, B. E., HU, M. Y. "Forecasting with artificial neural networks: The state of the art", **International Journal of Forecasting** v. 14, pp. 35-62, 1998.
- [38] WALD, A., **Statistical Decision Functions**. 2 ed. New York, Wiley, 1950.
- [39] CASTRO, L.N., IYODA, E.M., PINHEIRO, E., VON ZUBEN F., "Redes Neurais Construtivas: Uma Abordagem Comparativa". **IV Congresso Brasileiro de Redes Neurais**, pp. 102-107, Brasil, 1999.

- [40] OSÓRIO, F., BITTENCOURT, J.R., "Sistemas Inteligentes baseados em RNAs aplicados ao processamento de Imagens". **I Workshop de Inteligência Artificial**, Brasil, 2000.
- [41] PRECHELT, L. "Automatic Early Stopping Using Cross Validation: Quantifying the Criteria", **Neural Networks** v. 11, n. 4, pp. 761-767, 1998.
- [42] REED, R., "Pruning Algorithms – A Survey". **IEEE Transactions on Neural Networks**, v. 4, n. 5, pp. 740-747, March 1993.
- [43] HAYKIN, S., **Neural Networks – A Comprehensive Foundation**. 2 ed. New York, Macmillan College Publishing Company, Inc, 1998.
- [44] VALENÇA, M. J. S., **Análise e projeto de redes neurais com abordagem construtiva para modelagem de sistemas complexos**. Tese de D.Sc., UFPE, PE, Brasil, 1999.
- [45] Serviço Geológico do Brasil - CPRM, "Rede Hidrometeorológica e Caracterização Física da Bacia do Alto São Francisco - Sub-Bacia 40 Fase 1". **Programa Nacional de Hidrologia**, 1996.
- [46] Centro de Previsão do Tempo e Estudos Climáticos/ Instituto Nacional de Pesquisas Espaciais - CPTEC/ INPE, **Boletim de Monitoramento e Análise Climática – CLIMANÁLISE**, número especial, 1986.
- [47] Operador Nacional do Sistema Elétrico - ONS, **Hidroelétricas do SIN - Rede de Operação - Horizonte 2023**. Disponível em: <<http://www.ons.org.br/paginas/sobre-o-sin/mapas>>. Acesso em: Agosto 2018.
- [48] KOHAVI, R., "A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection". **Proceedings of the 14th International Joint Conference on Artificial Intelligence** v. 2, pp. 1137-1143, San Francisco, CA, USA, 1995.