

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
Programa de Pós-Graduação em Engenharia Química

Nº 011

P

P

E

Q



TESE DE DOUTORADO

**DESENVOLVIMENTO DE ESTRATÉGIAS PARA
UTILIZAÇÃO DE SISTEMAS INTELIGENTES NO
MONITORAMENTO DA QUALIDADE DA ÁGUA**

Helenice Leite Garcia

Orientadora: Profa. Dra. Valdinete Lins da Silva
Co-orientador: Prof. Dr. Frede de Oliveira Carvalho

Recife/PE

Agosto/2012

HELENICE LEITE GARCIA

**DESENVOLVIMENTO DE ESTRATÉGIAS PARA UTILIZAÇÃO DE
SISTEMAS INTELIGENTES NO MONITORAMENTO DA QUALIDADE
DA ÁGUA**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Pernambuco, como requisito parcial à obtenção do título de Doutor em Engenharia Química.

Área de concentração: Engenharia Ambiental

Orientadora: Profa. Dra. Valdinete Lins da Silva
Co-orientador: Prof. Dr. Frede de Oliveira
Carvalho

Recife/PE

2012

Catálogo na fonte
Biblioteca Margareth Malta, CRB-4 / 1198

G216d Garcia, Helenice Leite.
Desenvolvimento de estratégias para utilização de sistemas inteligentes no monitoramento da qualidade da água / Helenice Leite Garcia. - Recife: O Autor, 2012.
152 folhas, il., gráfs., tabs.

Orientadora: Profa. Dra. Valdinete Lins da Silva.
Co-Orientador: Prof. Dr. Frede de Oliveira Carvalho.
Tese (Doutorado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Química, 2012.
Inclui Referências e Anexos.

1. Engenharia Química. 2. Redes Neurais Artificiais. 3. Lógica Fuzzy. 4. Sistema Neurofuzzy. 5. Índice de Qualidade da Água. I. Silva, Valdinete Lins da. (Orientadora). II. Carvalho, Frede de Oliveira. (Orientador). III. Título.

UFPE

660.2 CDD (22. ed.)

BCTG/2013-026

HELENICE LEITE GARCIA

**DESENVOLVIMENTO DE ESTRATÉGIAS PARA UTILIZAÇÃO DE
SISTEMAS INTELIGENTES NO MONITORAMENTO DA
QUALIDADE DA ÁGUA**

Área de concentração: Engenharia Ambiental

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Pernambuco, defendida e aprovada em 02 de Agosto de 2012 pela banca examinadora constituída pelos seguintes membros:

Profa. Dra. Valdinete Lins da Silva
Orientador - UFPE

Prof. Dr. Mohand Benachour - UFPE

Profa. Dra. Marcia Maria Lima Duarte - UFRN

Prof. Dr. Gilson Lima da Silva - UFPE

Prof. Dr. Mauricio Alves da Motta Sobrinho - UFPE

*Dedico, em essência, este trabalho a minha
mãe e a meu pai (in memoriam), e aos meus
dois amores, filha Júlia e meu esposo Carlos.*

AGRADECIMENTOS

Agradeço a Deus por toda prova que nada sou sem Teu amor, sem Tua força, sem Tua presença em minha vida. Obrigada Deus por essa vitória, demorou, mas essa etapa foi cumprida!

Certa vez um mestre amigo me disse que a maior prova de amor e de respeito é o agradecimento. Amor à família, fonte de luz e de inspiração. Respeito aos amigos fonte de motivação e companheirismo. Sendo assim, quero agradecer a todos da minha família, em especial:

A minha doce filha, Júlia; quero agradecer por todo amor que Deus nos deu e te pedir desculpas pela dor de estarmos, muitas vezes, separadas não fisicamente, mas separadas pelas obrigações do meu dia a dia e pela força de querer conquistar algo melhor para nós duas. Te amo tanto, minha filha querida!

Ao meu esposo, pela superação dos momentos difíceis. Esse último ano foi incrivelmente difícil, mas fomos fortes para começarmos 2012 com a inspiração para construção de momentos felizes. Te amo, você sabe!

A minha mãe, Valdice, por toda luta travada nesses últimos anos, principalmente 2010 e 2011, por todos os ensinamentos, por nosso amor. A tua ausência física doe demais, porque a senhora é e sempre será meu grande amor, minha força, minha certeza de que Deus nos abençoou com um amor lindo. Te amo mãe!

Aos meus irmãos, pelas risadas, pelas brigas, pela união naqueles momentos que fomos muito mais do que irmãos, fomos uma força única de apoio um ao outro. Vocês são partes da minha vida.

Meus agradecimentos especiais a profa. Valdinete, além de meu pedido de desculpas por não ter sido a aluna que a senhora conheceu em Campinas no mestrado. Saiba que poucos professores são educadores como a senhora é. Obrigada por tudo, pela oportunidade, pela compreensão, pelo carinho, pelos ensinamentos. Quero que tenha certeza que a senhora é muito querida, mora realmente em meu coração, não somente pelo que tem feito por mim, mas pela força que sinto quando conversamos e trocamos confidências.

Há mais ou menos 16 anos, eu conheci o prof. Frede e desde então só tenho aprendido não somente sobre engenharia, como também sobre a vida. Quero te agradecer por tudo, tudo mesmo, e quero que você saiba que suas palavras, mesmo algumas vezes “toscas”, tiraram-me do fundo do poço. Obrigada, você também mora em meu coração.

Quero agradecer a Filipe Coelho, por toda ajuda no desenvolvimento deste trabalho e pela parceria nas publicações.

Agradeço a equipe de coleta de amostras de água do Laboratório de Química Analítica Ambiental, Departamento de Química – UFS, pela amizade e pela força quando somente os risos fizeram a diferença para vencer as dificuldades. Obrigada a todos!

Quero agradecer ao Departamento de Engenharia Química da UFPE, em particular, a Flavio Garrett por toda ajuda nos momentos acadêmicos e aos professores, em particular ao professor Benachour.

Sou grata ao CNPq pela bolsa concedida nos primeiros 18 meses do doutorado.

Por fim, a todos os meus familiares e amigos que com as orações foram essenciais para concretização de mais um sonho acadêmico.

*“A mente que se abre a uma nova idéia jamais
volta a sua condição original.”
(Albert Einstein)*

RESUMO

A disponibilidade de água no mundo inteiro tem se tornado pauta importante em todos os fóruns de discussão sobre as condições dos recursos hídricos no planeta. As avaliações sobre esta disponibilidade são, principalmente em relação à qualidade da água, em termos da definição de parâmetros físicos, químicos e biológicos da água, representando um conjunto de parâmetros extremamente importantes para o contexto da sustentabilidade ambiental. Neste cenário, as técnicas de Inteligência Computacional ou de Inteligência Artificial, têm se tornando alternativas de ampla aplicabilidade para modelagem e simulação da qualidade da água. Neste trabalho foi, então, desenvolvido um estudo para que as redes neurais, a lógica fuzzy e a análise de componente principal fossem utilizadas como estratégias para avaliação da qualidade da água em corpos hídricos do Estado de Sergipe, com vista à construção de interface fáceis de serem utilizadas em ambiente MATLAB. Neste estudo, foram coletados dados ambientais dos reservatórios Jacarecica, da Marcela e da bacia do Rio Poxim, em Sergipe. Para o desenvolvimento da modelagem em termos de redes neurais, foram utilizadas as redes Multi Layer Perceptron (MLP) e as redes Radial Basis Function (RBF) e um sistema neuro-fuzzy para modelar a qualidade da água utilizando como variável de saída a concentração de clorofila-a para caracterizar o fenômeno de eutrofização do sistema. Além dessa modelagem, foi desenvolvida uma equação com base na análise de componente principal em função das medidas de pH, oxigênio dissolvido, amônia, nitrito e nitrato, além das concentrações de ortofosfato, nitrogênio total e fósforo total, e clorofila-a. Quanto à aplicação da lógica fuzzy, foi calculado o índice de qualidade da água em função das concentrações de clorofila-a, nitrogênio total e fósforo total para classificação do sistema nos quatro níveis tróficos para que as variáveis linguísticas fossem identificadas. Para a análise fuzzy foram implementadas as regras fuzzy com base no conhecimento especialista do sistema hídrico, sendo o modelo fuzzy considerado representativo para classificar as condições ambientais dos reservatórios. É importante ressaltar que os resultados foram satisfatórios em termos da classificação e descrição do fenômeno de eutrofização entre os níveis de oligotrófico e hipertrófico para corpos hídricos em análise. Dessa forma, as técnicas de inteligência artificial, em particular as redes neurais e a lógica fuzzy, foram empregadas com sucesso para um conjunto de dados ambientais, mostrando a viabilidade numérica no que concerne a representação de fenômenos ambientais complexos e importantes para sustentabilidade ambiental dos corpos hídricos.

Palavras-chave: Redes Neurais Artificiais. Lógica Fuzzy. Sistema Neurofuzzy. Índice de Qualidade da Água.

ABSTRACT

The availability of water in the world has become important agenda in all forums on the conditions of water resources on the planet. The evaluations are available on this, especially in relation to water quality in terms of the definition of physical, chemical and biological water, representing a set of parameters extremely important for the context of environmental sustainability. In this scenario, the techniques of Computational Intelligence and Artificial Intelligence have become alternative broad applicability for modeling and simulation of water quality. This work was then carried out to the neural networks, fuzzy logic and principal component analysis were used as strategies for evaluation of water quality in water bodies of the State of Sergipe, with a view to building interface easily be used in MATLAB. In this study, data were collected Jacarecica environmental reservoirs of some of Marcela and Poxim River basin in Sergipe. For the development of modeling in terms of neural networks, the networks were used Multi-Layer Perceptron (MLP) and Radial Basis Function networks (RBF) and a neuro-fuzzy system for modeling water quality using as output variable concentration chlorophyll-a to characterize the phenomenon of eutrophication. In this model, we developed an equation based on principal component analysis on the basis of measurements of pH, dissolved oxygen, ammonia, nitrite and nitrate, and concentrations of orthophosphate, total nitrogen and total phosphorus and chlorophyll-a. Concerning the application of fuzzy logic, we calculated the index of water quality in the concentrations of chlorophyll a, total nitrogen and total phosphorus for the classification system in the four trophic levels for the linguistic variables were identified. To analyze fuzzy rules were implemented based on expert knowledge of the water system and the fuzzy model was representative to classify the environmental conditions of the reservoirs. Importantly, the results were satisfactory in terms of classification and description of the phenomenon of eutrophication levels for oligotrophic and hypertrophic water bodies under consideration. Thus, the artificial intelligence techniques, in particular neural networks and fuzzy logic have been successfully employed for a set of environmental data, showing the feasibility with regard to numerical representation of complex environmental phenomena and important for environmental sustainability of the bodies water.

Keywords: Artificial Neural Networks. Fuzzy Logic. Neuro Fuzzy System. Water Quality Index.

LISTA DE FIGURAS

Figura 1 - Classificação do corpo hídrico de acordo como o Índice de Qualidade da Água .	18
Figura 2 - Composição da arquitetura de uma rede neural.....	30
Figura 3 - Exemplos de arquiteturas de redes neurais artificiais.....	31
Figura 4 - Arquitetura <i>perceptron</i> de múltiplas camadas com uma camada oculta	31
Figura 5 - Rede Neural Feedforward (a) e a função de ativação (b).	33
Figura 6 - Fase de treinamento de uma RNA	34
Figura 7 - Rede neural com duas camadas.	35
Figura 8 - Rede neural de base radial (a) e rede neural de base radial com múltiplas saídas (b)	50
Figura 9 - Desenvolvimento das regras da lógica fuzzy.....	53
Figura 10 - Função do conjunto tradicional (a) e conjunto fuzzy (b).....	57
Figura 11 - Representação da função triangular.....	59
Figura 12 - Representação da função gaussiana.....	59
Figura 13 - Representação da função trapezoidal.....	60
Figura 14 - Função triangular do conjunto de temperatura de processamento de biscoitos...61	
Figura 15 - Sistema de inferência fuzzy	63
Figura 16 - Inferência Mamdani	65
Figura 17 - Arquitetura básica ANFIS	68
Figura 18 - ANFIS com 9 regras de inferência.	68
Figura 19 - Representação geométrica das componentes principais	70
Figura 20 - Estratégia metodológica do trabalho.....	72
Figura 21 - Reservatório da Marcela (SRH-SE, 2005).....	73
Figura 22 - Reservatório de Jacarecica II	74
Figura 23 - Sub-bacia do Rio Poxim	75
Figura 24 - Topologia para a rede neural MLP	77
Figura 25 - Topologia para RNA para os dados de Marcela.....	78
Figura 26 - Topologia para condições ótimas do modelo ANFIS.....	79
Figura 27 - Tela principal da ferramenta computacional	81
Figura 28 - Ambiente da Rede Neural (MLP, RBF e Neuro-Fuzzy)	81
Figura 29 - Rede MLP	82
Figura 30 - Treinamento da rede MLP	83

Figura 31 - Rede RBF.....	84
Figura 32 - Sistema neuro-fuzzy	85
Figura 33 - Sistema neuro-fuzzy para as etapas de treinamento, validação e teste.....	86
Figura 34 - Regras do sistema neuro-fuzzy e estimativa da clorofila-a	86
Figura 35 - Performance da rede neural com 10 neurônios na camada oculta.....	89
Figura 36 - Performance da rede neural com particionamento do conjunto de dados – 10 neurônios na camada oculta.....	89
Figura 37 - Performance da rede neural com 12 neurônios na camada oculta.....	90
Figura 38 - Performance da rede neural com particionamento do conjunto de dados - 12 neurônios na camada oculta.....	90
Figura 39 - Performance da rede neural com 14 neurônios na camada oculta.....	91
Figura 40 - Performance da rede neural com particionamento do conjunto de dados – 14 neurônios na camada oculta.....	91
Figura 41 - Aderência do modelo para a RBF sem PCA	92
Figura 42 - Resultado da simulação para a RBF sem PCA.....	93
Figura 43 - Aderência da rede RBF com PCA	93
Figura 44 - Resultado da simulação para a RBF com PCA.....	94
Figura 45 - Aderência do modelo para a MLP sem PCA.....	95
Figura 46 - Resultado da simulação para a MLP sem PCA	95
Figura 47 - Aderência do modelo para a MLP com PCA	95
Figura 48 - Resultado da simulação para a MLP com PCA.....	96
Figura 49 - Performance da rede neural no sistema neuro-fuzzy - rio Poxim.....	97
Figura 50 - Ilustração da tela do sistema ANFIS para o treinamento.....	98
Figura 51 - Ilustração da tela do sistema ANFIS para validação	99
Figura 52 - Ilustração da tela do sistema ANFIS com cinco regras para cada variável de entrada	99
Figura 53 - Comportamento da clorofila em função das concentrações de fósforo e nitrogênio.....	100
Figura 54 - Conjunto de quatro regras (a) dados para treinamento (b) dados para validação – teste 1	101
Figura 55 - Conjunto de cinco regras (a) dados para treinamento (b) dados para validação – teste 2.....	101
Figura 56 - Conjunto de seis regras (a) dados para treinamento (b) dados para validação – teste 3.....	102

Figura 57 - Conjunto de sete regras (a) dados para treinamento (b) dados para validação – teste 4.....	102
Figura 58 - Conjunto de treze regras (a) dados para treinamento (b) dados para validação – teste 5.....	103
Figura 59 - Ilustração do conjunto de regras fuzzy com menor erro no treinamento.....	104
Figura 60 - Interface principal para o índice do nível de trofia do reservatório.....	105
Figura 61 - Interface do assistente de calibração.....	106
Figura 62 - Interface do assistente de calibração para o carregamento de dados.....	107
Figura 63 - Processamento da rede.....	107
Figura 64 - Classificação do sistema	108
Figura 65 - Gráfico biplot da componente 1 pela componente 2 - PCA	109
Figura 66 - Tela principal para índice de estado trófico da água através da lógica fuzzy....	110
Figura 67 - Condições fuzzy – variável de entrada concentração de clorofila-a ($\mu\text{g/L}$)	111
Figura 68 - Condições fuzzy – variável de entrada nitrogênio.....	111
Figura 69 - Condições fuzzy – variável de entrada fósforo.....	111
Figura 70 - Condições fuzzy – variável de saída IET.....	112
Figura 71 - Regras fuzzy aplicadas.....	113

LISTA DE TABELAS

Tabela 1 - Índice de Estado Trófico.	23
Tabela 2 - Classificação do estado trófico.....	24
Tabela 3 - Funções de ativação.....	32
Tabela 4 - Notação adotada no algoritmo backpropagation.	34
Tabela 5 - Comparação de uma rede RBF típica com uma rede MLP convencional.....	51
Tabela 6 - Comparativo dos métodos de <i>defuzzificação</i>	65
Tabela 7 - Variáveis, métodos e referências utilizados para a caracterização da água dos reservatórios.....	75
Tabela 8 - Custos por campanha experimental.....	80
Tabela 9 - Aderência das redes neurais testadas para o reservatório da Marcela.....	96
Tabela 10 - Erro médio do teste para o sistema ANFIS	103

LISTA DE ABREVIACÕES E SIGLAS

ANA – Agência Nacional das Águas

ANFIS – Adaptive Neuro-Fuzzy Inference System

CETESB – Companhia de Tecnologia de Saneamento Ambiental

Chl – Concentração de clorofila-a

CONAMA – Conselho Nacional do Meio Ambiente

DBO – Demanda Bioquímica de Oxigênio

DS – Transparência de Disco Secchi

EIA – Estudo de Impacto Ambiental

IA – Inteligência Artificial

IBGE – Instituto Brasileiro de Geografia e Estatística

IC – Inteligência Computacional

IET – Índice de Estado Trófico

IQA – Índice de Qualidade de Águas

MLP – Multi Layer Perceptron

MMA – Ministério do Meio Ambiente

N-NH₄ – Nitrogênio na forma de amônia

N-NO₂ – Nitrogênio na forma de nitrito

N-NO₃ – Nitrogênio na forma de nitrato

N_T – Nitrogênio Total

OD – Oxigênio Dissolvido

PCA – Principal Component Analysis

pH – Potencial Hidrogeniônico

P_T – Fósforo Total

RBF – Radial Base Function

RIMA – Relatório de Impacto Ambiental

RNA- Redes Neurais Artificial

SUMÁRIO

1 INTRODUÇÃO	11
2 FUNDAMENTAÇÃO TEÓRICA	16
2.1 ÁGUA	16
2.2 FENÔMENOS DE EUTROFIZAÇÃO	18
2.2.1 Estado ou grau trófico	20
2.2.1.1 Parâmetros limnológicos	20
2.2.1.2 Índice de Estado Trófico	23
2.3 REDES NEURAIIS	28
2.3.1 Perceptron de Múltiplas Camadas (Multi Layer Perceptron - MLP)	31
2.3.2 Funções de ativação	32
2.3.3 Algoritmo de retro-propagação (<i>Backpropagation</i>)	33
2.3.4 Algoritmo de otimização para treinamento supervisionado	40
2.3.4.1 Método de Levenberg-Marquardt (<i>LM</i>)	41
2.3.5 Treinamento da rede MLP	42
2.3.5.1 Modo de treinamento	42
2.3.5.2 Topologia da rede	43
2.3.5.3 Conjunto de dados	44
2.3.5.4 Generalização	45
2.3.5.4.1 Parada antecipada (<i>early stopping</i>)	45
2.3.5.4.2 Regularização bayesiana	46
2.3.6 Redes com funções de ativação base radial (Radial Basis Function – RBF)	49
2.3.7 Redes neurais e qualidade de água	51
2.4 LÓGICA FUZZY	52
2.4.1 Características da lógica fuzzy	56
2.4.2 Conjunto fuzzy	57
2.4.3 Funções de pertinência	58
2.4.4 Variáveis Linguísticas no Sistema de Inferência Fuzzy	60
2.4.4.1 Regras de Inferência Fuzzy	62
2.4.4.2 Métodos de Inferência Mamdani	64
2.4.4.3 Métodos de Defuzzificação	65
2.4.5 Sistema Neuro-Fuzzy	66
2.5 ANÁLISE DE COMPONENTES PRINCIPAIS	69

3 METODOLOGIA	72
3.1 AMBIENTES DE ESTUDOS.....	73
3.1.1 Reservatório da Marcela.....	73
3.1.2 Reservatório Jacarecica	73
3.1.3 Rio Poxim	74
3.2 MATERIAIS E MÉTODOS.....	75
3.3 REDE NEURAL	76
3.3.1 Rede MLP.....	76
3.3.2 Rede RBF	77
3.3.3 Sistema Neuro-Fuzzy.....	79
3.4 FERRAMENTA COMPUTACIONAL	80
4 RESULTADOS E DISCUSSÃO	88
4.1 REDE NEURAL	88
4.1.1 Rede MLP.....	88
4.1.2 Redes MLP e RBF com o PCA	92
4.2 SISTEMA NEURO-FUZZY	97
4.3 ÍNDICE DE ESTADO TRÓFICO DA ÁGUA	104
4.3.1 PCA.....	105
4.3.2 Lógica Fuzzy	109
5 CONCLUSÕES	114
5.1 LIMITAÇÕES.....	115
5.2 SUGESTÃO DE TRABALHOS FUTUROS.....	115
REFERÊNCIAS	117
ANEXOS	131

1 INTRODUÇÃO

A disponibilidade de água no mundo inteiro tem se tornado pauta em muitos os fóruns de discussão sobre os recursos hídricos do planeta. As avaliações sobre esta disponibilidade são, principalmente, em relação à qualidade da água para os usos considerados nobres como abastecimento público, recreação, geração de energia, irrigação e pesca. Conforme comentam Srebotnjak *et al.* (2012), proteger a qualidade de água doce para as necessidades humanas e o equilíbrio ecológico é, portanto, um aspecto importante da gestão ambiental integrada ao desenvolvimento sustentável da humanidade.

O Brasil detém, aproximadamente, 12 % da água doce superficial do planeta, ou seja, tem-se água em abundância no país. Entretanto, o Brasil apresenta graves problemas quanto à distribuição e utilização dos recursos hídricos disponíveis. Esses problemas são mais graves na região Nordeste, que conta apenas com 3 % dessa disponibilidade. Agregada a esse fato, o crescimento e os problemas de abastecimento tornam o Nordeste a região mais carente do país quanto à quantidade e à qualidade da água, conforme relatórios apresentados pela Agência Nacional de Águas - ANA (ANA, 2012).

Neste cenário, a avaliação da qualidade da água é imprescindível para sustentabilidade dos recursos hídricos do país, pois sua disponibilidade vem se tornando limitada. Macedo (2004) comenta que 63% do descarte de lixo no país são feitos em rios, lagos e restingas (ecossistemas costeiros, fisicamente determinados pelas condições edáficas (solo arenoso) e pela influência marinha), agravando cada vez mais a qualidade e a quantidade de água em áreas populacionais carentes em abastecimento e tratamento de água. Em decorrência dessa condição socioeconômica, a avaliação da qualidade da água quanto aos fenômenos ambientais, a exemplo do fenômeno de eutrofização, vem se destacando em estudos nos principais centros de pesquisas no Brasil e no mundo.

Os estudos do comportamento dos corpos hídricos são desenvolvidos em termos da definição de parâmetros físicos, químicos e biológicos da água, que representam um conjunto de parâmetros ambientais necessário para classificação da qualidade desses recursos. Dessa forma, ressalta-se, por exemplo, que a qualidade da água de superfície em uma região depende muito da natureza e da magnitude das atividades antropogênicas industriais, agrícolas e outras nos corpos d'água (SINGH *et al.*, 2009).

Em geral, a poluição orgânica em um sistema hídrico pode ser medida e expressa, por exemplo, através da Demanda Bioquímica de Oxigênio (DBO) e da quantidade de Oxigênio Dissolvido (OD) indicando o nível de clorofila-a (principal parâmetro para

indicação do nível eutrófico do sistema). Sendo assim, é de extrema necessidade o desenvolvimento de métodos para determinação da clorofila-a, ou de parâmetros correlacionados, seja através de medidas empíricas seja através de medidas numéricas (estratégias numéricas).

É importante ressaltar que o interesse na análise e a previsão dos parâmetros da qualidade da água têm aumentado substancialmente nos últimos anos, devido à crescente disponibilidade de métodos de inteligência computacional (CHAU, 2006 apud WEST; DELLANA, 2011). No entanto, alguns autores comentam sobre a dificuldade na previsão dos parâmetros ambientais devido à natureza complexa e cumulativa dos processos biológicos e a necessidade de longos prazos na aquisição de informações sobre as condições ambientais do sistema hídrico.

Nesta abordagem, as técnicas de Inteligência Computacional (IC) ou de Inteligência Artificial (IA), têm sido desenvolvidas como alternativas consubstanciadas para avaliação da qualidade da água.

Atualmente, a área de Inteligência Computacional envolve as principais técnicas de redes neurais, lógica *fuzzy*, algoritmo genético ou modelos híbridos (*neurofuzzy*). O termo inteligência é atribuído aos desenvolvimentos computacionais, pois são sistemas inspirados no comportamento humano ou que tentam reproduzi-lo mesmo estando longe de ser autônomo em inteligência ou no pensar. Apesar desta ressalva, esses sistemas têm encontrado grande aceitação em diversas áreas do conhecimento, e muitas vezes apresentam desempenho superior quando comparados aos métodos convencionais utilizados para descrever os processos (WU; BANZHAT, 2010).

No entanto, um impedimento para uma maior aplicação de técnicas de Computação Inteligente na modelagem da qualidade da água é a falta de conhecimento dos profissionais desta área em relação às técnicas de Inteligência Artificial quando comparados, por exemplo, aos profissionais de informática que já vêm aplicando estas técnicas há muito tempo. Outro fato que pode ser observado como responsável por uma falta de uma maior utilização destas técnicas nessa área é a falta de estratégias para sua aplicação nas áreas de Química Analítica Ambiental, Engenharia Ambiental, ou mesmo em Química, que comecem com a caracterização do problema utilizando as técnicas de Inteligência Artificial, assim como uma avaliação de sua aplicabilidade e eficiência na apresentação da resposta ao problema.

Apesar da disponibilidade crescente de *softwares* para implementação dessas técnicas, tais como redes neurais e sistema *fuzzy*, ter facilitado sua disseminação nas diversas

áreas, há ainda necessidade de maior compreensão, utilização destas metodologias e sua validação na descrição do comportamento da qualidade da água.

Segundo Kuo *et al.* (2007), a dificuldade em prever o comportamento de corpos de água é devido a complexidade física, química e biológica envolvidas nos processos hídricos. Além disso, os sistemas de qualidade da água são processos biológicos complexos (CARLSSON; LINDBERG, 1998; SPALL; CRISTION, 1997 apud WEST; DELLANA, 2011), o que torna essa dificuldade ainda maior. Os balanços de massas para representar o comportamento fenomenológico, a exemplo a eutrofização de lagos e reservatórios, começaram aproximadamente no final da década de 70 do século passado. Com o aumento do entendimento do processo de eutrofização, assim como com o avanço da capacidade computacional, modelos multidimensionais de hidrodinâmica da água têm sido desenvolvidos e técnicas têm sido aplicadas para descrever a qualidade da água, porém essas técnicas são pouco difundidas no Brasil e até mesmo em outros países.

Dentre os trabalhos publicados, merecem destaques os de Lu e Lo (2002) que retrata o diagnóstico de reservatório de água utilizando a lógica *fuzzy* para representar o processo de eutrofização em termos de parâmetros como fósforo total e clorofila-a; o de Chaves e Kojiri (2007), que desenvolveram estratégias em *neuro-fuzzy* considerando os parâmetros da qualidade da água associados aos processos de aeração e remoção de sedimentos no reservatório; Singh *et al.* (2009) que descreveram o treinamento, validação e aplicação de Redes Neurais Artificiais (RNA) para calcular o oxigênio dissolvido e a demanda bioquímica de oxigênio em rios; o de Han *et al.* (2011), que apresenta a aplicação de uma Rede Neural Radial (Radial Basis Function - RBF) para previsão da qualidade da água; o de Carvalho *et al.* (2010), que utilizaram as redes neurais e a estatística multivariada para avaliar a qualidade da água em dois corpos d'água de Sergipe. Neste último, o fenômeno da eutrofização foi idealizado como um problema de entrada-saída, ou seja, os dados de clorofila-a (*output*) foram associados aos dados dos nutrientes (*input*) e como resultado foi possível classificar os sistemas oligotróficos e hipertróficos.

Além disso, em relação às técnicas de Sistemas Inteligentes duas delas se destacam, em função da crescente utilização na modelagem da qualidade da água: as Redes Neurais Artificiais e a Lógica *Fuzzy* (Sistemas Difusos). Merece destaque, também, a combinação entre estas duas técnicas, um sistema híbrido denominado de *Neuro-Fuzzy*. Inseridos nesta abordagem, a disponibilidade de estratégias utilizando essas técnicas torna-se imprescindível como ferramenta para monitoramento e controle da qualidade dos recursos hídricos, sendo, portanto, este o objetivo principal deste trabalho.

Inserido neste contexto, o presente trabalho tem como objetivo principal desenvolver estratégias para utilização de sistemas inteligentes no monitoramento da qualidade da água em reservatórios de Sergipe. E para o alcance deste objetivo, os seguintes objetivos específicos foram definidos:

1. Monitorar da qualidade dos recursos hídricos nos reservatórios de Jacarecica e Marcela e no Rio Poxim, no Estado de Sergipe, através da caracterização física, química, microbiológica e toxicológica.
2. Elaborar softwares com uso da ferramenta MATLAB para aplicação das técnicas de Redes Neurais, Lógica Fuzzy e Neuro-Fuzzy para caracterização desses reservatórios com bases nos parâmetros físicos, químicos, microbiológicos e toxicológicos.
3. Desenvolver um índice multivariável para o estado trófico da água utilizando a técnica de Análise de Componentes Principais com uso da ferramenta MATLAB.
4. Construir interfaces aceitáveis para o usuário, direcionadas para uma aplicação mais ágil destas técnicas na caracterização dos reservatórios.

Neste cenário, a maior dificuldade da aplicação das técnicas de Computação Inteligente em Engenharia é o pouco conhecimento dos profissionais desta área em relação às técnicas de Inteligência Artificial e a sua aplicabilidade a fenômenos específicos, como a eutrofização de um sistema hídrico. Esta ausência de familiarização se deve em muito à falta de preparo deste profissional que vai desde a caracterização do problema de engenharia como passível de solução pelas técnicas de Inteligência Artificial, assim como de estratégias para sua aplicação e formas de avaliação de sua aplicabilidade e eficiência.

Em geral, os Sistemas Inteligentes encontram aplicações em Modelagem, Otimização, Identificação, Operação e Controle de processos. Estes sistemas são alternativas ou complementos das técnicas já consagradas de estatística, pesquisa operacional e modelagem numérica utilizadas frequentemente na Engenharia.

No que concerne ao monitoramento ambiental, o processo de eutrofização traz consequências devastadoras ao meio ambiente e sua análise numérica não é simples, tendo em vista à complexidade física, química e biológica que dificulta a predição do

comportamento dos nutrientes nos lagos e reservatórios. Os modelos determinísticos e empíricos são desenvolvidos de forma mais geral e necessitam, frequentemente, de algum nível de calibração com dados específicos experimentais. E, importantes fatores que afetam o fenômeno de eutrofização relacionados à natureza do ecossistema, são, geralmente, desprezados em aproximações clássicas. No entanto, esses parâmetros podem ser relacionados mais facilmente através da aplicação dos sistemas inteligentes.

É importante ressaltar que em relação ao monitoramento da qualidade de um corpo hídrico, o desenvolvimento experimental constitui um dos aspectos mais impactante no sucesso desse monitoramento. A qualidade e a quantidade de dados experimentais necessários para classificar a qualidade de um recurso hídrico são influenciadas por variáveis que muitas vezes independem do executor do projeto, a exemplo, as condições climáticas, econômicas e até mesmo políticas institucionais.

Neste sentido, esta tese de doutorado propõe o desenvolvimento de estratégias utilizando as técnicas de Redes Neurais, Lógica Fuzzy e Neuro-fuzzy, prevendo o desenvolvimento de estratégias numéricas para aplicação destas técnicas com uso do MATLAB, com ênfase em aplicações na Engenharia Ambiental e na área de Química Analítica Ambiental, para monitoramento da qualidade dos recursos hídricos, em particular nos corpos hídricos do Estado de Sergipe.

2 FUNDAMENTAÇÃO TEÓRICA

Neste item do trabalho, será apresentado o levantamento bibliográfico sobre os assuntos relacionados ao desenvolvimento da tese. Primeiro, será caracterizado o objetivo de estudo, a água, suas características e sua importância como elemento fundamental para a sustentabilidade e desenvolvimento socioeconômico de um país. Segundo, os principais processos que distinguem a qualidade da água. Terceiro, os modelos que constituem as estratégias numéricas para avaliação da qualidade dos recursos hídricos.

2.1 ÁGUA

De todos os recursos naturais, a água é sem dúvida o mais importante e o mais precioso. No entanto, a água vem apresentando uma variedade de contaminantes, em decorrência de um grande número de práticas destrutivas e de má gestão ambiental. Estas são ameaças aos recursos hídricos, que, em escala mundial, restringem os diversos usos de água (MACEDO, 2004; ABBASI; ABBASI, 2012).

O uso da água exige diferentes padrões para a qualidade do sistema hídrico. Dentre estes, os principais usos e parâmetros são:

- a. Para a preservação da vida aquática: oxigênio dissolvido, pH, estado trófico, algas, nutrientes, turbidez, substâncias tóxicas (metais, agrotóxicos)
- b. Para o abastecimento doméstico: turbidez, demanda bioquímica de oxigênio, algas, nutrientes, salinidade, substâncias tóxicas, coliformes fecais.
- c. Para a recreação de contato primário: coliformes fecais, algas, óleos e graxas.

De acordo com a ANA, criada através da Lei 9.984/2000 e que tem como missão regulamentar o uso dos recursos hídricos da União, vem sendo implantado um sistema nacional de gerenciamento desses recursos para garantir o uso sustentável, monitorando a qualidade e a quantidade dos corpos de água. De forma que inserido no Artigo 4º dessa lei, tem-se a importância dos diversos estudos ambientais:

Art. 4º-Cabe à ANA:

“Promover a elaboração de estudos para subsidiar a aplicação de recursos financeiros da União em obras e serviços... de controle da poluição hídrica, em consonância com o estabelecido nos planos de recursos hídricos”.

“Fiscalizar os usos de recursos hídricos nos corpos de água de domínio da União”.

“Propor ao CNRH o estabelecimento de incentivos, inclusive financeiros, à conservação qualitativa e quantitativa de recursos hídricos”.

“Organizar, implantar e gerir o Sistema de Informações Sobre Recursos Hídricos (SNIRH)”.

Em 2007, no Seminário Macrorregional para consolidação da implementação do decreto nº 5.440/2005 (“Estabelece definições e procedimentos sobre o controle de qualidade da água de sistemas de abastecimento e institui mecanismos e instrumentos para divulgação de informação ao consumidor sobre a qualidade da água para consumo humano.”), a ANA divulgou que são mensurados os parâmetros temperatura, pH, condutividade, oxigênio dissolvido e turbidez trimestralmente em 630 pontos de monitoramento.

Em 2012, o relatório da Conjuntura dos Recursos Hídricos no Brasil – Informe Especial, elaborado pela ANA, mostrou o panorama da qualidade da água em todo território nacional. Esse relatório consta que mais de 100 corpos d’água estão em condições péssimas ou ruins. A Figura 1 mostra os pontos de monitoramento da qualidade da água, classificando as regiões hidrográficas em péssima, ruim, regular, boa e ótima de acordo com o Índice de Qualidade das Águas (IQA). É importante ressaltar, conforme consta nesse relatório, “o IQA é particularmente sensível à contaminação pelo lançamento de esgotos, sendo um índice de referência normalmente associado à qualidade da água bruta captada para o abastecimento público após tratamento”.

Dessa forma, a poluição das águas pode ser melhor compreendida através da análise de dois fenômenos: a bioacumulação e a eutrofização. No caso da bioacumulação trata-se de um fenômeno através do qual os organismos vivos retêm, dentro de si, certas substâncias tóxicas sem conseguir realizar sua eliminação naturalmente. A eutrofização pode ser definida como sendo a fertilização, aumento da concentração de nutrientes, das águas de rios, lagos e represas. Essa poluição ocorre, por exemplo, como consequência da falta de um sistema de coleta e tratamento de esgoto doméstico (MAGOSSÍ; BONACELLA, 2003).



Figura 1 - Classificação do corpo hídrico de acordo como o Índice de Qualidade da Água
Fonte: ANA (2012)

Nesse sentido, o Atlas de Saneamento de 2011, divulgado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), mostra que houve aumento na proporção de domicílios com acesso à rede de esgoto que passaram de 33,5%, em 2000, para 45,7%, em 2008. Nesse relatório, consta que apenas a Região Sudeste apresenta mais da metade dos domicílios (69,8%) com acesso à rede geral, em seguida a região Centro-Oeste (33,7%), com resultado próximo ao da Região Sul (30,2%). No entanto, as Regiões Nordeste (29,1%) e Norte (3,5%) estão bem abaixo desses valores (RODRIGUES, 2011).

2.2 FENÔMENOS DE EUTROFIZAÇÃO

O processo de eutrofização pode ser compreendido como o aumento de concentração de nutrientes, principalmente fósforo e nitrogênio, nos ambientes aquáticos. O fenômeno de eutrofização pode acontecer de forma natural, contribuindo para o equilíbrio ecológico, ou

de forma cultural (eutrofização antrópica – artificial), como sendo a fertilização provocada por atividades humanas. A eutrofização artificial é capaz de quebrar o equilíbrio entre a produção e a decomposição de matéria orgânica no ambiente aquático. Esse desequilíbrio é provocado, principalmente, pela utilização sem controle de fertilizantes agrícolas e pela descarga de esgotos industriais e domésticos sem tratamento no corpo hídrico (JI, 2008).

Dessa forma, Magossi e Bonacella (2003) comentam que no processo de eutrofização, o aumento da concentração de nutrientes favorecendo o crescimento e a multiplicação do *fitoplâncton*, o que provoca o aumento da turbidez da água e como consequência, a luz solar não chega às plantas que se encontram submersas, não ocorrendo o processo de fotossíntese, o que leva a deterioração da água para consumo humano.

Sob essa análise, o estabelecimento do comportamento de corpos d'água, através da medida de parâmetros ambientais, funciona com uma importante ferramenta no combate a problemas ambientais, como o processo de eutrofização de reservatórios. Entretanto, esta é uma tarefa difícil devido à complexidade dos processos físicos, químicos e biológicos causadores desses problemas (NIJBOER; VERDONSCHOT, 2004; KUO *et al.*, 2007).

Os reservatórios são fontes de água superficiais construídos artificialmente e, em geral, são usados para geração de energia, recreação, prática de esporte, pesca comercial, irrigação, abastecimento para consumo humano etc. No entanto, esses ambientes estão sujeitos ao processo de eutrofização, sendo, portanto, importante o monitoramento da qualidade da água (DALTRO FILHO; SANTOS, 2002).

Uma gama de problemas está relacionada aos efeitos da eutrofização em reservatórios. Dentre estes, ocorre, principalmente, a diminuição da diversidade de espécies, o aumento de biomassa de plantas e animais, o aumento da turbidez e o aumento da taxa de sedimentação, e a diminuição da concentração de oxigênio nos reservatórios. No que se refere aos problemas relacionados ao uso de determinado recurso hídrico, os efeitos da eutrofização estão relacionados à dificuldade no processo de tratamento da água potável, o odor e o gosto desagradáveis, presentes nas linhas de suprimento de água, várias doenças podem ser transmitidas, o aumento da vegetação aquática, que pode impedir o fluxo da água e a navegação. Além disso, problemas de gestão ambiental e econômica são agravados por ocorrer uma necessidade de maior investimento nos processos de tratamento e monitoramento da água em todo mundo (PARR; MASON, 2004; KUO *et al.*, 2008; ZANG *et al.*, 2008).

2.2.1 Estado ou grau trófico

O estado ou grau trófico de um sistema hídrico é determinado através do Índice de Estado Trófico (IET), que representa entradas externas de nutrientes, como, esgoto doméstico, resíduos industriais e agrícolas, e está associado a características específicas de cada reservatório, a exemplo, tempo de retenção, vazão, regime hidrológico. Dessa forma, o IET serve de base para a elaboração de sistema de monitoramento da eutrofização e definição do uso do corpo hídrico (MAIA, 2011).

Nesse contexto, determinar o estado trófico ou o IET é uma ação fundamental para classificação dos sistemas hídricos, ou seja, para avaliar a qualidade da água quanto ao processo de eutrofização, conforme ressaltam Sulis *et al.* (2011) quando desenvolveram um modelo linear para estimativa da qualidade da água com base no IET.

O estado trófico de um reservatório pode ser avaliado através de índices estabelecidos através de equações que utilizam parâmetros limnológicos. O Índice de Estado Trófico (IET) de Carlson (1977) é bastante utilizado na literatura, principalmente pela facilidade de aplicação numérica (equações simples) em função de parâmetros ambientais, como as concentrações de fósforo total e nitrogênio total (HAVENS, 1995; CHACON-TORRES; ROSAS-MONGE, 1998; MATTHEWS *et al.*, 2002).

2.2.1.1 Parâmetros limnológicos

Inicialmente definida como sendo a ciência que estuda os lagos, Esteves (2011) define a limnologia como sendo o “estudo ecológico de todas as massas de água continentais, independentemente de suas origens, dimensões e concentrações salinas”.

Tundisi e Tundisi (2008) definiram também a Limnologia como sendo o estudo das reações funcionais e produtividade das comunidades bióticas de lagos, rios, reservatórios e região costeira em relação aos parâmetros físicos, químicos e bióticos ambientais. Dentre os parâmetros limnológicos, destacam-se a clorofila-a, transparência da água, concentrações de fósforo e de nitrogênio por serem bastante utilizados nas equações dos índices de estado trófico.

A definição e determinação dos parâmetros limnológicos são extremamente importantes para diferentes áreas do conhecimento científico voltadas ao meio ambiente, como por exemplo, manejo de ecossistemas, com ênfase ao controle de eutrofização,

aspectos para elaboração de Estudos de Impactos Ambientais (EIA) e Relatório de Impacto Ambiental (RIMA) e outros, conforme comenta Lundberg (2012).

A seguir serão descritos os parâmetros limnológicos clorofila-a, transparência da água e concentrações de nitrogênio e fósforos, referindo-se aos principais parâmetros para caracterização do estado trófico da água.

a. Clorofila-a

Designa-se de Clorofila o grupo de pigmentos fotossintéticos produzidos nos cloroplastos das folhas e em outros tecidos vegetais. Estes pigmentos são responsáveis pela cor verde das plantas e atuam como fotorreceptor da luz visível utilizada no processo da fotossíntese. São quatro os tipos de clorofilas: clorofila-a, b, c e d. As clorofilas a e b estão presentes em plantas verdes e as clorofilas c e d são encontradas nas algas e cianobactérias. Os pigmentos envolvidos no processo de fotossíntese são as clorofilas a e b, os carotenóides e as ficobilinas. As diferenças aparentes nas cores dos vegetais são devidas à presença de outros pigmentos associados, como os carotenóides, que sempre acompanham as clorofilas (MARTINI *et al.* 2006; TUNDISI; TUNDISI, 2008; CAMPOS, 2010).

Dessa forma, a clorofila é considerada como sendo o principal parâmetro indicador do estado trófico, pois o acúmulo de plantas nas zonas hídricas indica forte enriquecimento de nutrientes aquáticos (FERREIRA, 2011; DIMBERG *et al.* 2012).

O processo utilizado para obtenção é a extração da clorofila-a, na qual, comumente se utiliza a acetona a 90%. De acordo com a Resolução CONAMA 357/2005, recomendam-se valores de clorofila-a de até 30 µg/L (ambientes lênticos) como adequados para a manutenção de uma boa qualidade da água e para evitar impactos nos corpos da água que recebem descargas de efluentes.

b. Transparência da água – profundidade de Secchi

A medida de penetração de luz na coluna da água é um dos indicadores da qualidade da água e pode ser obtida através de um método simples e de baixo custo, conforme comentam Tundisi e Tundisi (2008). Essa medida utiliza um disco, normalmente de 20 cm de diâmetro (denominado “disco de Secchi”) de material resistente, pode-se avaliar o quanto a luz incidente na água está sendo diminuída em decorrência da presença das algas.

Segundo Esteves (2011), a profundidade de Secchi corresponde à profundidade em que a.

O disco de Secchi é um equipamento indispensável a qualquer trabalho na área de limnologia. O sistema consiste em uma corda, graduada, que é colocada na água até

desaparecer. Assim que o padrão gráfico do disco não é mais observado, anote-se o quanto este desceu na coluna d'água, sendo que o valor adequado para a manutenção de uma boa qualidade da água situa-se ao redor de 40 cm (ESTEVES, 2011).

c. Concentração de Fósforo

De maneira geral, a eutroficação caracteriza-se pelo aumento da produção de fitomassa de algas, causada pelo aumento da disponibilidade de nutrientes na água. As concentrações de fósforo e nitrogênio são os principais fatores que provocam o processo de eutrofização (ESTEVES, 2011).

Sendo assim, a determinação das concentrações de fósforo na água é uma das principais etapas para o monitoramento e controle da eutrofização.

Inserido no desenvolvimento do *fitoplâncton* tanto o fósforo quanto o nitrogênio são nutrientes limitantes, uma vez que o fósforo é necessário para sínteses moleculares e transporte de íons e o nitrogênio é necessário para síntese de aminoácidos e proteínas (FERREIRA, 2011).

Do ponto de vista limnológico, o fósforo por ser um elemento a ser sempre monitorado, constituindo-se em ótimo indicador da qualidade da água. As fontes naturais de fósforo são as rochas da bacia de drenagem, precipitação atmosférica e a deposição de material particulado, como partículas de solo e rochas, de organismos vivos e em decomposição, compostos voláteis liberados de plantas etc (ESTEVES, 2011).

A Resolução CONAMA 357/2005 recomenda, para ambientes lênticos (água parada, lagos) e híbridos (entre lênticos e lóticos - barragens), valores de até 30 µg/L como valores de referência para concentração de fósforo.

d. Nitrogênio

O nitrogênio é um dos principais elementos químicos, sendo que é mais abundante na forma gasosa na atmosfera. Em particular, devido a sua participação na formação de proteínas, o nitrogênio é um dos elementos mais importantes no metabolismo de ecossistemas aquáticos, podendo atuar como fator limitante na produção primária desses ecossistemas e, em determinadas condições, tornar-se tóxico a organismos aquáticos. Dentre os compostos de nitrogênio dissolvidos na água, pode-se encontrar uma forma ionizada (NH_4^+ íon amônio), e outra não ionizada (NH_3 amônia). As duas formas juntas constituem a amônia total ou nitrogênio amoniacal total (ESTEVES, 2011).

As principais fontes de nitrogênio para os ecossistemas aquáticos continentais são a

fixação biológica de nitrogênio que ocorre com a transformação de gás nitrogênio (N₂) em nitrogênio biológico, chuvas, aporte de efluentes domésticos e industriais não tratados ou parcialmente tratados lançados nos corpos de água (FERREIRA, 2011; ESTEVES, 2011).

Quanto maior for o pH, maior será a porcentagem da amônia total presente na forma NH₃ não ionizada (forma tóxica). Por exemplo, na piscicultura intensiva, a principal fonte de compostos nitrogenados incorporados à água advém da alimentação. Os fertilizantes e rações utilizadas em viveiros contêm nitrogênio, principalmente na forma de amônia e nitrato. O excesso desses elementos pode prejudicar o crescimento dos peixes. Além disso, esse nutriente é consumido também pelo fitoplâncton, acarretando geralmente um crescimento excessivo das algas. De forma que, o manejo inadequado de fertilizantes químicos compromete bastante a qualidade da água (STOATE *et al.*, 2009; RIJN, 2012).

2.2.1.2 Índice de Estado Trófico

O Índice de Estado Trófico (IET) é índice numérico composto por parâmetros físicos, químicos, biológicos para indicar as condições ambientais de um sistema hídrico. O IET é dividido em cinco grupos com intervalos de 0-20, 20-40, 40-60, 60-80, 80-100 correspondentes a cinco estados tróficos, hiperoligotrófico, oligotrófico, mesotrófico, eutrófico e hipereutrófico, respectivamente (XU, 2008).

Carlson e Simpson (1996) definiram o Índice de Estado Trófico (IET) com base na biomassa fitoplanctônica presente em determinado corpo d'água, em local e tempo específicos. Estes autores definiram também que o estado trófico não é o mesmo que a qualidade da água, mas um aspecto importante da mesma. Nesse índice, estão inseridos as variáveis clorofila-a, transparência (disco de Secchi), nitrogênio total e fósforo total para estimar a concentração de biomassa de algas. Como resultado, estes autores propuseram as equações mostradas na Tabela 1 para determinar o IET.

Tabela 1 - Índice de Estado Trófico.

Equação do IET	Parâmetros
$IET (DS) = 60 - 14,41 \ln(DS)$	Transparência do disco de Secchi (DS), medida em m
$IET (Chl) = 9,81 \ln(Chl) + 30,6$	Clorofila (Chl), medida em $\mu g.L^{-1}$
$IET (P_T) = 14,42 \ln(P_T) + 4,15$	Fósforo total (P _T), medido em $\mu g.L^{-1}$
$IET (N_T) = 54,45 + 14,43 \ln(N_T)$	Nitrogênio total (N _T), medido em $mg.L^{-1}$

Fonte: Carlson; Simpson (1996)

Um corpo d'água, geralmente, pode ser classificado como oligotrófico, mesotrófico e eutrófico, (e termos associados às características dos sistemas hídricos), em função da

produtividade biológica e as condições de nutrientes, definindo assim seu estado trófico. De acordo com Ji (2008):

- O termo oligotrófico é empregado para descrever um corpo d'água com baixa atividade biológica e excelente qualidade de água, pois é um ambiente pobre em nutrientes e algas e com produção de biomassa limitada;
- O termo mesotrófico é empregado para descrever um corpo d'água com média atividade biológica e a qualidade da água é considerada boa;
- O termo eutrófico é empregado para descrever um corpo d'água com excessiva atividade biológica e a qualidade da água é considerada ruim.

Considerando, então, o estado trófico, os reservatórios podem ser classificados nas classes apresentadas na Tabela 2 e uma combinação das variações do IET, concentrações de clorofila e fósforo total e transparência do disco Secchi.

Tabela 2 - Classificação do estado trófico

Estado trófico	IET	Clorofila ($\mu\text{g.L}^{-1}$)	Transparência do disco de Secchi (m)	Fósforo total ($\mu\text{g.L}^{-1}$)	Características
Oligotrófico	< 30	< 0,95	> 8	< 6	Água limpa e oxigênio no hipolimnio durante todo o ano.
Hipolimnia de lagos rasos	30-40	0,95-2,60	8 – 4	6 – 12	Pode tornar-se anóxico (falta de oxigênio).
Mesotrófico	40-50	2,60-7,30	4 – 2	12 – 24	Água moderadamente limpa, maior probabilidade de anoxia no hipolimnio no verão.
Eutrófico	50-60	7,30-20	2 – 1	24 – 48	Anoxia no hipolimnio
Domínio de algas azuis	60-70	20-56	0,5 – 1	48 – 96	Podem surgir gosto e odor desagradáveis.
Hipereutrófico	70-80	56-155	0,25 – 0,5	96 – 192	Produtividade limitada pela luz.
Algas inferiores e poucas macrófitas	> 80	> 155	<0,25	192 – 384	Domínio de peixes resistentes e possível morte de peixes no verão.

Fonte: Carlson; Simpson (1996)

Nesse sentido, muitos trabalhos têm sido desenvolvidos a fim de estudar o processo de eutrofização em reservatórios ou a qualidade da água em todo mundo tomando como referência a classificação dos corpos hídricos apresentada na Tabela 2.

Havens (1995) estudou a quantificação histórica, de 1973 a 1993, do nutriente limitante do lago Okeechobee, na Flórida, Estados Unidos. O nível do nutriente limitante (nitrogênio/fósforo) foi estimado por desvios entre o índice de estado trófico, calculado a partir da rotina de monitoramento de dados. O estudo da variação do índice de estado trófico histórico indicou uma forte tendência da limitação do fósforo à limitação secundária do nitrogênio entre 1970 e 1980, que coincidiu com mudanças na saída de nutrientes e manejo hídrico. Essa documentação da limitação secundária do nitrogênio foi importante para o manejo do lago, porque confirmou que os esforços progressivos e dispendiosos para reduzir a concentração de fósforo e criar uma condição limitada pelo fósforo no lago, na realidade constitui uma restauração da condição natural prévia.

Soyupak *et al.* (1997) desenvolveram estudos no reservatório de Keban Dam, oeste da Anatólia, Turquia, no qual diferentes seções do reservatório têm diferentes estados tróficos, sendo a porção superior do lago eutrófica. Nesse sentido, avaliaram, estratégias alternativas a serem implementadas para controle do fósforo e da eutrofização. Técnicas de modelagem matemática são usadas para composição de várias alternativas, em termos de sua relativa eficácia, para controle de fósforo. O modelo usado simulou a concentração de oxigênio dissolvido e a clorofila *eufótica* sob condições correspondentes a várias estratégias propostas para o controle de fósforo. Os esforços da modelagem produziram informações valiosas que podem ser usadas para tomada de decisões na avaliação de diferentes alternativas de manejo, incluindo redução de entrada de carga de fósforo.

Chacon-Torres e Rosas-Monge (1998) realizaram estudos de monitoramento limnológico anual no lago tropical Zirahuen, no México, que sofre uma forte influência antropogênica. Os estudos foram desenvolvidos com base nas características da qualidade da água e no nível trófico do lago, para incentivar o desenvolvimento de um programa de manejo para uso sustentável e conservação do ecossistema do lago. Os modelos de estados tróficos, índice de estado trófico proposto por Shannon e Brezonik (1972) e o proposto por Carlson (1977), sugeriram que o lago Zirahuen é um sistema oligotrófico. Contudo, o aumento da erosão e o aporte de esgoto não tratado indicaram a necessidade de estratégias de manejo do lago para conservação e desenvolvimento sustentável do sistema.

Ibañez *et al.* (2000) estudaram as características físicas e químicas de três sistemas aquáticos na Baixada Maranhense: os rios Pindaré (lago Viana), Pericumã e Turiaçu. Nesse trabalho, observou-se que os corpos d'água diferiram quanto à condutividade elétrica, sólidos totais suspensos e nutrientes, com elevados valores (exceto para o nitrogênio total) no período seco.

Perkins e Underwood (2000) realizaram estudos no reservatório Alton Water, Suffolk, UK, durante três anos (1995 a 1997) e observaram uma variação entre o estado eutrófico a mesotrófico. No verão de 1995, a concentração da clorofila-a noroeste do reservatório era $95 \pm 34 \mu\text{g.L}^{-1}$, 24 vezes maior que na bacia. As concentrações de fósforo total, ortofosfato, nitrato e amônia reduziram de maneira significativa ao longo do reservatório, enquanto a transparência do disco de Secchi aumentou. Nesse trabalho, os autores verificaram também que a noroeste do reservatório havia alta concentração de sólidos suspensos e biomassa algal, enquanto a bacia principal apresentou água limpa com alta transparência no disco de Secchi.

Havens *et al.* (2001) compararam a dinâmica dos nutrientes em três lagos, situados no Japão (lago Kasamigaura), China (lago Danghu) e Estados Unidos (lago Okeechobee), fortemente influenciados por fonte de poluição pontual e difusa, todos grandes (área maior que 30 km²), rasos (profundidade menor que 4 m) e eutróficos. Neste trabalho, constatou-se, nos três lagos, fortes correlações entre os sedimentos, a coluna d'água e as atividades antrópicas. Processos importantes afetam a dinâmica dos nutrientes como fixação de nitrogênio e a limitação da luz devido à resuspensão de sedimentos. Os autores observaram que a dinâmica variável de nutrientes é típica de sistemas de lagos rasos.

Freitas (2001) observou que as águas do reservatório Marcela (Itabaiana, Sergipe) eram impróprias para irrigação de hortaliças. A classificação do estado trófico variou de mesotrófico a hipereutrófico, predominando o ambiente eutrófico, sendo que o nutriente limitante foi o nitrogênio.

Gulati e Van Donk (2002) descreveram a origem da eutrofização e das técnicas de restauração em lagos rasos, como o lago Dutch, na Holanda. Os estudos revelaram que a redução de fósforo a níveis baixos não foi garantia de que ocorresse o mesmo com a população de cianobactérias, pois elas podem resistir a grandes variações do fósforo e, portanto, a variações na razão carbono:fósforo.

Matthews *et al.* (2002) avaliaram durante onze anos parâmetros para definir o estado trófico e o nutriente limitante no lago Whatcom, caracterizado como oligotrófico, localizado em uma cadeia de lagos em Washington, nos Estados Unidos. Os autores perceberam que, apesar das concentrações do fósforo total e o fósforo reativo solúvel serem relativamente baixas em todo o lago, houve diferenças significativas entre os locais amostrados. Os coeficientes de correlação entre clorofila, profundidade do disco de Secchi, nitrogênio total e nitrogênio inorgânico dissolvido foram altos. No final do verão, a correlação da biomassa algal foi melhor com o nitrogênio inorgânico dissolvido e o fósforo total. O índice de estado trófico baseado no fósforo total, nitrogênio total, clorofila e profundidade do disco de Secchi revelou que apesar do crescimento algal ser limitado pelo fósforo durante o ano, o nordeste do lago pode ter tido, como co-limitante, o nitrogênio, durante o final do verão e outono. O modelo observado no lago foi mais semelhante ao descrito para lagos eutróficos.

Cavenaghi *et al.* (2003) caracterizaram a água e o sedimento em cinco reservatórios da bacia do rio Tietê e correlacionaram as informações obtidas com a ocorrência das principais plantas aquáticas. Nesse trabalho foi possível concluir que a ocorrência de plantas submersas foi a variável mais dependente da transparência da água e transmissão de luz. O deslocamento a jusante no rio Tietê provoca a redução dos teores de fosfato e nitrogênio, além de alterar a turbidez e a quantidade de sólidos em suspensão. Os sólidos em suspensão apresentaram grande efeito na transmissão de luz por colunas de água, determinando ou não a ocorrência de plantas aquáticas imersas. A grande infestação de plantas marginais e flutuantes está associada à sedimentação que ocorre na área de ingresso de tributários nos reservatórios.

Jones e Knowlton (2005) observaram 135 reservatórios de Missouri, Estados Unidos, e analisaram a relação entre os nutrientes e a bacia hidrográfica para explorar a variação entre os sólidos totais suspensos e suas duas frações: sólidos suspensos voláteis e sólidos suspensos não voláteis. Os sólidos suspensos, em particular os voláteis, correlacionaram com o estado trófico do lago e se relacionaram positivamente com a proporção de terra cultivada na bacia, e negativamente com a floresta coberta e fracamente com a área de terra coberta com grama. Efeitos de fatores da bacia sobre os sólidos totais suspensos no verão refletiram no crescimento de plânctons mediante a influência de nutrientes (afetando os sólidos suspensos voláteis) sobre a entrada direta de sedimentos (afetando os sólidos suspensos não voláteis).

Souza *et al.* (2007) avaliaram as condições ambientais do reservatório de Acarape do Meio - Ceará em função dos parâmetros clorofila-a, fosfato, ortofosfato solúvel e a média entre estes parâmetros. Nesse trabalho foi observada uma tendência à hipereutrofia do reservatório, mostrando um agravamento da qualidade da água.

Guedes *et al.* (2011) realizam um estudo para caracterizar a qualidade da água e estimar o estado trófico do reservatório Lago do Amor, em Campo Grande - Mato Grosso do Sul. Nesse trabalho, os autores mostraram que, de acordo com os elevados valores dos parâmetros ambientais, como DBO, OD, fósforo, nitrogênio total, clorofila-a, o reservatório foi classificado como águas pertencentes à classe 4 (Resolução do Conselho Nacional do Meio Ambiente - CONAMA nº 430/2011 – água destinadas a navegação e harmonia paisagística). Os autores constataram, ainda, que o mesmo encontra-se eutrofizado e em degradação, em decorrência do processo de poluição e contaminação dos córregos tributários que deságuam no reservatório.

Akkoyunlu e Akiner (2012) analisaram a qualidade da água do lago Sapanca (Marmara - Turquia) em relação a quinze parâmetros ambientais (por exemplo, temperatura, OD, DBO, nitrato, nitrito e clorofila-a) e propuseram um índice de eutrofização (WQI_{eut}) que levou em conta um número menor de parâmetros. Os autores constatam que o índice proposto estabelecia uma boa relação com a análise obtida a partir de todos os parâmetros, e concluíram que o lago está em processo de eutrofização.

Assim, é importante ressaltar que o processo de eutrofização resulta do excesso de algas e macrófitas nos reservatórios, provocando a deterioração da qualidade da água para uso humano e decréscimo nos níveis do oxigênio dissolvido, com efeitos adversos para os peixes. Nesse contexto, a eutrofização é um fenômeno extremamente grave, conforme comentam Magossi e Bonacella (2003), pois além de comprometer a qualidade das águas, impõe condições que praticamente impedem a reversão natural, ou seja, um sistema eutrofizado tende a permanecer naturalmente assim. É possível observar que, na literatura da área, dentre os parâmetros, a quantidade de clorofila-a é utilizada amplamente como indicador do processo de eutrofização.

2.3 REDES NEURAIS

De acordo com Haykin (2001), uma Rede Neural Artificial (RNA) é um processador numérico constituído de unidades de processamento simples, que tem como finalidade armazenar conhecimentos empíricos e torná-los disponíveis para aplicação ou uso em futuros experimentos. A RNA se assemelha ao cérebro humano em dois aspectos: quando o conhecimento é adquirido pela rede através de um processo de aprendizagem e quando as forças de conexão entre neurônios (pesos sinápticos) são utilizadas para armazenar o conhecimento adquirido.

Dessa forma, as redes neurais artificiais possuem duas fases de processamento, a de aprendizagem e a de utilização, que seria a própria aplicação da rede. A primeira fase consiste no ajuste dos pesos das conexões, em resposta ao estímulo apresentado à rede neural (histórico de dados). A segunda fase consiste na resposta da rede a um estímulo de entrada, sem que ocorram modificações em sua estrutura de aprendizagem (OLIVEIRA JUNIOR *et al.*, 2007).

O conhecimento é adquirido pela rede por meio de dados do ambiente, em um processo de aprendizagem. O processo de treinamento é denominado de Algoritmo de Aprendizagem, e tem como finalidade ajustar os pesos sinápticos da rede de uma forma ordenada para alcançar um objetivo desejado (ZHANG *et al.*, 2011).

As conexões entre os neurônios, designadas de pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido. Então, uma rede neural funciona com vários neurônios, organizados em grupos (ou camadas) sequenciados. Os dados são alimentados na camada de entrada e a resposta da rede ao seu efeito (perturbação) é observada na saída. Podem existir uma ou mais camadas, denominadas de camada oculta, entre a camada de entrada e a camada de saída da rede. O número de camadas ocultas é dado pelas características de cada sistema.

As Redes Neurais podem, ainda, ser sintetizadas por sistemas de equações, em geral não-lineares, interconectadas, em que o resultado de uma equação é o valor de entrada para várias outras na rede. Esta estrutura foi concebida a partir de estudo sobre a fisiologia do cérebro humano e do seu componente primário, o neurônio. As redes neurais são aproximadores universais de funções, ou seja, conseguem reproduzir o comportamento de qualquer função matemática. Esta capacidade as popularizou, rapidamente, como uma ferramenta tipo Caixa Preta para modelar relações entre variáveis, de modo semelhante aos métodos estatísticos de regressão. (GARCIA *et al.*, 2004a; 2004b; CARVALHO *et al.*, 2007)

A arquitetura de uma rede neural artificial depende diretamente do problema que será tratado pela rede. Como parte da definição da arquitetura da rede e com o objetivo de classificá-las, tem-se: quantidades de camadas, números de neurônios em cada camada e tipo de conexão entre os neurônios e sua conectividade, conforme ilustra a Figura 2 (BRAGA *et al.*, 2007).

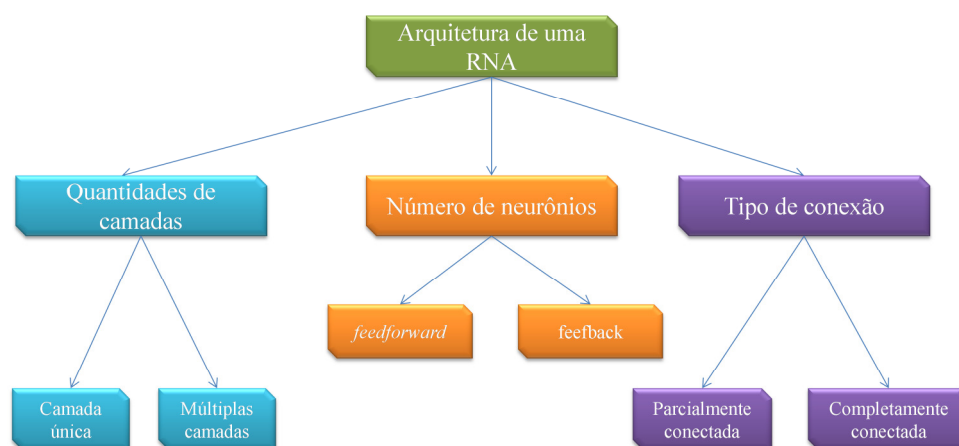


Figura 2 - Composição da arquitetura de uma rede neural

A Figura 3 mostra as redes de acordo com a composição da arquitetura.

Quanto ao número de camadas, pode-se ter:

- ✓ Redes de camada única: essa é a forma mais simples de uma rede em camadas e surge quando possui uma camada de entrada que se projeta para a camada de saída, mas não vice-versa, como mostrado na Figura 3 (a) e (d);
- ✓ Redes com múltiplas camadas: essas redes se distinguem das redes com camada única pela presença de uma ou mais camadas ocultas, Figura 3 (b) e (c). Nestas, a função das camadas ocultas é extrair informações das amostras.

Quanto aos tipos de conexões entre os neurônios, têm-se:

- ✓ Feedforward ou acíclica: neste tipo de rede, a saída do neurônio na i -ésima camada não pode ter entradas com neurônios em camadas de índice menor ou igual a i , como mostrado na Figura 3 (a), (b) e (c);
- ✓ Feedback ou cíclica: para esse tipo de rede, a saída do neurônio na i -ésima camada tem entradas com neurônios em camadas de índice menor ou igual a i , como mostrado na Figura 3 (d).

E quanto a sua conectividade, têm-se:

- ✓ Rede fracamente (ou parcialmente) conectada, como na Figura 3 (c);
- ✓ Rede completamente conectada, como mostrado na Figura 3 (a), (b) e (d).

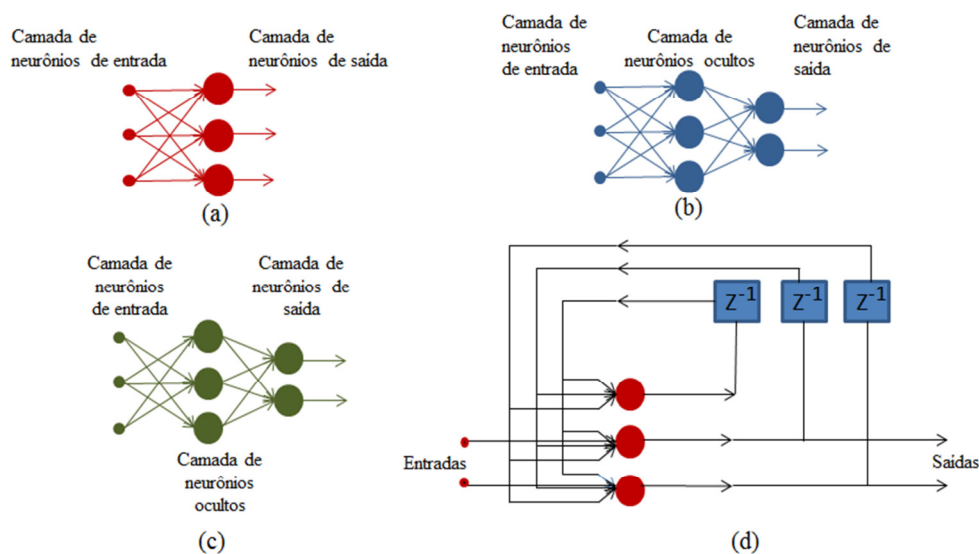


Figura 3 - Exemplos de arquiteturas de redes neurais artificiais
Fonte: Adaptada de Haykin (2001)

2.3.1 Perceptron de Múltiplas Camadas (Multi Layer Perceptron - MLP)

Os *perceptrons* de múltiplas camadas com retropropagação de erro têm sido aplicados com sucesso para resolver diversos problemas difíceis, através do seu treinamento de forma supervisionada com um algoritmo *backpropagation*. Este algoritmo é baseado na regra de aprendizagem por correção de erro (HAYKIN, 2001).

A Figura 4 mostra a arquitetura de uma rede *perceptron* de múltiplas camadas com uma camada de entrada, uma oculta e uma de saída, totalmente conectadas. Isso significa que um neurônio em qualquer camada da rede está conectado a todos os neurônios da camada anterior.

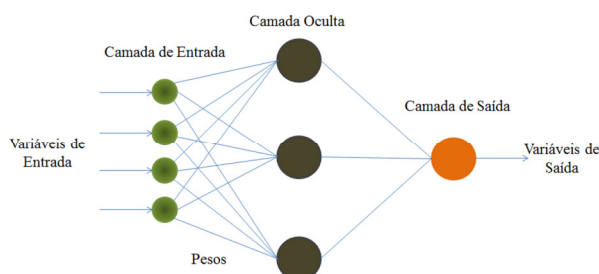


Figura 4 - Arquitetura *perceptron* de múltiplas camadas com uma camada oculta
Fonte: Adaptada de Haykin (2001)

Nos próximos itens serão mostradas as funções de ativações que foram empregadas na presente tese e que são importantes no desenvolvimento do algoritmo mais utilizado para as redes MLP - o algoritmo *backpropagation*.

2.3.2 Funções de ativação

A função de ativação define a saída (em geral não linear) de um neurônio, após o processamento da informação recebida pelo neurônio (HAYKIN, 2001). A Tabela 3 mostra as três funções de ativação mais utilizadas na literatura.

Tabela 3 - Funções de ativação.

Função	Comentário	Equação e sua derivada
Linear	Este tipo de função de ativação é muito utilizado nas unidades que compõem a camada de saída das arquiteturas MLP.	$f(x) = p.x \quad (1)$ $f'(x) = p.x$
	A origem deste tipo de função está vinculada à preocupação em limitar o intervalo de variação da função (0, 1) pela inclusão de um efeito de saturação. Sua derivada também é uma função contínua.	$f(x) = \frac{1}{1 + e^{-px}} \quad (2)$ $f'(x) = p.f(x).(1 - f(x))$
Tangente hiperbólica	Pelo fato da função sigmóide apresentar valores de ativação no intervalo (0, 1), em muitos casos ela é substituída pela função tangente hiperbólica, que preserva a forma sigmoidal da função sigmóide, mas assume valores positivos e negativos (-1, 1).	$f(x) = \frac{e^{px} - e^{-px}}{e^{px} + e^{-px}} = \tanh(px) \quad (3)$ $f'(x) = p.(1 - f(x)^2)$

Fonte: Braga *et al.* (2007)

A título de exemplo, na Figura 5 estão representados os principais elementos de uma típica Rede Neural Feedforward, sendo que x e y representam entrada e saída (variáveis do processo), respectivamente; w representa os pesos e f(x), a função de ativação. Nesta rede, a função de ativação é do tipo Sigmóide.

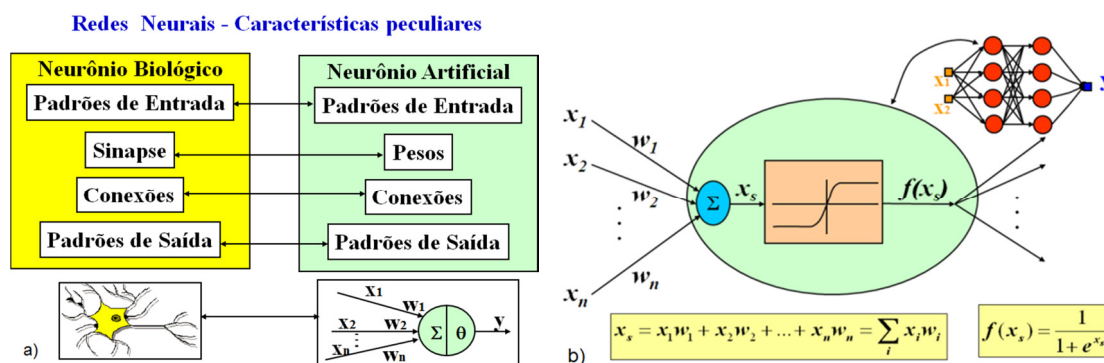


Figura 5 - Rede Neural Feedforward (a) e a função de ativação (b).

Fonte: Carvalho *et al.* (2010)

2.3.3 Algoritmo de retro-propagação (*Backpropagation*)

O *backpropagation* padrão é um algoritmo de gradiente descendente, no qual os pesos da rede são movidos ao longo do negativo do gradiente da função de desempenho. O termo *backpropagation* refere-se à maneira como o gradiente é computado para redes de múltiplas camadas não lineares. Existem diversas variações do algoritmo básico que são baseados em outras técnicas de otimização, tais como o gradiente conjugado e os métodos de Newton (HAYKIN, 2001).

Durante o treinamento com o algoritmo *backpropagation*, a rede opera em uma sequência de dois passos. Primeiro, um conjunto de variáveis é apresentado à camada de entrada da rede. A atividade flui através da rede, camada por camada, até que a resposta seja produzida pela camada de saída. No segundo passo, a saída fornecida pela rede é comparada com a saída desejada para esse conjunto particular. Se esta não estiver correta, o erro é calculado. O erro é propagado a partir da camada de saída para a camada de entrada, e os pesos das conexões das unidades das camadas internas vão sendo modificados até que o erro retro-propagado atinja o valor estabelecido (OLIVEIRA, 2000).

O treinamento ou aprendizado supervisionado de uma RNA visa a obtenção dos pesos das conexões que minimizam uma função erro dada pela diferença entre saída calculada pela rede e um valor de saída conhecido. Este processo de aprendizado ocorre através dos ajustes sucessivos dos pesos e, em alguns algoritmos, por meio dos resíduos de ativação, de forma a se obter dados de saída o mais próximo possível dos desejados. Esta é a etapa que exige maior demanda computacional. Um dos algoritmos de treinamento mais utilizados é o *backpropagation*, apesar da literatura trazer vários outros (TORRECILLA *et al.*, 2007).

A título de exemplo, a Figura 6 ilustra a fase de treinamento supervisionado clássico de uma RNA, sendo x o vetor de entrada de dados; Y_d o vetor de saída desejado e Y_c o vetor de saída calculado; W é o vetor de pesos e n é o número da iteração.

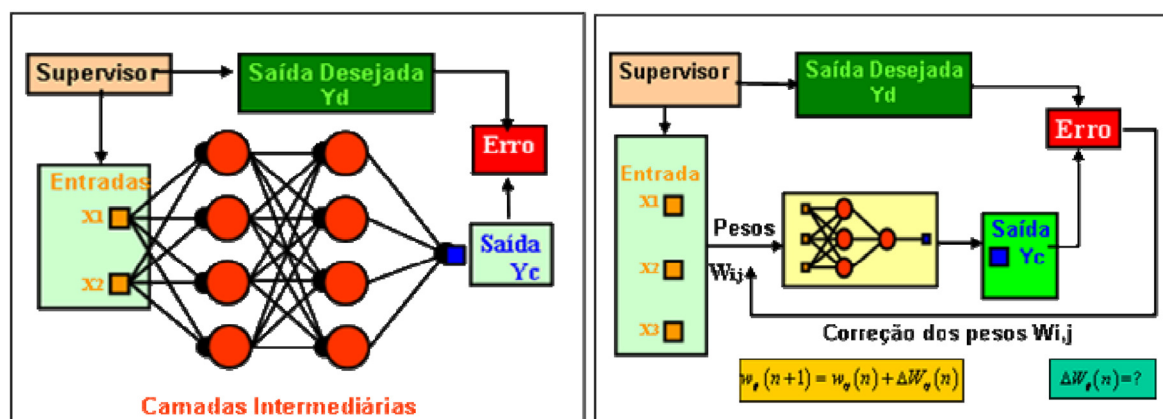


Figura 6 - Fase de treinamento de uma RNA
Fonte: Carvalho *et al.* (2010)

Como o algoritmo *backpropagation* (retro-propagação) foi historicamente um dos primeiros algoritmos desenvolvidos para o tratamento das redes neurais, este foi utilizado neste trabalho e será descrito a seguir.

Para facilitar a derivação do algoritmo *backpropagation*, será adotada a notação apresentada na Tabela 4.

Tabela 4 - Notação adotada no algoritmo *backpropagation*.

Simbologia	Significado
i, j	índices referentes a diferentes neurônios da rede
n	n -ésimo vetor de entrada (iteração)
N	número de amostras (padrões de treinamento)
M	número de camadas
$y_j(n)$	sinal de saída da unidade j na iteração n
$e_j(n)$	sinal de erro da unidade de saída j na iteração n
$w_{i,j}(n)$	peso sináptico conectando a saída da unidade i à entrada da unidade j na iteração n
$u_j(n)$	ativação da unidade j na iteração n ; sinal a ser aplicado à não-linearidade
$f_j(.)$	função de ativação associada à unidade j
X	matriz de dados de entrada (amostras de treinamento)
S	matriz de dados de saída (saídas desejadas)

Continuação...

$x_i(n)$	i-ésimo elemento do vetor de entradas
$s_j(n)$	j-ésimo elemento do vetor de saídas
α	taxa de aprendizagem
letras minúsculas em negrito (a , b , c)	vetores
letras maiúsculas em negrito (A , B , C)	matrizes
letras a, b, c	escalares

As matrizes \mathbf{W}^m (para $m = 0, 1, \dots, M - 1$; sendo M é o número de camadas da rede) possuem dimensão $S^{m+1} \times S^m$, em que $S^0 =$ número de entradas da rede; e os vetores \mathbf{b}^m possuem dimensão $S^{m+1} \times 1$.

Para simplificar o desenvolvimento do algoritmo backpropagation, utiliza-se a notação abreviada para uma arquitetura genérica com duas camadas (NARENDRA; PARTHASARATHY, 1990; HAGAN; MENHAJ, 1994; DEMULTH; BEALE, 2002;) e tomou-se como base a Figura 7.

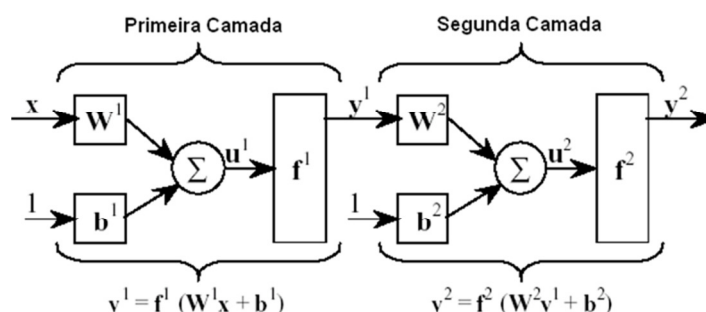


Figura 7 - Rede neural com duas camadas.

Fonte: Haykin (2001)

Na Figura 7 tem-se uma rede neural artificial com uma camada intermediária e uma camada de saída ($M = 2$); as unidades na primeira camada (camada oculta) recebem as entradas externas agrupadas em um vetor na forma:

$$\mathbf{y}^0 = \mathbf{x} \quad (4)$$

O vetor de saída da camada oculta da rede é dado por:

$$\mathbf{u}^1 = \mathbf{W}^1 \mathbf{x} + \mathbf{b}^1 \quad (5)$$

$$\mathbf{y}^1 = \mathbf{f}^1(\mathbf{u}^1) = \mathbf{f}^1(\mathbf{W}^1 \mathbf{x} + \mathbf{b}^1) \quad (6)$$

O vetor de saída da camada de saída da rede é dado por:

$$\mathbf{u}^2 = \mathbf{W}^2 \mathbf{y}^1 + \mathbf{b}^2 \quad (7)$$

$$\mathbf{y}^2 = \mathbf{f}^2(\mathbf{u}^2) = \mathbf{f}^2(\mathbf{W}^2 \mathbf{y}^1 + \mathbf{b}^2) \quad (8)$$

Logo, a saída da rede é dada em função do vetor de entrada \mathbf{x} , das matrizes de pesos \mathbf{W}^1 e \mathbf{W}^2 e dos vetores de limiares \mathbf{b}^1 e \mathbf{b}^2 . A expressão é:

$$\mathbf{y} = \mathbf{y}^2 = \mathbf{f}^2(\mathbf{W}^2 \mathbf{f}^1(\mathbf{W}^1 \mathbf{x} + \mathbf{b}^1) + \mathbf{b}^2) \quad (9)$$

Pode-se representar as equações anteriores de uma forma geral para um total de M camadas na rede. Assim:

$$\mathbf{u}^{m+1} = \mathbf{W}^{m+1} \mathbf{y}^m + \mathbf{b}^{m+1} \quad (10)$$

$$\mathbf{y}^{m+1} = \mathbf{f}^{m+1}(\mathbf{u}^{m+1}) = \mathbf{f}^{m+1}(\mathbf{W}^{m+1} \mathbf{y}^m + \mathbf{b}^{m+1}) \quad (11)$$

Sendo $m = 0, 1, \dots, M-1$.

O algoritmo *backpropagation* para as redes de múltiplas camadas é uma generalização do método dos quadrados mínimos (LS –*Least Squares*) e utiliza como medida de desempenho o erro quadrático médio (**MSE** –*Mean Squared Error*) (MUELLER, 1996). Inicialmente, é apresentado um conjunto de exemplos:

$$\{(\mathbf{x}_1, \mathbf{s}_1), (\mathbf{x}_2, \mathbf{s}_2), \dots, (\mathbf{x}_N, \mathbf{s}_N)\} \quad (12)$$

Sendo \mathbf{x}_n a n -ésima entrada para a rede e \mathbf{s}_n a saída desejada correspondente ($n = 1, \dots, N$).

Após cada entrada ser aplicada à rede, a saída produzida pela rede é comparada com a saída desejada, \mathbf{s} . O algoritmo deve ajustar os parâmetros da rede (pesos e limiares), com o objetivo de minimizar o erro quadrático médio na iteração n . Logo,

$$\mathbf{J}(n) = \mathbf{e}(n)^T \mathbf{e}(n) = (\mathbf{s}(n) - \mathbf{y}(n))^T (\mathbf{s}(n) - \mathbf{y}(n)) \quad (13)$$

Para $\mathbf{e}(n)$ sendo o erro $(\mathbf{s}(n) - \mathbf{y}(n))$ na iteração n .

A lei de ajuste, denominada de *steepest descente*, para minimizar o erro quadrático é dada por:

$$w_{i,j}^m(n+1) = w_{i,j}^m(n) - \alpha \frac{\partial J(n)}{\partial w_{i,j}^m} \quad (14)$$

$$b_i^m(n+1) = b_i^m(n) - \alpha \frac{\partial J(n)}{\partial b_i^m} \quad (15)$$

Na qual α é a taxa de aprendizagem.

Como o erro é função indireta dos pesos nas camadas intermediárias, a regra da cadeia deverá ser usada para o cálculo das derivadas. O conceito de regra da cadeia será utilizado na determinação das derivadas das Equações 14 e 15:

$$\frac{\partial J}{\partial w_{i,j}^m} = \frac{\partial J}{\partial u_i^m} \frac{\partial u_i^m}{\partial w_{i,j}^m} \quad (16)$$

$$\frac{\partial J}{\partial b_i^m} = \frac{\partial J}{\partial u_i^m} \frac{\partial u_i^m}{\partial b_i^m} \quad (17)$$

Porém,

$$u_i^m = \sum_{j=1}^{S^{m-1}} w_{i,j}^m y_j^{m-1} + b_i^m$$

Logo,

$$\frac{\partial u_i^m}{\partial w_{i,j}^m} = y_j^{m-1}, \quad \frac{\partial u_i^m}{\partial b_i^m} = 1.$$

Definindo agora a sensibilidade de J, a mudança no i-ésimo elemento da ativação da rede na camada m como:

$$\delta_i^m = \frac{\partial J}{\partial u_i^m} \quad (18)$$

As Equações 16 e 17 podem ser simplificadas para:

$$\frac{\partial J}{\partial w_{i,j}^m} = \delta_i^m y_j^{m-1} \quad (19)$$

$$\frac{\partial J}{\partial b_i^m} = \delta_i^m \quad (20)$$

Agora, é possível aproximar as Equações 14 e 15 através de

$$w_{i,j}^m(n+1) = w_{i,j}^m(n) - \alpha \delta_i^m y_j^{m-1} \quad (21)$$

$$b_i^m(n+1) = b_i^m(n) - \alpha \delta_i^m \quad (22)$$

Em notação matricial, as duas equações anteriores tornam-se:

$$\mathbf{W}^m(n+1) = \mathbf{W}^m(n) - \alpha \delta^m (\mathbf{y}^{m-1}) \quad (23)$$

$$\mathbf{b}^m(n+1) = \mathbf{b}^m(n) - \alpha \delta^m \quad (24)$$

Sendo:

$$\delta^m = \frac{\partial J}{\partial \mathbf{u}^m} = \begin{bmatrix} \frac{\partial J}{\partial u_1^m} \\ \frac{\partial J}{\partial u_2^m} \\ \vdots \\ \frac{\partial J}{\partial u_{s^m}^m} \end{bmatrix} \quad (25)$$

Ainda é necessário calcular as sensibilidades δ^m , que requerem outra aplicação da regra da cadeia. É este processo que dá origem ao termo retro-propagação (*backpropagation*), pois descreve a relação de recorrência na qual a sensibilidade na camada m é calculada a partir da sensibilidade na camada $m+1$.

Para derivar a relação de recorrência das sensibilidades, utiliza-se a seguinte matriz jacobiana:

$$\frac{\partial \mathbf{u}^{m+1}}{\partial \mathbf{u}^m} = \begin{bmatrix} \frac{\partial u_1^{m+1}}{\partial u_1^m} & \frac{\partial u_1^{m+1}}{\partial u_2^m} & \cdots & \frac{\partial u_1^{m+1}}{\partial u_{s^m}^m} \\ \frac{\partial u_2^{m+1}}{\partial u_1^m} & \frac{\partial u_2^{m+1}}{\partial u_2^m} & \cdots & \frac{\partial u_2^{m+1}}{\partial u_{s^m}^m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_{s^{m+1}}^{m+1}}{\partial u_1^m} & \frac{\partial u_{s^{m+1}}^{m+1}}{\partial u_2^m} & \cdots & \frac{\partial u_{s^{m+1}}^{m+1}}{\partial u_{s^m}^m} \end{bmatrix} \quad (26)$$

Em seguida encontra-se uma expressão para esta matriz. Considerando o elemento i, j da matriz:

$$\frac{\partial u_i^{m+1}}{\partial u_j^m} = w_{i,j}^{m+1} \frac{\partial y_j^m}{\partial u_j^m} = w_{i,j}^{m+1} \frac{\partial f^m(u_j^m)}{\partial u_j^m} w_{i,j}^{m+1} \dot{f}^m(u_j^m) \quad (27)$$

Sendo:

$$\dot{f}^m(u_j^m) = \frac{\partial f^m(u_j^m)}{\partial u_j^m} \quad (28)$$

Entretanto, a matriz jacobiana pode ser escrita como:

$$\frac{\partial \mathbf{u}^{m+1}}{\partial \mathbf{u}^m} = \mathbf{W}^{m+1} \dot{\mathbf{F}}^m(\mathbf{u}^m) \quad (29)$$

Para a qual, tem-se:

$$\dot{\mathbf{F}}^m(\mathbf{u}^m) = \begin{bmatrix} \dot{f}^m(u_1^m) & 0 & \dots & 0 \\ 0 & \dot{f}^m(u_2^m) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dot{f}^m(u_{S^m}^m) \end{bmatrix} \quad (30)$$

Agora, pode-se escrever a relação de recorrência para a sensibilidade utilizando a regra da cadeia em forma matricial:

$$\delta^m = \frac{\partial J}{\partial \mathbf{u}^m} = \left(\frac{\partial \mathbf{u}^{m+1}}{\partial \mathbf{u}^m} \right)^T \frac{\partial J}{\partial \mathbf{u}^{m+1}} = \dot{\mathbf{F}}^m(\mathbf{u}^m) (\mathbf{W}^{m+1})^T \frac{\partial J}{\partial \mathbf{u}^{m+1}} \quad (31)$$

$$\delta^m = \dot{\mathbf{F}}^m(\mathbf{u}^m) (\mathbf{W}^{m+1})^T \delta^{m+1} \quad (32)$$

Observa-se que as sensibilidades são propagadas da última para a primeira camada através da rede:

$$\delta^M \rightarrow \delta^{M-1} \rightarrow \dots \rightarrow \delta^2 \rightarrow \delta^1 \quad (33)$$

Ainda existe um último passo a ser executado para que o algoritmo de retro-propagação fique completo. Precisa-se do ponto de partida, δ^M , para a relação de recorrência da Equação (32). Este ponto é obtido na última camada:

$$\delta_i^M = \frac{\partial J}{\partial u_i^M} = \frac{\partial (\mathbf{s} - \mathbf{y})^T (\mathbf{s} - \mathbf{y})}{\partial u_i^M} = -2(s_i - y_i) \frac{\partial y_i}{\partial u_i^M} \quad (34)$$

Como:

$$\frac{\partial y_i}{\partial u_i^M} = \frac{\partial y_i^M}{\partial u_i^M} = \dot{f}^M(u_j^M) \quad (35)$$

Pode-se escrever:

$$\delta_i^M = -2(s_i - y_i) \dot{f}^M(u_j^M) \quad (36)$$

A Equação 36 pode ser colocada em forma matricial, resultando na Equação 37.

$$\delta^M = -2 \mathbf{F}^M(\mathbf{u}^M)(\mathbf{s} - \mathbf{y}) \quad (37)$$

2.3.4 Algoritmo de otimização para treinamento supervisionado

A maior parte dos algoritmos de treinamento é baseada nos métodos de gradientes descendentes e de Newton. As abordagens baseadas nos métodos de Newton apresentam, em geral, melhores resultados pelo fato de serem métodos de segunda ordem, apresentando uma convergência quadrática próxima ao mínimo. No entanto, estes métodos são limitados pelo grande espaço de memória necessário e pelo volume de cálculos matriciais envolvidos, o que os torna praticamente inviáveis para redes de grande dimensão (HAYKIN, 2001).

Um método simplificado para o treinamento de RNA é o algoritmo de Lavenberg-Marquart, desenvolvido para se obter uma rápida velocidade de treinamento (TORRECILLA *et al.*, 2007).

Portanto, descreve-se o método de otimização não-linear irrestrito para treinamento de redes multicamadas. O treinamento de redes neurais com várias camadas pode ser entendido como um caso especial de aproximação de funções, no qual não é levado em consideração nenhum modelo dos dados específico (SAMPAIO, 2006).

Na presente tese, o algoritmo escolhido foi o de Levenberg-Marquardt (LM). A escolha deste método foi feita com base na capacidade que o mesmo possui de conseguir convergências mais rápidas em relação aos outros algoritmos nas mais variadas aplicações, como reconhecimento de padrões e em problemas de aproximação de funções (DEMULTH; BEALE, 2002; HAGAN; MENHAJ, 1994; SAMPAIO, 2006).

2.3.4.1 Método de Levenberg-Marquardt (LM)

O método de Levenberg-Marquardt é um algoritmo bastante eficiente quando se trata de redes que não possuem mais do que algumas centenas de conexões a serem ajustadas (MARQUARDT, 1963; HAGAN; MENHAJ, 1994). Isto se deve, principalmente, ao fato de que estes algoritmos necessitam armazenar uma matriz quadrada cuja dimensão é da ordem do número de conexões da rede.

Considerando como funcional de erro a soma dos erros quadráticos, e ainda levando em conta que o problema pode ter múltiplas saídas, obtém-se a seguinte expressão para o funcional de erro:

$$J(\theta) = \sum_{i=1}^N \sum_{j=1}^m \left(g_{ij}(\mathbf{x}) - \hat{g}_{ij}(\mathbf{x}, \theta) \right)^2 = \sum_{k=1}^q r_k^2 \quad (38)$$

Sendo $J(\theta)$ o funcional de erro, $\hat{g}_{ij}(\mathbf{x}, \theta)$ o modelo que procura aproximar a função $g_{ij}(\mathbf{x})$, N o número de amostras, l o número de unidades intermediárias, r o erro residual, m o número de saídas, e q o produto $N \times m$.

Seja \mathbf{J} o Jacobiano (matriz das derivadas primeiras) do funcional J dado pela Equação 38. Esta matriz pode ser escrita da seguinte forma:

$$\mathbf{J} \equiv \begin{bmatrix} \nabla r_1^T \\ \nabla r_2^T \\ \vdots \\ \nabla r_q^T \end{bmatrix} \quad (39)$$

Sendo r denominado erro residual.

Diferenciando a Equação 38 obtém-se:

$$\nabla J = 2\mathbf{J}^T \mathbf{r} = 2 \sum_{k=1}^q r_k \nabla \mathbf{r}_k \quad (40)$$

$$\nabla^2 J = 2 \left(\mathbf{J}^T \mathbf{J} + \sum_{k=1}^q r_k \nabla^2 \mathbf{r}_k \right) \quad (41)$$

A matriz de derivadas segundas do funcional de erro é chamada de matriz hessiana. Quando os erros residuais são suficientemente pequenos, a matriz hessiana pode ser aproximada pelo primeiro termo da Equação 41, resultando em:

$$\nabla^2 J \approx 2\mathbf{J}^T \mathbf{J} \quad (42)$$

Esta aproximação geralmente é válida em um mínimo de J para a maioria dos propósitos, e é a base para o método de Gauss-Newton (HAGAN; FORESSE, 1997). A lei de atualização torna-se então:

$$\Delta\theta = [\mathbf{J}^T \mathbf{J}]^{-1} \mathbf{J}^T \mathbf{r} \quad (43)$$

A modificação de Levenberg-Marquardt para o método de Gauss-Newton é:

$$\Delta\theta = [\mathbf{J}^T \mathbf{J} + \mu \mathbf{I}]^{-1} \mathbf{J}^T \mathbf{r} \quad (44)$$

O efeito da matriz adicional $\mu \mathbf{I}$ é adicionar μ a cada autovalor de $\mathbf{J}^T \mathbf{J}$. Uma vez que a matriz $\mathbf{J}^T \mathbf{J}$ é semidefinida positiva e, portanto, o autovalor mínimo possível é zero, qualquer valor positivo, pequeno, mas numericamente significativo, de μ , será suficiente para restaurar a matriz aumentada e produzir uma direção descendente de busca.

Os valores de μ podem ser escolhidos de várias maneiras, sendo a mais simples escolher zero, a menos que a matriz hessiana encontrada na iteração i seja singular. Quando isso ocorrer, um valor pequeno como $\mu = 10^{-4} \sum_i (\mathbf{J}^T \mathbf{J})_{ii}$ pode ser usado. Outras formas de determinação do parâmetro μ são sugeridas por Hagan e Menhaj (1994).

É importante observar que, quanto maior for o valor de μ , menor é a influência da informação de segunda ordem e mais este algoritmo se aproxima de um método de primeira ordem.

2.3.5 Treinamento da rede MLP

O treinamento de uma rede MLP envolve os seguintes passos: modo de treinamento, topologia da rede, divisão dos dados e avaliação da eficiência do treinamento. Ressalta-se que será adotada neste trabalho a função hiperbólica como função de ativação na camada oculta e, na camada de saída, a função linear. Além disso, é importante comentar que o objetivo é obter uma rede neural com o melhor desempenho ou o menor erro quadrático médio de previsão das amostras de validação.

2.3.5.1 Modo de treinamento

O programa usado para treinar e testar as redes neurais neste trabalho foi o Neural Networks Toolbox para uso com MATLAB. Neste *Toolbox*, os algoritmos foram

desenvolvidos para treinamento em lote. Segundo Demuth e Beale (2002), existem duas diferentes formas de treinamento para o algoritmo *backpropagation*: o treinamento na forma sequencial e o treinamento na forma de lote. Na forma sequencial, a atualização dos pesos é realizada após a apresentação de cada exemplo, ou amostra, de treinamento. Já na forma em lote, todos os exemplos, ou amostras, são aplicados a rede antes da atualização dos pesos. Uma apresentação completa de todo o conjunto de treinamento é denominada de época.

2.3.5.2 Topologia da rede

Konderla e Mokanek (2000) e Swingler (1996) apud Palma e Portugal (2009) argumentaram que para a topologia da rede, número de camadas e de neurônios da rede, na resolução de problemas com o uso de redes neurais artificiais, têm sido utilizadas redes com apenas uma, e algumas vezes, duas camadas ocultas. Neste trabalho foram utilizadas redes contendo uma única camada oculta. Segundo Oliveira (2000), o uso de uma única camada interna tem se mostrado suficiente na modelagem de processos químicos, visto que quando há necessidade de modelos mais complexos o ajuste do número de neurônios na camada oculta geralmente é suficiente.

O número de neurônios na camada de entrada é, em geral, igual ao número de variáveis de entrada do processo. Entretanto, este número pode ser reduzido através do uso de técnicas estatísticas de redução de dimensionalidades (variáveis), como a análise dos componentes principais (PCA). Para o número de neurônios na camada oculta não há ainda uma regra que indique o número necessário para se obter resultados satisfatórios no treinamento da rede. Turner *et al.* (1996) apresentam algumas observações gerais para determinação da topologia da rede consistindo em:

- a. A rede deve ter a estrutura mais simples possível, para evitar sobre-parametrização;
- b. Pode ser demonstrado que qualquer função contínua não linear pode ser modelada utilizando uma camada oculta;
- c. O número de neurônios na camada oculta deve ser inicialmente igual ao número de entradas. Do ponto de vista prático, este procedimento funciona de maneira satisfatória e tende a manter um número relativamente pequeno de pesos necessários para a rede. Se a rede falhar para modelar as relações de entrada e saída, o número de neurônios na camada oculta pode ser aumentado.

Da mesma forma que na camada de entrada, o número de neurônios na camada de saída é igual ao número de variáveis de saídas (variáveis a serem preditas) do processo. Segundo Oliveira-Esquerre (2003), é recomendável que cada modelo apresente uma única resposta (um neurônio) na camada de saída, o que diminui o número de parâmetros a serem ajustados e, conseqüentemente, a carga computacional exigida. Uma exceção a esta regra é para situações em que se deseja prever diversas respostas correlacionadas, como as concentrações de diferentes constituintes de uma mistura em um sistema fechado. As redes neurais podem ser utilizadas de duas formas:

- a. Um modelo que forneça todas as variáveis de saída a partir das entradas fornecida e;
- b. Um modelo de RNA para cada saída da rede, lembrando que para isto é necessário que se tenha o mesmo conjunto de dados padrões para treinamento.

2.3.5.3 Conjunto de dados

Em geral, o conjunto de dados disponível para análise de determinados fenômenos é, geralmente, imposto ou limitado em problemas práticos. Segundo Oliveira-Esquerre (2003), é possível obter excelentes resultados para a modelagem de sistemas utilizando um número limitado de dados durante o treinamento. Entretanto, se for validar o modelo para um conjunto independente de dados, geralmente, uma significativa degradação dos resultados pode ser observada devido ao sobre-ajuste (ou *overfitting*) dos parâmetros e, conseqüentemente, perda da habilidade de generalização.

Dentro desse contexto, um importante passo no desenvolvimento de um modelo está na divisão do conjunto de dados disponíveis em dois ou três subconjuntos, quais sejam:

1. Treinamento – utilizado para estimar os parâmetros do modelo;
2. Validação – utilizada para verificar a habilidade de generalização do modelo frente a amostras independentes do conjunto de treinamento;
3. Teste – utilizado para validar o modelo usando novas amostras.

Ressalta-se que, dependendo da quantidade de dados disponível, pode-se ter apenas o conjunto de dados de treinamento e validação.

De acordo com Despagne e Massart (1998), o ideal para um conjunto de dados considerado grande é dividir este conjunto em 40% das amostras para treinamento, 20 % para validação e 40% para teste. A performance da rede não deve ser julgada pelo ajuste dos

dados de treino, pois estes podem ser ajustados perfeitamente. Os resultados podem ser apresentados tanto pelo conjunto de validação como pelo conjunto de teste. Outros autores (HAYKIN, 2001; OLIVEIRA *et al.*, 2007; LIMA *et al.*, 2009; NAZÁRIO *et al.*, 2009) sugeriram ou utilizaram como particionamento do conjunto a seguinte forma: 50% dos dados para treinamento, 25% para validação e 25% para teste.

No processo de aprendizagem e validação deve-se observar com atenção à escolha dos conjuntos de dados, pois a rede deve ser treinada sobre o mais amplo domínio possível, de forma que o conjunto de validação esteja contido no conjunto de aprendizagem. Neste contexto, uma das limitações das redes reside na dificuldade de extrapolar dados para os quais a rede não foi treinada.

2.3.5.4 Generalização

Quando a rede é treinada para atingir um erro mínimo, esta, na maioria dos casos, é incapaz de prever bem amostras que não foram usadas no conjunto de treinamento. A este fato é dado o nome de sobre-ajuste (*overfitting*), pois a rede se especializou nos dados de treinamento e perdeu sua capacidade de generalizar para novas situações. A seguir são apresentados dois métodos para melhorar a generalização dos dados: a parada antecipada e a regularização.

2.3.5.4.1 Parada antecipada (*early stopping*)

Quando é feito o treinamento de uma rede neural, geralmente deseja-se obter uma rede com a melhor capacidade de generalização possível, ou seja, a maior capacidade de responder corretamente a dados que não foram utilizados no processo de treinamento. As arquiteturas convencionais, totalmente conectadas, como o MLP, estão sujeitas a sofrerem sobre-treinamento (*overtraining*): quando a rede parece estar representando o problema cada vez melhor, ou seja, o erro do conjunto de treinamento continua diminuindo, mas, em algum ponto deste processo, a capacidade da rede em responder a um novo conjunto de dados piora. Para combater o sobre-treinamento pode-se utilizar os procedimentos de parada antecipada que são largamente utilizados por serem de fácil entendimento e implementação (HAYKIN, 2001; SILVA NETTO, 2006; RIBEIRO, 2007).

Na parada antecipada, o conjunto de treinamento é usado para computar o gradiente e atualizar os pesos da rede. O erro do conjunto de validação é monitorado durante o

processo de treinamento. No entanto, quando a rede inicia a sobre ajustar os dados, o erro no grupo de validação irá aumentar. Quando o erro de validação aumenta para um número específico de iterações, o treinamento é parado e os pesos no erro mínimo de validação são retornados. Esta regra é conhecida como parada antecipada e está ilustrada na Figura 6. A parada antecipada foi utilizada no presente trabalho. A raiz quadrada do erro quadrático médio (RMSE – *Root Mean Squared Error*) do conjunto de validação foi usada para avaliar o desempenho dos modelos empíricos usados.

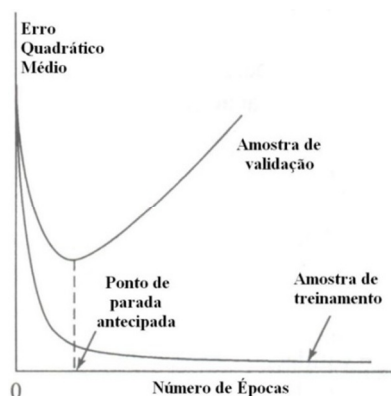


Figura 6 - Regra de parada antecipada baseada na validação cruzada

Fonte: Haykin (2001)

2.3.5.4.2 Regularização bayesiana

Um modelo desejado de rede neural deve produzir pequeno erro não somente nos dados de treinamento, mas também nos dados que não pertencem ao conjunto de treinamento (conjunto de validação ou de teste). Para produzir uma rede com a melhor capacidade de generalização, Mackay (1992) propôs um método para restringir os valores dos parâmetros da rede através da regularização. A técnica de regularização direciona a rede a responder suavemente e, então, é pouco provável que o sobreajuste (*overfitting*) aconteça (HAGAN; FORESEE, 1997).

Na técnica de regularização, a função de custo F é definida como:

$$F_{\text{reg}} = \alpha F + \beta F_w \quad (45)$$

Na Equação 45 F é uma função típica para treinar redes neurais do tipo MLP, que é a soma dos erros quadráticos, dada por:

$$F = \frac{1}{N} \sum_{i=1}^N (s_i - y_i)^2 \quad (46)$$

E F_w é a soma dos quadrados dos parâmetros (pesos e bias), dada por:

$$F_w = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (47)$$

Sendo que na Equação 45 α e β são os parâmetros da função objetivo.

O problema da regularização é a dificuldade em determinar um valor adequado para a taxa. Se o valor for muito grande poderá levar a um sobreajuste e, se for muito pequeno, a rede poderá não ajustar adequadamente os dados de treinamento (SAMPALIO, 2006). É desejável determinar esses parâmetros de uma forma automatizada e, sendo uma destas abordagens o processo que usa a estrutura Bayesiana.

Na estrutura Bayesiana, os pesos são considerados aleatórios e variáveis. Depois os dados são tomados e a função densidade para os pesos podem ser antecipados de acordo com a regra de Bayes:

$$P(\mathbf{w} | D, \alpha, \beta, M) = \frac{P(D | \mathbf{w}, \beta, M) P(\mathbf{w} | \alpha, M)}{P(D | \alpha, \beta, M)} \quad (48)$$

Na Equação 48, D representa o conjunto de dados, M o modelo de rede neural usado e \mathbf{w} o vetor de pesos da rede. $P(\mathbf{w} | \alpha, M)$ é a densidade anterior, que representa o conhecimento dos pesos antes de qualquer dado ser coletado. $P(D | \mathbf{w}, \beta, M)$ é a função de probabilidade dos dados, quando indicados os pesos \mathbf{w} . $P(D | \alpha, \beta, M)$ é o fator de normalização, que garante que a probabilidade total seja igual a 1.

Se for assumido que o ruído no conjunto de treinamento é Gaussiano e a distribuição anterior para os pesos é Gaussiana, a função probabilidade dos dados pode ser escrita como:

$$\begin{aligned} P(D | \mathbf{w}, \beta, M) &= \frac{1}{Z_D(\beta)} \exp(-\beta E_D) \\ P(\mathbf{w} | D, \alpha, \beta, M) &= \frac{1}{Z_w(\alpha)} \exp(-\alpha E_w) \end{aligned} \quad (49)$$

Sendo $Z_D(\beta) = \left(\frac{\pi}{\beta}\right)^{n/2}$ e $Z_w(\alpha) = \left(\frac{\pi}{\alpha}\right)^{N/2}$. Substituindo estas probabilidades na

Equação (48), obtém-se:

$$P(\mathbf{w} | D, \alpha, \beta, M) = \frac{\frac{1}{Z_D(\beta)} \frac{1}{Z_w(\alpha)} \exp(-(\beta E_D + \alpha E_w))}{\text{fator de normalização}} = \frac{1}{Z_F(\alpha, \beta)} \exp(-F(\mathbf{w})) \quad (50)$$

Nesta estrutura Bayesiana, os pesos ótimos deverão maximizar a probabilidade posterior $P(\mathbf{w} | D, \alpha, \beta, M)$. Maximizar a probabilidade posterior é equivalente a minimizar a função objetiva regularizada $F = \alpha E_D + \beta E_w$.

Para otimizar os parâmetros da função objetivo α e β , considerar-se a aplicação da regra de Bayes. Agora, tem-se:

$$P(\alpha, \beta | D, M) = \frac{P(D | \alpha, \beta, M) P(\alpha, \beta | M)}{P(D | M)} \quad (51)$$

Assumindo ser uniforme a densidade anterior $P(\alpha, \beta | D, M)$ para os parâmetros α e β , então a maximização da posterior é realizada pela maximização da função de probabilidade $P(D | \alpha, \beta, M)$. Nota-se que esta função de probabilidade é o fator de normalização da Equação (48). Desde que todas as probabilidades tenham a forma Gaussiana, pode-se conhecer a forma da densidade posterior da Equação (48). Isto é mostrado na Equação (50). Agora, pode-se resolver a Equação (48) para o fator de normalização.

$$\begin{aligned} P(D | \alpha, \beta, M) &= \frac{P(D | \mathbf{w}, \beta, M) P(\mathbf{w} | \alpha, M)}{P(\mathbf{w} | D, \alpha, \beta, M)} = \\ &= \frac{\left[\frac{1}{Z_D(\beta)} \exp(-\beta E_D) \right] \left[\frac{1}{Z_w(\alpha)} \exp(-\alpha E_w) \right]}{\frac{1}{Z_F(\alpha, \beta)} \exp(-F(\mathbf{w}))} = \\ &= \frac{Z_F(\alpha, \beta)}{Z_D(\beta) Z_w(\alpha)} \cdot \frac{\exp(-\beta E_D - \alpha E_w)}{\exp(-F(\mathbf{w}))} = \frac{Z_F(\alpha, \beta)}{Z_D(\beta) Z_w(\alpha)} \end{aligned} \quad (52)$$

As constantes $Z_D(\beta)$ e $Z_w(\alpha)$ são conhecidas da Equação 49, e $Z_F(\alpha, \beta)$ pode ser estimada por expansão com série de Taylor. Desde que a função objetivo tenha a forma quadrática em uma pequena área ao redor do ponto mínimo, pode-se expandir $F(\mathbf{w})$ em volta do ponto mínimo da densidade posterior \mathbf{w}^{MP} , para o gradiente zero. Resolvendo para a constante de normalização obtém-se:

$$Z_F = (2\pi)^{N/2} \left(\det \left(\left(\mathbf{H}^{MP} \right)^{-1} \right) \right)^{1/2} \exp \left(-F(\mathbf{w}^{MP}) \right) \quad (53)$$

Em que $H = \beta \nabla^2 E_D + \alpha \nabla^2 E_w$ é a matriz Hessiana da função objetivo. Colocando este resultado na Equação 52, pode-se resolvê-la para os valores ótimos de α e β no ponto de mínimo, ou seja, deriva-se em relação a cada logaritmo de Equação 52 e iguala-se a zero. Obtém-se, então:

$$\alpha^{MP} = \frac{\gamma}{2E_w(\mathbf{w}^{MP})} \quad \text{e} \quad \beta^{MP} = \frac{n - \gamma}{2E_D(\mathbf{w}^{MP})} \quad (54)$$

Sendo $\gamma = N - 2\alpha^{MP} \text{tr}(\mathbf{H}^{MP})$ o número efetivo de parâmetros e N o número total de parâmetros. O parâmetro γ é a medida de quantos parâmetros da rede são efetivamente usados na redução da função erro. Este parâmetro pode variar de 0 até N .

Hagan e Foresse (1997) propuseram aplicar a aproximação Gauss-Newton à matriz Hessiana, que pode ser convenientemente implementada se o algoritmo de otimização Levenberg-Marquardt for usado para a localização do ponto mínimo. Este minimiza a computação adicional requerida para a regularização.

2.3.6 Redes com funções de ativação base radial (Radial Basis Function – RBF)

De acordo com Garg *et al.* (2007) apud Araujo *et al.* (2010), a RBF apresenta na camada de entrada (camada de base radial) neurônios de base radial e apresenta na camada de saída (camada linear simples) neurônios com função de ativação linear. Então, as funções de base radial são funções não-lineares que podem ser utilizadas como funções-base em qualquer tipo de modelo de regressão não-linear (linear ou não-linear nos parâmetros) e, particularmente, como função de ativação de qualquer tipo de rede multicamada.

As redes com funções de ativação de base radial (RBF) são redes neurais cuja função de ativação é a distância entre os vetores de entrada e de peso (funções radiais), diferentemente das redes MLP, nas quais se tem como argumento o produto escalar entre o vetor de entrada e o vetor de pesos. Na Figura 8 é possível observar dois exemplos de rede neural de base radial

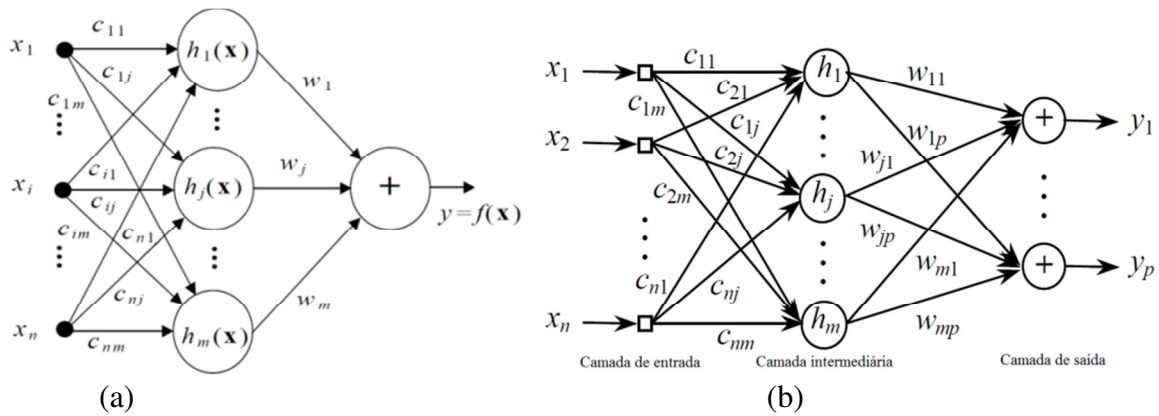


Figura 8 - Rede neural de base radial (a) e rede neural de base radial com múltiplas saídas (b)
 Fonte: BROOMHEAD; LOWE (1988)

As funções radiais representam, então, uma classe especial de funções cujo valor diminui ou aumenta em relação à distância de um ponto central. A função radial mais utilizada é a gaussiana, expressa pela Equação 55.

$$f(u) = \exp\left(-\frac{v^2}{2\sigma^2}\right) \quad (55)$$

Sendo $v = x_n - c_{jn}$ para a distância em relação ao centro (c_{jn})

Semelhante às redes MLP, as redes RBF contêm no mínimo 3 camadas: uma camada de entrada, uma camada escondida e uma camada de saída e é bastante difundida a utilização de apenas uma camada escondida. Cada camada desempenha um papel específico na rede, Braga *et al.* (2007):

- Camada de entrada – distribui as variáveis de entrada para a camada escondida;
- Camada escondida – agrupa os dados de entrada em *clusters*, transformando um conjunto de padrões de entrada não-linearmente separáveis em um conjunto de saída linearmente separáveis. Os neurônios dessa camada têm apenas funções de base radial como função de ativação, ao invés de funções sigmoidais ou outras.
- Camada de saída – procura classificar os padrões recebidos da camada escondida.

Apesar de serem aproximadores universais de funções, as redes MLP e RBF apresentam algumas diferenças, destacadas na Tabela 5.

Tabela 5 - Comparação de uma rede RBF típica com uma rede MLP convencional

MLP	RBF
Pode ter uma ou mais camadas intermediárias	Geralmente possui apenas uma camada intermediária
Neurônios das camadas intermediárias e de saída têm funções semelhantes	Neurônios das camadas intermediárias têm funções diferentes dos neurônios da camada de saída
Entrada da função de ativação é o produto interno dos vetores de entrada e de pesos	Entrada da função de ativação é a distância euclidiana entre os vetores de entrada e de pesos
Separa padrões de entrada com hiperplanos	Separa padrões de entrada com hiperelipsóides
Melhor em problemas complexos	Melhor em problemas bem-definidos
Constrói aproximadores globais para mapeamento entrada-saída	Constrói aproximadores locais para mapeamento entrada-saída

Fonte: Carvalho *et al.* (2010)

É importante ressaltar que, na presente tese, foram aplicadas os dois tipos de redes, a MLP e a RBF para avaliar a qualidade de água em relação aos parâmetros ambientais.

2.3.7 Redes neurais e qualidade de água

A degradação dos recursos hídricos tem aumentado a necessidade de desenvolvimentos de projetos relacionados à qualidade da água. A predição do comportamento de corpos d'água, através da medida de parâmetros ambientais, funciona como uma importante ferramenta no combate a problemas ambientais, como processo de eutrofização de reservatórios. Entretanto, esta é uma tarefa difícil devido à complexidade dos processos físico-químicos e biológicos causadores desses problemas (KUO *et al.*, 2007).

O processo de eutrofização resulta no excesso de algas e macrófitas nos reservatórios, provocando a deterioração da qualidade da água para uso humano e decréscimo nos níveis do oxigênio dissolvido com efeitos adversos para os peixes. Dentre os parâmetros, a quantidade de clorofila é o indicador preliminar da eutrofização. Em alguns casos, a clorofila-a é linearmente dependente de algumas variáveis como profundidade e quantidade de sólidos suspensos (KARUL *et al.*, 2000).

Os modelos clássicos fazem aproximações de vários processos envolvendo características da água, de modo a representá-los por equações lineares. O comportamento não linear do processo de eutrofização tem sido modelado, então, usando redes neurais e sistemas fuzzy, com sucesso considerável (KARUL *et al.*, 2000; LU; LO, 2002).

Dentre os trabalhos publicados que mostram a utilização de redes neurais artificiais para estimativa da qualidade da água, merecem destaque os de Lu e Lo (2002) que retrata o diagnóstico de reservatório de água utilizando a lógica fuzzy para representar o processo de eutrofização em termos de parâmetros como fósforo total e clorofila-a; o de Strobl *et al.* (2007) que utilizaram diferentes redes neurais para classificar o grau de eutrofização de um lago; o de Lo e Wu (2008) que utilizaram uma RNA do tipo MLP para determinação em tempo real da dosagem de coagulante a ser utilizada no tratamento da água; e o de Vilas *et al.* (2011) que estudaram a qualidade da água em rios da Espanha utilizando a combinação entre os dados de espectrometria (MERIS) e aplicação de redes neurais para determinação da clorofila-a, ressaltando a importância desse parâmetro como um bioindicador da qualidade da água.

Predizer o comportamento de corpos d'água, através da medida de parâmetros físico-químicos consiste em um poderoso instrumento de combate e controle da eutrofização. Dentre os parâmetros de qualidade da água, a concentração de clorofila-a é o indicador preliminar da eutrofização. Em alguns casos, a clorofila-a é linearmente dependente de variáveis como profundidade do sistema e quantidade de sólidos suspensos embora, geralmente, não exista relação direta entre qualquer variável medida, clorofila-a e a maioria das espécies de algas encontradas em ambientes eutrofizados. (CARVALHO et al, 2007).

2.4 LÓGICA FUZZY

A lógica como ciência teve origem com o filósofo Aristóteles na Grécia Antiga. Este filósofo atribuiu um conjunto de regras que classificavam as sentenças como verdadeiras ou falsas.

Em 1847, o matemático George Boole, com a publicação do Livro *The Mathematical Analysis of Logic* fez nascer a lógica booleana. Essa lógica classificou as sentenças atribuindo valores numéricos de 1 para sentenças verdadeiras e 0 para sentenças falsas. No entanto, esse conceito dificilmente representava as respostas aos processos reais. Assumir somente duas possibilidades deixa de lado fatores que influenciam na resposta real do processo.

Em 1960, Zadeh publicou *Fuzzy sets, information and control*, que mostrava que os recursos tecnológicos eram incapazes de automatizar processos complexos utilizando a

bivalência booleana. Isto provocou o surgimento da lógica *fuzzy* ou lógica difusa (BARBALHO, 2001).

A lógica fuzzy introduziu os conceitos de parcialmente falso ou parcialmente verdadeiro, considerando um grau de incerteza que varia entre 1 (verdadeiro) e 0 (falso).

Segundo Vieira *et al.* (1999), os sistemas difusos (lógica *fuzzy*) foram responsáveis, nas ultimas três décadas, por mudanças significativas nas aplicações industriais da Inteligência Artificial, especialmente os sistemas de controle, tendo sua aplicação na avaliação de riscos e impactos econômicos e ambientais, automação e controle de processos, e em sistemas de apoio de decisão.

De acordo com Barbalho (2001), a lógica *fuzzy* é uma técnica de inteligência artificial baseada no conhecimento heurístico, geralmente representado por um conjunto de expressões condicionantes para representar os fenômenos.

Angulo *et al.* (2012) aplicaram técnicas matemáticas avançadas para auxiliar no gerenciamento e monitoramento da qualidade da água. Nesse trabalho foi desenvolvido uma ferramenta *software* para apoio à decisão, com base na aplicação de técnicas de lógica fuzzy, que podem detectar episódios de qualidade a partir do comportamento das variáveis contínuas medidas no monitoramento.

A Figura 9 mostra o processo de desenvolvimento para elaboração das regras da lógica fuzzy, esquematizado por Angulo *et al.* (2012). O sistema desenvolvido nesse trabalho foi capaz de gerar, em tempo real, um conjunto de indicadores para definir a qualidade da água com base em: dados da rede e conhecimento especializado (empírico) sobre a qualidade da água, expresso através de regras. O objetivo, então, foi de traçar um perfil da evolução da qualidade da água de uma bacia hidrográfica, permitindo a detecção de dois eventos (fenômenos) ambientais: pontos de descargas urbanas, ou causadas por uma estação de tratamento de águas residuais, e condições ambientais (eutrofização).



Figura 9 - Desenvolvimento das regras da lógica fuzzy.
Fonte: Adaptada de Angulo *et al.* (2012)

A lógica nebulosa foi desenvolvida a partir da teoria de conjuntos nebulosos, para tratar do aspecto vago da informação (SANDRI; CORREA, 1999; ALTUNKAYNAK *et al.*, 2005; ICAGA, 2007; PEREIRA *et al.*, 2009), ou seja, com o objetivo de fornecer uma ferramenta matemática para o tratamento de informações de caráter impreciso ou vago e auxiliar no controle e na tomada de decisão. Esta teoria generaliza o conceito clássico dos conjuntos, que pode ser visto como um conjunto nebuloso específico, denominado geralmente de *crisp*, permitindo que objetos possuam graus de pertinência a determinados conjuntos, o que possibilita a representação de conceitos imprecisos, sem perder a precisão matemática no tratamento.

A lógica fuzzy possibilita classificar em números reais uma determinada realidade ou situação que está atrelada a variáveis linguísticas, incertas ou vagas. Sendo assim é possível a modelagem realista de sistemas tendo como objeto de análise, variáveis qualitativas, quantidades imprecisas, conceitos vagos e mal definidos (FONSECA, 2003).

A estrutura da lógica fuzzy foi inicialmente construída a partir dos conceitos já estabelecidos da lógica clássica ou tradicional, a qual foi fundamentada na teoria dos conjuntos e suas regras, como união, intersecção, além de regras expressas através de implicações lógicas. Com a evolução da informática, e associado ao desenvolvimento de técnicas relacionadas aos sistemas especialistas, novos operadores foram sendo definidos, permitindo que problemas de maior complexidade pudessem ser modelados sob a ótica da lógica fuzzy.

De forma mais objetiva e preliminar, define-se Lógica Fuzzy como sendo uma ferramenta capaz de capturar informações vagas, em geral descritas em uma linguagem natural e convertê-las para um formato numérico, de fácil manipulação pelos computadores de hoje em dia.

A representação destas informações vagas é desenvolvida através do uso de conjuntos nebulosos. Devido a esta propriedade e a capacidade de realizar inferências, a Lógica Fuzzy tem encontrado grandes aplicações nas seguintes áreas: Sistemas Especialistas; Computação com Palavras; Raciocínio Aproximado; Linguagem Natural; Controle de Processos; Robótica; Modelamento de Sistemas Parcialmente Abertos; Reconhecimento de Padrões; Processos de Tomada de Decisão (*decision making*).

Os conjuntos Fuzzy constituem uma ligação entre o raciocínio humano ao da lógica executada por uma máquina. Tradicionalmente, tem-se, em um conjunto convencional, limites bruscos (por exemplo, o conjunto dos números maiores que 2), na qual a transição dos membros (elementos que pertencem ao conjunto) para não – membros (elementos que não pertencem ao conjunto) é abrupta e finita. O grau de associação do conjunto é especificado, por exemplo, por um número **1** para os membros e **0** para os não - membros. Por exemplo um controlador de temperatura pode considerar 40°C como quente mas 39,9°C como frio, não fazendo uso de uma percepção de quanto afasta ou se aproxima do calor de referência. Esta percepção seria considerar aspectos de pertinência aproximada utilizando um linguístico adequado (GOMIDE; ROCHA, 1992).

Então, um conceito relacionado com conjuntos nebulosos é o de variável linguística. Entende-se por variável um identificador que pode assumir um dentre vários valores. Deste modo, uma variável linguística pode assumir um valor linguístico dentre vários outros em um conjunto de termos linguísticos. Formalmente, uma variável linguística é caracterizada pela quintupla $\{X, T(X), U, G, M\}$, sendo X o nome do conjunto de termos, U o universo de discurso, G uma gramática para gerar os termos $T(X)$ e M o significado dos termos linguísticos, representado através de conjuntos nebulosos. Variáveis linguísticas podem também conter modificadores (também linguísticos) que alteram seu valor. Exemplos de modificadores válidos são: muito, pouco, não muito, mais ou menos (OLIVERIA JUNIOR *et al.*, 2007).

Existem também conectivos que podem ser aplicados a estas variáveis, *e* e *ou*. Assim, um valor válido para a variável linguística altura seria não muito alto e não muito baixo. Os modificadores linguísticos podem ser definidos matematicamente, como no exemplo dos conjuntos, baixo e muito baixo, no qual o modificador muito é caracterizado por elevar cada ponto da função de pertinência à segunda potência. Os conectivos E (and) e OU (or) são equivalentes às operações de união e intersecção de conjuntos, respectivamente, podendo dar origem a conjuntos complexamente definidos, porém representados linguisticamente de maneira simples (CAMPOS FILHO, 2004).

A Lógica Fuzzy pode ser utilizada para a implementação de controladores nebulosos, aplicados nos mais variados tipos de processos. A utilização de regras nebulosas e variáveis linguísticas confere ao sistema de controle algumas vantagens, incluindo:

- a. Simplificação do modelo do processo;

- b. Melhor tratamento das imprecisões inerentes aos sensores utilizados;
- c. Facilidade na especificação das regras de controle, em linguagem próxima à natural;
- d. Satisfação de múltiplos objetivos de controle;
- e. Facilidade de incorporação do conhecimento de especialistas humanos.

Entretanto, visto que tanto as leituras de sensores quanto os sinais esperados pelos atuadores do sistema de controle, não são nebulosos, são necessários elementos adicionais entre o controlador nebuloso e o processo a ser controlado. Estes elementos são denominados fuzzificador e defuzzificador, e estão posicionados na entrada e saída do sistema de controle, respectivamente. Estes elementos são responsáveis por transformar as medidas obtidas dos sensores, em conjuntos nebulosos (fuzzificador), e em transformar os conjuntos nebulosos obtidos na saída do controlador, em valores não nebulosos de controle para o processo (defuzzificador) (CAMPOS; SAITO, 2004).

2.4.1 Características da lógica fuzzy

De forma geral, a lógica *fuzzy* apresenta as seguintes características (CAMPOS; SAITO, 2004; OLIVEIRA JUNIOR *et al.*, 2007):

- a. A Lógica Fuzzy está baseada em palavras e não em números, ou seja, os valores verdades são expressos linguisticamente. Por exemplo: quente, muito frio, verdade, longe, perto, rápido, vagaroso, médio, etc.
- b. Possui vários modificadores de predicado como por exemplo: muito, mais ou menos, pouco, bastante, médio, etc.
- c. Possui também um amplo conjunto de quantificadores, como por exemplo : poucos, vários, em torno de, usualmente.
- d. Faz uso das probabilidades linguísticas, como por exemplo: provável, improvável, que são interpretados como números fuzzy e manipulados pela sua aritmética.
- e. Manuseia todos os valores entre 0 e 1, tomando estes, como um limite apenas.

Com base em Oliveira Junior *et al.* (2007) e Campos e Kaito (2004), é possível citar as seguintes vantagens e perspectivas da lógica fuzzy:

- a. Requer poucas regras, valores e decisões;
- b. Mais variáveis observáveis podem ser valoradas;

- c. O uso de variáveis linguísticas nos deixa mais perto do pensamento humano;
- d. Simplifica a solução de problemas;
- e. Proporciona um rápido protótipo dos sistemas;
- f. Simplifica a aquisição da base do conhecimento.

A seguir serão apresentados os conceitos básicos para utilização da lógica fuzzy em geral e no controle de processos.

2.4.2 Conjunto fuzzy

A lógica fuzzy é baseada na teoria dos conjuntos fuzzy e esta é, em grande parte, uma extensão da teoria dos conjuntos tradicionais (OLIVEIRA JUNIOR *et al.*, 2007).

A Figura 10 exemplifica os dois tipos de conjuntos, tradicional e fuzzy, considerando que na teoria de conjuntos clássica, um elemento pertence ou não a um dado conjunto e a teoria de conjuntos fuzzy existe um grau de pertinência de cada elemento a um determinado conjunto (faixa de pertinência). Ou seja, conforme a definição formal de um conjunto tradicional tem-se:

$$f_A(x): X \rightarrow 0,1 \quad (56)$$

Sendo que se o elemento pertence ao conjunto A é atribuído o valor 1 e se não pertence é atribuído o valor 0: $f_A(x) = \begin{cases} 1, & \text{se } x \in A \\ 0, & \text{se } x \notin A \end{cases}$

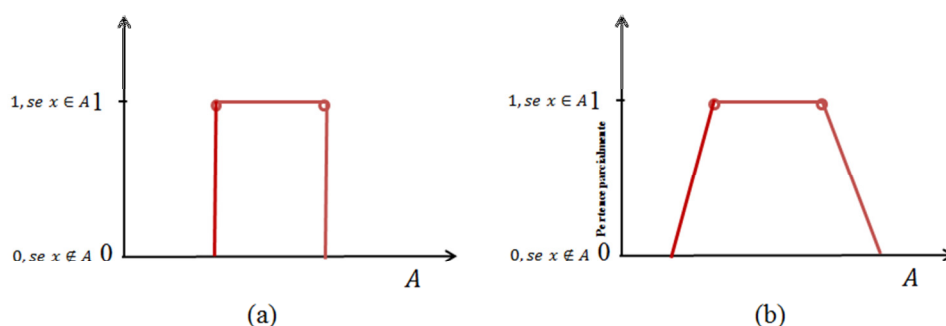


Figura 10 - Função do conjunto tradicional (a) e conjunto fuzzy (b)

A representação do grau de pertinência (o quanto pertence, parcialmente, fortemente, fracamente, etc...) é definida por meio de uma função característica generalizada denominada de função de pertinência $\mu_A(x)$:

$$\mu_A(x): X \rightarrow [0,1] \quad (57)$$

Sendo que:

$\mu_A(x): 1$ se $X \rightarrow [0,1]$ indica que x é completamente compatível com A

$\mu_A(x): 0$ se $X \rightarrow [0,1]$ indica que x é completamente incompatível com A

$0 < \mu_A(x) < 1$ se $X \rightarrow [0,1]$ indica que x é parcialmente compatível com A

Sendo que X é o universo de discurso e A é um subconjunto fuzzy de X . Essa função associa a cada elemento x de X o grau $\mu_A(x)$, com o qual x pertence a A .

A representação de pertinência anterior indica o grau com que um elemento x pertence ao subconjunto A , grau este que pode assumir infinitos valores no intervalo $[0,1]$. A representação formal como um conjunto é:

$$A = \{x(\mu_A(x)) \mid x \in X, \mu_A(x): X \rightarrow [0,1]\} \quad (58)$$

Um conjunto fuzzy definido no universo de discurso A é caracterizado por uma função de pertinência μ_A , a qual mapeia os elementos de X para o intervalo $[0,1]$. Assim, a função de pertinência associa a cada elemento x pertencente a X um número real $\mu_A(x)$ no intervalo $[0,1]$, que representa o grau de pertinência do elemento x ao conjunto A .

2.4.3 Funções de pertinência

Os conjuntos *fuzzy* são definidos através de sua função de pertinência que mostra a intensidade com que o objeto pertence ao conjunto fuzzy. Existem várias formas de representar uma função *fuzzy* de pertinência, sendo que, as mais usuais são a triangular, gaussiana, trapezoidal, sigmóide bipolar, S e quadrática, sendo todas definidas no intervalo de pertinência de 0 a 1. As funções mostradas a seguir (triangular, gaussiana e trapezoidal) são as mais utilizadas na literatura (CAMPOS; SAITO, 2004; OLIVEIRA JUNIOR *et al.*, 2007).

2.4.3.1 Função de pertinência triangular

A função triangular é caracterizada por uma terna (a, b, c) sendo que a e c determinam o intervalo para o qual a função arrume valores diferentes de zero, e b

representa o ponto no qual a função de pertinência é máxima. Ou seja, os números fuzzy começam a subir a partir de zero $x = a$; atingem um máximo de 1 em $x = b$; e retornam a zero em $x = c$. Em seguida, a função $\mu(x)$ de um número *fuzzy* triangular é representada na equação (59) e na Figura 11.

$$\text{trimf}(x; a, b, c) = \begin{cases} 0, & x \leq a \\ (x - a)/(b - a), & a < x \leq b \\ (c - x)/(c - b), & b < x \leq c \\ 0, & x > c \end{cases} \quad (59)$$

$$\text{Ou: } \text{trimf}(x; a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right)$$

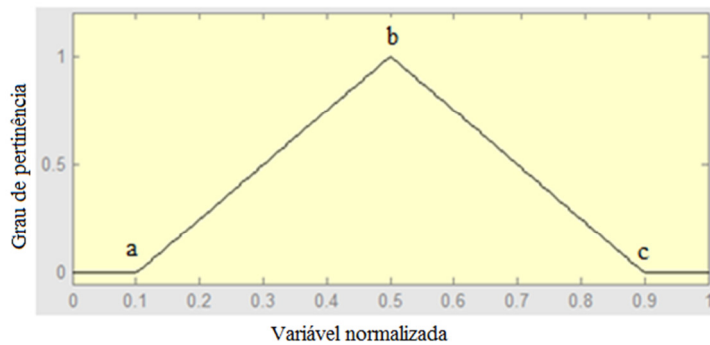


Figura 11 - Representação da função triangular
Fonte: MATLAB (2011)

2.4.3.2 Função de Pertinência Guassiana

A função de pertinência gaussiana é caracterizada pela sua média e pelo desvio padrão. A função $\mu(x)$ de um número *fuzzy* gaussiana é representada na Equação 60 e na Figura 12.

$$\text{gaussmf}(x; a, b, c) = ae^{\frac{-(x-b)^2}{2c^2}} \quad (60)$$

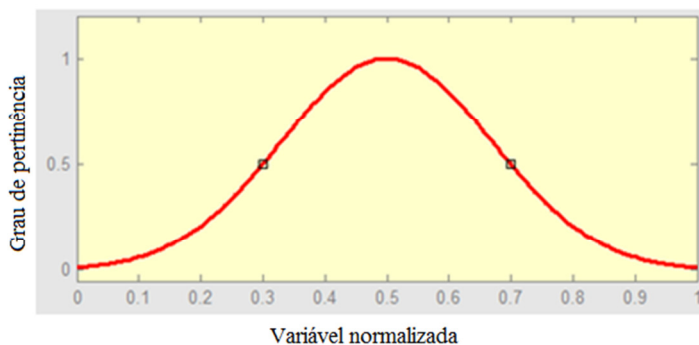


Figura 12 - Representação da função gaussiana
Fonte: MATLAB (2011)

2.4.3.3 Função de Pertinência Trapezoidal

A função de pertinência trapezoidal é caracterizada por um conjunto de valores de a , b , c , e d , sendo que a e d determinam o intervalo dentro do qual a função assume valores diferentes de zero, e b e c determinam o intervalo para o qual a função é máxima e igual a 1. Ou seja, na função trapezoidal os números *fuzzy* começam a subir a partir de zero $x = a$; atinge um máximo de 1 em $x = b$, se mantém em 1 no intervalo $\{b, c\}$ e retorna a zero em $x = d$. A função $\mu(x)$ de um número fuzzy trapezoidal é representada na Equação 61 e na Figura 13.

$$\text{trapmf}(x; a, b, c) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1, & b < x < c \\ \frac{d-x}{d-c}, & c \leq x < d \\ 0, & x \geq d \end{cases} \quad (61)$$

Ou:

$$\text{trapmf}(x; a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right)$$

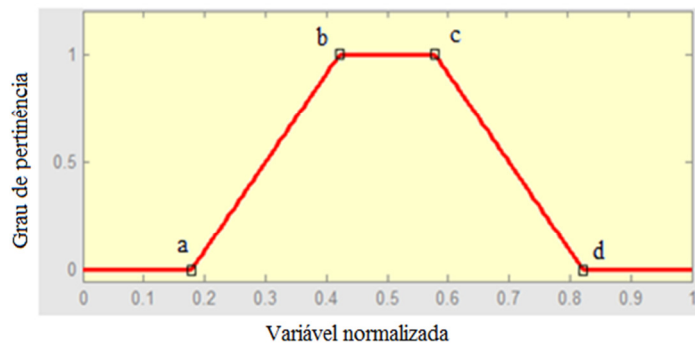


Figura 13 - Representação da função trapezoidal

Fonte: MATLAB (2011)

Dessa forma, as funções das Figuras 11, 12 e 13, refletem o grau de pertinência do elemento x para o conjunto *fuzzy*.

2.4.4 Variáveis Linguísticas no Sistema de Inferência Fuzzy

Na representação *fuzzy* do conhecimento o sistema não pode ser caracterizado de maneira definitiva, do tipo *é* ou *não é*, *pertence* ou *não pertence*. Dessa forma, de acordo

com Campos Filho (2004), uma variável linguística pode ser considerada como sendo o nome dado a um conjunto fuzzy, cabendo variações associadas às expressões linguísticas como é, mais ou menos, não é, ou alto, médio, baixo. Essas variações são definidas com base no conhecimento do sistema por especialistas da área.

A forma de utilização das variáveis linguísticas depende das propriedades do sistema de inferência fuzzy. Essas propriedades podem ser sintáticas e semânticas. Para as propriedades sintáticas as informações linguísticas são armazenadas, gerando a criação de uma base de conhecimento do sistema em análise. Para as propriedades semânticas é especificado o modo como é extraído e processado o conhecimento, contido na estrutura definida pelas propriedades sintáticas e pela interferência das regras estabelecidas, conforme comenta, ainda, Campos Filho (2004).

Por exemplo, o conjunto de temperatura de sistema de processamento de biscoitos, poderia ser representado através das variáveis “baixa”, “média” e “alta”, conforme mostra a Figura 14.

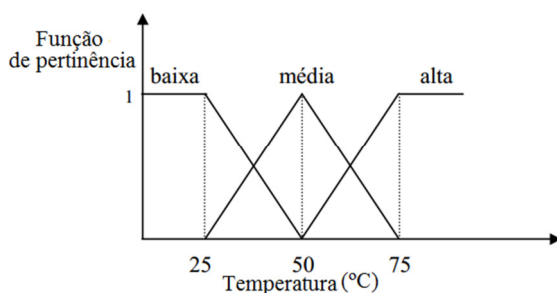


Figura 14 - Função triangular do conjunto de temperatura de processamento de biscoitos

Na construção do modelo fuzzy, para cada variável linguística é atribuído um intervalo numérico contínuo de entrada. Conforme o exemplo mostrado na Figura 14, referente ao conjunto de temperatura, será analisado o grau de pertinência de cada temperatura a cada conjunto fuzzy de temperatura. Neste conjunto, tem-se:

- a. O conjunto temperatura baixa: tem seu grau de pertinência 100% quando o valor de X varia entre 0 e 25. Esse grau de pertinência vai diminuindo até chegar em 0 quando $x = 50$.
- b. O conjunto temperatura média: de 25 até 50 vai aumentando o grau de pertinência. Quando a temperatura é 50°C tem grau de 100%, e começa a cair até chegar em 0, quando X for 75.

c. O conjunto temperatura alta: a partir de 62,5°C vai aumentando o seu grau de pertinência. Quando chega a 75°C, o grau de pertinência atinge 100% e se mantém alta.

No sistema de inferência fuzzy é através das variáveis linguísticas que ocorre o relacionamento entre os conjuntos, para gerar a resposta do modelo. Esse relacionamento, feito mediante as regras de inferência, que serão comentadas nos próximos itens.

2.4.4.1 Regras de Inferência Fuzzy

Define-se inferência como a passagem, através de regras válidas, do antecedente (SE) ao consequente (ENTÃO) de um objeto de estudo. Na lógica *fuzzy*, essa passagem é realizada mediante a interação, determinada pelas regras de inferência, entre as variáveis linguísticas de entrada (SE), gerando um conjunto de dados de saída (ENTÃO). Essas regras são aplicadas aos conjuntos fuzzy através das variáveis linguísticas e são construídas mediante a operação entre os conjuntos (CAMPOS; SAITO, 2004; OLIVEIRA JUNIOR *et al.*, 2007).

No procedimento de inferência é analisado o grau de pertinência, associado àquele mesmo valor numérico no universo de discurso, relacionando-os a uma base de regras conforme a condicional se – então.

O tipo de inferência ocorre:

$$\text{If}(\text{Se}) < \text{antecedente} > \text{then}(\text{Então}) < \text{consequente} >$$

Ou

$$\text{SE} < \text{situação} > \text{ ENTÃO} < \text{ação} >$$

Na lógica clássica a inferência é dada pela comparação, sendo que se permite apenas uma compatibilidade exata, ou seja, é ou não é, pertence ou não pertence. No raciocínio difuso é possível adotar um valor aproximado dependendo da pertinência ou *fuzzificação* dessa variável ao conjunto fuzzy. (WESTPHAL, 2003; SIVANANDAM *et al.*, 2007)

O processo de *defuzzificação* é utilizado quando a inferência é realizada em um sistema do tipo clássico com o objetivo de obter um valor de saída numérico (SAADE, 2000; SIVANANDAM *et al.*, 2007). Em um sistema de inferência fuzzy, como no método de Mamdani, a característica de saída é obtida a partir de valores defuzzificados de produção de conjuntos *fuzzy*. Estes originados da agregação de diferentes resultantes de cada regra

(fornecidas após a fuzzyficação) da base de regras de inferência distribuídas no universo de discurso.

Sendo assim, após a entrada das variáveis numéricas precisas, são ativadas as regras (*fuzzyficação*), em seguida o sistema de inferência determina como as regras (determinadas por especialistas) são combinadas. Como resultado dessa combinação tem-se uma agregação entre as respostas das regras e, após a escolha do tipo de resposta em relação à distribuição dos dados agregados (*defuzzyficação*), tem-se a resposta do modelo no domínio das variáveis de saída num correspondente universo de discurso. Dessa forma, as entradas e saídas do sistema são denominadas respectivamente, *fuzzyficação* e *defuzzyficação* e correspondem às etapas principais de modelos de inferência fuzzy (MALUTTA, 2004). Esses passos estão descritos na Figura 15.

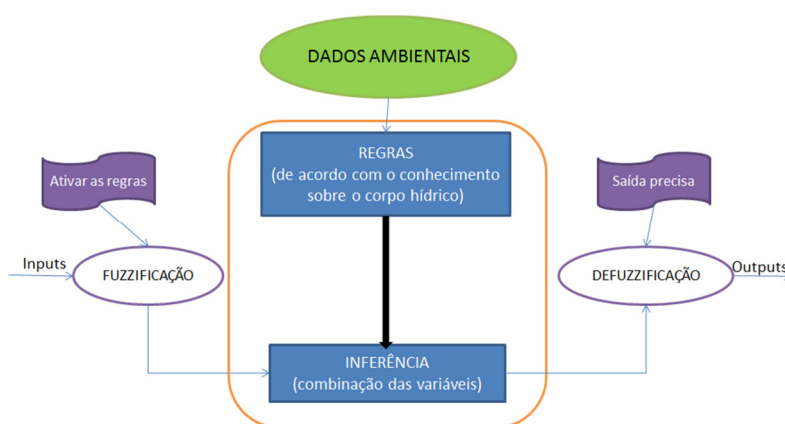


Figura 15 - Sistema de inferência fuzzy

Fonte: Adaptada de Malutta (2004)

Neste sentido, em resumo, são estabelecidas as etapas para metodologia fuzzy:

- Etapa de fuzzyficação: etapa na qual se modela matematicamente a informação das variáveis de entrada do sistema por meio de conjuntos *fuzzy*. Nesta etapa, o papel do especialista ou o conhecimento do sistema é notado, pois, para cada variável de entrada, termos linguísticos devem ser atribuídos para representá-la e cada termo linguísticos deve ser associado ao conjunto *fuzzy* através da função de pertinência definida.
- Etapa de base de regras: é o cerne do conhecimento, ou seja, nesta etapa são inseridas as variáveis e suas classificações linguísticas.
- Etapa de inferência: nesta, define-se os conectivos lógicos para modelar as regras estabelecidas.
- Etapa de defuzzyficação: nesta etapa é realizada transformação da variável de saída *fuzzy* em um valor numérico.

2.4.4.2 Métodos de Inferência Mamdani

Campos Filho (2004) cita que a primeira aplicação da lógica fuzzy em controle de processos foi realizada por Mamdani e Assilam em 1975, com base as teorias propostas por Zadeh em 1965. Neste trabalho, o desenvolvimento do sistema de controle foi voltado para o funcionamento de máquinas a vapor através da modelagem matemática das atividades (ações) dos operadores das máquinas, considerados os especialistas do sistema, tornando as tarefas automatizáveis.

A partir de então, o método de inferência proposto por Mamdani e Assilam tornou-se uma referência padrão para a utilização da lógica fuzzy em processamento de conhecimento em diversas áreas.

Ainda de acordo com Campos Filho (2004), o método de inferência de Mamdani e Assilam possui módulos de interface que transformam as variáveis de entrada baseadas em grandezas numéricas, em conjuntos fuzzy equivalentes e, posteriormente, as variáveis fuzzy geradas em variáveis numéricas proporcionais, adequadas para os sistemas de atuação existentes.

No método de inferência fuzzy de Mamdani, a regra semântica tradicionalmente usada para o processamento de inferência é denominada de Máx-Min, utilizando as operações de união e interseção entre conjuntos por meio de operadores de máximo e mínimo, respectivamente (ALMEIDA; EVSUKOFF, 2003 apud LANDMANN; ERDMANN, 2011). Gomide (2007), através da Figura 16, mostra o método Mamdani de inferência.

Tosun *et al.* (2011) comentam que, comumente, no modelo de Mamdani, tanto o antecedente e consequente são proposições *fuzzy* e os termos linguísticos fuzzy if (se) e then (então) atribuem a seguinte forma geral para as regras *fuzzy*:

$$R_i: \text{Se } x \text{ é } A_i \text{ Então } y \text{ é } B_i, \quad i = 1, 2, \dots, k$$

Sendo R_i é o número de regra *fuzzy*, A_i e B_i são os conjuntos *fuzzy*, x é a variável antecedente que representa a entrada no sistema *fuzzy*, e y é a variável consequente relacionada à saída do sistema *fuzzy*.

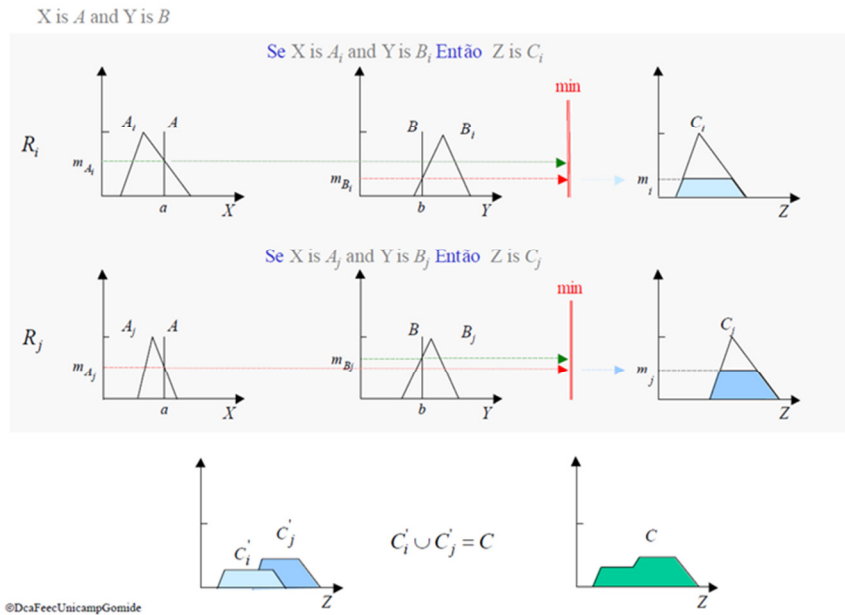


Figura 16 - Inferência Mamdani
Fonte: Gomide (2007)

2.4.4.3 Métodos de Defuzzificação

Na etapa de defuzzificação, a variável difusa produzida pela etapa de inferência é transformada em variável numérica (determinística) que proporcionará a análise da resposta do sistema. Os métodos de defuzzificação mais utilizados são: o primeiro máximo, a média dos máximos e o centro da área ou centróide. Esses métodos estão representados Tabela 5.

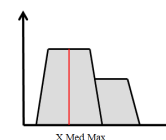
- Método do primeiro máximo o valor de saída corresponde ao ponto em que o grau de pertinência da distribuição atinge o primeiro valor máximo;
- No método da média dos máximos o valor de saída corresponde ao ponto médio entre os valores que tem maior grau de pertinência;
- No método centróide, mais utilizado, o valor de saída corresponde ao centro da gravidade da função de distribuição.

Tabela 6 - Comparativo dos métodos de defuzzificação.

Método	Fórmula	Gráfico
Primeiro máximo	$X_{Max} = x_{max_1} + x_{max_2}$	

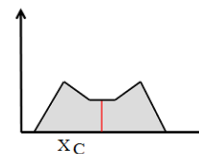
Média dos máximos

$$X_{MedMax} = \frac{(x_{max_1} + x_{max_2})}{2}$$



Centróide

$$X_C = \frac{\sum x_i w_i}{\sum w_i}$$



No que concerne a avaliação e monitoramento de sistemas hídricos, vários autores utilizaram a lógica *fuzzy* para classificar ou avaliar os sistemas quanto a sua qualidade para determinado uso do recurso hídrico Altunkaynak *et al.* (2005), Icaga (2007), Pereira *et al.* (2009), Lermontov *et al.* (2009), Peche e Rodríguez (2012) e Angulo *et al.* (2012). Dentre estes, no trabalho de Angulo *et al.* (2012) merece destaque a aplicação de técnicas matemáticas avançadas para auxiliar no gerenciamento e monitoramento da qualidade da água, com o objetivo de detectar episódios (fenômenos) de qualidade, a partir do comportamento das variáveis ambientalmente representativas desses fenômenos.

2.4.5 Sistema Neuro-Fuzzy

A modelagem utilizando os sistemas *neuro-fuzzy* é, atualmente, uma das técnicas mais modernas e poderosas em engenharia, principalmente quando aplicada a sistemas fortemente não-lineares e de comportamento transiente.

De acordo com Oliveira Júnior *et al.* (2007), os sistemas *neuro-fuzzy* consistem na representação do sistema *fuzzy* na forma de redes passíveis de treinamento, por técnicas semelhantes às usadas em redes neurais. O processo de treinamento na verdade é o ajuste de parâmetros, com o objetivo de minimizar a função erro entre as saídas desejadas e as apresentadas pela rede. Os sistemas neurofuzzy têm como objetivo, então, conjugar a capacidade de aprendizagem das redes neurais à interpretação característica dos sistemas *fuzzy*.

Uma das redes *neuro-fuzzy* mais conhecidas e aplicadas é a estrutura ANFIS (*Adaptive Neuro-Fuzzy Inference System*) proposta em Jang (1993). Esta estrutura é composta por 5 camadas e exemplificada nas Figuras 17 e 18:

a. Camada 1: esta camada recebe as variáveis de entrada e as “fuzzyfica”, ou seja, associa cada entrada x com um termo linguístico A_i através da seguinte equação:

$$N_i^1 = \mu_{A_i}(x) \quad (62)$$

Sendo que x é o valor da entrada no nó i e A_i é o valor linguístico associado a esta função; N_i^1 é a função de pertinência de A_1 e especifica o grau com que o valor da entrada satisfaz o valor linguístico A_i . Para representar esta função é usada, comumente, a função gaussiana.

b. Camada 2: cada neurônio desta camada representa uma regra de inferência *fuzzy* que, ao receber as variáveis de entrada “fuzzyficadas”, define o grau de aplicabilidade da regra, ou a força da regra, executando uma função fixa:

$$w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(x) \quad (63)$$

Sendo w_i a saída da unidade i , que representa o grau de aplicabilidade de uma regra.

c. Camada 3: esta camada é definida pela normalização do grau de pertinência de cada regra, considerando-se como um pré-processamento da defuzzyficação; nesta, para cada unidade i , calcula-se a razão entre a aplicabilidade da i -ésima regra e a soma de todas as aplicabilidades do sistema de regra através da equação:

$$\bar{w}_i = \frac{w_i}{w_1 + w_2} \quad (64)$$

Sendo \bar{w}_i são as saídas dessa camada, denominadas de aplicabilidades normalizadas.

d. Camada 4: nesta camada adaptativa efetua-se o produto entre o grau de aplicabilidade da regra normalizado e os valores processados em cada regra, correspondentes aos *singletons* ou a combinação linear das variáveis de entrada, conforme a Equação 65.

$$N_i^4 = \bar{w}_i \cdot f_i = \bar{w}_i \cdot (p_i x + q_i y + r_i) \quad (65)$$

Sendo \bar{w}_i a saída da terceira camada e $\{p_i, q_i, r_i\}$ o conjunto de parâmetros das consequências.

e. Camada 5: nesta camada calcula-se a variável de saída de cada multiplexador, ou seja, realiza-se o processo de defuzzyficação através da soma dos produtos obtidos na Camada 4:

$$N_i^5 = \sum_i \bar{w}_i \cdot f_i = \frac{\sum_i \bar{w}_i \cdot f_i}{\sum_i \bar{w}_i} \quad (66)$$

As Figuras 17 e 18 apresentam a relação entre as equações 62 a 66.

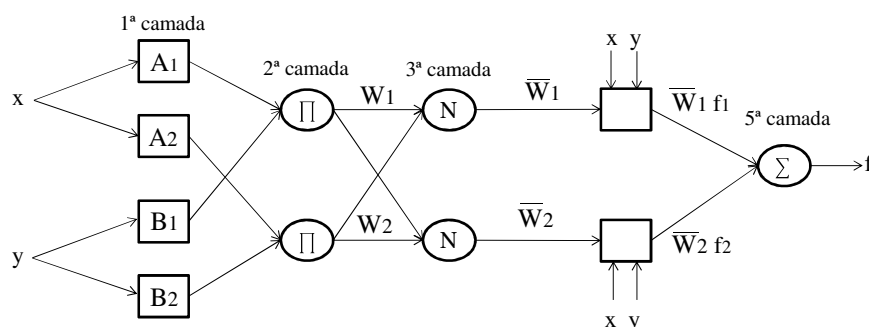


Figura 17 - Arquitetura básica ANFIS

Fonte: Jang (1993)

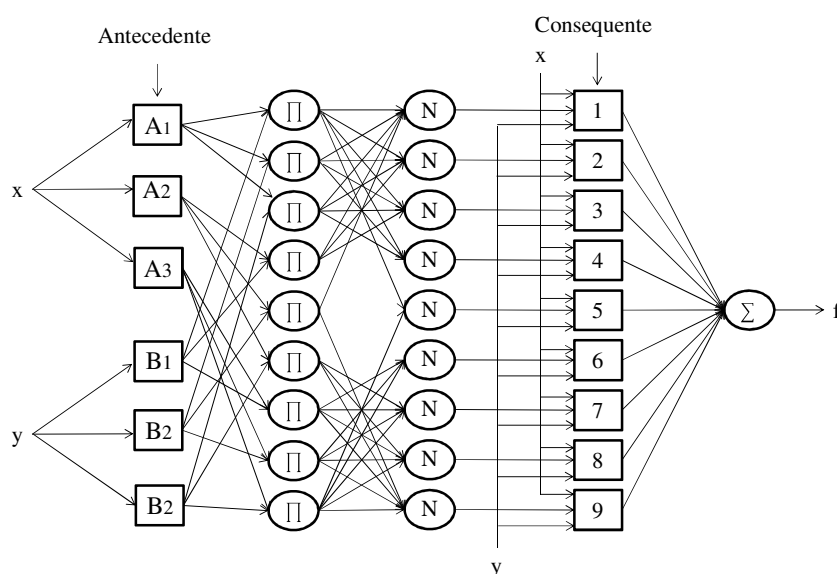


Figura 18 - ANFIS com 9 regras de inferência.

Fonte: Jang (1993)

Para diversas áreas, alguns trabalhos publicados utilizam o sistema de inferência adaptativa neuro-fuzzy (ANFIS), merecendo destaque, por exemplo:

- Taylan e Darrab (2011) apresentam uma abordagem sistemática para um sistema de inferência adaptativa neuro-fuzzy de controle de produção de tapetes em relação a quantidade de látex presente para minimizar as variações indesejáveis na qualidade do produto. Neste trabalho, o sistema ANFIS apresentou resultados melhores do que as outras ferramentas de controle estatístico do processo, fornecendo uma distribuição mais eficiente e uniforme do peso do látex.
- Yetilmezsoy *et al.* (2011) realizaram um estudo para prever a estabilidade da mistura (emulsões) água e óleo utilizando o sistema de inferência adaptativa neuro-fuzzy. Neste trabalho foram utilizados fatores básicos, como viscosidade, densidade e fração de SARA (saturados, aromáticos, resinas e asfaltenos) para desenvolver um índice de estabilidade da

mistura. Os resultados deste estudo mostraram que a modelagem neuro-fuzzy pode ser usada com sucesso para prever a estabilidade dessa mistura água e óleo.

c. Sucena *et al.* (2012) propõem um modelo para um Indicador de Sustentabilidade Sistêmica (ISS) de apoio ao Sistema de Gestão Ambiental (SGA). Para composição desse índice, foram utilizados os conceitos conjugados das teorias fuzzy e de redes neurais associados aos indicadores ambientais, propostos no SGA, do sistema MagLev-Cobra, para gestão de transporte urbano. Os autores deste trabalho ressaltaram que o sistema neuro-fuzzy permitiu desenvolver, de forma simplificada e mais barata, bancos de dados para armazenar as variáveis de entrada e os resultados, bem como as telas para sua gestão, quando comparado com outros sistemas da área (plataformas de *Business Intelligence* (BI) e softwares especialistas).

No entanto, dentre os poucos artigos publicados sobre análise da qualidade da água aplicando o sistema neuro-fuzzy, merece destaque o trabalho de Yan *et al.* (2010). Neste trabalho, os autores desenvolveram um modelo baseado na ANFIS para avaliar a aplicação de um sistema neuro-fuzzy para classificar a qualidade da água e comparar com a performance obtida por um modelo de redes neurais. Neste trabalho, os parâmetros de qualidade da água selecionados foram a concentração de oxigênio dissolvido, a demanda química de oxigênio e a concentração de amônia no sistema. Para estes dados, o modelo obteve 89,59% de precisão quanto ao nível de qualidade do rio estudado, resultado este satisfatório. Além disso, o modelo apresentou melhor desempenho que a redes neurais.

2.5 ANÁLISE DE COMPONENTES PRINCIPAIS

A Análise de Componentes Principais (*Principal Component Analysis* - PCA) é uma técnica estatística multivariada, em que novas variáveis (componentes principais) são formadas através da combinação linear das variáveis originais, mantendo a máxima variância possível e sujeitas à duas condições: a variância total deve se manter a mesma e as componentes devem ser não-correlacionadas (MIDDLETON, 2000).

Parinet *et al.* (2004), Ouyang (2005), Çamdevýren *et al.* (2005), Karydis *et al.* (2007 e 2009), Zimmermann *et al.* (2008), Primpas *et al.* (2010) e Kitsiou e Karydis (2011) apresentam a utilização da estatística multivariada, a exemplo do PCA, na classificação do nível de eutrofização de corpos d'água.

As primeiras componentes mantêm a maior parte da variância do sistema original (Figura 19), que contém m variáveis, de modo que é possível compactar grande parte da

informação linear desse sistema em apenas k novas variáveis, sendo $k < m$. Assim, a i -ésima componente principal de um conjunto de m variáveis é definida segundo a Equação 62, sendo Z_i é a componente principal, os a_{ij} são os *loadings* e X_j são as variáveis originais.

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{im}X_m \quad (62)$$

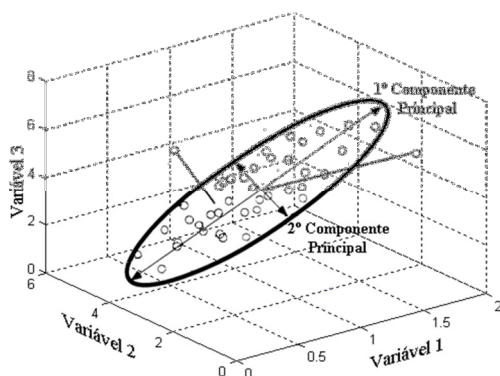


Figura 19 - Representação geométrica das componentes principais
Fonte: Carvalho *et al.* (2010)

De acordo com Çamdevýren *et al.* (2005), o PCA oferece um método objetivo para lidar com um grande conjunto de dados bióticos e abióticos e ajuda a reduzir a complexidade de sistemas multidimensionais através da maximização da variância dos *loadings* das componentes e da eliminação de componentes pouco representativas.

Um índice tem como objetivo sintetizar em uma única variável a informação de todas as variáveis que foram medidas sobre o fenômeno, sendo que seus valores podem ser analisados por métodos de estatística univariada. As técnicas como análise de componentes principais, análise fatorial e análise de correlações canônicas são úteis na construção desses índices (MINGOTI, 2005).

Com o objetivo de quantificar as concentrações de nutrientes e o estado do fitoplâncton em águas marinhas, diversos índices foram propostos e buscando uma nova abordagem nesse campo. Karydis *et al.* (2007) propuseram um índice multivariado baseado em PCA para classificar a água do mar em oligotrófica, mesotrófica e eutrófica utilizando variáveis já conhecidas como sendo de grande influência no fenômeno da eutrofização: concentrações de PO_4 , NO_3 , NO_2 , NH_3 e clorofila-a.

Em outro trabalho, Karydis *et al.* (2009) aplicaram com sucesso a metodologia para classificação do nível trófico do mar Egeu e o compara com a classificação das Diretrizes da União Européia para Gestão das Águas (*European Water Framework Directive*),

demonstrando que o índice poderia ser utilizado para classificar a qualidade da água em alta, boa, moderada, pobre e ruim.

O índice de trofia utilizado nesse trabalho foi proposto por Karydis *et al.* (2007 e 2009). Os autores desse trabalho justificam o uso dessa técnica estatística na construção do índice pelo fato do PCA ser capaz de realizar uma redução de dimensionalidade de modo a sintetizar em algumas poucas componentes, grande parte da informação contida no sistema original.

A construção do índice foi feita a partir da utilização da primeira componente principal utilizando a matriz de correlação das variáveis originais para evitar que a discrepância das unidades de medidas influencie na determinação da direção dos novos eixos ortogonais. O índice, conforme Karydis *et al.* (2007 e 2009) é dado pela Equação 63.

$$IE = a[PO_4] + b[NO_3] + c[NO_2] + d[NH_3] + e[Cl-a] \quad (63)$$

Sendo que os coeficientes *a*, *b*, *c*, *d* e *e* são os *loadings* da primeira componente principal. Esse índice utiliza somente as concentrações de PO_4 , NO_3 , NO_2 , NH_3 e clorofila-a como variáveis para sua construção, visando a classificação de águas. O índice desenvolvido nesse trabalho, por outro lado, abre a possibilidade para que outras variáveis, como pH, oxigênio dissolvido, sólidos solúveis e outros, possam ser utilizadas pois essas variáveis físico-químicas estão em maior ou menor grau ligadas às causas ou efeitos da eutrofização (PARINET *et al.*, 2004).

A definição dos limites para a classificação no nível de trofia foi feita utilizando-se análise de distribuição de frequência (KARYDIS *et al.*, 2009), na qual os subconjuntos de dados, com exemplos para um corpo eutrófico, mesotrófico e oligotrófico, são normalizados por transformação logarítmica e os limites do índice são calculados usando $\mu - \sigma$ e $\mu + \sigma$ em que μ é a média e σ é o desvio padrão dos valores do índice antes da transformação.

3 METODOLOGIA

A presente tese constitui um estudo de caso de base empírica, pelo fato de ser uma abordagem metodológica com ênfase nas características do fenômeno (processo) e tem como base as características associadas aos processos, para a obtenção de dados e da melhor estratégia para análise dos mesmos (MARKONI; LAKATOS, 2010).

O fenômeno, neste trabalho, foi a avaliação da qualidade da água em corpos hídricos através da utilização de redes neurais, lógica fuzzy e modelo híbrido neurofuzzy. Avaliou-se, também, a qualidade da água através da construção de um índice de qualidade da com base na análise de componentes principais.

Neste processo, as análises das amostras de água foram realizadas no Laboratório de Química Analítica Ambiental (LQA), bem como foram utilizados os dados ambientais disponibilizados pela Secretária de Recursos Hídricos de Sergipe (SRHSE). Os dados estão disponibilizados no ANEXO A. É importante ressaltar que a metodologia adotada para caracterização físico-química dos reservatórios foi a análise dos parâmetros de acordo com a Standard Methods for the Examination of Water and Wastewater, American Public Health Association (APHA, 1998).

O fluxograma apresentado na Figura 20 mostra a estratégia empírica e numérica aplicadas para desenvolvimento do trabalho.

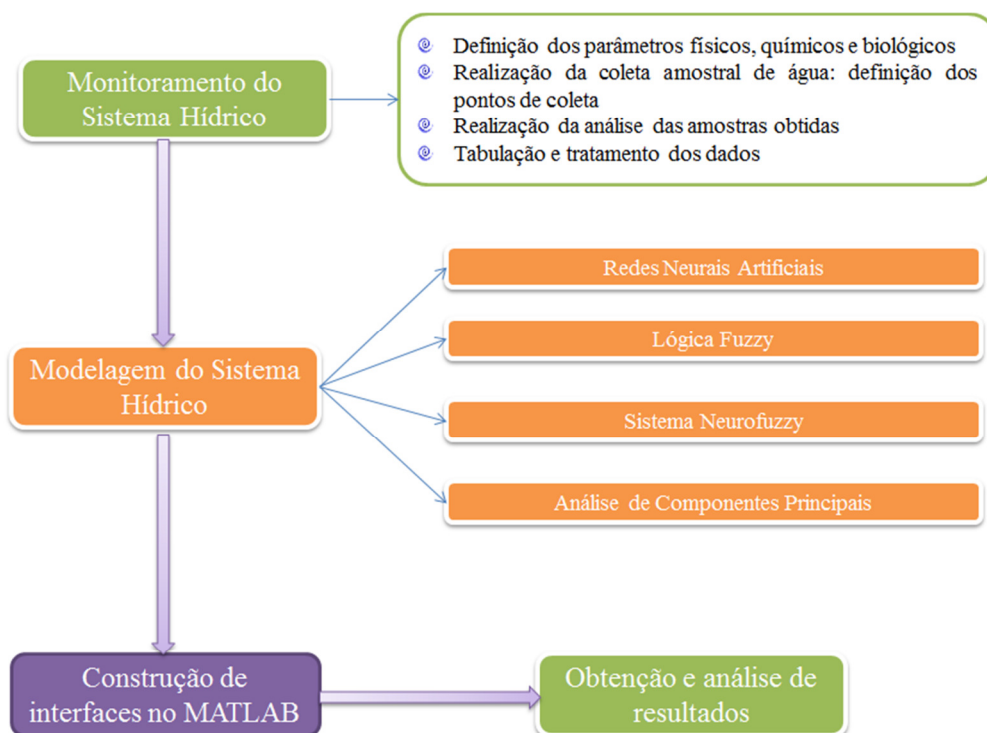


Figura 20 - Estratégia metodológica do trabalho

3.1 AMBIENTES DE ESTUDOS

Na presente tese foram avaliados os dados ambientais dos reservatórios da Marcela e de Jacarecica e do Rio Poxim, em Sergipe.

3.1.1 Reservatório da Marcela

O reservatório da Marcela (Figura 21), situado na cidade de Itabaiana-SE, foi construído no período 1953-1957 pelo barramento do riacho Fuzil. A obra teve a parceria dos Governos Federal e Estadual, tendo como objetivo maior a manutenção do homem no campo, ou seja, a diminuição do êxodo rural. Esse reservatório tem uma capacidade de $2,7 \times 10^6 \text{ m}^3$, ocupa uma área de 14 km^2 e foi projetado para fornecer água para irrigação de 156 hectares, com extensão de 1,0 km (ALVES; GARCIA, 2006).



Figura 21 - Reservatório da Marcela (SRH-SE, 2005)

3.1.2 Reservatório Jacarecica

O reservatório de Jacarecica II (Figura 22) foi construído pelo barramento do Rio Jacarecica, afluente da margem direita do Rio Sergipe, está situado na divisa do município de Malhador e Areia Branca, distante 55 km da cidade de Aracaju, capital do Estado. O reservatório teve sua fase de enchimento iniciado em novembro de 2000, tem capacidade máxima de $29,7 \times 10^6 \text{ m}^3$ de água, com uma cota, prevista no projeto, de 70 m e aproveitamento para irrigação de 1600 hectares e uso doméstico, além de uma extensão de 4,5 km.



Figura 22 - Reservatório de Jacarecica II

Fonte: Google Earth (Acesso em 2007)

3.1.3 Rio Poxim

A área de drenagem da sub-bacia hidrográfica do rio Poxim compreende 381,5 km² e faz parte da bacia hidrográfica do rio Sergipe que drena uma superfície de cerca de 3.670 km². De formato alongado, no sentido noroeste - sudeste, esta unidade de planejamento é limitada ao sul pela bacia hidrográfica do Rio Vaza-Barris e, ao norte, pelo Rio Sergipe. As suas principais nascentes localizam-se a oeste, limite final da Serra dos Cajueiros e sua foz a leste, no rio Sergipe, próxima ao Oceano Atlântico. Ressalta-se que desde 1958 o Rio Poxim tem sido uma das fontes de suprimento de água para Aracaju, sendo, portanto, vital seu monitoramento (ALVES; GARCIA, 2006).

A Figura 23 mostra a sub-bacia do Rio Poxim e identifica a presença de possíveis aportes de contaminantes (Conjunto habitacional e a Universidade Federal de Sergipe – SE). Nesta figura também são mostrados o ponto de captação de água e a localização de uma Estação de Tratamento de Água (ETA) da Companhia de Saneamento de Sergipe (DESO).



Figura 23 - Sub-bacia do Rio Poxim

Fonte: Google Earth (2012)

3.2 MATERIAIS E MÉTODOS

No monitoramento ambiental, os dados experimentais foram obtidos utilizando os procedimentos analíticos consagrados na literatura da área. Nesse caso, a metodologia adotada para caracterização físico-química dos ambientes de estudo foi a análise dos parâmetros de acordo com a APHA (1998). A Tabela 7 mostra os parâmetros (variáveis), o método empregado e a referência metodológica APHA.

Tabela 7 - Variáveis, métodos e referências utilizados para a caracterização da água dos reservatórios

Variável	Metodologia	Referência
pH	Peagâmetro marca Digimed, modelo DMPH-2	
Transparência	Disco de Secchi, equipamento de campo	
Temperatura	Termômetro digital	
Condutividade Elétrica	Condutivímetro, marca Digimed, modelo CD-21	APHA 2510
Amônia	Espectrofotômetro, método do feno,	APHA 4500-NH ₃ F
Nitrito	Espectrofotômetro, método colorimétrico	APHA 4500-NO ₂ ⁻ B
Nitrato	Espectrofotômetro, com redução em coluna automática de cádmio	APHA 4500-NO ₃ ⁻ F
Nitrogênio Total	Espectrofotômetro, com redução em coluna automática de cádmio	APHA 4500-N _{org} D
Fósforo Total	Espectrofotômetro	APHA 4500-P E

Continuação...

Oxigênio Dissolvido	Método de Winkler modificado, com utilização de azida sódica	APHA 4500-O C
Clorofila a	Espectrofotômetro	APHA 10200 H
Cor	Espectrofotômetro	
Profundidade	Equipamento de campo	
Sólidos totais suspensos	Gravimétrico	APHA 2540 C
Sólidos totais dissolvidos	Gravimétrico	APHA 2540 D
Alcalinidade	Titulométrico	
Demanda bioquímica de oxigênio	Titulométrico	APHA 5210 B

3.3 REDE NEURAL

Na presente tese as redes MLP e RBF foram construídas para os dados dos reservatórios da Marcela e de Jacarecica. É importante ressaltar que, em decorrência da disponibilidade dos dados, foram utilizados diferentes parâmetros para os dois reservatórios. Essa disponibilidade era limitada pelo fato de ser o LQA um prestador de serviços e a depender do projeto, alguns parâmetros não eram exigidos ou necessários.

3.3.1 Rede MLP

A rede neural artificial MLP foi construída com 9 neurônios de entrada e 1 neurônio de saída por vez. Os parâmetros escolhidos para os neurônios de entrada foram: temperatura da água, oxigênio dissolvido (OD), pH, alcalinidade, condutividade elétrica, sólidos suspensos, nitrogênio, fosfato e demanda bioquímica de oxigênio (DBO₅).

A alcalinidade, pH, DBO₅ e a condutividade elétrica foram incluídos para simular as propriedades químicas do ambiente, os quais podem ter efeitos indiretos no ecossistema. A radiação solar fornece a energia essencial para a produção primária, logo, para simular os efeitos dessa energia, também foi usada a temperatura como parâmetro de entrada. A concentração de OD foi usada por ser vital para os organismos e para algumas reações químicas. As concentrações de fósforo e nitrogênio são os nutrientes limitantes da eutrofização. Os dados de clorofila-a, indicadores da eutrofização, foram definidos como parâmetro de saída da RNA.

Com base na metodologia descrita anteriormente, as redes neurais foram treinadas e validadas de acordo com os dados disponíveis. Primeiramente foram utilizados os dados dos reservatórios da Marcela e de Jacarecica utilizando a rede MLP.

A rede MLP inicialmente foi construída para os dados da bacia de Jacarecica II com as 3 camadas, sendo inicialmente 9 neurônios na camada de entrada, e tentativas de 6, 8, 10 e 12 neurônios na camada intermediária e 1 neurônio de saída, conforme representação na Figura 24. A função de transferência utilizada nas camadas intermediária e de saída foi a função tangente hiperbólica.

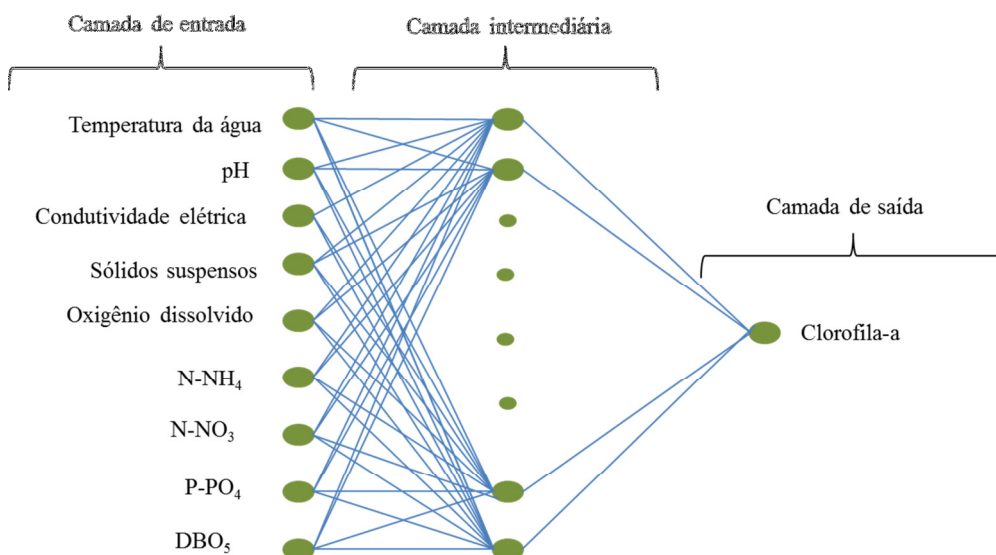


Figura 24 - Topologia para a rede neural MLP

Dentre as muitas variações do método de treinamento *backpropagation*, o algoritmo utilizado foi o Levenberg-Marquadt, que utiliza a informação da Hessiana para a busca do mínimo global do erro e apresenta uma convergência mais rápida para redes que contêm até centenas de neurônios.

3.3.2 Rede RBF

Para o reservatório da Marcela foi possível explorar mais as ferramentas disponíveis no *ToolBox* de redes neurais do MATLAB para o pré-tratamento de dados antes da etapa de treinamento da RNA, tendo em vista que para esse reservatório se tinha uma maior número parâmetros. O pré-tratamento, separando uma parte dos dados, poderia melhorar a eficiência da rede neural. Então, procurou-se, também, explorar a utilização de outro tipo de rede neural disponibilizada pelo MATLAB, a rede com funções de ativação de base radial (RBF).

Neste estudo, os parâmetros escolhidos para os neurônios de entrada foram: profundidade, temperatura da água, temperatura do ar, transparência, condutividade, cor, pH, sólidos totais, sólidos suspensos, oxigênio dissolvido, N-NH₄, N-NO₂, N-NO₃, P-PO₄,

Nitrogênio total, Fósforo total perfazendo um total de 16, ou seja, a camada de entrada consistiu de 16 neurônios. Deve-se observar que os dados de clorofila-a, indicador da eutrofização, foram definidos como parâmetro de saída da RNA, assim como no reservatório de Jacarecica. A Figura 25 mostra o arranjo de entradas e saída da rede neural construída para os dados do reservatório de Marcela.

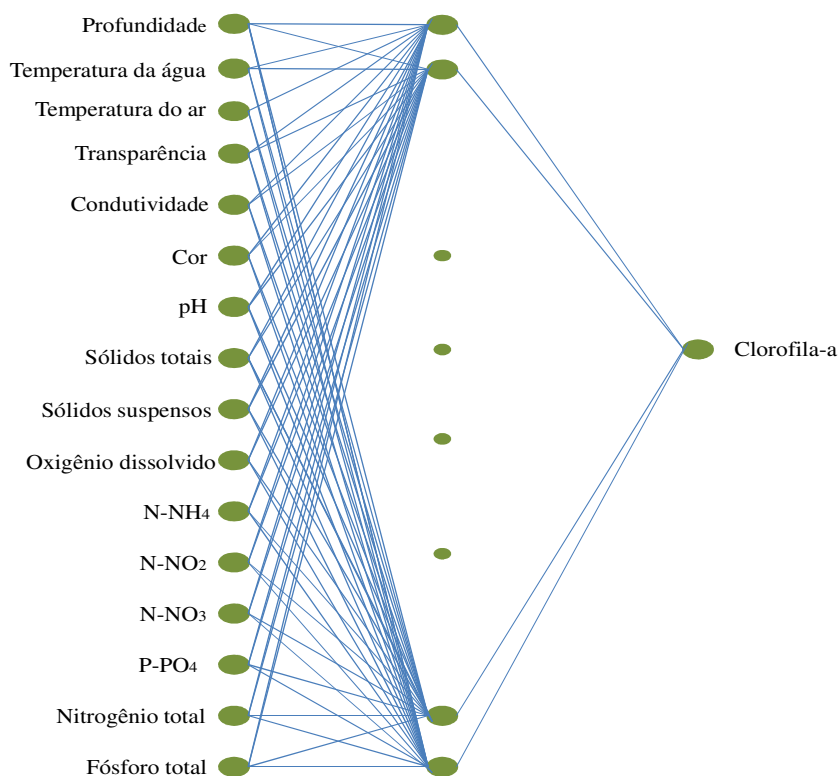


Figura 25 - Topologia para RNA para os dados de Marcela

No pré-tratamento foi utilizado a análise dos componentes principais (PCA – Principal Component Analysis), que é uma ferramenta para compressão de dados e extração de informações, conforme argumentado anteriormente. A técnica de PCA encontra combinações de variáveis, ou fatores, que descrevem a maior tendência nos dados.

A técnica de PCA consiste em uma manipulação da matriz de dados com o objetivo de representar alterações presentes em muitas variáveis, através de um número menor de fatores. Nesta, constrói-se um novo sistema de eixos, denominados de fatores, componentes principais, variáveis latentes ou autovetores, para representar as amostras, no qual a natureza multivariada dos dados pode ser visualizada em poucas dimensões.

3.3.3 Sistema Neuro-Fuzzy

Como já argumentado anteriormente, o sistema neuro-fuzzy mostra a integração entre duas abordagens de inteligência artificial: lógica fuzzy e redes neurais. No sistema ANFIS o algoritmo de inferência constrói o FIS (Fuzzy Inference System), cuja função de partição é ajustada usando o algoritmo backpropagation ou combinação com métodos de mínimos quadrados. Portanto, o sistema ANFIS apresenta técnica de aprendizagem Neuro-Adaptativo (*Neuro-Adaptive*).

Neste trabalho foi utilizado o sistema ANFIS do *ToolBox* do MATLAB para os dados do rio Poxim utilizando a rede MLP. Os parâmetros escolhidos para os neurônios de entrada foram: temperatura da água, condutividade, cor, pH, sólidos totais, sólidos suspensos, oxigênio dissolvido, DBO, N-NH₄, N-NO₂, N-NO₃, P-PO₄, nitrogênio total e fósforo total, perfazendo um total de 13, ou seja, a camada de entrada apresenta 13 neurônios. Deve-se observar que os dados de clorofila-a, o indicador da eutrofização, foram, mais uma vez, definidos como parâmetro de saída da rede neural e do sistema ANFIS. A Figura 26 mostra a topologia do modelo ANFIS com a identificação das variáveis.

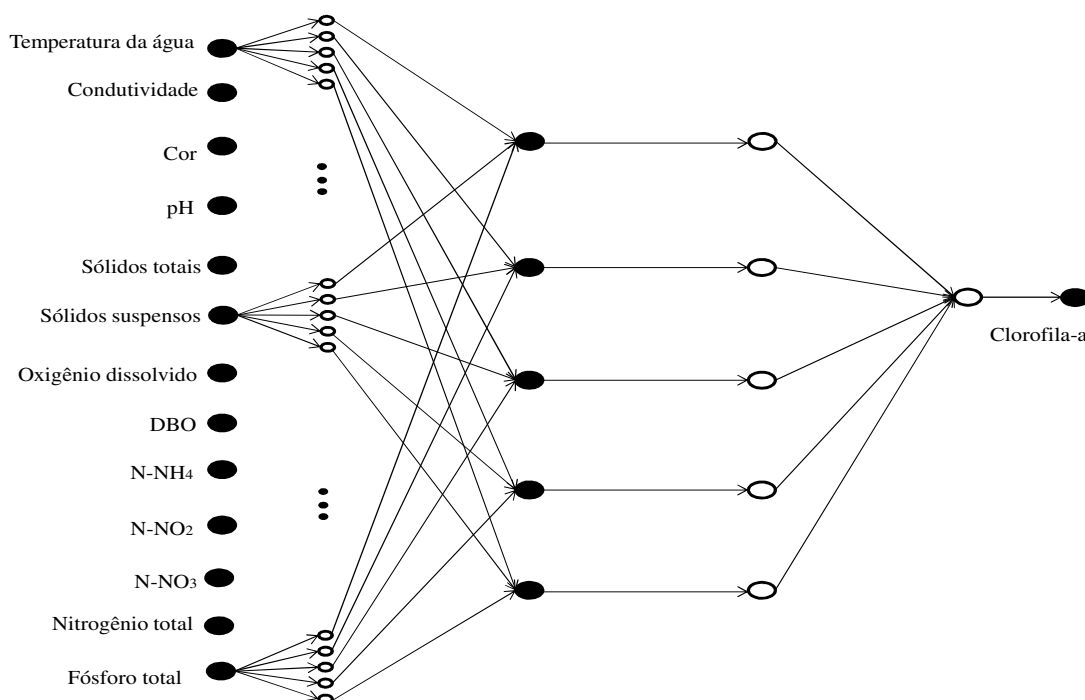


Figura 26 - Topologia para condições ótimas do modelo ANFIS

É importante salientar que a utilização de diferentes parâmetros para ambientes hídricos é consequência da dificuldade de obtenção desses dados experimentais. Alguns desses dados foram obtidos de projetos e subprojetos de pesquisa do Laboratório de Química Analítica Ambiental da UFS. Além disso, o custo de cada campanha fazia com que somente

os dados exigidos pelos projetos, não sendo, portanto, objeto da tese a coleta experimental e sim a avaliação numérica do recurso hídrico em análise.

A Tabela 8 mostra, a título de exemplo, os custos para obtenção dos dados experimentais.

Tabela 8 - Custos por campanha experimental

Aluguel de barco	R\$ 300,00
Aluguel de canoa	R\$ 50,00
Aluguel de veículo	R\$ 250,00
Alimentação	R\$ 200,00
Diária para pessoal de apoio (duas pessoas)	R\$ 200,00
Material de coleta (isopor, gelo, frascos, etc...)	R\$ 200,00
Diária do pessoal de coleta	R\$ 100,00
Total de gastos com mobilização e coleta	R\$ 1.300,00
Custo de análise de uma amostra de água (parâmetros físico-químicos, microbiológicos)	R\$ 450,00
Reservatório da Marcela	R\$ 5.350,00
Reservatório de Jacarecica I	R\$ 9.400,00
Reservatório de Jacarecica II	R\$ 47.650,00
Poxim	R\$ 6.250,00
Total	R\$ 68.650,00

3.4 FERRAMENTA COMPUTACIONAL

Neste trabalho foi utilizado o ambiente MATLAB, versão 2008 e versão 2011, através de seus *toolboxes* para redes neurais e lógica *fuzzy*. No MATLAB foi construída uma ferramenta computacional, denominada de Modelagem da Qualidade da Água, para análise em termos de redes neurais, determinação do índice trófico da água e de lógica *fuzzy*, conforme mostra a Figura 27. Clicando em cada ícone das etapas, são obtidos os resultados da Modelagem da Qualidade da Água.

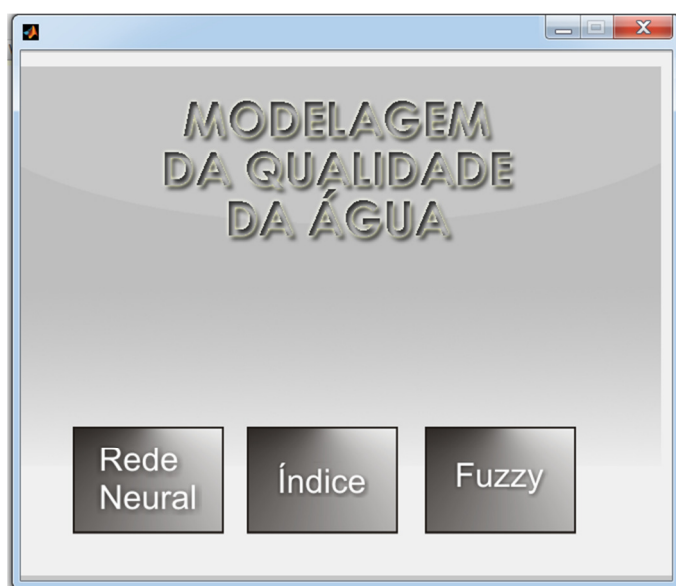


Figura 27 - Tela principal da ferramenta computacional

Clicando no ícone Rede Neural da Figura 27, obtém-se a tela mostrando a modelagem para os dois tipos de redes (MLP, RBF) e para o sistema neuro-fuzzy exemplificada na Figura 28.

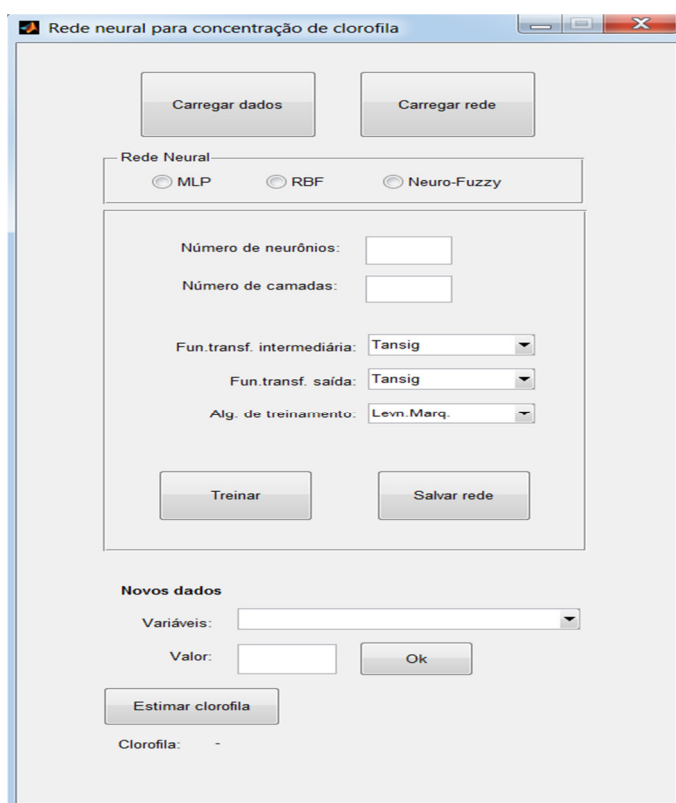


Figura 28 - Ambiente da Rede Neural (MLP, RBF e Neuro-Fuzzy)

Para obtenção dos resultados, a seguinte metodologia foi empregada:

- Primeiro passo: Na Figura 28 clicar no ícone *Carregar os dados*, caso sejam novos dados, ou *Carregar rede*, caso os dados já tenham sido rodados.
- Segundo passo: Escolher o tipo de rede a ser aplicada (MLP e RBF) ou escolher aplicar o sistema neuro-fuzzy.
- Terceiro passo: Para rede MLP definir o número de neurônio e o número de camadas (Figura 29) e clicar no ícone treinar resultando na Figura 30. Após o treinamento da rede os gráficos são gerados e é possível estimar o valor da clorofila-a para novos dados experimentais.

Rede neural para concentração de clorofila

Carregar dados Carregar rede

Rede Neural

☒ MLP ☐ RBF ☐ Neuro-Fuzzy

Número de neurônios:

Número de camadas:

Fun.transf. intermediária: Tansig

Fun.transf. saída: Tansig

Alg. de treinamento: Levn.Marq.

Treinar Salvar rede

Novos dados

Variáveis:

Valor: Ok

Estimar clorofila

Clorofila: -

Figura 29 - Rede MLP

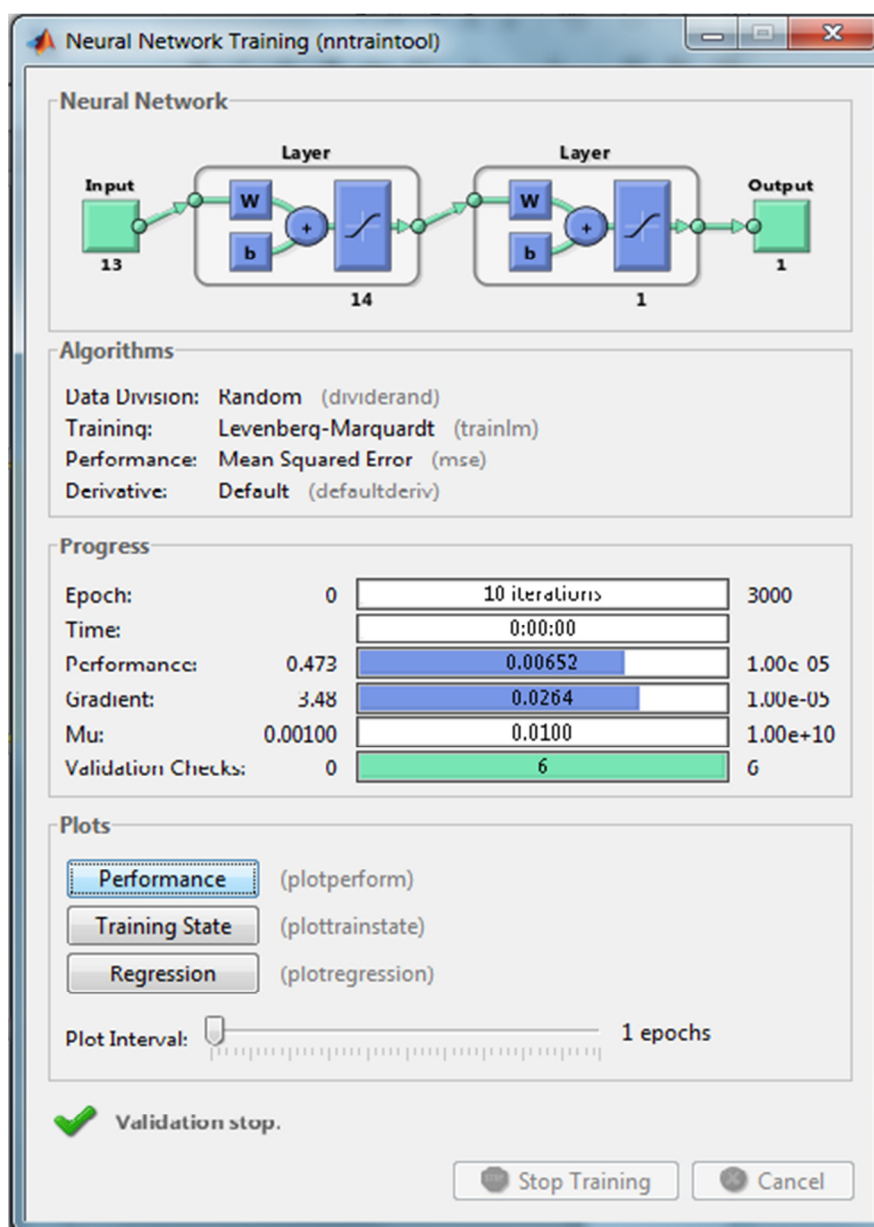


Figura 30 - Treinamento da rede MLP

- d. Quarto passo: Para rede RBF definir o *spread* (taxa de espalhamento dos dados) e o número máximo de neurônios, conforme indicado na Figura 31. Esses valores são fornecidos com base nos resultados da rede MLP ou com base na confiabilidade dos dados experimentais. Após o treinamento da rede os gráficos são gerados e é possível estimar o valor da clorofila para novos dados experimentais.

Rede neural para concentração de clorofila

Carregar dados Carregar rede

Rede Neural

☐ MLP ☒ RBF ☐ Neuro-Fuzzy

Spread: 0.55

Número max. de neurônios: 10

Treinar Salvar rede

Novos dados

Variáveis: Ptotal

Valor: Ok

Estimar clorofila

Clorofila: 0.545578

Figura 31 - Rede RBF

- e. Quinto passo: Para o sistema neuro-fuzzy (Figura 32), clicando no ícone neuro-fuzzy resultará nas Figuras 33 e 34. Essas duas figuras corresponde às etapas necessárias para obtenção dos resultados: selecionar o ícone *Training* (Seta A - arquivo em xls com os dados separados para o treinamento) e selecionar o ícone *Checking* (Seta B - arquivo em xls com os dados separados para a validação). Depois de selecionados os dados, clicar no ícone *Test now* (Seta C – obtenção dos resultados). Após a etapa do test now é possível obter a tela de saída das regras para estimativa da clorofila-a. Ressalta-se que é possível estimar o valor da

clorofila-a para novos dados experimentais inseridos no campo passível de inserção mostrado na Figura 34.

Rede neural para concentração de clorofila

Carregar dados Carregar rede

Rede Neural

☐ MLP ☐ RBF ☒ Neuro-Fuzzy

Spread: 0.55

Número max. de neurônios: 10

Treinar Salvar rede

Novos dados

Variáveis: Ptotal

Valor: Ok

Estimar clorofila

Clorofila: 0.545578

Figura 32 - Sistema neuro-fuzzy

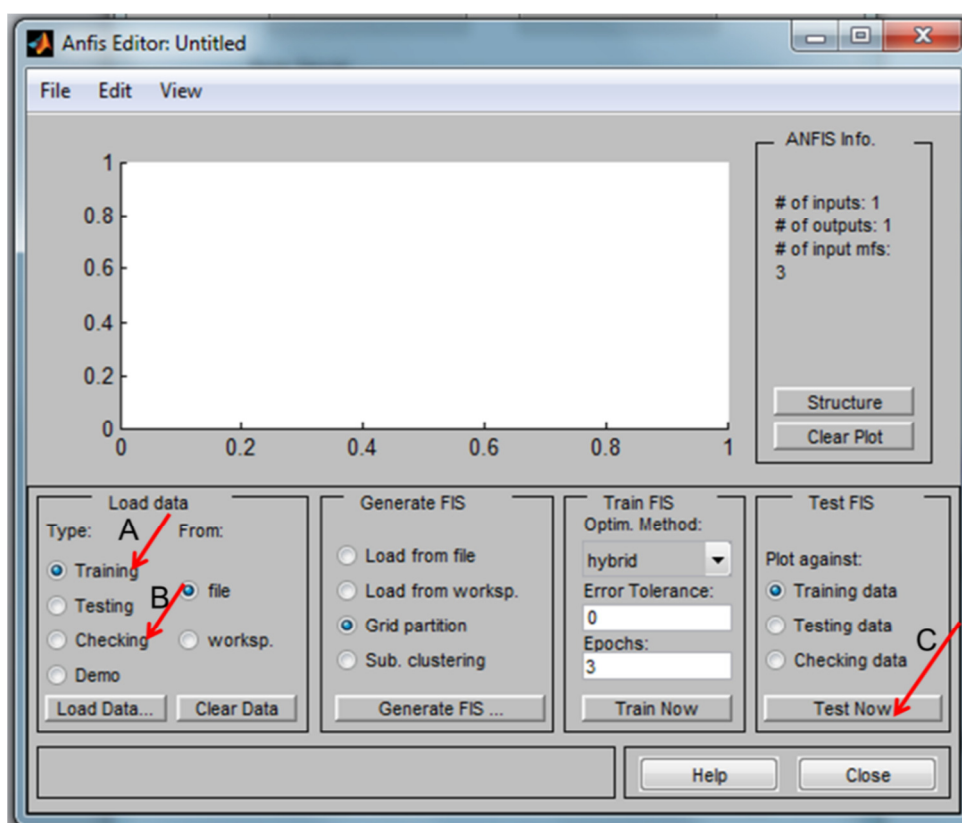


Figura 33 - Sistema neuro-fuzzy para as etapas de treinamento, validação e teste.

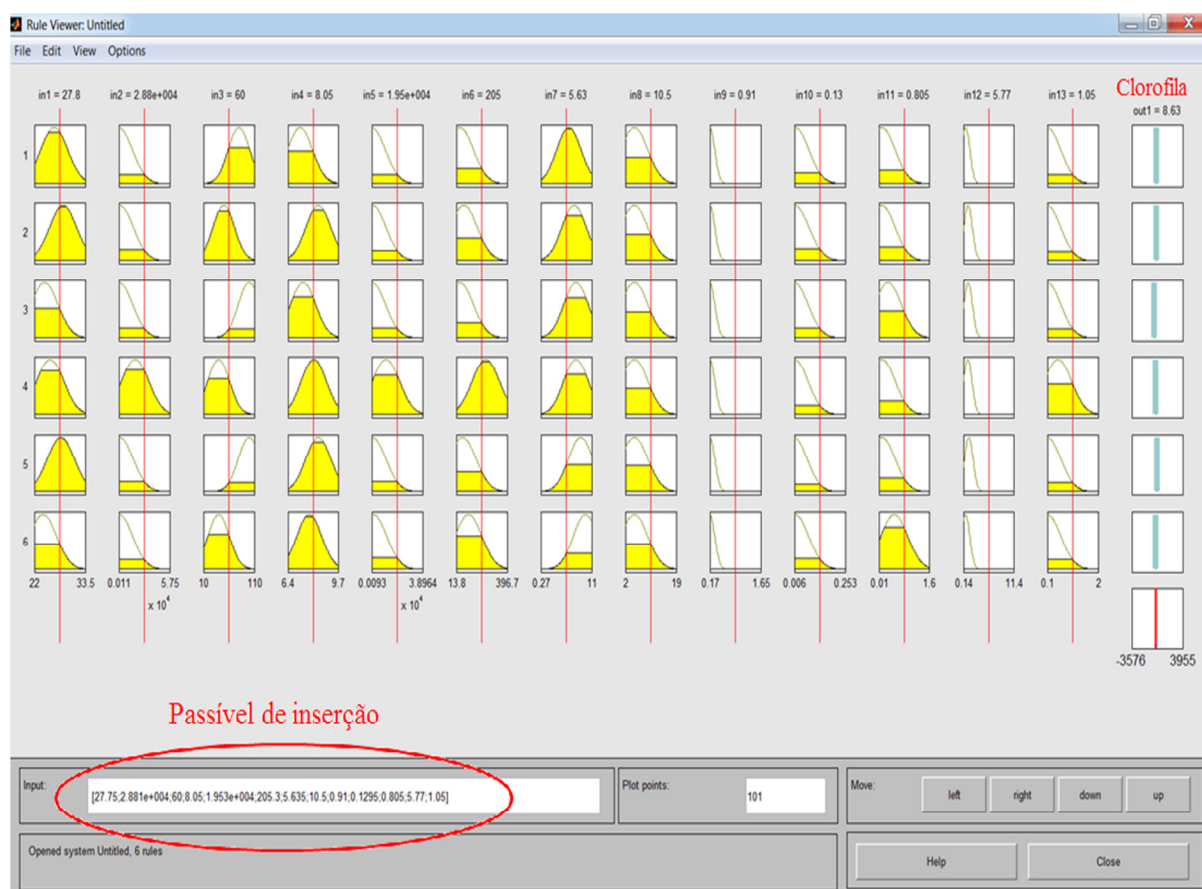


Figura 34 - Regras do sistema neuro-fuzzy e estimativa da clorofila-a

A construção da ferramenta computacional para criação de interfaces amigáveis, ou seja, a metodologia utilizada para aplicação mais fácil e mais rápida das técnicas de redes neurais e lógica fuzzy está descrita no ANEXO B.

4 RESULTADOS E DISCUSSÃO

O fenômeno da eutrofização, assim como qualquer outro problema a ser resolvido por um modelo de rede neural ou lógica *fuzzy*, foi idealizado como um problema de entrada-saída, ou seja, os dados de clorofila (*output*) são associados aos dados dos nutrientes (*input*). Para o sucesso da modelagem e simulação foi necessário que os fatores físico, químico ou biológico envolvidos no fenômeno estivessem direta ou indiretamente envolvidos no sistema.

Neste item, então, os resultados foram organizados de acordo com a aplicação da ferramenta computacional desenvolvida para a Modelagem da Qualidade da Água, ou seja, os resultados para cada ícone mostrado na Figura 27: Rede Neural, Índice e Fuzzy.

4.1 REDE NEURAL

4.1.1 Rede MLP

Para a modelagem empírica utilizando as redes neurais artificiais, foi utilizado o MatLab versão 2008 e adaptada a versão 2011, através do toolbox de redes neurais artificiais. Neste toolbox constam programas específicos para a utilização das redes neurais. Vale salientar que a modelagem através das redes neurais exige várias escolhas de parâmetros como arquitetura da rede neural, número de camadas, números de neurônios em cada camada, funções de ativação dos neurônios por camada e algoritmo de treinamento. Essas definições foram realizadas originalmente no toolbox de redes neurais através de modificações em linhas de programação MATLAB se caracterizando como um processo de tentativa e erro.

De acordo com a metodologia apresentada no item 3.3.1, as figuras 35 a 40 mostram os resultados obtidos pela rede MLP utilizada na modelagem da clorofila-a para diferentes números de neurônios na camada oculta para o reservatório de Jacarecica.

As Figuras 35 e 36 mostram uma comparação entre os dados experimentais e os dados simulados com o objetivo de ajustar os dados para estimativa da concentração de clorofila-a. Nestas figuras a rede MLP foi treinada com 10 neurônios na camada oculta e o melhor ajuste entre os dados foi a equação $A = 0,98 T + 0,33$, na qual A representa a concentração da clorofila-a experimental e T a concentração da clorofila-a experimental, com o coeficiente de correlação (R) entre os dados igual a 0,979.

A Figura 36 representa, também, uma comparação entre os dados experimentais e os dados simulados de clorofila-a, estratificando os dados utilizados para validação, teste e treinamento. É importante ressaltar que, conforme recomendado na literatura da área, o particionamento do conjunto de dados foi 50% dos dados para treinamento, 25% para validação e 25% para teste.

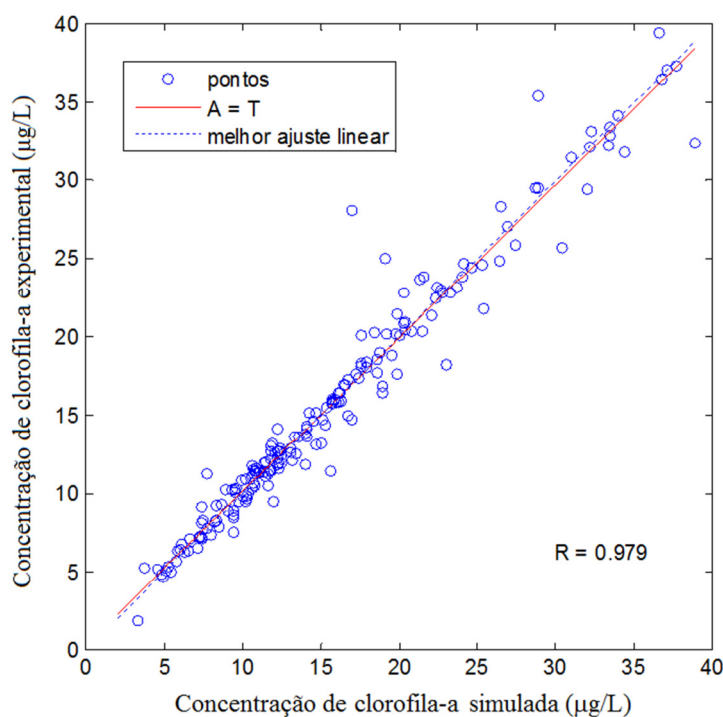


Figura 35 - Performance da rede neural com 10 neurônios na camada oculta

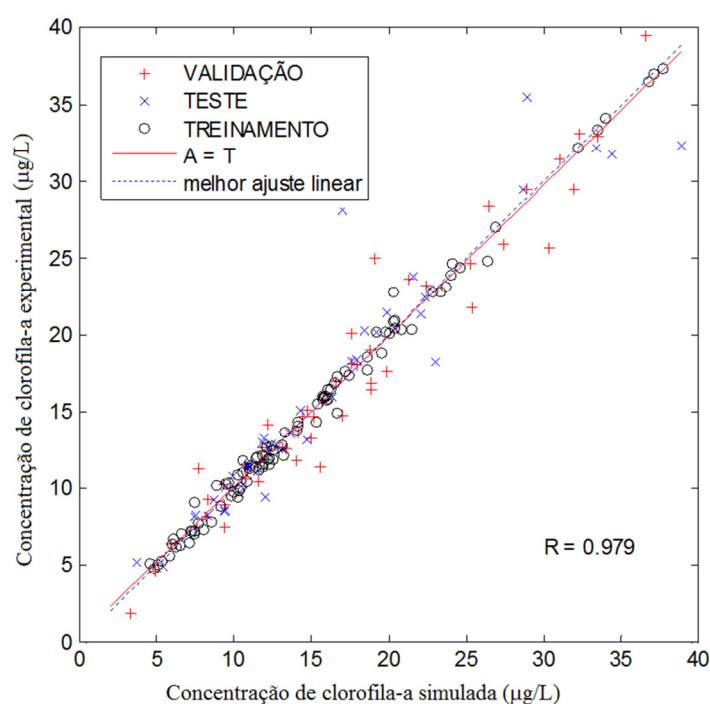


Figura 36 - Performance da rede neural com particionamento do conjunto de dados – 10 neurônios na camada oculta

As Figuras 37 e 38 mostram uma comparação entre os dados experimentais e os dados simulados com o objetivo de ajustar os dados para estimativa da concentração de clorofila-a. Nestas figuras a rede MLP foi treinada com 12 neurônios na camada oculta e o melhor ajuste entre os dados foi a equação $A = 0,918 T + 0,528$, na qual A representa a concentração da clorofila-a experimental e T a concentração da clorofila-a experimental, com o coeficiente de correlação (R) entre os dados igual a 0,886.

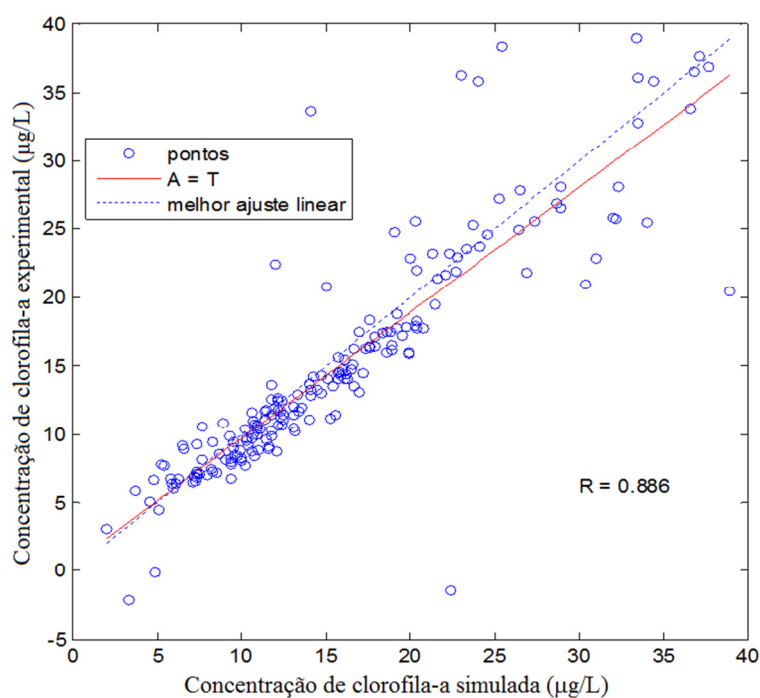


Figura 37 - Performance da rede neural com 12 neurônios na camada oculta

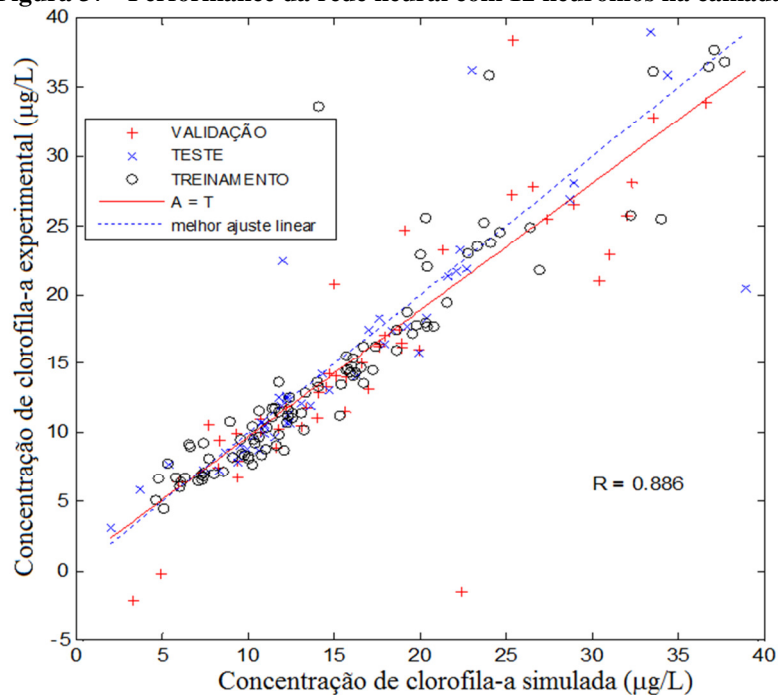


Figura 38 - Performance da rede neural com particionamento do conjunto de dados - 12 neurônios na camada oculta

As Figuras 39 e 40 mostram uma comparação entre os dados experimentais e os dados simulados com o objetivo de ajustar os dados para estimativa da concentração de clorofila-a. Nestas figuras a rede MLP foi treinada com 14 neurônios na camada oculta e o melhor ajuste entre os dados foi a equação $A = 0,729 T + 4,32$, na qual A representa a concentração da clorofila-a experimental e T a concentração da clorofila-a simulada, com o coeficiente de correlação (R) entre os dados igual a 0,978.

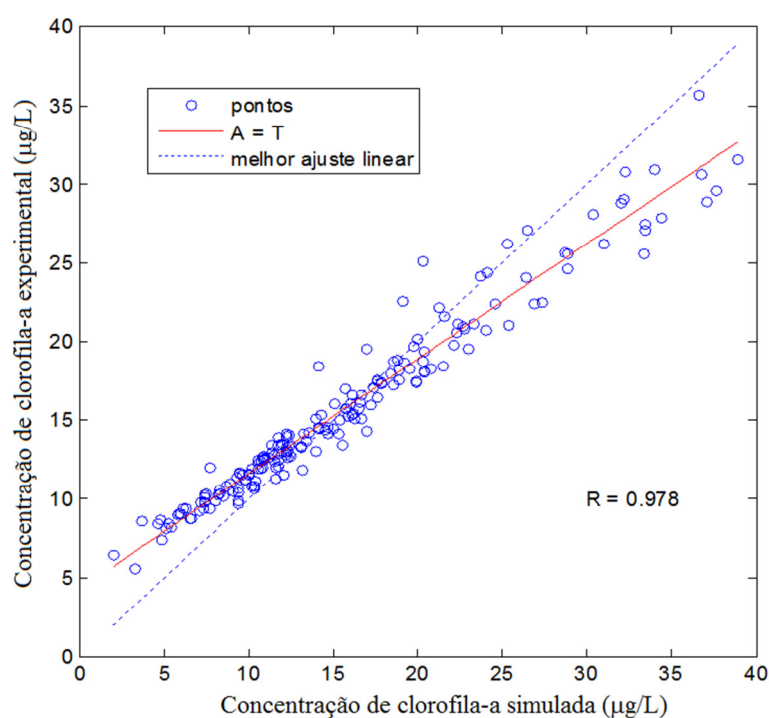


Figura 39 - Performance da rede neural com 14 neurônios na camada oculta

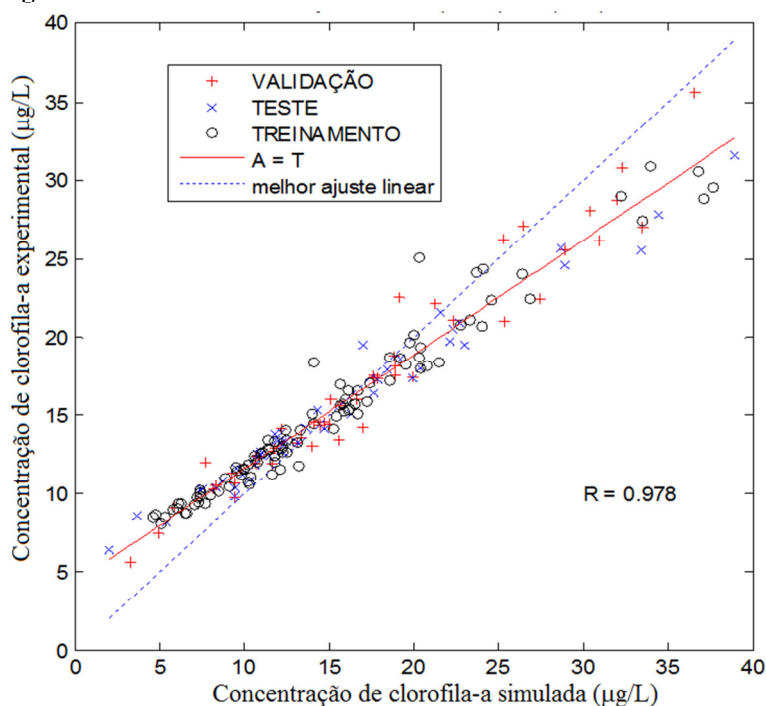


Figura 40 - Performance da rede neural com particionamento do conjunto de dados – 14 neurônios na camada oculta

Através da análise das figuras 35 a 40, é possível identificar que o número de neurônios na camada oculta que apresentou o melhor desempenho para a rede MLP foi aquela com 10 neurônios. É importante comentar que os resultados em termos de coeficiente de ajuste linear são bem próximos para o número de neurônios implementado na simulação.

4.1.2 Redes MLP e RBF com o PCA

As Figuras 41 a 48 ilustram os resultados obtidos com o treinamento das redes MLP e RBF para o reservatório da Marcela, com e sem o uso do PCA para pré tratamento de dados. Nestas figuras, a rede MLP foi treinada com 12 neurônios na camada oculta e a rede RBF foi treinada com 25 neurônios na camada oculta, pois foram identificados como sendo os números de neurônios que representavam o melhor ajuste entre os dados de concentração de clorofila-a.

A Figura 41 mostra o melhor ajuste da rede RBF sem realizar o pré-tratamento usando o PCA. O melhor ajuste entre as concentrações de clorofila-a experimental e simulada foi a equação $A = 0,865 T + 6,16$ e um coeficiente de correlação de 0,932. Com esse ajuste, é possível admitir que a rede simulou com boa precisão as concentrações de clorofila-a, conforme mostra a Figura 42.

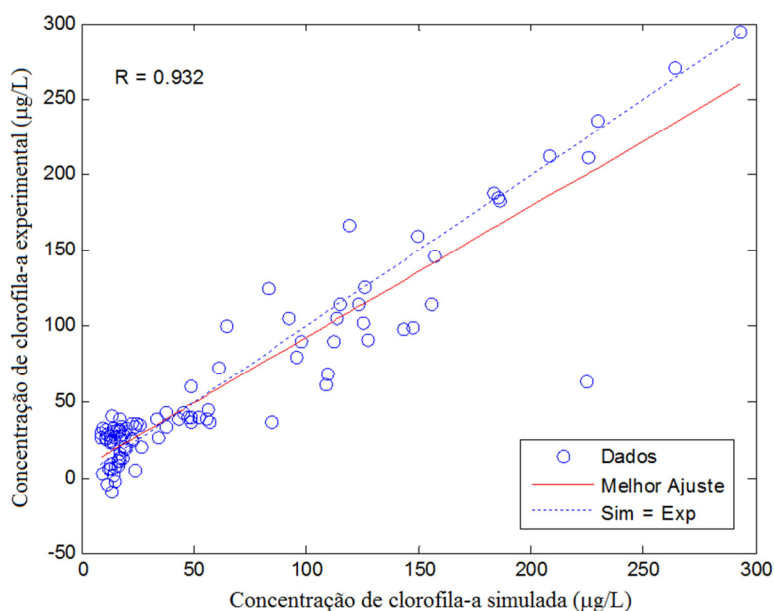


Figura 41 - Aderência do modelo para a RBF sem PCA

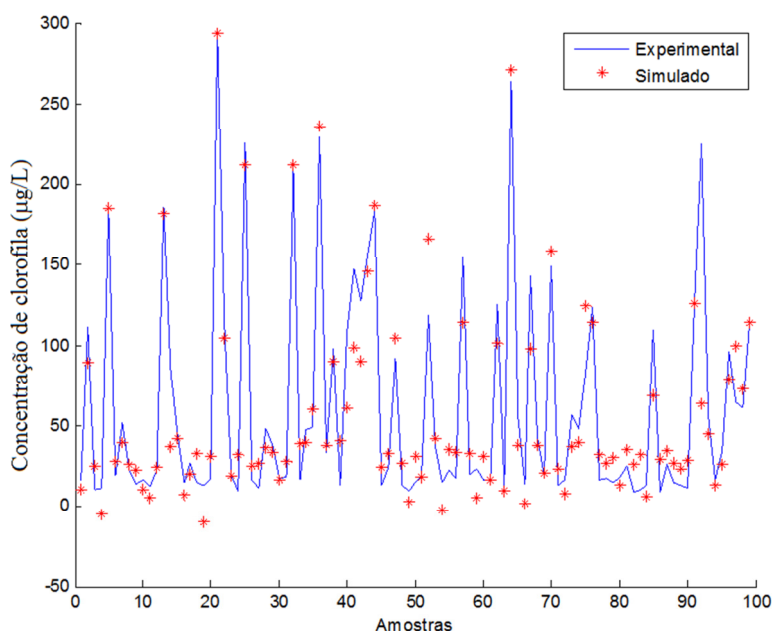


Figura 42 - Resultado da simulação para a RBF sem PCA

A Figura 43 mostra o melhor ajuste da rede RBF utilizando o pré-tratamento através da técnica de PCA. O melhor ajuste entre as concentrações de clorofila-a experimental e simulada neste caso foi a equação $A = 0,869 T + 6,28$ e um coeficiente de correlação de 0,93. Observa-se que para a rede RBF a utilização do PCA não influencia n o ajuste dos dados, pois a diferença entre os coeficientes de correlação é de apenas 0,2145%. Com esse ajuste, também, admite-se que a rede conseguiu simular com boa precisão a concentração de clorofila-a.

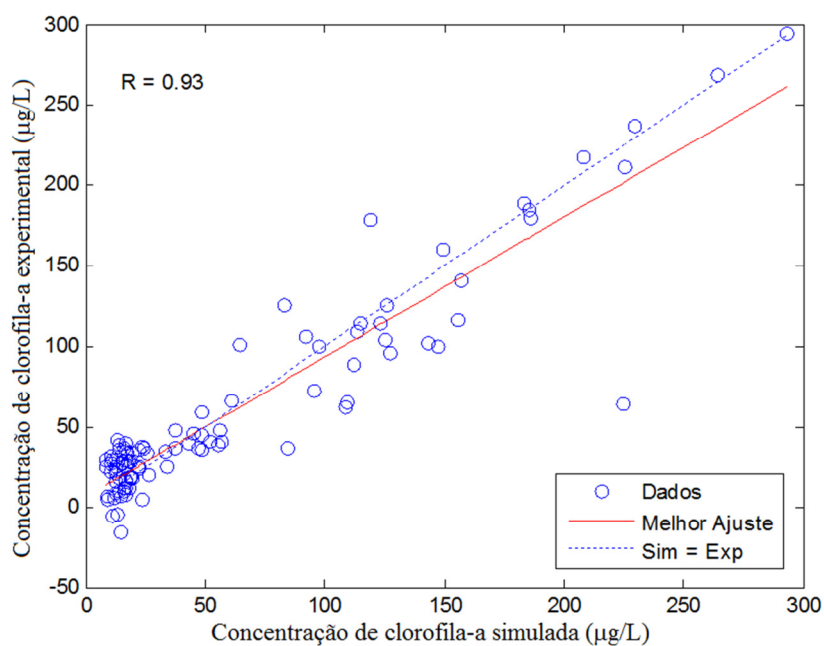


Figura 43 - Aderência da rede RBF com PCA

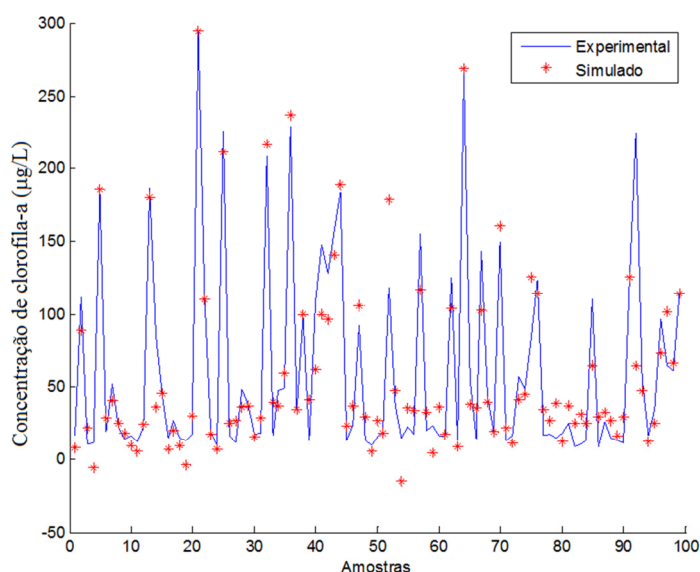


Figura 44 - Resultado da simulação para a RBF com PCA

As Figuras 45 a 48 mostram a rede MLP sem e com o pré-tratamento através da técnica de PCA.

A Figura 45 mostra o melhor ajuste da rede MLP sem realizar o pré-tratamento usando o PCA. O melhor ajuste entre as concentrações de clorofila-a experimental e simulada neste caso foi $A = 0,957 T + 7,05$ e um coeficiente de correlação de 0,944. Para a rede MLP ocorre desempenho similar ao obtido pela rede RBF, simulando com boa precisão as concentrações de clorofila-a, conforme mostra a Figura 46.

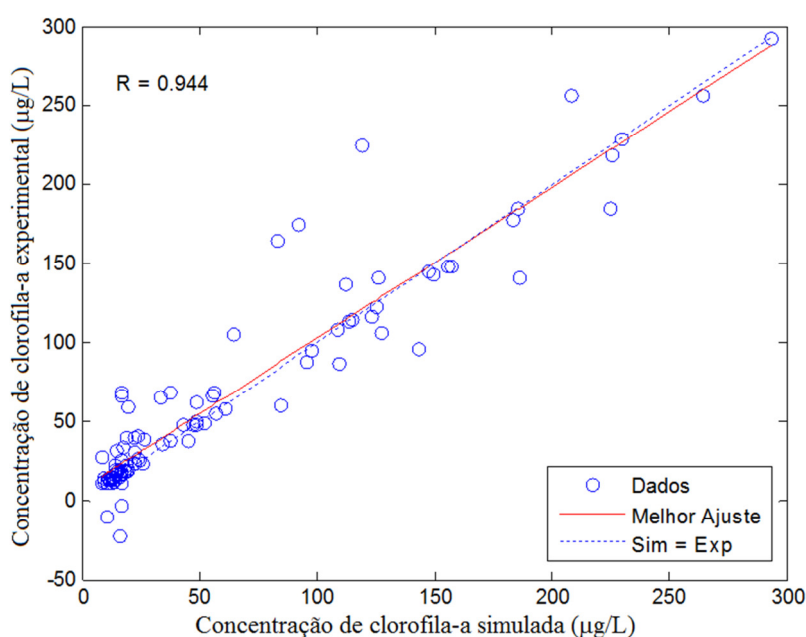
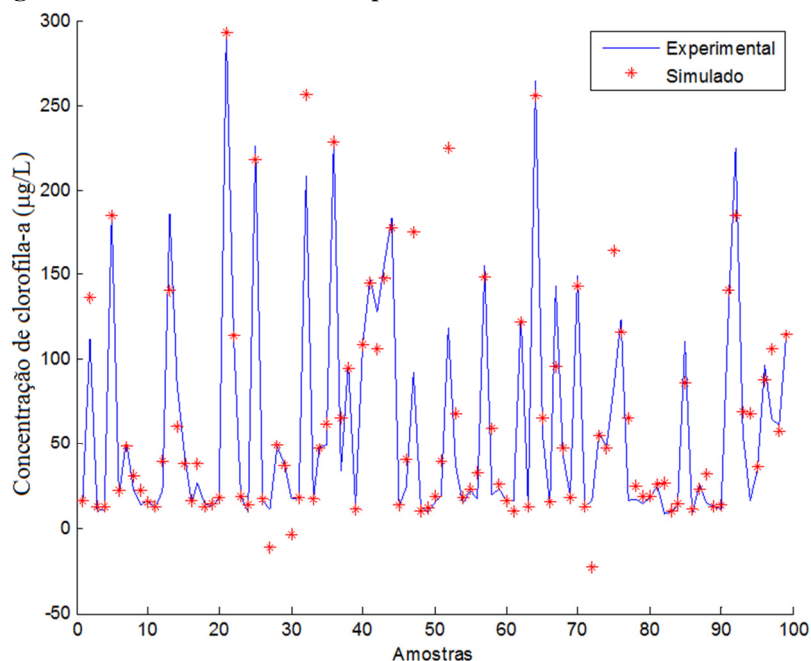
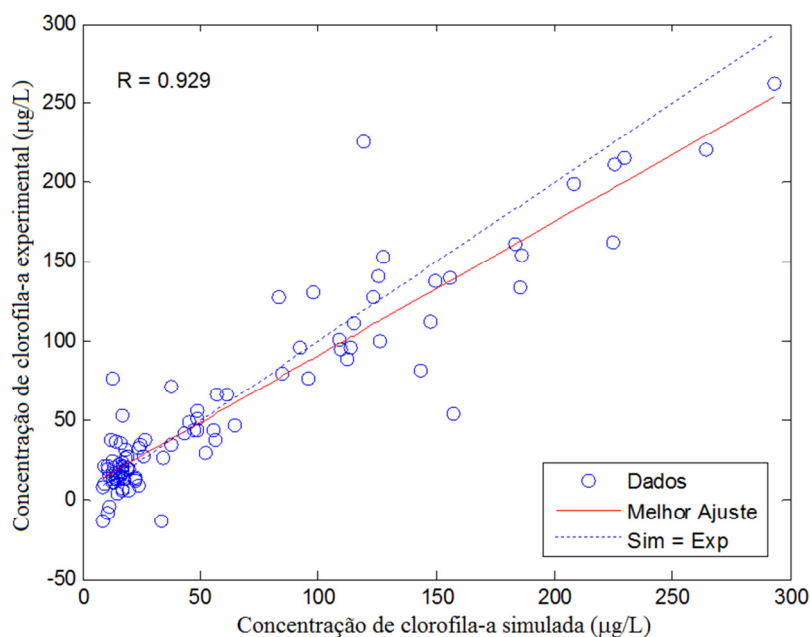


Figura 45 - Aderência do modelo para a MLP sem PCA**Figura 46 - Resultado da simulação para a MLP sem PCA**

A Figura 47 mostra o melhor ajuste da concentração de clorofila usando a rede MLP com pré-tratamento através da técnica de PCA. O melhor ajuste entre as concentrações de clorofila-a experimental e simulada, neste caso, foi $A = 0,845 T + 6,38$ e um coeficiente de correlação de 0,929. Para esse ajuste, a Figura 48 mostra que a rede conseguiu simular com boa precisão as concentrações de clorofila-a.

**Figura 47 - Aderência do modelo para a MLP com PCA**

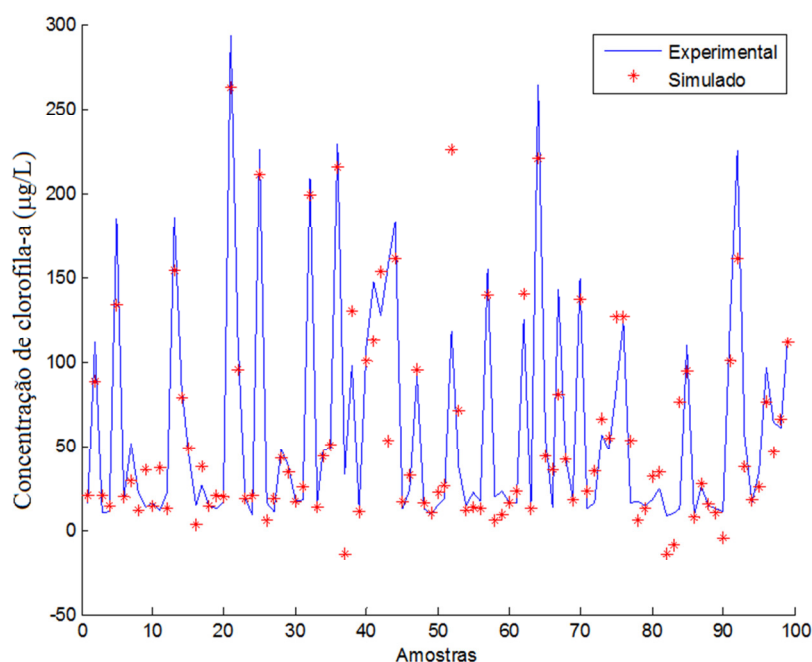


Figura 48 - Resultado da simulação para a MLP com PCA

A Tabela 9 resume os melhores valores dos ajustes (aderência) de cada tipo de rede com e sem o pré-tratamento através do PCA. Nesta tabela, observa-se que não houve diferença significativa entre os modelos (redes) e que o uso do PCA não resultou em uma melhoria da qualidade dos resultados. Porém, deve-se ressaltar que o fato do PCA não fornecer melhoria para os dados utilizados no treinamento, não implica que no futuro, com um conjunto de dados mais representativo, seu desempenho não possa ser significativo. Ressaltando mais uma vez que a dificuldade na obtenção dos dados experimentais pode justificar o fato de não ocorrer essa melhoria, levando em consideração que o número de medidas foi bem menor quando comparados com o número de dados apresentados na literatura.

Tabela 9 - Aderência das redes neurais testadas para o reservatório da Marcela

Tipo de rede	Neurônios ocultos	Com PCA	Sem PCA
RBF	25	0,930	0,932
MLP	12	0,929	0,944

Para os conjuntos de dados utilizados nesta tese, as redes neurais do tipo MLP e RBF apresentaram resultados semelhantes e produziram melhores resultados se comparados a modelos com estatística multivariada encontrados na literatura sobre o mesmo tema. A modelagem RNA da eutrofização, através da determinação da clorofila-a no reservatório de Marcela Sergipe/Brasil, realizada com um número maior de dados foi capaz de simular as concentrações de clorofila-a com um coeficiente de regressão médio de aproximadamente

0,93 tanto para a rede MLP como para RBF . A avaliação dos resultados induz a constatação de que o processo de eutrofização pode ser modelado com razoável precisão usando as RNA. Cabe ressaltar que, pelo fato dos dados apresentados para o reservatório da Marcela não representarem uma série histórica, o modelo não tem nenhum componente temporal, possibilitando, desta forma, apenas um diagnóstico atual da qualidade das águas do reservatório.

4.2 SISTEMA NEURO-FUZZY

O sistema neuro-fuzzy foi aplicado aos dados experimentais do rio Poxim, e, lamentavelmente, somente 70 medidas foram possível para os mesmos parâmetros das redes treinadas para os dois reservatórios. No entanto, com os dados da topologia apresentada no item 3.3.3 e na Figura 28, que são utilizados para o treinamento do sistema ANFIS, obteve-se um razoável ajuste para 14 neurônios na camada intermediária, como mostra a Figura 49. Nesta figura o melhor ajuste entre as concentrações de clorofila-a experimental e simulada, foi $A = 0,819 T - 0,102$ e um coeficiente de correlação de 0,839.

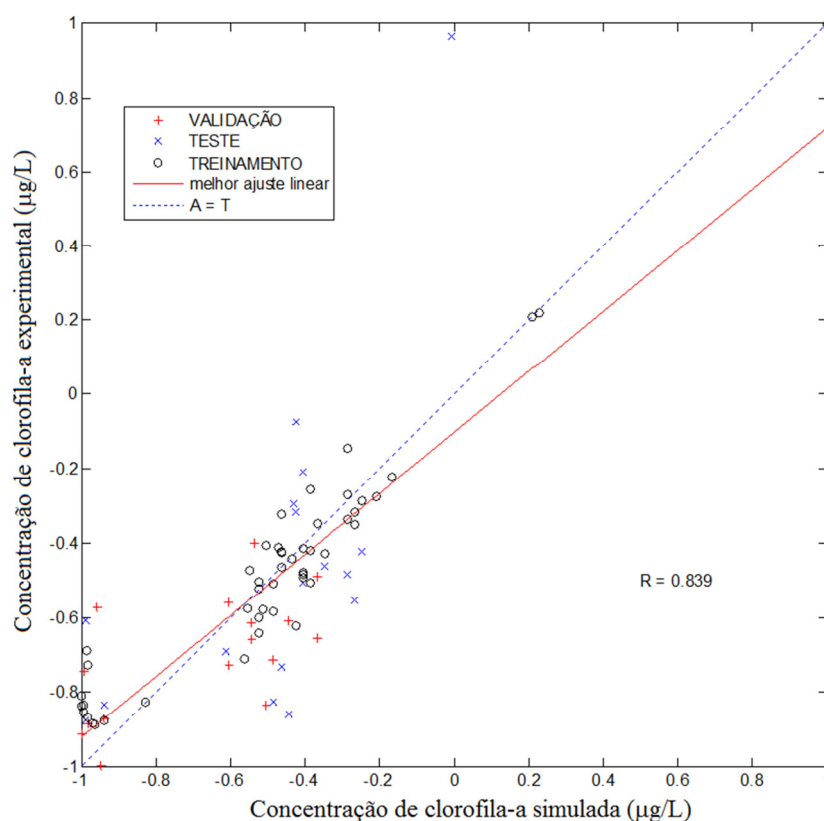


Figura 49 - Performance da rede neural no sistema neuro-fuzzy - rio Poxim

Ressalta-se que estes dados foram extraídos da topologia dada pela Figura 28, que foram utilizados para o treinamento e validação do sistema neuro-fuzzy para avaliar o rio Poxim. As Figuras 50 e 51 mostram as telas do sistema ANFIS do Matlab sem customização para a concentração de clorofila-a. A Figura 50 para o conjunto de dados destinado ao treinamento da rede e a Figura 51 para o conjunto de dados utilizado na validação da rede. Nestas figuras, conforme metodologia, foram utilizados 13 parâmetros de entrada (# of inputs) e o parâmetro de saída foi a concentração de clorofila-a (# of outputs). O número de regras para cada parâmetro foi 3 (# of input mfs) e 65 dados foram utilizados na etapa de treinamento e 17 na etapa de validação. É importante ressaltar que o próprio sistema particiona os dados para fornecer o melhor desempenho do sistema neuro-fuzzy e que não necessariamente, quando somados os números de dados de cada etapa seja o número total de dados fornecidos.

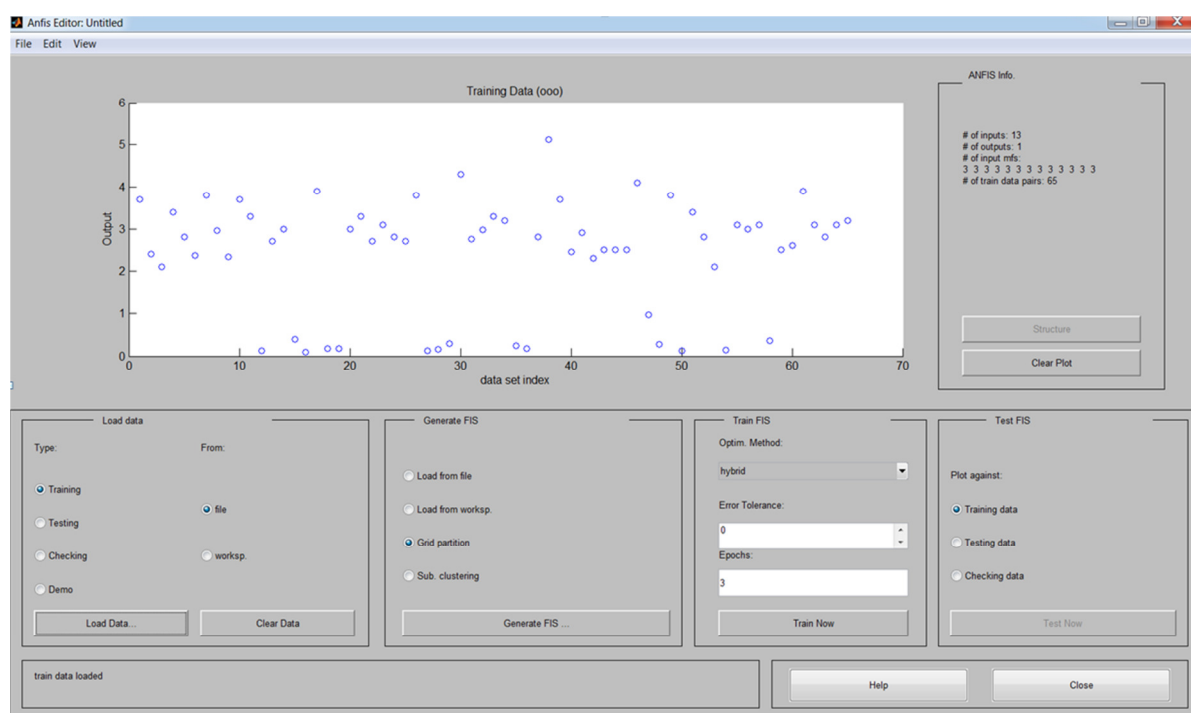


Figura 50 - Ilustração da tela do sistema ANFIS para o treinamento.

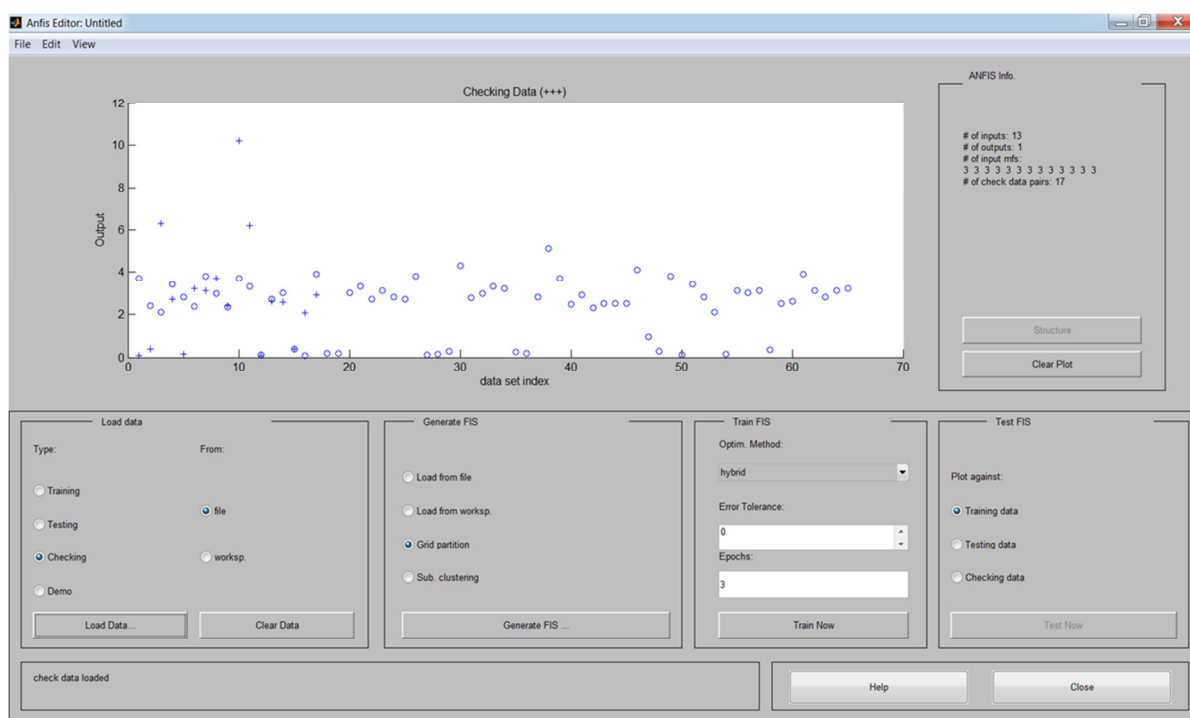


Figura 51 - Ilustração da tela do sistema ANFIS para validação

Depois de realizado o treinamento e a validação da rede, realizou-se o teste do sistema (*test now*) conforme descrito no item 3.4. Para cada variável de entrada, um conjunto de regras foi otimizado no sistema ANFIS. Para ilustrar essa otimização, a Figura 52 mostra um conjunto com cinco regras para a concentração de nitrogênio (12º parâmetro de entrada).

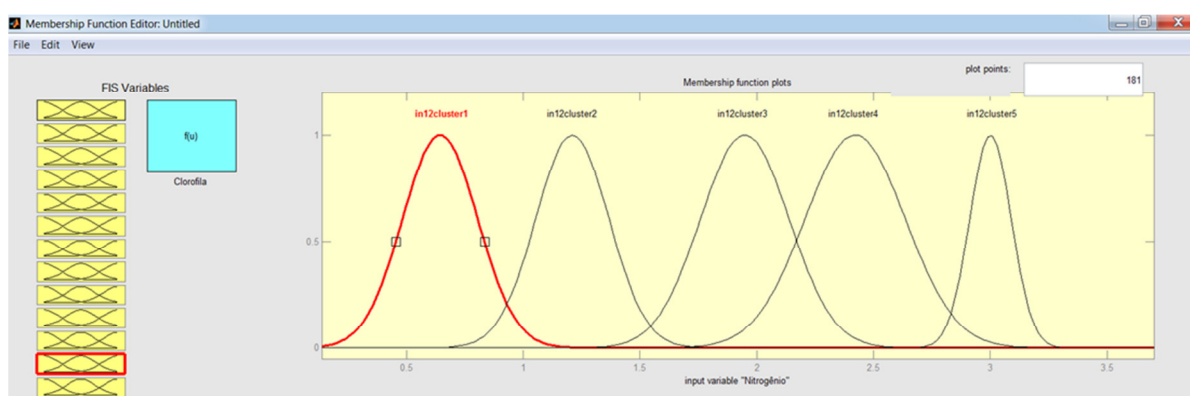


Figura 52 - Ilustração da tela do sistema ANFIS com cinco regras para cada variável de entrada

A Figura 53 ilustra o comportamento da variável de saída (clorofila) em função das variáveis de entrada concentrações de fósforo e nitrogênio. Observa-se um aumento da concentração de clorofila com a maior concentração de nitrogênio e com uma menor influência da concentração de fósforo. Esta observação pode não refletir a real relação entre

as concentrações de fósforo e de nitrogênio como influência sobre o perfil da clorofila-a no sistema hídrico; pois a relação estequiométrica entre as concentrações de nitrogênio e de fósforo, que permitiria concluir qual o nutriente limitante no processo de eutrofização não foi feita.

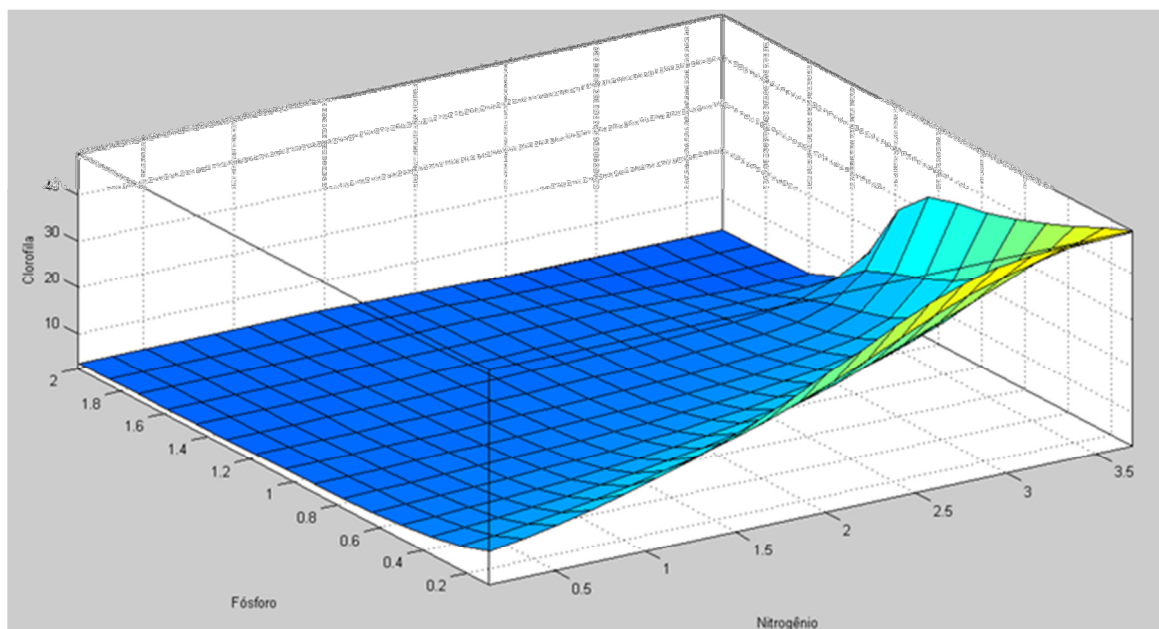


Figura 53 - Comportamento da clorofila em função das concentrações de fósforo e nitrogênio

Neste trabalho foi avaliada a influência do número de “épocas” para o treinamento e a validação do sistema ANFIS. O procedimento consistiu em gerar a melhor faixa de influência das variáveis de entrada e determinar o número de regras que melhor respondesse a simulação da concentração de clorofila-a. As Figuras 54 a 58 mostram esta avaliação para qual foi elaborada a Tabela 7. Nestas figuras são apresentadas as etapas de treinamento e validação do sistema como teste para determinação do melhor número de regras, ou seja, a melhor combinação entre os parâmetros de entrada que produz o menor erro entre a concentração de clorofila-a experimental e simulada.

As Figuras 54 a 58 correspondem aos cinco testes numéricos realizados para obtenção do número de regras que melhor representava os dados experimentais. As setas nas figuras indicam o número de regras que o programa apresentou em cada simulação.

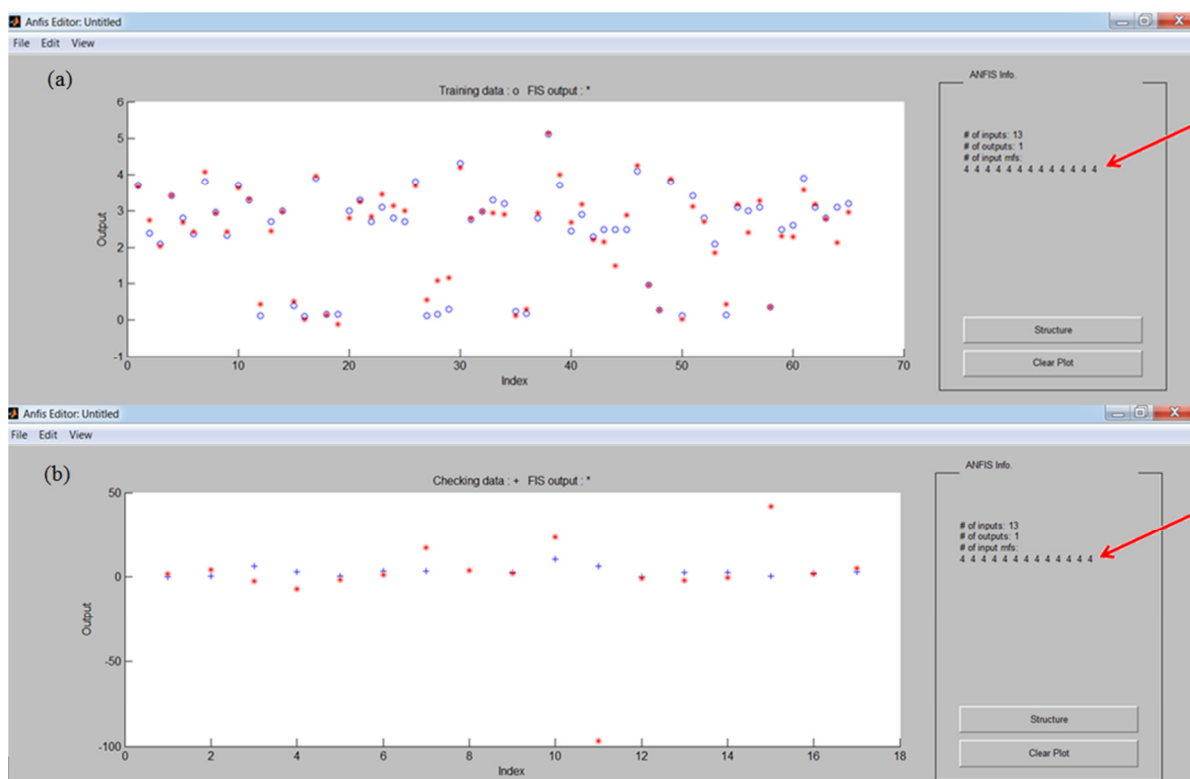


Figura 54 - Conjunto de quatro regras (a) dados para treinamento (b) dados para validação – teste 1

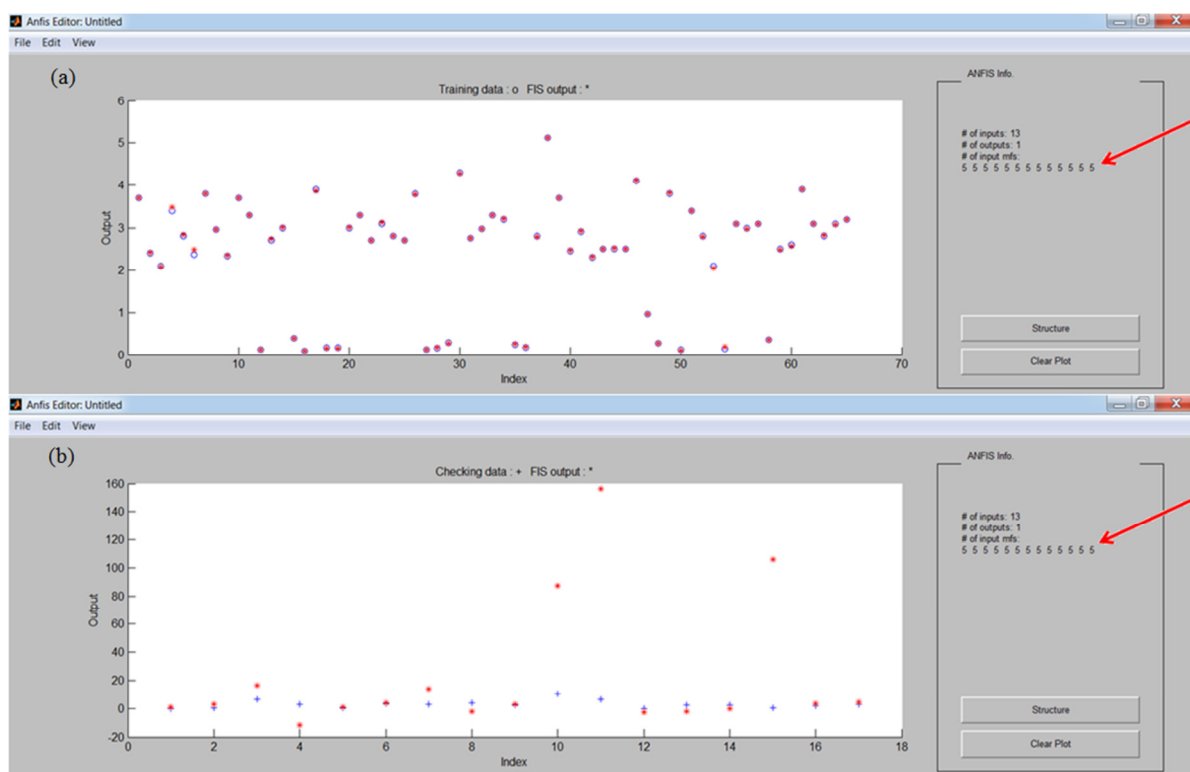


Figura 55 - Conjunto de cinco regras (a) dados para treinamento (b) dados para validação – teste 2

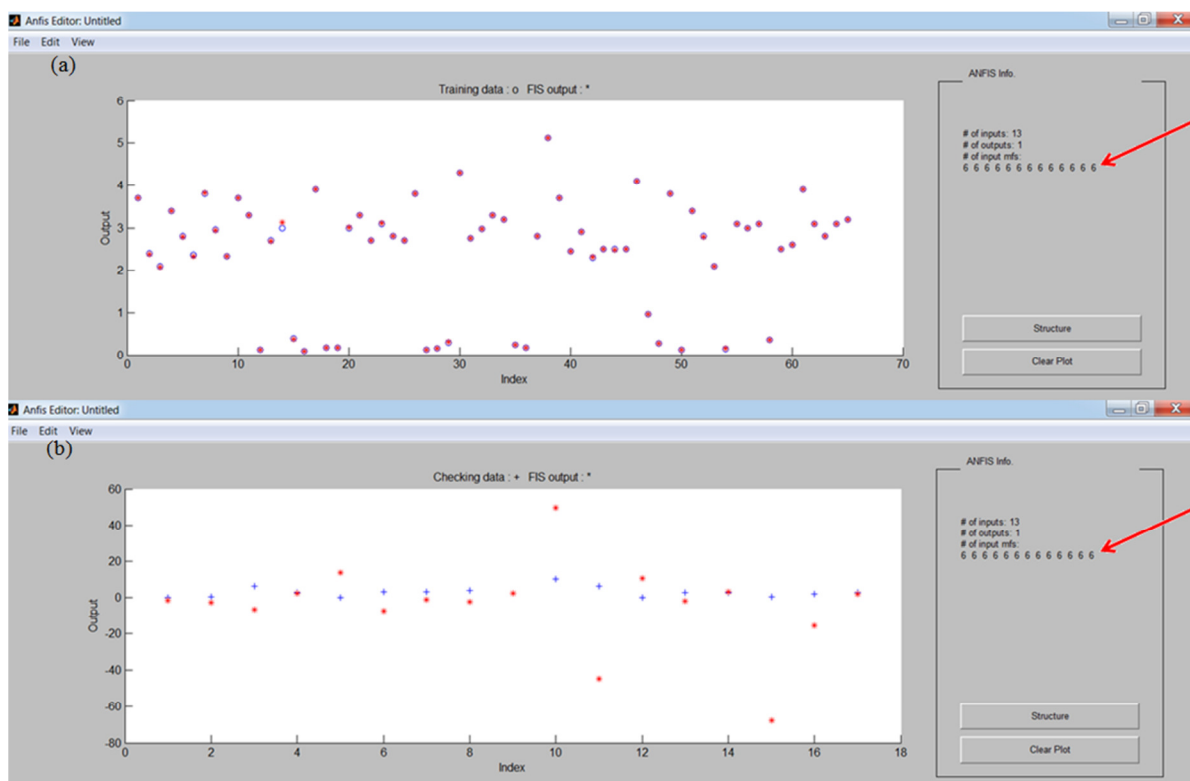


Figura 56 - Conjunto de seis regras (a) dados para treinamento (b) dados para validação – teste 3

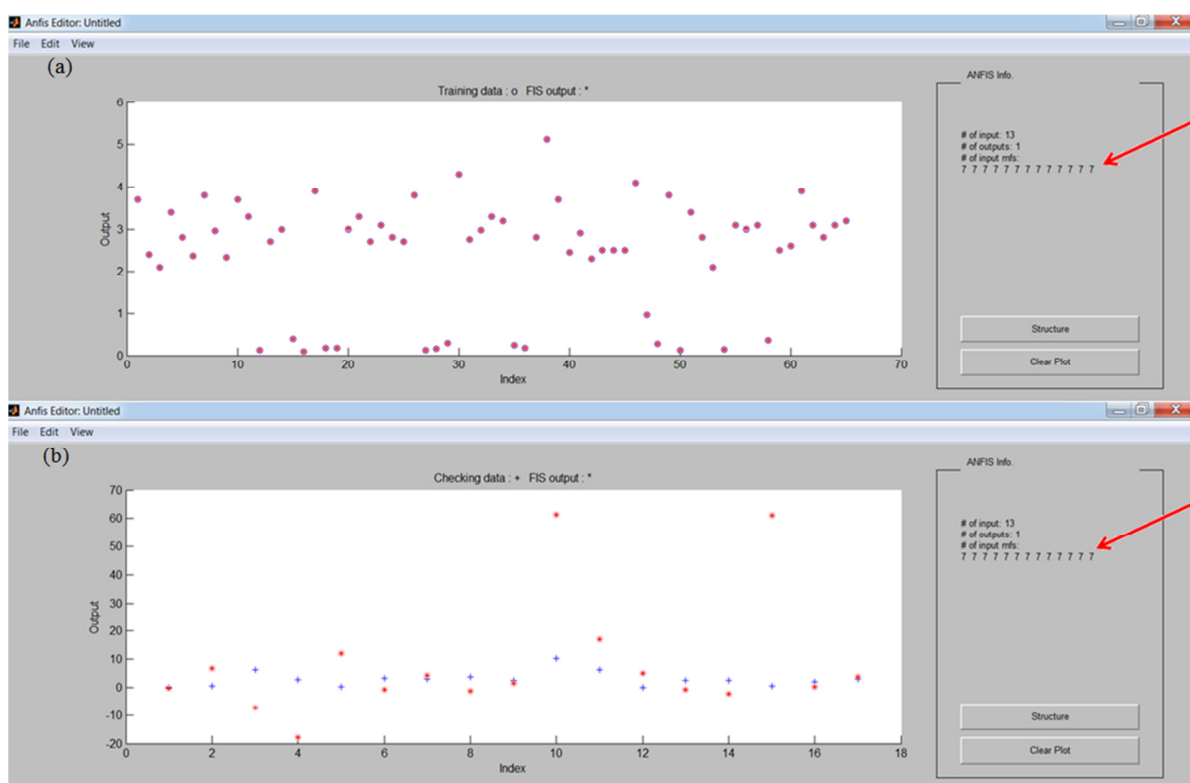


Figura 57 - Conjunto de sete regras (a) dados para treinamento (b) dados para validação – teste 4

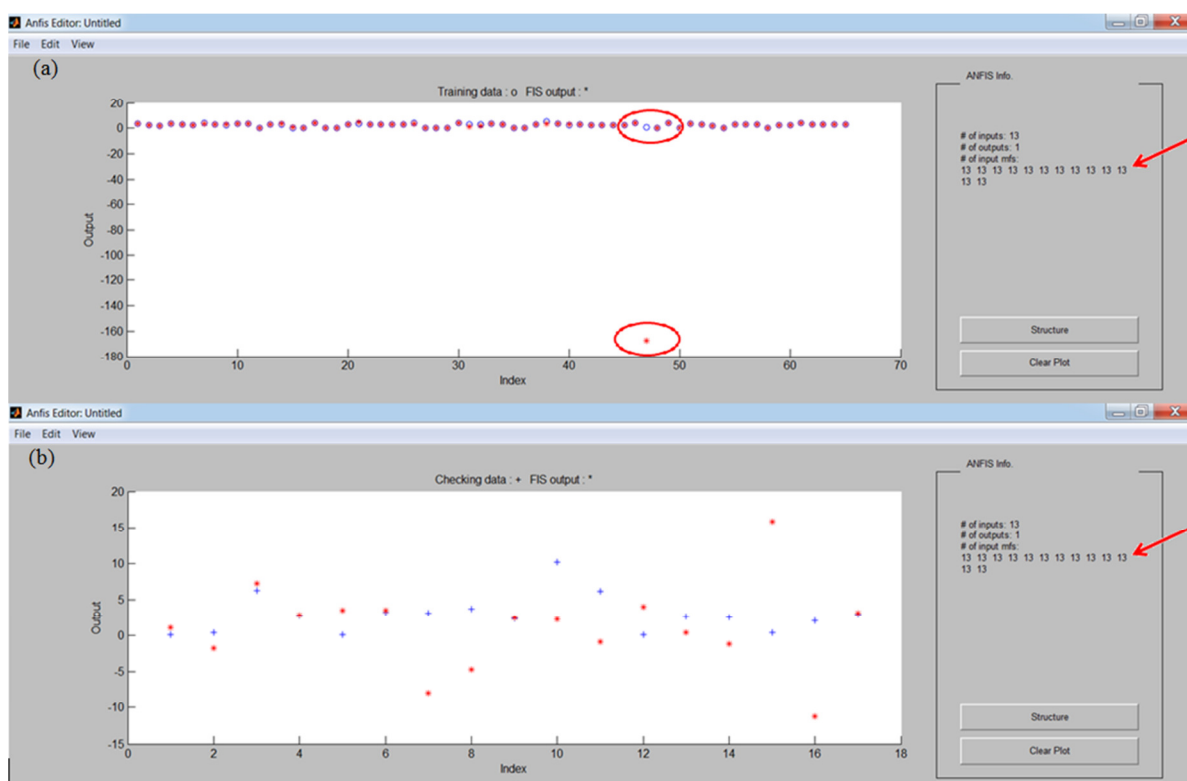


Figura 58 - Conjunto de treze regras (a) dados para treinamento (b) dados para validação – teste 5

A Tabela 10 foi elaborada a partir das informações dos testes mostrados anteriormente nas figuras 54 a 58. Nesta tabela, para cada etapa (treinamento e validação) o erro médio foi identificado correspondendo ao número de regras simulado de cada teste. Observa-se que o melhor conjunto de regras é o que apresenta seis regras, pois produziu o menor erro no treinamento. No entanto, no que se refere à validação o menor erro corresponde ao conjunto com 13 regras, o que teria consequentemente uma melhor representatividade do processo já que mostraria uma maior interação entre as variáveis de entrada. Mas, entre os erros do treinamento e da validação, foi escolhido o menor erro para o treinamento, tendo em vista que, de acordo com a literatura da área, neste conjunto concentra-se o maior número de valores experimentais para treinamento da rede (50%).

Tabela 10 - Erro médio do teste para o sistema ANFIS

Testes	Número de regras obtido	Erro médio no treinamento	Erro médio na validação
1	4	0,31010	27,6093
2	5	0,02489	48,6202
3	6	0,01596	23,9894
4	7	0,01886	20,7180
5	13	20,8566	6,74660

Definido o melhor conjunto de regras como sendo seis regras para cada parâmetro de entrada, foi obtida a Figura 59. Esta figura mostra um conjunto de dados de entrada que apresenta como resposta a concentração de clorofila igual a $8,63 \mu\text{g.L}^{-1}$. Além disso, esta figura ilustra uma tela pela qual é possível realizar a simulação da clorofila-a com novos dados de entrada. Esta simulação é realizada inserindo novos dados no campo *input* destacado (Passível de inserção). É possível ainda, nesta tabela, avaliar a influência de cada variável de entrada sobre o perfil da concentração de clorofila-a, mantendo as outras constantes.

É importante ressaltar a ausência da variável temporal, que pode ter considerável influência nos resultados, visto que o período do ano influencia no valor da clorofila-a nos reservatórios de água. Ressalta-se ainda que as variáveis de entrada in1 a in13 correspondem a temperatura da água, condutividade, cor, pH, sólidos totais, sólidos suspensos, oxigênio dissolvido, DBO, N-NH₄, N-NO₂, N-NO₃, nitrogênio total e fósforo total, respectivamente.

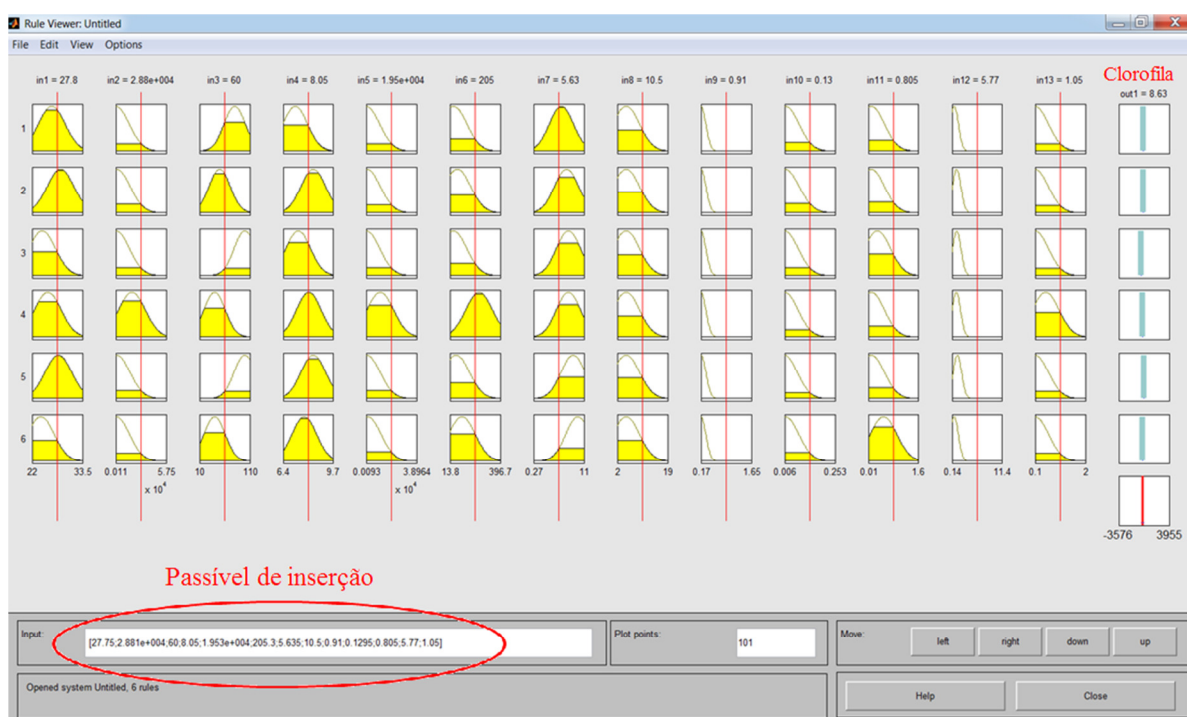


Figura 59 - Ilustração do conjunto de regras fuzzy com menor erro no treinamento.

4.3 ÍNDICE DE ESTADO TRÓFICO DA ÁGUA

A construção do índice de estado trófico foi desenvolvida a partir da utilização da primeira componente principal, utilizando a matriz de correlação das variáveis originais, para evitar que a discrepância das unidades de medidas influenciasse na determinação da direção dos novos eixos ortogonais, conforme comentado anteriormente.

4.3.1 PCA

Para esta etapa do trabalho, foi utilizado o toolbox *Statistics* do Matlab, que permite o cálculo automatizado da matriz de covariância e dos *loadings* do PCA através do comando *princomp*. Para a matriz de correlação, usou-se o comando *prestd* que efetua uma transformação nos dados de modo a normalizá-los com média nula e variância unitária, antes de usar o comando *princomp*.

A construção da interface foi feita através do GUIDE e foi dividida em duas etapas: interface principal (Figura 60) e assistente de calibração (Figura 61). O assistente de calibração serve para simplificar e automatizar a tarefa de construção do índice, cabendo ao usuário apenas a tarefa de fornecer os dados divididos de forma apropriada para as classes oligotrófica, mesotrófica, eutrófica e hipertrófica (Figura 62). É importante ressaltar que para estabelecer o índice de estado trófico, foram utilizados os dados do reservatório da Marcela e do rio Poxim.

De forma resumida, o procedimento consiste nos seguintes passos:

- a. Clicar no ícone índice da Figura 27 (Tela da Modelagem da Qualidade da Água), para obtenção da Figura 60.

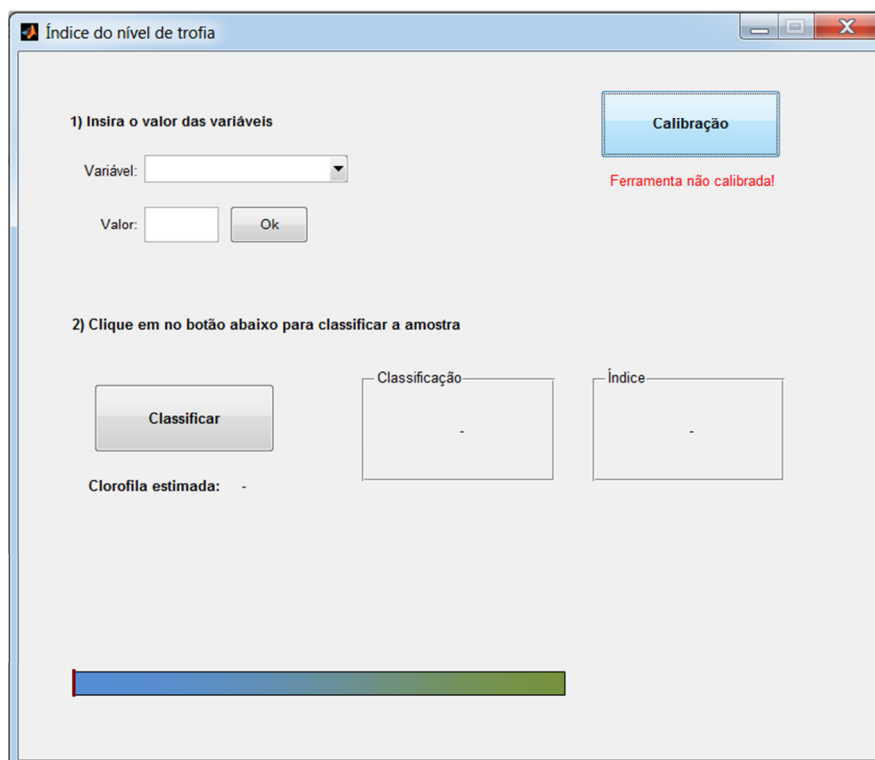


Figura 60 - Interface principal para o índice do nível de trofia do reservatório

- b. Na Figura 60, clicar no ícone Calibração para obtenção da Figura 61.

- c. Na Figura 61, clicar no ícone próximo para calibrar novos dados ou em carregar perfil para um perfil já calibrado.
- d. Clicando em próximo na Figura 61, obtém-se a Figura 62 para carregar os dados já divididos nas quatro classes tróficas, e clica-se em próximo e esperar o tempo de processamento da rede (Figura 63).
- e. Após carregar os dados, é possível inserir os valores para o conjunto de variáveis e classificar os novos dados, calcular o índice de estado trófico e estimar a concentração de clorofila-a (Figura 64).

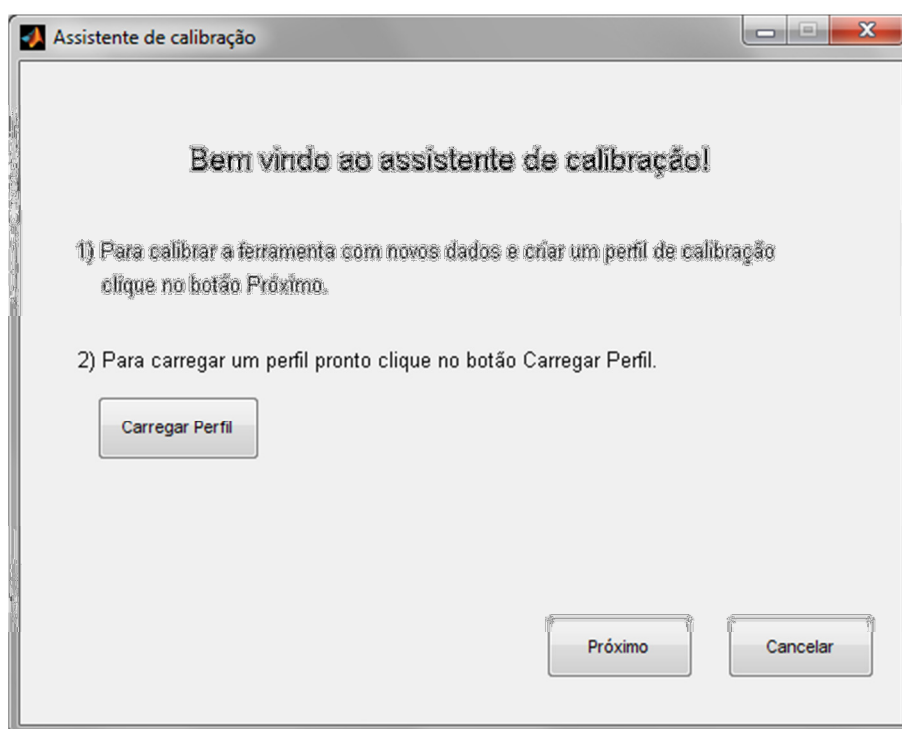


Figura 61 - Interface do assistente de calibração.

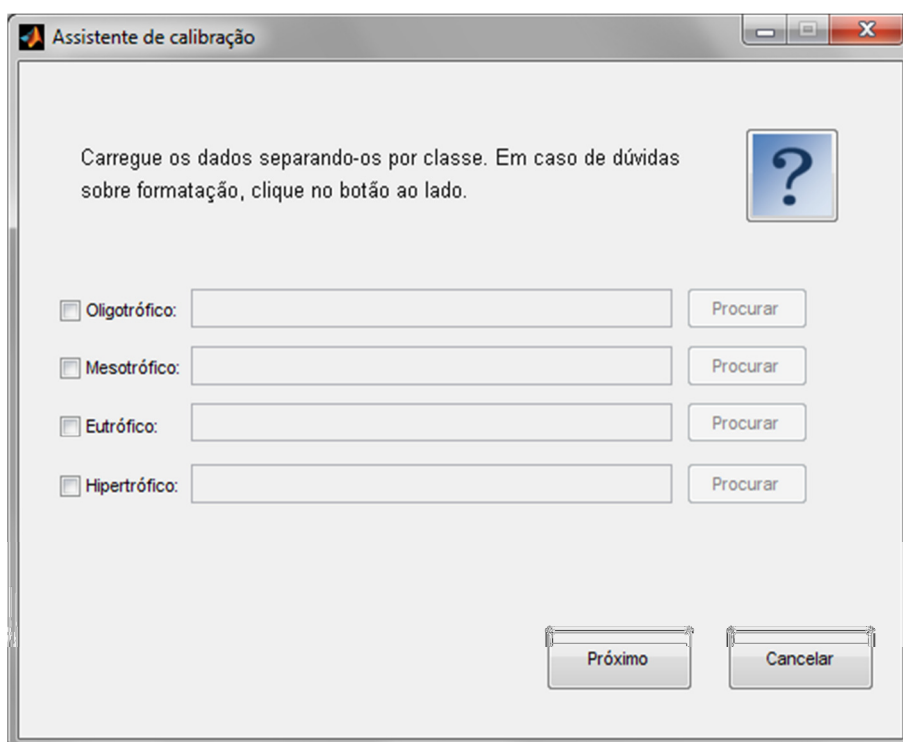


Figura 62 - Interface do assistente de calibração para o carregamento de dados.

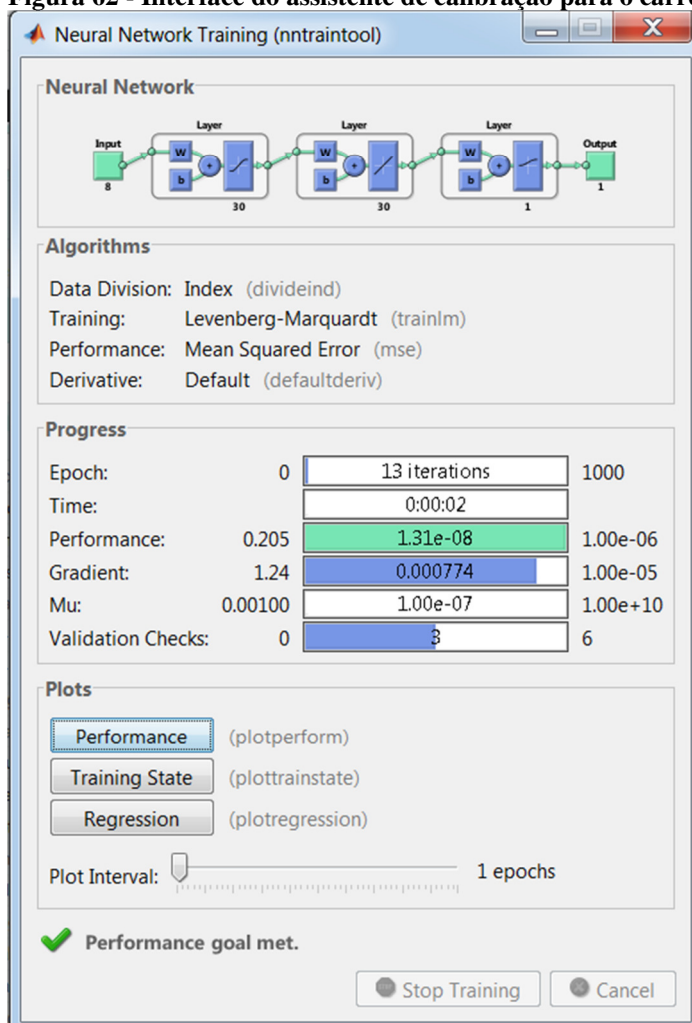


Figura 63 - Processamento da rede

Figura 64 - Classificação do sistema

É importante comentar que a ferramenta utiliza somente dados de planilhas com extensão .xls. Nas planilhas apresentadas no ANEXO C, as variáveis devem estar dispostas em colunas e a primeira linha deve conter os nomes das variáveis.

Os dados foram testados na ferramenta e a 1ª componente principal do PCA explicou 57,2% da covariância do sistema, permitindo a construção do índice que originou a Equação 69. Observa-se nessa equação que o índice aponta para um crescimento do nível de eutrofização conforme as variáveis aumentam, com exceção do oxigênio dissolvido, que diminui conforme aumenta a quantidade de algas e estas passam a consumir o oxigênio para crescer.

$$\begin{aligned} \text{I.E.} = & 0.3177[\text{pH}] - 0.2703[\text{OD}] + 0.2889[\text{N-NH}_4] + 0.2372[\text{N-NO}_2] + 0.3663[\text{N-NO}_3] \\ & + 0.3991[\text{P-PO}_4] + 0.3218[\text{Ntotal}] + 0.3811[\text{Ptotal}] + 0.3794[\text{Cl-a}] \end{aligned} \quad (69)$$

A análise de distribuição de frequências mostrou que o ponto de separação entre as classes é de 1,4157. Isso significa que amostras com índice superior a esse valor serão consideradas hipertróficas e abaixo desse valor, oligotróficas.

A Figura 65, por meio de uma representação biplot entre as componentes 1 e 2, mostra a separação das classes oligotrófica e hipertrófica, correspondente aos dados do rio Poxim e da Bacia de Marcela, respectivamente, utilizados nesta etapa do trabalho.

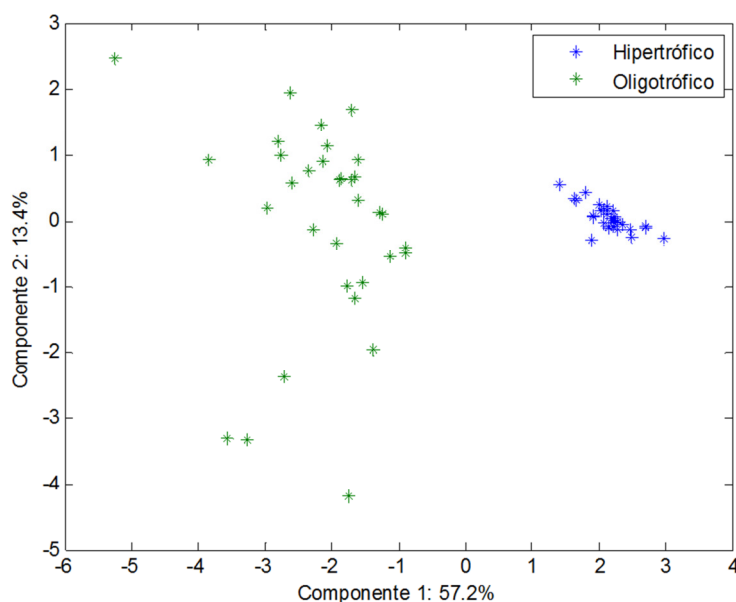


Figura 65 - Gráfico biplot da componente 1 pela componente 2 - PCA

Ao classificar os mesmos dados utilizados para a construção do índice, verificou-se um acerto de 100% na classificação. A visível separação das classes observadas NA Figura 65 foi o que permitiu a relativa facilidade do na classificação dos sistemas hídricos.

Desse modo, a ferramenta desenvolvida neste trabalho conseguiu classificar de modo satisfatório os dados de águas oligotróficas e hipertróficas do rio Poxim e do reservatório da Marcela, respectivamente.

4.3.2 Lógica Fuzzy

Com base no conhecimento dos sistemas, as funções das variáveis de entrada e da variável de saída foram definidas utilizando o Fuzzy Logical Toolbox do MATLAB. Nesta etapa, o método de inferência fuzzy foi o método de Mamdani e para a etapa de defuzzificação foi utilizado o método do centro de gravidade (*Centroid*), conforme se identifica, em vermelho) na Figura 66. A função de pertinência trapezoidal foi utilizada para todas as variáveis.

Para implementação da estratégia fuzzy foram utilizados os dados experimentais o reservatório de Jacarecica e considerou-se que:

- a. As três variáveis de entrada foram definidas a partir dos resultados da aplicação do modelo de redes neurais e da avaliação do índice de qualidade da água, como aquelas que mais influenciam na qualidade da água, segundo especialistas: clorofila-a, fósforo e nitrogênio.

- b. A variável de saída foi o índice de estado trófico da água, classificando o sistema em quatro níveis tróficos do sistema.
- c. A função de pertinência triangular foi a mesma para cada variável de entrada e para a variável de saída.
- d. Foram implementadas 23, regras fuzzy de acordo com o grau de importância, ou de pertinência, para a resposta do sistema, em conformidade com as definições dos especialistas.

A Figura 65 mostra a tela obtida no MATLAB para identificação das três variáveis de entrada e da variável de saída.

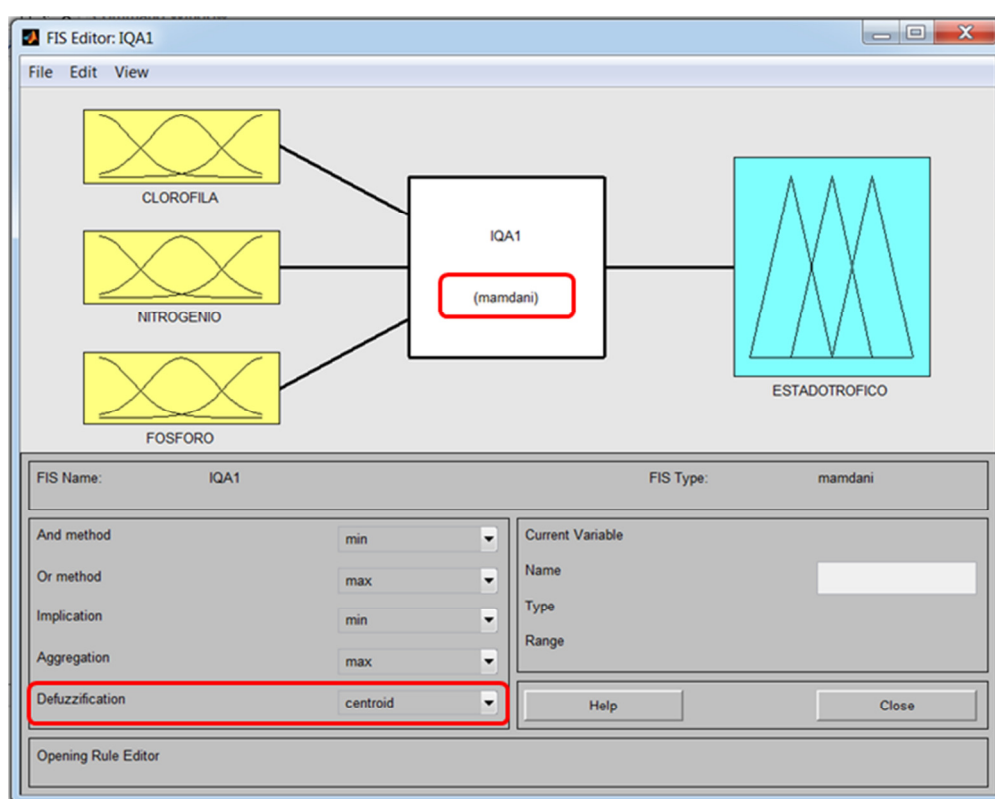


Figura 66 - Tela principal para índice de estado trófico da água através da lógica fuzzy

As Figuras 67, 68 e 69 apresentam as condições *fuzzy* implementadas para as variáveis de entrada do modelo. Para definir as classes linguísticas foram calculados o índice de estado trófico (IET) através das equações mostradas na Tabela 1 e as faixas de trofia apresentadas na Tabela 2 no item 2.2.1.2 da fundamentação.

A partir das figuras 67 a 69 é possível inferir sobre o estado trófico do sistema em relação a cada parâmetro ambiental. Em decorrência dessa inferência, obtém-se a Figura 70

que mostra essa combinação entre as concentrações de clorofila-a, de nitrogênio e de fósforo para se definir o estado trófico do sistema.

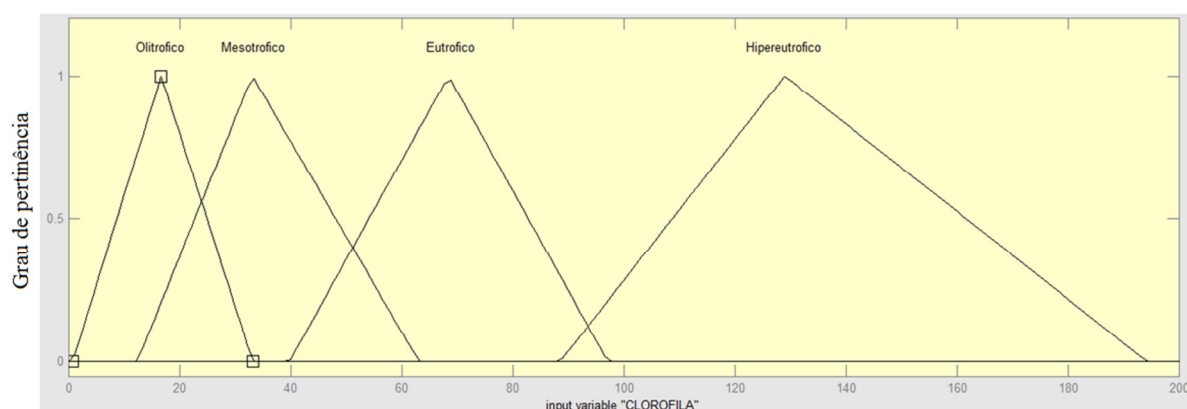


Figura 67 - Condições fuzzy – variável de entrada concentração de clorofila-a (µg/L)

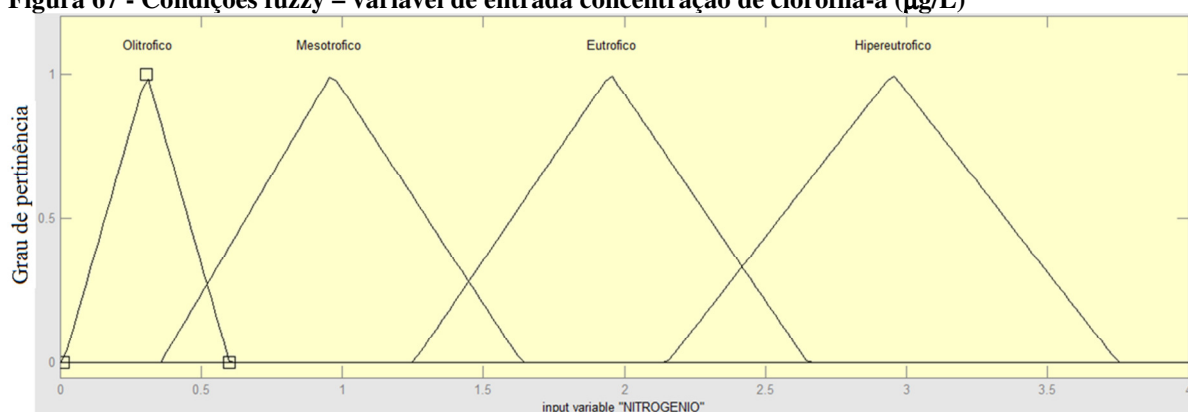


Figura 68 - Condições fuzzy – variável de entrada nitrogênio

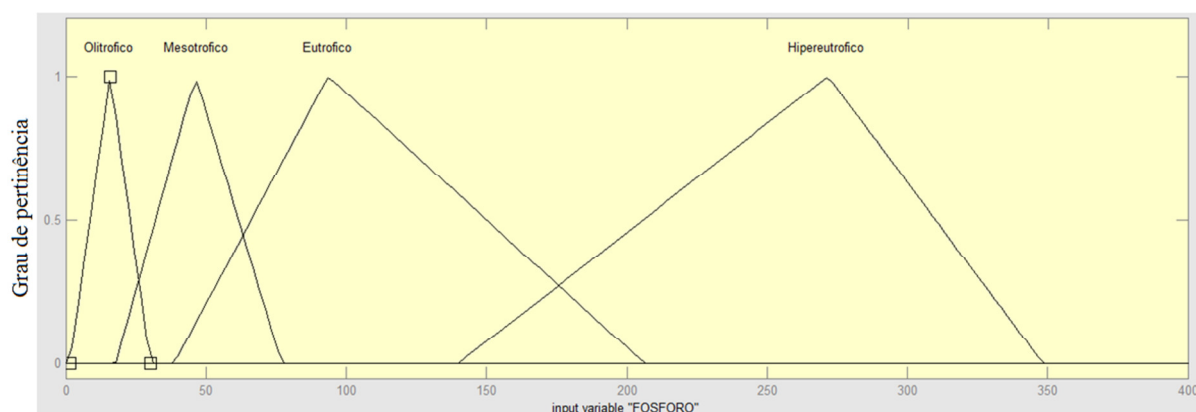


Figura 69 - Condições fuzzy – variável de entrada fósforo

A Figura 70 mostra, então, o conjunto de saída para o estado trófico do sistema hídrico. Para os subconjuntos deste índice, tem-se que o subconjunto OLIGOTRÓFICO é representado pela terna (0,00414; 0,1141; 0,234); o subconjunto MESOTRÓFICO é representado pela terna (0,0959; 0,26; 0,4013); o subconjunto EUTRÓFICO é representado pela terna (0,3395; 0,4205; 0,5095); e o subconjunto HIPEREUTRÓFICO é representado

pela terna (0,444; 0,626; 0,85). É importante destacar, então, que acima de 0,85 o sistema poderia ser classificado como hiperhipereutrófico, podendo ser um novo termo linguístico a ser aplicado.

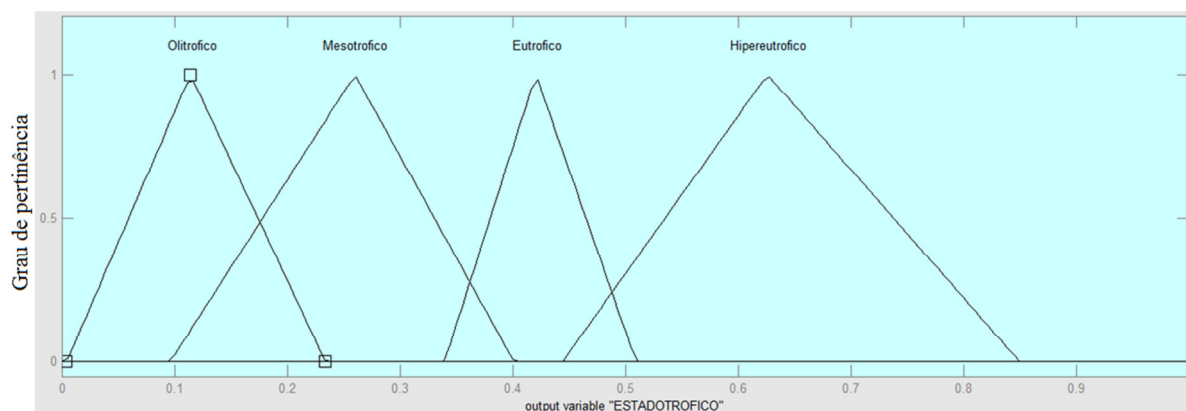


Figura 70 - Condições fuzzy – variável de saída IET

A Figura 71 apresenta o conjunto de 23 regras *fuzzy* aplicado ao sistema hídrico. Diferentemente do sistema ANFIS (neuro-fuzzy), esse conjunto de regras foi definido para ser inserido no programa. É possível observar na Figura 49 que para cada valor atribuído às variáveis de entrada, por exemplo, clorofila 80,5 $\mu\text{g.L}^{-1}$, nitrogênio 2 mg.L^{-1} e fósforo 200 mg.L^{-1} , como resposta o sistema gera IET de 0,747 que representa 74,7 % de segurança na resposta referente às condições do reservatório. Na literatura, uma resposta acima de 70% é bom grau de acerto na análise da variável de saída (MELO, 2009). É importante ressaltar que quanto maior o número de variáveis de entrada maior o número de regras, o que torna o sistema mais dependente do conhecimento dos especialistas na integração entre as regras. Observa-se, então, que nesta simulação hipotética, o IET igual a 0,747 indica que o sistema pode ser enquadrado no termo linguístico hipereutrófico, conforme mostra anteriormente a Figura 70.

É importante comentar que em termos limnológicos, esses resultados mostram que o reservatório de Jacarecica I, apresenta características de sistemas oligotróficos a eutróficos, tornando-se mais eutrofizado logicamente a depender do período da coleta. Além disso, pela análise dos índices de estado trófico e de seus desvios, o nutriente limitante do reservatório no período de estiagem é o nitrogênio e, durante o período chuvoso, a limitação passa a ser do fósforo.

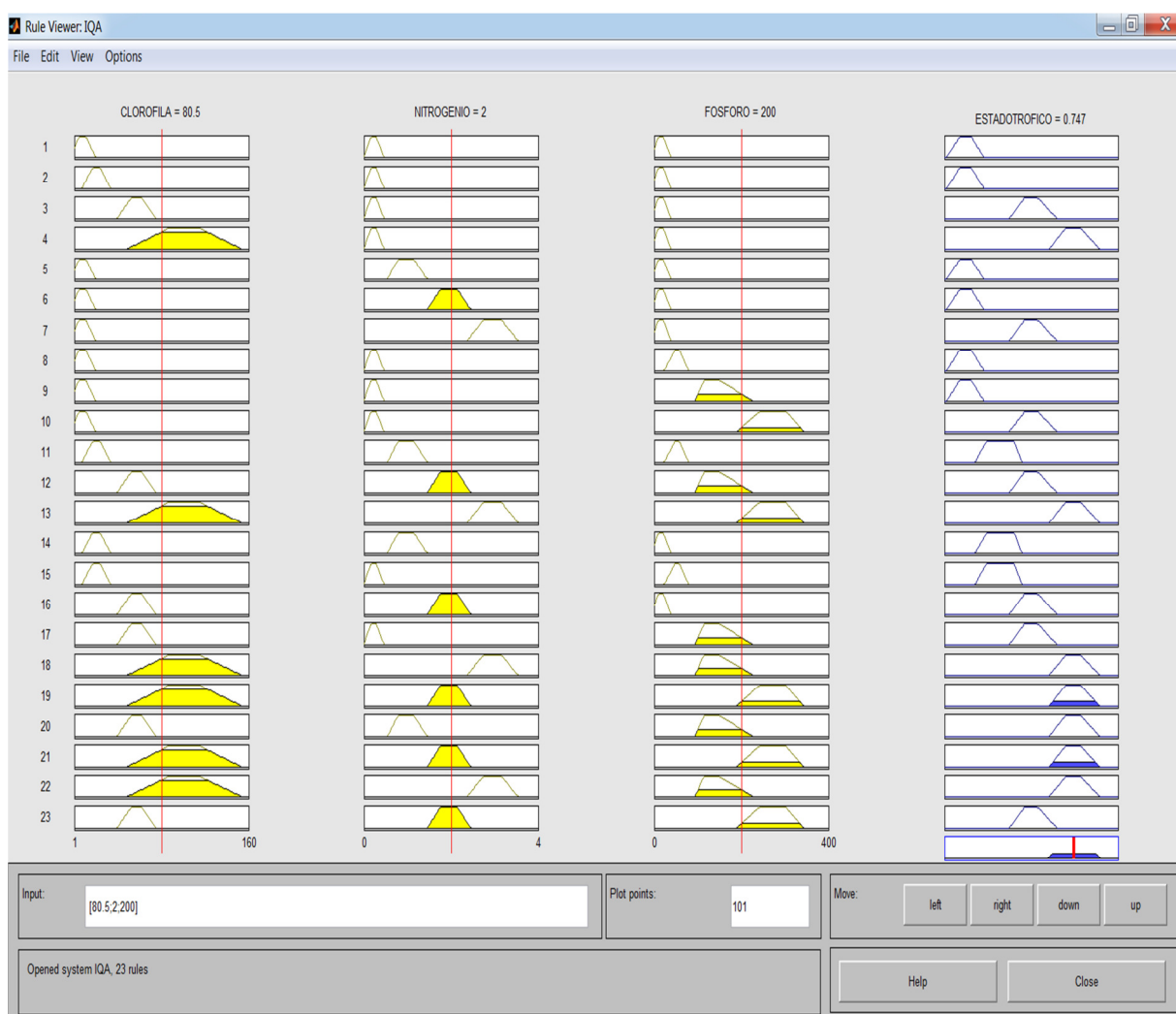


Figura 71 - Regras fuzzy aplicadas

5 CONCLUSÕES

Neste trabalho, foram avaliadas as condições de eutrofização dos reservatórios da Marcela e de Jacarecica, e do rio Poxim em Sergipe através do desenvolvimento de estratégias numéricas e construção de ferramentas computacionais utilizando redes neurais, lógica fuzzy e o sistema neuro-fuzzy.

Para essa avaliação ambiental foram analisadas 226 amostras coletadas durante o monitoramento dos reservatórios no Laboratório de Química Analítica Ambiental (LQA) da Universidade Federal de Sergipe e 70 dados cedidos pela Secretária de Recursos Hídricos de Sergipe (SRH-SE). Cabe, neste momento, ressaltar a dificuldade na obtenção de dados sobre a qualidade da água, principalmente em relação à quantidade de dados exigidos para implementação de ferramentas computacionais.

A avaliação dos resultados para os reservatórios foi desenvolvida através da aplicação de redes neurais do tipo MLP e RBF e um sistema neuro-fuzzy. Para as redes MLP e RBF o ajuste para os dois reservatórios permitiu concluir que o processo de eutrofização pode ser modelado com razoável precisão com redes neurais, ou seja, os resultados mostram que essas arquiteturas de redes neurais foram capazes de verificar com razoável precisão a tendência dos dados experimentais simulando a concentração de clorofila-a.

Para os dados disponíveis, a lógica fuzzy permitiu que o conhecimento empírico dos especialistas fosse corroborado matematicamente e foi possível o utilizar o IET como parâmetro de gestão indicando a influência dos problemas de eutrofização dos corpos hídricos quando o grau de pertinência fosse, em termos linguísticos, superior ao estado de mesotrofia do sistema. Além disso, a classificação do IET foi desenvolvida com base nos valores de concentrações de fósforo, nitrogênio e clorofila-a e os resultados fuzzy mostraram uma combinação satisfatória entre os indicadores ambientais que classificam o sistema entre níveis oligotrófico e hipereutrófico a depender da inferência dos dados. Neste estudo, 23 regras foram aplicadas e constatou-se que o rio Poxim, nos pontos de coleta, foi classificado como oligotrófico e os reservatórios da Marcela e de Jacarecica foram classificados como sendo hipereutrófico e oligotrófico, respectivamente.

Em relação ao sistema neuro-fuzzy, apesar desse sistema combinar as vantagens da lógica fuzzy e das redes neurais, uma limitação da técnica é o número de variáveis de entrada que deve ser o menor possível para que não haja uma elevada combinação de regras. Neste sentido, no presente trabalho foram utilizados treze parâmetros, ou seja, um número

de variáveis de entrada considerado elevado, promovendo assim, matematicamente, um maior esforço computacional.

Em termos limnológicos, considerando os limites estabelecidos para estes parâmetros analisados para qualidade da água de acordo com a Resolução nº 375/05 de Conama, a água dos reservatórios e do rio Poxim foi considerada de boa qualidade. Os parâmetros de qualidade da água utilizados numericamente neste trabalho (como por exemplo, sólidos suspensos, sólidos totais dissolvidos, oxigênio dissolvido, condutividade elétrica, concentrações de amônia, de fósforo, de nitrogênio e de clorofila) servem como indicadores de qualidade da água, visando o seu monitoramento ambiental, possibilitando, por exemplo, inferir sobre o uso e ocupação do solo da região e a construção de um índice de qualidade de água.

Neste sentido, os resultados mostram que é possível avaliar o fenômeno de eutrofização de corpos hídricos através do desenvolvimento de estratégias e construção de ferramentas computacionais, em particular as redes neurais e a lógica fuzzy, utilizando os principais parâmetros limnológicos, mostrando a viabilidade numérica no que concerne a representação de fenômenos ambientais complexos e importantes para sustentabilidade ambiental dos corpos hídricos.

5.1 LIMITAÇÕES

Atualmente, os sistemas de gestão possibilitam propor modelos de melhoria através da indicação dos pontos fracos de um processo/projeto. Sendo assim, neste trabalho, os pontos fracos na análise do fenômeno de eutrofização estão relacionados com a obtenção dos dados experimentais que foram utilizados na simulação. O elevado custo das campanhas de coleta e a heterogeneidade dos parâmetros a serem mensurados dificultaram a quantidade e qualidade dos dados experimentais, constituindo as principais limitações no desenvolvimento da presente tese.

Como consequência dessas limitações, o número de dados experimentais desta tese foi considerado pequeno quando comparado com alguns trabalhos apresentados na literatura.

5.2 SUGESTÃO DE TRABALHOS FUTUROS

Como consequência das limitações da presente tese, tem-se a necessidade de se desenvolver trabalhos futuros no sentido de se aplicar a utilização de ferramentas de

Inteligência Artificial na análise de fenômenos complexos com maior desempenho e com melhores resultados. Sendo assim, os seguintes trabalhos são sugestões consubstanciadas nos resultados da tese e na análise do fenômeno de eutrofização:

1. Aplicação do Índice de Qualidade da Água (IQA) e do Índice de Estado Trófico para elaboração de um modelo de gestão ambiental.
2. Customização dos sistemas fuzzy e neuro-fuzzy no ambiente MATLAB para representação do fenômeno de eutrofização.

REFERÊNCIAS

ABBASI, T.; ABBASI, S. A.. **Water Quality Indices**. UK: Elsevier, 2012.

AKKOYUNLU, A.; AKINER, M.. **Pollution evaluation in streams using water quality indices: A case study from Turkey's Sapanca Lake Basin**. Ecological Indicators v 18 p 501–511, 2012.

ALMEIDA, P. E. M.; EVSUKOFF, A. G. **Sistemas fuzzy**. In: REZENDE, S. O. (Org.). Sistemas inteligentes – fundamentos e aplicações. São Paulo: Manole, 2003, apud LANDMANN, R.; ERDMANN, R. H.. **Uma abordagem heurística para a programação da produção na indústria de fundição com utilização da lógica fuzzy**. Gestão & Produção, São Carlos, v. 18, n. 1, p. 119-130, 2011.

ALTUNKAYNAK, A.; OZGER, M.; ÇAKMAKCI, M.. **Fuzzy logic modeling of the dissolved oxygen fluctuations in Golden Horn**. Ecological Modelling, 189, 436-446, 2005.

ALVES, J. P. H. ; GARCIA, C. A. B. . **O rio Sergipe no entorno de Aracaju: qualidade da água e poluição orgânica**. In: José do Patrocínio Hora Alves. (Org.). Rio Sergipe: Importância, vulnerabilidade e preservação. 1 ed. Aracaju: v. 1, p. 87-109, 2006.

American Public Health Association (APHA). **Standard Methods for the examination of water and wastewater**. 20 ed. Washington: APHA, 1998.

ANGULO, C.; CABESTANY, J.; RODRIGUEZ, P.; BATLLE, M.. **Fuzzy expert system for the detection of episodes of poor water quality through continuous measurement**. Expert Systems with Applications, 39, 1011-1020, 2012.

ARAÚJO, K. A. M; VITTI, B. C.; LOTUFO, A. D. P.; LOPES, M. L. M.. **Previsão de aflúências utilizando redes neurais artificiais com o auxílio das TOOLBOXES do MATLAB**. In: XVII Congresso Brasileiro de Automática, Bonito-MS, 2010.

BARBALHO. V. M. S.. **Sistemas Baseados em Conhecimento e Lógica Difusa para Simulação do Processo Chuva-Vazão**. Tese de Doutorado. COPPE/UFRJ, 2001.

BRAGA, A. P.; CARVALHO, A. C. P. L.; LUDERMIR, T. B.. Redes Neurais Artificiais – Teoria e Prática, Rio de Janeiro: LCT. 2007.

BRASIL. Agência Nacional de Água. **Conjuntos dos Recursos Hídricos no Brasil – Informe 2012**. <http://arquivos.ana.gov.br/imprensa/arquivos/Conjuntura2012.pdf> acesso em junho de 2012.

BRASIL. Ministério do Meio Ambiente. Conselho Nacional do Meio Ambiente – CONAMA **Resolução nº 430**, de 13 de maio de 2011. Dispõe sobre as condições e padrões de lançamento de efluentes, complementa e altera a Resolução nº 357, de 17 de março de 2005, do Conselho Nacional do Meio Ambiente-CONAMA. Disponível http://www.mma.gov.br/port/conama/res/res11/propresol_lanceflue_30e31mar11.pdf.

Acesso em: 17 de setembro de 2012.

BROOMHEAD, D.S. e LOWE, D.. **Multivariate functional interpolation and adaptive networks**. Complex Systems, 2: 321-355, 1988.

ÇAMDEVÝREN H.; DEMÝR, N.; KANIK, A.; KESKÝN, S.; **Use of principal component scores in multiple linear regression models for prediction of Chlorophyll – a in reservoirs**. Ecological Modelling, 181, 581-589, 2005.

CAMPOS FILHO. Pio. **Método para apoio à decisão na verificação da Sustentabilidade de uma Unidade de Conservação, usando Lógica Fuzzy**. 2004, 210f. Dissertação (Doutorado em Engenharia de Produção). Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina, Florianópolis, 2004.

CAMPOS, M. C. M. M., SAITO, K. **Sistemas Inteligentes em controle e automação de processos**. Rio de Janeiro: Editora Ciência Moderna, 2004, 235p.

CAMPOS, M. L. A. M.. **Introdução à Biogeoquímica de ambientes aquáticos**. Campinas: Ed. Átomo. 2010.

CARLSON, R. E. e SIMPSON, J. **A coordinator's guide to volunteer lake monitoring methods**. North American Lake Management Society, p. 96, 1996.

CARLSON, R. E.. **A trophic state index for lakes** Limnology and Oceanography, v. 22, n. 2, p. 361-369, 1977.

CARLSSON, B.; LINDBERG, C. F.. **Some control strategies for the activated sludge process.** <http://www.cheric.org/ippage/p/ipdata/2000/07/file/control-Ofwwt.pdf>. 1998. Acesso em 26 de maio de 2011.

CARVALHO, F. O., GARCIA, C. A. B., GARCIA, H. L., ALVES, J. P. H., PIMENTEL, W. R. O., SILVA, J. L.. **Modelagem do processo de eutrofização no reservatório Jacarecica II – Sergipe/Brasil utilizando redes neurais artificiais e estatística multivariada** In: Congresso Ibero Latino-Americano sobre Métodos Computacionais em Engenharia - CILAMCE 2007, Porto, 2007.

CARVALHO, F. O.; COELHO, F. A.; SILVA, A. N.; GARCIA, H. L.; SILVA, V. L.; GARCIA, C. A. B.. **Redes Neurais e Estatística Multivariada na qualidade da água.** In: XVIII Congresso Brasileiro de Engenharia Química, Foz do Iguaçu, 2010.

CARVALHO, F. O.; GARCIA, C. A. B.; GARCIA, H. L.; COELHO, F. A.; WANDERLEY, H. S.; SANTOS, R. R. C.; CAVALCANTI, J. R. A.. **Sistemas Inteligentes (Redes Neurais) e Estatística Multivariada (PCA) aplicados a qualidade da água.** In: XVIII Simpósio Brasileiro de Recursos Hídricos, Campo Grande, 2009.

CAVENAGHI, A.L.; VELINI, E.D.; GALO, M.L.B.T.; CARVALHO, F.T.; NEGRISOLI, E.I; TRINDADE, M.L.B.; SIMIONATO, J.L.A.. **Caracterização da qualidade de água e sedimento relacionados com a ocorrência de plantas aquáticas em cinco reservatórios da Bacia do rio Tietê.** Planta Daninha, Viçosa, v. 21, p. 43-52, 2003.

CHACON-TORRES, A. e ROSAS-MONGE, C. **Water quality characteristics of high altitude oligotrophic Mexican lake.** Aquatic Ecosystem Health & Management, v. 1, p. 237-243, 1998.

CHAU, K. W.. **A review on integration of artificial intelligence into water quality modelling.** Marine Pollution Bulletin, 52, 726–733, 2006, *apud* WEST, D.; DELLANA, S.. **An empirical analysis of neural network memory structures for basin water quality forecasting.** International Journal of Forecasting, 27, 777–803, 2011.

CHAVES, P. E KOJIRI, T.. **Deriving reservoir operational strategies considering water quantity and quality objectives by stochastic fuzzy neural networks.** Advances in Water Resources, 30, 1329–1341, 2007.

COMPANHIA DE TECNOLOGIA E SANEAMENTO AMBIENTAL – CETESB. **Relatório de qualidade das águas interiores do estado de São Paulo 2008** / CETESB. São Paulo: Série Relatórios-Secretaria de Estado do Meio Ambiente, 2009. v. 1, 472 p. Relatório anual. Disponível em <http://www.cetesb.sp.gov.br/agua/aguas-superficiais/35-publicacoes/-relatorios>. Acesso em 27 de agosto de 2012.

DALTRO FILHO, J.; SANTOS, D.C.G. **Avaliação da qualidade da água do rio Poxim, no entorno da captação da deso, na Grande Aracaju - Sergipe – Brasil**. VI Simpósio Ítalo Brasileiro de Engenharia Sanitária e Ambiental, 2002.

DEMUTH, H., BEALE, M., **Neural Network Toolbox – For use with Matlab**, version 4, The MathWorks Inc., 2002.

DESPAGNE F., MASSART, D. L.. **Variable selection for neural networks in multivariate calibration**. Chemometrics and Intelligent Laboratory Systems 40, 145-163, 1998.

DIMBERG, P. H.; HYTTEBORNA, J. K.; BRYHNB, A. C.. **Predicting median monthly chlorophyll-a concentrations**. Limnologica. In Press, 2012.

ESTEVES, F. A.. Fundamentos de Limnologia. 3ª Ed. Rio de Janeiro: Interciência. 2011.

FERREIRA, M. S.. **Inferência espacial de clorofila a por redes neurais artificiais aplicadas a imagens multiespectrais e medidas tomadas in situ**. Dissertação DE Mestrado. Universidade Estadual Paulista. 2011.

FONSECA, O. L. H. Aplicação de métodos de análise espacial e da teoria dos conjuntos nebulosos em estudos sobre pobreza. 2003 149f. Dissertação (Mestrado em Engenharia de Computação). Faculdade de Engenharia da Universidade do Estado do Rio de Janeiro – UERJ, Rio de Janeiro, 2003.

FREITAS, S. S. **Eutrofização no reservatório Marcela em Itabaiana –SE, e suas implicações ambientais**. Universidade Federal de Sergipe, Monografia – Especialização em Gestão de Recursos Hídricos e Meio Ambiente, 50 p., 2001.

GARCIA, H. L.; ALVES, J. P. H.; CARVALHO, F. O.; GARCIA, C. A. B.. **Modelagem de processos de eutrofização em reservatórios usando a técnica de redes neurais** In: XV

Congresso Brasileiro de Engenharia Química – XV COBEQ – Curitiba, 2004a.

GARCIA, H. L., CARVALHO, F. O., PEREIRA, J. A F R, SANTOS, J. P. L. **Utilização de redes neurais artificiais na solução de modelos neurais-híbridos de não-equilíbrio para processos de absorção com reação química complexa** In: CILAMCE 2004, Recife, 2004b.

GARG, S.; PAL, S. K.; CHAKRABORTY, D. (2007). **Evaluation of the Performance of Backpropagation and Radial Basis Function Neural Networks** in Predicting the Drill Flank Wear. *Neural Comput & Applic*, Vol. 16, pp. 407-417, *apud* ARAUJO, K. A. M; VITTI, B. C.; LOTUFO, A. D. P.; LOPES, M. L. M.. **Previsão de aflúências utilizando redes neurais artificiais com o auxílio das TOOLBOXES do MATLAB**. In: XVII Congresso Brasileiro de Automática, Bonito-MS, 2010.

GOMIDE, F. A. C.. **Computação com Regras**. Departamento de Engenharia da Computação e Automação Industrial (DCA/UNICAMP). Notas de Aulas, 2007. Disponível em http://www.dca.fee.unicamp.br/~gomide/courses/IA861/transp/Notas_IA_861_Computacao_com_Regras.pdf. Acesso em 15 de junho de 2010.

GOMIDE, F., ROCHA, A. **Neurofuzzy Components Based on Threshold** – IFAC SICICA Symposium, Malaga, Espanha, pp.425-430, 1992.

GUEDES, T. L.; ROCHE, K. F.; MARTINS, R. C.; GUIMARÃES, J. R.. **Monitoramento da Qualidade da Água e do Processo de Eutrofização do Reservatório “Lago do Amor” (Campo Grande, MS)** XIX Simpósio Brasileiro de Recursos Hídricos. 2011.

GULATI, R. D. e VAN DONK, E. **Lakes in Netherlandds, their origin, eutrophication and restoration: state-of-art review**. *Hydrobiologia*, v. 478, p. 73-106, 2002.

HAGAN, M. T. e MENHAJ, M. B., **Training Feedforward Networks with the Marquardt Algorithm**, *IEEE Transactions on Neural Networks*, V.5 (6), p.989 ,1994.

HAGAN, M. T., FORESSE, F. D.. **Gauss-Newton Approximation to Bayesian Learning**. In: *Proceedings of the 1997 International Joint Conference on Neural Networks*, URL: <http://elecengr.okstate.edu/mhagan>, 1997.

HAN, H.; CHEN, Q.; QIAO, J.. **An efficient self-organizing RBF neural network for water quality prediction.** *Neural Networks*, v. 24, p. 717-725, 2011.

HAVENS, K. E. **Secondary nitrogen limitation in a subtropical lake impacted by non-point source agricultural pollution.** *Environmental Pollution*, v. 89, n. 3, p. 241-246, 1995.

HAVENS, K. E., *et al.* **Nutrient dynamics and the eutrophication of shallow lakes Kasamigaura (Japan), Donghu (PR China), and Okeechobee (USA).** *Environmental Pollution*, v. 111, p. 263-272, 2001.

HAYKIN, S., **Redes Neurais, Princípios e Práticas**, 2. ed., Bookman, Porto Alegre (2001).

IBÁÑEZ, M. S. R., *et al.* **Limnological characteristics of three aquatic systems of the pre-amazonian floodplain, Baixada Maranhense (Maranhão, Brazil).** *Aquatic Ecosystem Health & Management*, v. 3, p. 521-531, 2000.

ICAGA, Y.. **Fuzzy evaluation of water quality classification.** *Ecological Indicators*, 7, 710-718, 2007.

JANG, J.S.R.. **ANFIS – Adaptive-network-based fuzzy inference system**, *IEEE Trans. Syst. Man Cybern.* 23, 665–685, 1993.

Jl, Z.. **Hydrodynamics and Water Quality: Modeling Rivers, Lakes, and Estuaries.** USA: John Wiley & Sons. Chapter 5, p. 247, 2008.

JONES, J. R. e KNOWLTON, M. F. **Suspended solids in Missouri reservoirs in relation to catchment features and internal processes.** *Water Research*, v. 39, p. 3629-3635, 2005.

KARUL, C., SOYUPAK, S., ÇİLESİZ, A. F., AKBAY, N. e GERMEN, E. **Case Studies on the Use of Neural Networks in Eutrophication Modeling.** *Ecological Modelling*, 134, 145-152, 2000.

KARYDIS, M.; TSIRTSIS, G.; PRIMPAS, I. **Quantitative assessment of eutrophication: a proposed multivariate index.** *Proceedings of the 10th International Conference of Environmental Science and Technology*, 2007.

KARYDIS, M.; TSIRTIS, G.; PRIMPAS, I; KOKKORIS, G.D. **Principal component analysis: development of a multivariate index for assessing eutrophication according to the European water framework directive.** Ecological Indicators 10, 2009.

KITSIOU, D.; KARYDIS, M.. **Coastal marine eutrophication assessment: A review on data analysis.** Environment International, 37, 778–801, 2011.

KONDERLA, P.; MOKANEK, T.. **Comparison of two methods for the analysis of composite material.** Journal of Materials Processing Technology, 106, 87-93, 2000

KUO, J. WANG, I. LUNG, W. **A hybrid neural–genetic algorithm for reservoir water quality management.** Water Research 40 1367 – 1376, 2007.

KUO, J.; HSIEHB, P.; JOU, P.. **Lake eutrophication management modeling using dynamic programming.** Journal of Environmental Management, 88, 677–687, 2008.

LANDMANN, R.; ERDMANN, R. H.. **Uma abordagem heurística para a programação da produção na indústria de fundição com utilização da lógica fuzzy.** Gestão & Produção, São Carlos, v. 18, n. 1, p. 119-130, 2011.

LERMONTOV, A.; YOKOYAMA, L.; LERMONTOV, M.; MACHADO, M. A. S.. **River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed, Brazil.** Ecological Indicators, 9, 1188-1197, 2009.

LIMA, F. G.; PERERA, L. C..J.; KIMURA, H.; SILVA FILHO, A. C.. **Aplicação de redes neurais na análise e na concessão de crédito ao consumidor.** Revista de Administração:São Paulo. USP, 44, 34-45, 2009.

LO, S.; WU, G.; **Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system.** Engineering Applications of Artificial Intelligence. 21, 1189 – 1195, 2008.

LU, R. E LO, S. **Diagnosing reservoir water quality using self-organizing maps and fuzzy theory.** Water Research, 36, 2265–2274, 2002.

LUNDBERG, C.. **Eutrophication, risk management and sustainability. The perceptions of different stakeholders in the northern Baltic Sea.** Marine Pollution Bulletin, in press, 2012.

MACÊDO, J. A. B. **Águas e Águas**. 2. ed. Belo Horizonte, MG: CRQ-MG, 977p. 2004.

MACKAY, D. J. C.. **Bayesian interpolation**. Neural Computation 4(3),415–447, 1992.

MAGOSSI, L. R.; BONACELLA, P. H.. **Poluição das águas**, São Paulo:Ed. Moderna, p 27-36, 2003.

MAIA, A. A. D.. **Determinação do grau de trofia no Baixo São José dos Dourados por meio da comparação entre dois diferentes índices de estado trófico**. Dissertação de Mestrado. Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira. 2011.

MALUTTA C. **Método De Apoio à Tomada de Decisão Sobre a Adequação de Aterros Sanitários Utilizando a Lógica Fuzzy**. 2004, 221f. Dissertação (Doutorado em Engenharia de Produção). Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina, Florianópolis 2004.

MARCONI, M. A.; LAKATOS, E. M.. **Fundamentos de Metodologia Científica**.7. Ed. rev. São Paulo: Atlas, 2010.

MARQUARDT D. W. **An Algorithm for Least-Squares Estimation of Nonlinear Parameters**. Journal of the Society for Industrial and Applied Mathematics, v. 11, n. 2, p. 431 - 441, 1963.

MARTINI, L. C. P.; MATTOS, D. S.; BARBOSA, D. F. P.; RECH, A. I. B.. Uso de sensoriamento remoto orbital para avaliação da distribuição espacial de Clorofila_a na Lagoa da Conceição – Florianópolis, SC. Revista Engenharia Sanitária E Ambiental/Associação Brasileira De Engenharia Sanitária E Ambiental, Rio de Janeiro, Vol. 11, n 4, 318-324, out/dez 2006

MATTHEWS, R., HILLES, M. e PELLETIER, G. **Determining trophic state in Lake Whatcom, Washington (USA), a soft water lake exhibiting seasonal nitrogen limitation**. Hydrobiologia, 468, p. 107-121, 2002.

MELO, M. N.. **Mapeamento difuso no auxílio da redução dos impactos ambientais em uma usina hidrelétrica**. Dissertação de Mestrado. Universidade Federal de Itajubá, Departamento de Engenharia de Energia. 2009.

MIDDLETON, G.V. **Data analysis in the earth sciences using MATLAB**. Prentice-Hall, 2000.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

MUELLER, A.. **Uma aplicação de redes neurais artificiais na previsão do mercado acionário**. Dissertação de Mestrado. Departamento de Engenharia de Produção. UFSC, Florianópolis, 1996.

NARENDRA, K.; PARTHASARATHY, K.. **Identification and Control of Dynamical Systems using Neural Networks**, IEEE Transactions on Neural Networks 1, 4–27, 1990.

NAZÁRIO, S. L. S.; BUIOCHU, F.; KITANO, C.; ISEPON, J. S.; ADMAOWSKI, J. C.; HIGUTI, R. T.. **Caracterização de leite bovino utilizando ultra-som e redes neurais artificiais**. Revista Controle & Automação, 20, 4, 2009.

NIJBOER, R. C., VERDONSCHOT, P. F.M.. **Variable selection for modelling effects of eutrophication on stream and river ecosystems**. Ecological Modelling 177, 17–39, 2004.

OLIVEIRA JUNIOR, H. A., CALDEIRA, A. M., MACHADO, M. A., SOUZA, R. C. e TANSCHKEIT, R., **Inteligência Computacional aplicada à Administração, Economia e Engenharia em MATLAB**. São Paulo: Thomson Learning, 2007.

OLIVEIRA, K. P. S., **Aplicação das técnicas de Redes Neurais e de análise de componentes principais na modelagem de uma lagoa aerada da RIPASA S/A**, Dissertação de Mestrado, Faculdade de Engenharia Química, Universidade Estadual de Campinas (2000).

OLIVEIRA, M. M. F.; EBECKEN, N. F. F.; SANOTS, I. A.; NEVES, C. F., CALOBA, L. P.; OLIVEIRA, J. L. F.. **Modelagem da maré meteorológica utilizando redes neurais artificiais: uma aplicação para a baía de Paranaguá – PR**. Revista Brasileira de Meteorologia, v.22, n.1, 53-62, 2007.

OLIVEIRA-ESQUERRE, K. P. S.. **Aplicação de técnicas estatísticas multivariadas e de redes neurais na modelagem de um sistema de tratamento de efluentes industriais**. Tese de Doutorado. FEQ/UNICAMP, 2003.

TURNER, P.; MONTAGUE, G.; MORRIS, J.. **Dynamic neural networks in non-linear predictive control (an industrial application)**. Computers & Chemical Engineering, v 20, 937–S942, 1996.

OUYANG, Ying. **Evaluation of river water quality monitoring stations by principal component analysis**. Water Research 39, 2005.

PALMA, A. A.; PORTUGAL, M. S.. **Análise empírica da formação de expectativas de inflação no Brasil: uma aplicação de redes neurais artificiais a dados em painel**. Revista Econ. Contemp., Rio de Janeiro, v. 13, n. 3, p. 391-437, 2009.

PARINET, B.; LHOTE, A.; LEGUBE, B.; **Principal Components Analysis: An Appropriate Tool For Water Quality Evaluation And Management Application To A Tropical Lake System**. Ecological Modelling, 178, 295-311, 2004.

PARR, L.B.; MASON, C. F.. **Causes of low oxygen in a lowland, regulated eutrophic river in Eastern England**. Science of the Total Environment. 321, 273–286, 2004.

PECHE, R.; RODRIGUEZ, E.. **Development of environmental quality indexes based on fuzzy logic. A case study**. Ecological Indicators, 23, 555-565, 2012.

PEREIRA, G. C.; EVSULKOFF, A.. EBECKEN, N. F. F.. **Fuzzy Modelling of Chlorophyll production in a Brazilian upwelling system**. Ecological Modelling, 220, 1506-1512, 2009.

PERKINS, R. G. e UNDERWOOD, G. J. C. **Gradients of chlorophyll A and water chemistry along an eutrophic reservoir with determination of the limiting nutrient by in situ nutrient addition**. Water Resources, v. 34, n. 3, p. 713-724, 2000.

PRIMPAS, I; TSIRTSIS, G; KARYDIS, M; KOKKORIS, GD. **Principal component analysis: development of a multivariate index for assessing eutrophication according to the European water framework directive**. Ecol Indicat, 10:178–83, 2010.

RIBEIRO, R. N.. **Utilização de redes neurais artificiais e tecnologia ft-nir para predição do número *kappa* em um processo *kraft* de cozimento de madeira em indústria de celulose**. Dissertação de Mestrado. UNILESTE-MG, Coronel Fabriciano-MG, 2007.

RIJN, J. V.. **Waste treatment in recirculating aquaculture systems**. Aquacultural Engineering, in press, 2012.

RODRIGUES, I. O.. Abrangência dos Serviços de Saneamento – Atlas de Saneamento 2011. Disponível em http://www.ibge.gov.br/home/estatistica/populacao/atlas_saneamento/default_zip.shtm. Acesso em 05 de setembro de 2012.

SAADE, J.J. **A Defuzzification Based New Algorithm for the design of Mamdani-Type Fuzzy Controllers**, Mathware & Soft Computing. Granada, vol 7, n.2-3, 2000.

SAMPAIO, P. T. **Modelagem Térmica de um Forno Panela utilizando Redes Neurais Artificiais**. Dissertação de Mestrado. Escola de Engenharia – UFMG, 2006.

SANDRI, S; CORREA, C.. **Lógica Nebulosa**. V Escola de Redes Neurais, Conselho Nacional de Redes Neurais, ITA, São José dos Campos-SP. 73-090, 1999.

Shannon, E.E.; Brezonik P.L.. **Eutrophication analysis: a multivariate approach**. Journal of Sanitary Engineering Division, ASCE v.981, 37-57, 1972

SILVA NETTO, R.. **Deteccão de Intrusão Utilizando Redes Neurais Artificiais no Reconhecimento de Padrões de Ataque**. Dissertação de Mestrado. UNIFEI-Itajubá – MG, 2006.

SINGH, K.P.; BASANT, A.; MALIK, A.; JAIN, G. **Artificial neural network modeling of the river water quality—A case study**. Ecological Modelling 220, 2009.

SIVANANDAM, S.N.; SUMATHI, S.; DEEPA, S.N. **Introduction to fuzzy logic using MATLAB**. Heidelberg: Springer, 430 p. 2007.

SMITH, V.H., SCHINDLER, D.W.. Eutrophication science: where do we go from here? Trends Ecol. Evol. 24, 201–207, 2009.

SOUZA, I. V. A.; SOUZA, R. O.; PAULINO, W. D. **Cálculo do índice de estado trófico em reservatório com estudo de caso no reservatório Acarape do Meio**. XVII Simpósio Brasileiro de Recursos Hídricos. 2007.

SOYUPAK, S. et al.. **Evaluation of eutrophication control strategies for the Keban Dam reservoir**. Ecological Modelling, v. 97, p. 99-110, 1997.

SPALL, J. C., CRISTION, J. A.. **A neural network controller for systems with unmodeled dynamics with applications to wastewater treatment.** IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, 37, 369–375, 1997, *apud* WEST, D.; DELLANA, S.. **An empirical analysis of neural network memory structures for basin water quality forecasting.** International Journal of Forecasting, 27, 777–803, 2011.

SREBOTNJAK, T.; CARR, G.; SHERBININ, A.; RICKWOOD, C.. **A global Water Quality Index and hot-deck imputation of missing data.** Ecological Indicators, 17, 108–119, 2012.

STOATE, C; BÁLDI, A.; BEJA, P.; BOATMAN, N. D.; HERZON, I.; VAN DOORN, A.; DE SNOO, G. R.; RAKOSY, L.; RAMWELL, C.. **Ecological impacts of early 21st century agricultural change in Europe – A review.** Journal of Environmental Management, 91, 22–46, 2009.

STROBL, R.O.; FORTE, F.; PENNETTA, L.. **Application of artificial neural networks for classifying lake eutrophication status.** Lakes & Reservoirs: Research and Management. 12, 15–25, 2007.

SUCENA, M. P.; STEPHAN, R. M.; ALMEIDA JUNIOR, V.. **Modelo integral-neuro-fuzzy para avaliação holística de sistema de transporte urbano por levitação magnética.** VII Congresso Nacional de Excelência em Gestão. Rio de Janeiro-RJ, 2012.

SULIS, A.; BUSCARINU, P.; SECHI, G. M.. **Using reservoir trophic-state indexes in optimisation modelling of water-resource systems.** Environmental Modelling & Software, v 26, p. 731–738, 2011.

SWINGLER, K. **Applying neural networks: a practical guide.** Londres: Academic Press, 1996, *apud* PALMA, A. A.; PORTUGAL, M. S.. **Análise empírica da formação de expectativas de inflação no Brasil: uma aplicação de redes neurais artificiais a dados em painel.** Revista Econ. Contemp., Rio de Janeiro, v. 13, n. 3, p. 391–437, 2009.

TAYLAN, O.; DARRAB, I. A.. **Determining optimal quality distribution of latex weight using adaptive neuro-fuzzy modeling and control systems.** Computers & Industrial Engineering, 61, 686–696, 2011.

TORRECILLA, J. S.; OTERO, L.; SANZ, P. D.. Optimization of an artificial neural network for thermal/pressure food processing: Evaluation of training algorithms. *Computers and Electronics in Agriculture*, 56, p. 101–110, 2007.

TOSUN, M., DINCER, K., BASKAYA, S. **Rule-based Mamdani-type fuzzy modelling of thermal performance of multi-layer precast concrete panels used in residential buildings in Turkey.** *Expert Systems with Applications*, 38, p. 5553–5560, 2011.

TUNDISI, J. G.; TUNDISI, T. M. *Limnologia*. São Paulo: Oficina de Textos, 2008.

VIEIRA, V. P.; DINIZ, L.S.; LACERDA, E.G.M. ; CARVALHO, A.C.P.; LUDEMIR, T.B. **Sistemas Inteligentes: aplicações a Recursos Hídricos e Ciências Ambientais.** Porto Alegre. Editora da Universidade / UFRGS 1999.

VILAS, L. G.; SPYRAKOS, E.; PALENZUELA, J. M. T.. **Neural network estimation of chlorophyll a from MERIS full resolution data for the coastal waters of Galician rias (NW Spain).** *Remote Sensing of Environment* 115 524–535, 2011.

WEST, D.; DELLANA, S.. **An empirical analysis of neural network memory structures for basin water quality forecasting.** *International Journal of Forecasting*, 27, 777–803, 2011.

WESTPHAL, J. T. **Modelagem Difusa de um Sistema Especialista Médico: Avaliação dos Fatores de Internação em Crianças Queimadas.** 2003, 115f. Dissertação (Mestrado em Ciência da Computação). Programa de pós-graduação em ciência da Computação, Universidade Federal de Santa Catarina, Florianópolis, 2003.

WU, S. X.; BANZHAF, W.. The use of computational intelligence in intrusion detection systems: A review. *Applied Soft Computing*, v 10, p. 1-35, 2010.

XU, F. L.. **Trophic Classification for Lakes.** *Encyclopedia of Ecology*, 3594–3601, 2008.

YAN, H.; ZOU, Z.; WANG, H.. **Adaptive neuro fuzzy inference system for classification of water quality status.** *Journal of Environmental Sciences*, 22(12), 1891–1896, 2010.

YETILMEZSOY, K.; FINGAS, M.; BEN FIELDHOUSE, B.. **An adaptive neuro-fuzzy approach for modeling of water-in-oil emulsion formation.** *Colloids and Surfaces A: Physicochem. Eng. Aspects*, 389, 50– 62, 2011.

ZADEH, L. A., 1965. **Fuzzy sets. Information and Control** 8, 338-353, apud BARBALHO. V. M. S.. **Sistemas Baseados em Conhecimento e Lógica Difusa para Simulação do Processo Chuva-Vazão**. Tese de Doutorado. COPPE/UFRJ, 2001.

ZANG, M.; SHEN, Y.; GUO, Y.. **Development and application of a eutrophication water quality model for river networks**. Journal of Hydrodynamics, 20, 719-726, 2008.

ZHANG,Y.; SU, A.; JIANG, G.. **Understanding data center network architectures in virtualized environments: A view from multi-tier applications**. Computer Networks, 55, p. 2196–2208, 2011.

ZIMMERMANN, C.M.; GUIMARÃES, O.M.; PERALTA-ZAMORA, P.G. **Avaliação da qualidade do corpo hídrico do rio Tibagi na região de ponta grossa utilizando análise de componentes principais (PCA)**. Química Nova, Vol. 31, No. 7, 1727-1732, 2008.

ANEXOS

ANEXO A – Tabelas com os dados ambientais dos corpos hídricos avaliados

Tabela 1 - Valores médios, mínimos, máximos e desvios padrões dos parâmetros do reservatório da Marcela

Estações	Prof.	T água	T ar	Trans.	Cond.	Cor	pH	Stotal	SS	OD	N-NH4	N-NO2	N-NO3	P-PO4	Ntotal	Ptotal	Clorofila.a
	m	oC	oC	m	mS/cm	Pt-Co		mg/L	mg/L	mg/L	ug/L	ug/L	ug/L	ug/L	mg/L	mg/L	ug/L
1S	0	27	25	0,55	1,5	18,0	8,1	908	26	6,7	16,3	167,1	1815,7	569,8	4,2	0,6	37,4
1M	1,9	27	25	0,55	1,5	17,8	8,2	898	26	6,5	12,4	166,4	1311,5	565,3	4,5	0,6	48,7
2S	0	26	27	0,5	1,5	19,4	8,3	898	28	6,4	10,0	130,6	1403,3	553,4	3,9	0,7	64,3
2M	2,85	27	27	0,5	1,5	17,4	8,1	940	28	6,5	9,0	138,7	1265,8	569,8	3,2	0,7	61,2
3S	0	27	28	0,55	1,4	18,3	8,3	910	26	5,2	6,8	105,6	1297,2	563,9	3,7	0,7	157,1
3M	3,2	27	28	0,55	1,5	17,8	8,2	892	48	7,0	9,0	101,5	1315,5	554,9	2,7	0,7	95,9
1S	0	23	23	0,5	1,31	19,8	8,1	820	28	6,6	159,4	217,2	1201,1	593,6	3,2	0,6	92,3
1M	2,07	23	23	0,5	1,31	19,8	8,1	824	32	6,5	162,3	227,3	1378,5	577,3	3,4	0,6	126,4
3S	0	26	27	0,5	1,9	15,7	8,15	716	32	9,1	15,6	413,2	2663,8	304,6	2,5	0,3	84,8
3M	3	25	27	0,5	1,9	15,6	7,95	746	2	5,4	14,9	284,7	1782,6	322,3	3,8	0,1	56,3
1S	0	29,8	32	0,4	1,0	25,4	8,1	860	42	4,6	91,4	17,7	1876,6	409,3	14,2	0,6	224,8
1M	2,3	30	32	0,4	1,0	24,4	8,18	844	16	5,2	35,3	16,3	1315,5	407,8	4,7	0,5	112,0
2S	0	30	31	0,3	0,9	23,1	8,19	802	34	2,8	22,7	34,6	1614,5	412,2	2,2	0,6	83,0
2M	2,5	30	31	0,3	1,0	21,3	7,97	880	26	4,2	29,0	10,9	1436,3	410,7	2,9	0,5	147,7
3S	0	29,5	31	0,25	0,9	23,3	8,08	908	18	2,2	25,4	106,9	1080,6	416,6	6,1	0,6	186,2
3M	2,5	30	31	0,25	0,9	22,2	7,94	924	24	3,4	40,5	140,1	1537,9	429,9	10,1	0,7	114,9
1S	0	28	28,5	0,3	1,0	25,4	8,31	918	54	6,4	20,5	23,7	1394,6	341,4	4,5	0,6	113,8
1M	2	28	28,5	0,3	1,0	23,7	8,09	894	40	4,9	22,2	96,8	1985,0	375,3	2,5	0,6	125,3
2S	0	29	28	0,4	1,0	24,8	8,45	878	40	7,8	12,0	12,9	2241,9	353,2	1,3	0,6	109,0
2M	2,5	28	28	0,4	1,0	23,7	8,29	880	48	4,6	14,4	51,5	1842,8	356,2	10,6	0,3	123,7
3S	0	28	28	0,3	1,0	25,0	8,33	896	48	5,2	10,5	21,0	2320,4	345,9	5,5	0,3	155,7
3M	2,2	28	28	0,3	1,0	24,4	8,22	898	42	3,6	13,2	21,0	2709,8	366,5	9,0	0,6	149,6

Continuação...

1S		31	34	0,3	1,6	38,9	8,66	966	92	14,1	29,7	67,7	1567,0	70,1	3,6	0,6	167,6
1M		30	34	0,3	1,7	38,3	8,71	966	104	13,3	28,0	64,3	1786,7	89,3	1,2	0,6	140,0
2S		31	34	0,25	1,6	34,8	8,68	916	64	14,7	30,7	72,4	1504,5	174,8	0,9	0,6	65,4
2M		30,5	34	0,25	1,6	34,4	8,67	926	82	14,4	39,2	70,4	1982,5	74,6	9,7	0,6	82,4
3S		30	32,5	0,3	1,6	33,1	8,54	964	76	9,6	58,0	174,6	2383,2	135,0	5,3	0,6	77,4
3M		29	32,5	0,3	1,5	31,7	8,39	944	76	4,4	85,0	208,4	2320,5	120,3	5,4	0,6	83,3
1S	0	27,6	31,3	0,25	1,1	38,5	8,58	1072	92	9,1	36,1	7,1	552,9	309,0	3,9	1,0	98,0
1M	1,5	27,9	31,3	0,25	1,1	38,9	8,98	1078	10	9,0	29,7	6,8	351,0	319,3	2,7	1,0	109,9
2S	0	27,3	29,3	0,3	1,1	34,1	8,8	1074	78	7,7	27,1	9,5	769,2	320,8	4,4	0,7	143,2
2M	2	26,3	29,3	0,3	1,1	34,4	9,47	1072	12	7,3	31,7	12,5	1014,4	313,4	9,4	0,8	183,7
3S	0	26,8	30,8	0,3	1,1	34,8	8,67	1088	88	5,0	33,9	77,1	977,7	375,3	9,5	0,9	127,6
3M	2	26	30,8	0,3	1,1	34,4	9,25	1058	88	3,1	37,8	136,0	676,6	384,2	28,0	1,2	185,3
1S	0	27,7	28	0,2	1,6	63,5	8,3	1040	78	6,5	28,8	1,7	693,2	196,9	3,6	0,9	264,4
1M	1,75	27,2	28	0,2	1,6	57,8	8,4	974	54	5,3	22,7	2,4	1430,4	183,7	4,3	1,0	225,7
2S	0	26,8	28	0,2	1,5	65,5	8,5	996	26	5,5	31,9	3,8	1577,8	196,9	5,5	0,7	119,0
2M	2,3	25,4	28	0,2	1,6	66,8	8,5	1014	74	4,4	24,6	1,7	1198,0	194,0	4,6	0,9	229,6
3S	0	25	31,7	0,2	1,6	73,9	8,4	964	52	4,0	29,7	2,1	1105,1	254,4	8,1	0,6	293,7
3M	1,5	24,5	31,7	0,2	1,6	70,9	8,4	1002	66	3,9	27,5	0,7	481,7	279,5	6,6	0,4	208,4
1S	0	28,7	26,7	0,25	1,4	8,9	8,65	1040	6	4,4	87,5	96,1	447,7	474,1	3,2	1,8	9,5
1M	1,4	28,7	26,7	0,25	1,3	7,6	7,95	1100	5	3,6	97,7	83,9	649,8	480,0	3,1	1,0	9,3
1F	2,8	28,9	26,7	0,25	1,3	5,9	8,85	1135	6	3,0	64,6	92,0	3197,8	452,0	2,1	1,2	10,8
2S	0	28,1	30,5	0,4	1,3	6,1	8,79	1160	7	4,6	13,2	59,6	1991,7	488,9	3,1	1,1	12,7
2M	1,9	27,7	30,5	0,4	1,3	7,4	8,93	1030	11	4,0	14,6	54,8	1142,5	483,0	3,2	1,4	10,4
2F	3,7	27,2	31,2	0,4	1,3	6,7	8,69	1025	8	0,6	19,8	46,7	1046,5	516,9	2,4	0,8	12,7
3S	0	28,2	31,2	0,25	1,3	6,3	8,703	1115	7	4,6	16,8	48,1	2836,6	500,7	3,5	1,0	11,3
3M	2	27,9	31,2	0,25	1,4	7,0	8,724	1780	6	0,9	21,7	55,5	2203,4	506,6	4,5	1,2	11,0

Continuação...

3F	4	27,5	31,2	0,25	1,4	7,8	8,649	1100	7	0,1	25,1	52,1	2567,2	503,6	4,9	1,1	11,1
1S		30	32	0,25	1,3	32,6	9,55	1175	13	9,0	29,7	264,4	835,4	166,0	9,5	0,7	17,3
1M	3,8	30	32	0,25	1,3	34,1	9,50	1175	14,5	6,2	28,3	257,7	1721,6	170,4	9,8	0,7	18,1
1F		31	32	0,25	1,3	36,3	9,43	1200	18,5	5,3	59,0	291,5	1227,9	177,8	8,4	1,0	23,5
2S		29,7	31	0,3	1,3	29,8	9,41	1120	11	8,6	20,7	311,8	1107,7	177,8	6,7	0,8	49,5
2M	4,5	29	31	0,3	1,3	30,0	9,30	1115	12,5	4,6	20,0	345,6	1594,8	183,7	9,6	1,0	48,5
2F		31	31	0,3	1,3	27,4	9,45	1165	10	0,9	58,2	480,8	1365,4	208,7	9,3	0,9	47,9
3S		29	31	0,3	1,3	28,9	9,30	1145	13	5,6	18,5	453,8	1419,9	189,6	7,5	0,8	63,1
3M	4,5	29	31	0,3	1,3	29,6	9,16	1185	13,5	2,4	17,8	507,9	1426,2	198,4	6,5	0,7	55,4
3F		30	31	0,3	1,4	29,4	9,09	1245	11	0,9	18,0	575,5	1275,4	204,3	9,5	0,9	82,5
1S	0	29,5	29	0,3	1,3	10,4	8,88	1150	49	9,4	21,2	4,1	1201,9	505,1	3,9	0,8	13,2
1M	1,6	28	29	0,3	1,3	9,4	8,67	1170	12	4,2	26,6	100,8	899,1	593,6	4,6	0,6	16,3
1F	3,2	29	29	0,3	1,4	10,6	8,71	1135	30	2,1	17,8	48,1	1274,1	575,9	4,1	0,9	13,4
2S	0	30	27	0,3	1,3	2,8	8,77	1100	3	6,5	12,2	9,5	1562,5	592,1	3,5	0,7	16,9
2M	2,5	29	27	0,3	1,3	5,9	8,6	1175	9	5,3	23,9	105,6	610,6	543,4	3,3	0,7	16,4
2F	5	29	27	0,3	1,4	10,6	8,4	1140	29	1,2	78,2	308,5	495,2	627,5	3,5	0,8	16,9
3S	0	23,5	30	0,3	1,4	4,3	8,71	1150	9	8,5	9,0	38,6	726,0	561,1	4,6	0,8	16,8
3M	2,9	30	30	0,3	1,3	5,6	8,58	1190	10	1,4	57,0	167,8	826,9	620,1	4,6	0,9	16,4
3F	5,8	29	30	0,3	1,4	6,5	8,43	1135	9	0,9	114,3	336,2	134,6	599,5	4,8	0,8	16,7
1S		28,2	26,3	0,3	1,2	15,4	8,98	2280	9	9,3	59,0	< 0,14	1317,3	490,4	2,7	0,8	18,9
1M	2	27,1	26,3	0,3	1,3	17,0	8,5	2350	44	3,0	51,7	34,6	974,6	516,9	2,9	1,0	16,6
2S		30	32,1	0,4	1,3	17,6	8,86	2145	6	10,0	56,8	28,5	599,7	590,6	3,0	0,5	15,7
2M	3	26,9	32,1	0,4	1,2	17,2	8,55	2275	6	3,6	54,8	10,9	1372,6	415,2	2,9	1,0	15,1
2F		27,2	32,1	0,4	1,2	14,8	8,26	2220	30	0,9	93,6	114,4	447,3	605,4	2,8	0,8	17,0
3S		26,1	30,4	0,45	1,3	16,9	8,78	2325	5	6,3	32,9	27,1	1160,7	266,2	2,9	0,8	16,6
3M	4	25	30,4	0,45	1,3	15,4	8,45	2180	11	2,2	57,5	25,1	978,0	370,9	2,7	0,8	16,6

Conitnuação...

3F		24,9	30,4	0,45	1,3	16,7	8,24	1045	5	0,7	106,7	112,3	221,3	549,3	3,6	1,0	17,6
1S	0	31,7	31,6	0,4	2,3	17,2	8,38	1090	3	5,8	88,5	6,8	1382,3	825,3	2,2	0,9	19,2
1M	2,2	31,2	31,6	0,4	2,9	18,7	8,4	1080	4	4,1	59,3	8,2	1643,5	899,6	2,9	0,9	22,4
1F	4,4	31,3	31,6	0,4	2,7	20,2	8,23	1070	6	3,9	73,9	6,2	1331,0	902,6	2,2	1,0	26,6
2S	0	30,4	34	0,5	2,2	16,9	8,32	1145	2	5,7	168,9	6,8	1295,7	956,1	3,2	1,0	17,5
2M	3	32,5	34	0,5	2,8	16,7	8,29	1055	3	3,6	165,2	5,5	1827,8	962,1	2,9	0,9	22,2
2F	6	28	34	0,5	2,7	17,2	8,18	1235	3	2,8	223,7	6,8	1367,8	861,0	3,2	1,0	18,9
3S	0	29,7	32,1	0,5	2,3	16,7	8,46	1060	4	4,5	94,3	5,5	1106,6	941,2	4,3	0,9	19,6
3M	2,9	32,2	32,1	0,5	2,6	13,5	8,3	1055	3	5,0	107,9	6,8	1079,4	881,8	3,2	0,9	23,6
3F	5,8	32,6	32,1	0,5	3,0	15,0	8,31	1070	4	7,0	121,4	6,2	1446,4	887,7	2,7	0,9	17,1
1S	0	30	28	0,4	2,8	19,1	8,15	495	12	7,0	41,0	263,1	1650,3	322,3	3,9	1,5	57,2
1M	2,8	29	28	0,4	2,8	17,2	8,21	615	4	6,4	28,2	272,2	1433,5	285,4	4,8	1,4	16,6
1F	5,5	27	28	0,4	2,7	15,6	8,25	1140	1	5,9	6,2	281,0	1361,0	257,4	4,6	1,1	33,3
2S	0	30	29,5	0,5	2,8	17,2	7,8	1055	7	5,3	23,6	271,2	1390,3	334,1	5,1	1,5	45,0
2M	2,8	29,5	29,5	0,5	2,7	15,0	7,78	1100	11	4,1	79,3	253,0	1376,8	247,1	4,0	1,5	43,1
2F	5,5	29,5	29,5	0,5	2,8	13,3	8,09	1050	5	4,3	28,7	291,2	1072,0	233,8	4,6	1,1	47,6
3S	0	30	28	0,45	2,6	15,9	7,93	705	3	5,2	23,1	239,8	1593,8	171,9	4,9	1,5	48,6
3M	2,5	28	28	0,45	2,8	17,6	8,05	960	8	2,9	16,4	295,6	1403,5	462,3	5,0	1,4	37,7
3F	5	29	28	0,45	2,8	15,4	8,18	920	6	2,8	7,7	296,6	1735,2	763,1	4,3	1,6	52,0
1S	0	26,6	28	0,45	1,5	19,1	8,79	795	4	9,4	55,8	286,1	2190,0	437,3	5,1	0,8	14,2
1M	1	27	28	0,45	1,5	18,3	8,78	835	2	8,8	52,2	340,9	1251,1	416,6	1,9	1,0	14,8
1F	2	27,4	28	0,45	1,5	16,9	8,79	815	38	9,0	35,8	337,5	1352,2	381,2	3,6	0,8	13,1
2S	0	27,3	28,4	0,45	1,5	15,2	8,7	765	2	8,2	35,3	161,7	1298,9	432,8	4,3	0,5	14,4
2M	1,5	27,1	28,4	0,45	1,5	19,3	8,67	840	2	7,9	41,5	188,8	1976,1	421,1	5,8	1,0	15,8
2F	3	27,4	28,4	0,45	1,5	19,6	8,53	860	5	5,1	46,6	361,2	1611,2	429,9	6,6	0,8	14,7
3S	0	27,1	28,3	0,45	1,5	15,4	8,39	850	12	6,6	53,8	272,6	1151,9	437,3	2,5	0,8	14,2

Conitnuação...

3M	2	27,4	28,3	0,45	1,5	14,6	8,57	800	1	3,9	70,6	332,1	1611,9	447,6	3,8	1,7	12,9
3F	4	27,3	28,3	0,45	1,5	19,1	8,43	1005	27	3,7	55,3	134,0	1314,0	418,1	3,7	1,0	14,5
1S	0	28,1	26,9	0,3	1,5	15,2	8,69	900	22	8,2	20,5	33,2	346,6	276,6	3,6	0,5	16,9
1M	0,9	28,2	26,9	0,3	1,5	13,5	8,83	830	31	8,0	31,8	47,4	318,0	227,9	4,8	0,5	19,3
1F	1,8	28,9	26,9	0,3	1,6	14,8	8,75	860	21	5,7	49,2	6,2	272,7	185,1		0,4	20,6
2S	0	27,8	27,1	0,4	1,5	17,8	8,61	835	23	5,1	69,7	4,8	331,7	220,5	0,2	0,5	24,1
2M	1,6	27,7	27,1	0,4	1,5	16,7	8,61	800	21	4,8	29,7	5,5	345,5	205,8		0,5	25,3
2F	3,2	27,4	27,1	0,4	1,5	18,3	8,34	940	25	1,3	142,3	1,7	262,7	183,7		0,5	28,3
3S	0	27,8	26,7	0,4	1,5	13,0	8,52	885	10	2,8	37,9	4,1	318,0	248,5		0,6	27,6
3M	1,4	27,9	26,7	0,4	1,5	15,4	8,28	1060	12,9	2,0	111,6	14,9	321,6	238,2	1,3	0,6	24,6
3F	2,8	27,9	26,7	0,4	1,5	15,7	8,46	860	15	3,2	92,2	36,6	256,7	245,6	3,6	0,6	26,0
1S	0	27	25	0,3	3,0	17,2	8,83	890	2	7,5	74,8	270,7	729,4	416,6	1,3	0,8	34,4
1M	1,3	28	25	0,3	3,0	15,2	9,06	820	1	6,3	69,7	343,7	367,9	410,7	1,4	0,8	8,3
1F	2,5	28	25	0,3	3,0	17,8	9,3	880	2	4,8	87,0	347,7	335,0	431,4	1,2	0,7	8,6
2S	0	28	26	0,4	2,9	18,1	9,14	1370	5	6,1	23,6	418,1	841,6	424,0	1,7	0,7	22,4
2M	1,5	27	26	0,4	2,9	18,0	9,11	895	4	4,9	50,2	345,0	986,7	425,5	1,7	0,8	18,5
2F	3	27,5	26	0,4	2,9	18,0	9,08	735	5	3,4	62,5	354,5	1020,5	435,8	1,7	1,1	12,0
3S	0	28	31	0,5	2,9	15,4	9,03	830	3	4,8	28,7	407,3	953,3	437,3	1,7	1,4	12,7
3M	1,7	28	31	0,5	2,9	13,3	8,92	830	2	1,9	72,7	449,2	695,1	452,0	1,3	0,8	13,1
3F	3,3	29	31	0,5	2,9	18,1	8,97	855	5	1,9	34,9	395,1	734,7	438,7	0,9	0,7	19,4
Mínimo	0	23	23	0,2	0,9	2,8	7,78	495	1	0,1	6,2	0,7	134,6	70,1	0,2	0,1	8,3
Máximo	6	32,6	34	0,5	3,0	73,9	9,6	2350	104	14,7	223,7	575,5	3197,8	962,1	28	1,8	293,7
Média	1,782	28,239	29,28	0,3667	1,674636	21,47695	8,563	1058,025	23,1	5,324	48,19	146,2	1256	415,1	4,5164	0,829	59,2837
Dpadrão	1,696	1,8231	2,476	0,0971	0,636051	13,27792	0,399	339,6461	24,8	2,852	39,66	146,32	615,88	200,3	3,3154	0,296	63,2058

S – medida na superfície M – medida no meio do corpo hídrico F - medida no fundo do corpo hídrico

Tabela 2 - Valores médios, mínimos, máximos e desvios padrões dos parâmetros do reservatório de Jacarecica

Estações	Prof.	Tágua	T ar	Trans.	Cond.	Cor	pH	Stotal	SS	OD	N-NH4	N-NO2	N-NO3	P-PO4	Ntotal	Ptotal	Clorof.a
	m	°C	°C	m	mS/cm	Pt-Co		mg/L	mg/L	mg/L	µg/L	µg/L	µg/L	µg/L	mg/L	mg/L	µg/L
2	0,0	29,1	30,6	2,00	0,37		7,4	190,0	2,0	3,25	< 2,80	< 0,14	58,17	< 0,91		0,68	18,55
2	5,0	28,5		2,00	0,37		7,3	225,0	3,0	2,45	< 2,80	8,86	49,76	< 0,91		0,41	20,11
4	0,0	28,7	31,4	2,30	0,37		7,5	170,0	2,0	4,10	< 2,80	< 0,14	59,62	< 0,91		0,39	23,55
4	5,0	28,3		2,30	0,41		7,4	215,0	2,0	2,70	< 2,80	12,24	< 0,70	< 0,91		1,57	19,12
5	0,0	28,9	30,0	2,0	0,39		7,7	210,0	1,0	4,40	< 2,80	< 0,14	< 0,70	< 0,91		0,66	11,70
5	5,0	28,2		2,0	0,37		7,6	155,0	2,0	3,60	< 2,80	4,80	< 0,70	< 0,91		0,87	25,86
7	0,0	28,1	29,7	2,0	0,38		7,7	210,0	2,0	4,28	< 2,80	3,45	174,6	< 0,91		0,08	6,05
7	5,0	27,8		2,0	0,39		7,4	230,0	3,0	2,60	< 2,80	10,89	< 0,70	< 0,91		0,26	1,76
8	0,0	26,8	29,8		0,50		7,4	295,0	15,0	9,34	< 2,80	< 0,14	199,5	< 0,91		0,84	4,66
1	0,0	26,5	30,4	3,4	0,34	2,97	7,8	225,0	6,0	3,56	< 2,80	10,18	20,86	< 0,91	0,64	1,18	8,72
1	5,0	29,2			0,37	24,26	8,0	215,0	5,3	0,00	772,9	10,85	28,15	< 0,91	0,85	0,25	6,63
2	0,0	29,5	27,4	3,0	0,35	2,78	8,2	200,0	1,0	4,30	< 2,80	9,84	11,81	< 0,91	0,64	0,25	5,86
2	5,0	28,7			0,35	2,97	8,2	200,0	0,7	0,00	< 2,80	10,18	14,37	< 0,91	0,08		4,97
3	0,0	28,7	27,5	3,0	0,35	2,60	8,4	200,0	0,7	5,04	6,37	9,84	34,88	472,7		0,52	5,39
3	5,0	26,6			0,35	2,78	8,3	240,0	1,0	3,70	6,37	10,18	46,82	< 0,91	0,28	0,49	5,89
4	0,0	28,2	27,1	2,8	0,35	2,78	8,3	195,0	1,3	5,41	< 2,80	10,18	36,73	394,5	0,09	0,43	5,47
4	5,0	27,5			0,35	2,78	8,2	205,0	1,0	5,38	1,97	10,85	38,25	335,5	0,06	0,38	5,39
5	0,0	27,9	30,4	2,4	0,35	2,97	8,4	215,0	1,0	6,52	< 2,80	9,84	40,65	307,5		0,35	4,88
5	5,0	26,6			0,35	2,78	8,4	190,0	2,7	5,33	1,72	11,19	40,81	281,0		0,41	4,92
6	0,0	27,1	31,5	2,8	0,35	2,97	8,1	210,0	1,0	6,08	11,22	9,84	28,39	245,6		0,26	7,11
6	5,0	26,8			0,35	2,78	8,1	190,0	1,0	4,89	< 2,80	11,19	33,60	176,3		0,59	5,38
7	0,0	28,0	27,5	2,5	0,35	2,60	8,2	195,0	0,0	6,22	44,36	9,84	33,44	362,1		0,36	4,93
7	5,0	26,0			0,35	3,15	8,1	200,0	2,0	5,48	72,13	10,85	36,08	341,4		0,42	7,53

Continuação...

8	0,0	27,8	30,4	0,0	0,46	3,52	8,3	305,0	0,0	6,34	32,66	9,84	26,95	0,84		0,0402	5,47
1	0,0	29,0	26,2	3,5	0,41	3,15	7,7	150,0	0,0	1,48	< 2,80	< 0,14	6,01	< 0,91		0,00	0,94
1	5,0	26,8			0,39	6,30	7,6	215,0	2,7	2,67	10,6	12,21	26,87	< 0,91			1,53
2	0,0	29,0	24,0	3,0	0,37	2,78	7,7	255,0	0,0	3,26	26,8	13,56	8,27	< 0,91			0,98
2	5,0	28,9			0,41	2,78	7,7	265,0	0,0	3,11	53,6	10,52	14,05	< 0,91			2,00
3	0,0	29,4	27,0	3,7	0,42	2,78	7,9	255,0	0,0	5,04	51,52	4,77	31,77	< 0,91			1,45
3	5,0	29,0			0,45	2,78	7,8	235,0	0,0	5,04	23,91	6,46	50,36	< 0,91		0,03	2,46
4	0,0	30,0	29,0	3,0	0,40	2,60	7,9	260,0	1,0	5,33	95,0	7,47	39,30	< 0,91		0,02	2,18
4	5,0	25,9			0,42	23,33	7,7	265,0	4,0	0,00	123,1	14,91	34,39	52,4		0,10	0,98
5	0,0	28,7	28,4	2,7	0,41	2,41	8,0	215,0	0,0	5,48	14,15	1,72	7,26	< 0,91		0,00	0,74
5	5,0	28,6			0,41	2,23	8,0	210,0	0,0	5,33	15,85	1,72	28,17	< 0,91			3,70
6	0,0	29,1	27,5	2,7	0,39	2,78	8,0	235,0	0,0	6,08	<2,8	6,46	24,39	< 0,91			0,94
6	5,0	28,9			0,39	2,41	8,0	240,0	0,0	5,78	14,63	7,13	30,24	< 0,91			0,70
7	0,0	29,3	29,0	3,0	0,39	2,23	8,0	235,0	0,0	5,19	20,8	3,41	25,84	< 0,91		0,02	1,54
7	5,0	28,9			0,41	2,60	8,0	240,0	0,0	5,63	46,4	5,44	26,80	< 0,91		0,02	2,00
8	0,0	26,8	27,0	0,0	0,55	3,90	8,0	340,0	0,0	5,93	496,6	3,75	55,49	< 0,91		0,05	0,94
1	0,0	28,0	26,0	1,5	0,36	7,41	6,6	185,0	2,0	6,19	430,1	7,47	77,52	692,4			0,70
	5,0	26,6			0,32	7,23	6,3	215,0	7,0	3,40	471,0	9,16	21,82	696,8			4,26
2	0,0	23,0	26,5	1,4	0,39	6,49	6,6	230,0	2,0	5,57	348,3	5,78	24,32	742,5			0,98
2	5,0	29,6			0,31	9,45	6,3	175,0	3,7	2,94	368,7	11,87	8,92	718,9			0,51
3	0,0	23,1	28,9	1,7	0,39	6,30	7,5	245,0	3,3	5,41	169,2	7,81	49,07	742,5			4,90
3	5,0	22,1			0,78	12,04	7,3	260,0	12,3	3,09	245,9	10,85	48,34	774,9			2,18
4	0,0	23,3	27,3	1,6	0,38	4,82	7,5	190,0	5,0	5,80	87,3	10,52	44,34	738,1			12,73
4	5,0	22,5			0,38	9,08	7,2	230,0	5,3	1,24	174,3	14,24	31,72	770,5			4,81
5	0,0	22,0	23,1	1,5	0,73	5,01	7,4	205,0	2,0	4,72	138,5	9,16	46,34	757,2			6,52
5	5,0	21,3			0,71	7,78	7,0	250,0	8,7	0,00	194,8	29,79	27,47	786,7			1,34

Continuação...

6	0,0	21,2	24	1,4	0,37	4,27	6,9	235,0	2,3	4,48	61,8	8,15	43,70	751,3			3,24
6	5,0	21,3			0,68	5,75	6,9	230,0	7,3	2,78	87,3	30,80	28,67	746,9			3,45
7	0,0	28,0	24,6	1,6	0,73	4,45	7,3	220,0	6,0	3,79	102,7	12,88	26,22	741,0			0,74
7	5,0	27,9			0,42	11,85	7,0	255,0	8,7	0,00	133,4	26,75	6,39	693,8			0,74
8	0,0	28,0	25,9	0,0	0,38	4,45	7,8	285,0	4,0	7,19	384,1	8,83	24,31	5,26			0,8211
1	0,0	28,1	28,1	2,0	0,61	7,04	7,9	285,0	0,3	5,72	261,3	4,43	72,48	< 0,91			0,79
	5,0	26,1			0,61	7,41	7,8	210,0	1,3	3,22	128,3	12,54	14,60	< 0,91			3,87
2	0,0	26,0	29,1	2,0	0,65	7,04	7,9	205,0	0,3	5,72	164,1	3,41	25,12	< 0,91			4,59
2	5,0	25,1			0,60	6,49	7,8	?	2,0	3,59	506,9	4,77	15,18	< 0,91			4,22
3	0,0	27,0	29,1	2,0	0,69	6,49	8,0	115,0	1,7	5,64	31,1	2,74	49,56	< 0,91			2,68
3	5,0	25,1			0,68	6,67	7,9	145,0	0,3	3,08	77,1	3,75	54,37	< 0,91			6,34
4	0,0	27,0	27,0	1,9	0,71	6,67	8,1	180,0	0,7	6,16	9,3	2,74	46,19	< 0,91			6,72
4	5,0	26,0			0,61	6,86	7,8	230,0	6,0	2,13	9,3	4,77	53,17	< 0,91			4,84
5	0,0	28,0	29,0	2,0	0,65	6,49	8,0	170,0	1,3	6,82	6,4	1,72	47,64	< 0,91			4,10
5	5,0	24,0			0,60	6,12	7,8	165,0	1,7	2,49	63,5	3,08	51,41	< 0,91			4,08
6	0,0	26,1	30,1	1,8	0,63	6,49	7,1	200,0	0,7	7,04	< 2,80	5,11	45,15	8,2			5,12
6	5,0	25,1			0,60	5,75	7,9	215,0	2,0	2,57	31,1	12,21	44,17	< 0,91			3,14
7	0,0	26,0	26,0	2,0	0,64	6,30	8,0	165,0	1,7	4,69	5,6	3,75	33,21	< 0,91			3,79
7	5,0	26,0			0,65	6,86	7,7	205,0	4,0	1,32	7,1	4,77	36,34				3,35
8	0,0	29,0	31,0	0,0	0,94	3,90	8,0	178,0	10,0	7,11	82,2	4,09	29,53	< 0,91			4,83
Mínimo	0	21,2	23,1	0	0,31	2,23	6,3	115	0	0	1,7	1,7	6,01	0,84	0,06	0	0,51
Máximo	5	30,0	31,5	3,7	0,78	24,26	8,4	340	15	9,34	772,9	30,8	199,5	786,7	0,85	1,57	25,86
Média	2,3	27,0	28,0	2,1	0,5	5,5	7,7	217,3	2,6	4,3	131,8	8,9	39,2	494,0	0,4	0,4	5,1
Dpadão	2,51	2,27	2,16	0,91	0,14	4,16	0,48	39,05	3,07	1,98	168,74	5,72	30,80	283,35	0,33	0,37	5,29

S – medida na superfície

M – medida no meio do corpo hídrico

F - medida no fundo do corpo hídrico

Tabela 3 - Valores médios, mínimos, máximos e desvios padrões dos parâmetros Rio Poxim

Campa nha	Tágu a	pH	Cond.	Turbi dez	Cor verda deira	STD	O ₂	Alcali nidad e	Durez a	Na	K	Ca	Mg	Cl	SO ₄	N- NH ₄	N- NO ₂	N- NO ₃	P- PO ₄	Ntotal	Ptotal	Cloro fila-a
	°C		mS/c m	NTU	Pt-Co	mg/L	mg/L	mg/L(CaC O ₃)	mg/L(CaC O ₃)	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	µg/L
1	30,0	7,2	0,12	9,20	7,41	21,0	7,10	43,00	46,18	15,00	1,00	7,39	1,30	19,02	2,50	0,00	0,01	0,54	< 0,001	0,72	0,00	0,06
2	32,0	6,9	0,11	9,70	7,41	27,0	7,41	59,30	48,09	16,00	0,57	5,90	1,56	22,60	6,50	0,01	0,01	0,77	< 0,001	1,04	0,02	0,05
3	27,5	7,2	0,14	9,33	10,37	57,0	7,34	43,61	75,96	10,00	0,91	10,39	1,59	18,33	5,00	0,00	0,01	0,88	0,00	1,39	0,01	0,05
4	26,5	7,3	0,15	11,00	14,63	72,0	7,63	43,37	85,25	11,30	0,95	11,07	1,59	19,14	5,00	0,01	0,01	0,95	0,01	0,50	0,01	0,05
5	28,0	7,3	0,16	14,20	15,56	75,0	7,78	43,02	83,61	11,30	1,16	12,29	1,65	18,82	7,00	0,01	0,01	0,30	0,01	1,33	0,04	0,05
6	30,0	7,0	0,17	11,57	11,11	62,0	6,99	66,44	106,5 7	13,00	1,00	13,18	2,28	23,83	8,00	0,00	0,01	0,71	0,01	0,96	0,04	0,07
7	27,0	7,1	0,24	7,56	9,08	112,0	7,3	67,14	108,7 5	19,00	1,00	13,50	2,42	19,81	4,00	0,01	0,01	0,63	0,00	1,14	0,03	0,05
8	26,0	7,9	0,19	10,20	11,11	39,0	7,56	65,65	58,80	12,31	1,06	16,36	1,64	19,25	10,00	0,01	0,01	0,43	0,00	1,26	0,08	0,06
9	27,0	7,0	0,11	14,20	14,63	66,0	7,20	69,29	65,86	15,40	0,98	14,76	2,98	27,96	15,00	0,01	0,01	0,52	0,01	0,74	0,05	0,06
10	28,0	7,3	0,18	10,80	7,78	74,0	6,82	47,78	94,82	12,11	0,51	18,04	2,05	23,10	5,00	0,01	0,01	0,71	< 0,001	0,32	0,03	0,05
11	30,0	7,7	0,21	11,96	7,41	82,0	7,50	46,05	107,3 3	17,00	0,89	13,82	2,66	23,21	5,00	0,00	0,01	0,27	0,00	0,25	0,06	0,05
12	27,0	7,2	0,14	11,30	11,30	35,0	7,30	67,18	47,00	11,00	0,79	15,99	0,57	18,00	7,50	0,01	0,01	0,58	0,01	0,47	0,05	0,06
13	27,0	7,1	0,11	13,50	9,82	56,0	6,99	66,95	44,54	11,00	0,89	5,99	0,63	21,14	7,50	<0,00 2	0,00	0,68	< 0,001	1,80	0,04	0,06
14	27,0	7,4	0,14	14,10	8,30	70,0	6,59	48,87	57,66	19,00	0,36	9,37	1,21	29,83	7,50	0,00	0,01	0,78	< 0,001	0,91	0,03	0,05
15	30,4	7,9	0,16	11,40	17,96	102,0	7,40	121,1 2	46,12	14,73	1,11	13,05	1,30	2,82	5,60	0,00	0,01	0,57	0,00	0,64	0,08	0,05

Continuação...

16	27,2	5,3	0,09	11,90	9,10	40,0	7,30	77,63	41,31	11,00	1,00	12,65	0,52	14,91	6,00	<0,002	<0,001	0,31	< 0,001	0,86	0,02	0,06
17	25,1	6,8	0,11	9,50	5,93	45,0	6,80	84,28	45,39	12,00	1,00	11,43	0,56	11,09	8,50	<0,002	<0,001	0,81	< 0,001	1,74	0,03	0,07
18	26,0	7,6	0,14	9,40	8,34	66,1	7,84	52,32	75,48	19,00	0,95	11,43	1,59	20,78	5,50	<0,002	<0,001	0,45	< 0,001	1,22	0,03	0,07
19	28,0	7,6	0,17	11,03	10,19	76,5	7,84	41,97	72,31	19,00	0,95	13,06	1,59	20,36	7,50	<0,002	0,00	0,19	0,01	0,88	0,04	0,05
20	27,1	6,8	0,13	11,10	7,40	77,0	7,88	43,86	71,91	25,78	1,10	11,71	1,21	29,80	9,00	<0,002	0,00	0,08	0,00	0,68	0,03	0,05
21	29,0	7,5	0,13	7,25	7,60	71,4	6,89	37,23	95,37	24,00	1,00	13,06	0,96	25,34	6,50	<0,002	0,00	0,08	0,01	0,67	0,03	0,06
22	31,3	7,1	0,25	9,90	7,41	66,9	6,45	47,47	98,43	23,00	1,00	11,63	2,42	29,23	6,00	<0,002	<0,001	0,12	0,01	0,72	0,03	0,05
23	29,5	8,0	0,17	10,11	15,74	81,2	7,40	49,90	65,99	17,70	1,21	11,50	3,80	18,59	7,20	0,00	0,00	0,65	0,01	1,00	0,04	0,07
24	30,2	8,7	0,11	12,20	6,12	53,6	8,94	72,64	46,80	9,61	0,95	12,65	2,80	14,80	4,98	<0,002	0,00	0,13	0,05	0,63	0,07	0,05
25	24,0	8,3	0,16	4,24	4,76	89,0	6,82	87,28	58,65	27,00	1,00	14,49	2,05	24,72	6,00	<0,002	<0,001	0,69	0,00	1,56	0,03	0,05
26	25,1	7,7	0,12	7,48	6,49	102,3	6,16	39,97	90,27	20,00	0,90	18,16	1,49	23,30	5,50	<0,002	<0,001	0,39	0,01	1,19	0,02	0,05
27	29,0	8,6	0,14	10,80	25,18	101,0	6,08	78,54	84,64	15,00	0,97	15,10	1,58	43,92	8,32	0,00	<0,001	0,39	0,03	0,99	0,07	0,07
28	26,1	6,4	0,12	10,72	12,00	98,0	7,50	53,20	45,39	16,00	0,98	13,47	1,22	26,20	7,50	0,00	0,01	0,30	0,05	1,16	0,10	0,09
29	28,2	5,7	0,12	10,80	10,93	75,0	6,90	70,99	42,84	17,84	1,00	12,65	3,25	19,57	6,50	0,01	<0,001	0,14	0,01	0,76	0,08	0,08
30	27,5	8,3	0,20	10,41	12,97	84,0	6,70	87,87	87,70	13,30	1,02	14,70	2,30	41,80	8,40	0,00	0,01	0,13	0,08	0,73	0,05	0,06
31	25,5	5,8	0,14	2,10	8,00	31,0	6,90	54,58	34,78	12,00	0,67	12,05	1,00	12,03	9,50	0,00	0,00	0,26	0,02	0,74	0,04	0,07
32	25,5	7,3	0,17	7,20	9,45	84,0	7,18	39,79	46,04	13,00	0,71	13,10	0,93	14,08	11,00	0,00	0,01	0,36	0,01	0,95	0,01	0,08
33	24,0	7,3	0,15	12,00	11,11	75,0	6,74	39,65	56,78	19,00	0,98	12,89	1,41	19,68	12,00	0,00	0,01	0,33	0,01	0,91	0,04	0,05
34	24,5	7,5	0,17	11,40	11,30	92,0	7,80	52,14	98,21	15,00	0,80	10,49	1,49	14,95	10,00	<0,002	0,01	0,31	0,01	0,93	0,02	0,05
35	23,0	7,1	0,14	7,10	10,19	65,0	6,74	42,80	40,92	30,00	1,15	9,73	1,39	24,28	12,50	0,00	0,01	0,90	0,03	0,88	0,07	0,09
36	22,0	6,8	0,12	5,80	9,82	63,0	8,52	50,62	52,94	16,88	1,00	7,98	1,43	17,01	19,00	0,00	0,00	0,44	0,03	0,69	0,60	0,09

Continuação...

mínima	22,0	5,3	0,09	2,10	4,76	21,0	6,08	37,23	34,78	9,61	0,36	5,90	0,52	2,82	2,50	0,00	0,00	0,08	0,00	0,25	0,00	0,05
maximo	32,0	8,7	0,25	14,20	25,18	112,0	8,94	121,12	108,75	30,00	1,21	18,16	3,80	43,92	19,00	0,01	0,01	0,95	0,08	1,80	0,60	0,09
média	27,3	7,3	0,15	10,07	10,39	69,1	7,23	58,43	67,46	16,23	0,93	12,36	1,68	21,48	7,74	0,00	0,01	0,47	0,02	0,93	0,06	0,06
dpadrão	2,3	0,7	0,04	2,65	3,94	22,4	0,60	18,33	22,66	5,01	0,18	2,83	0,77	7,64	3,19	0,00	0,00	0,25	0,02	0,35	0,10	0,01

Tabela 4 - Conjunto de dados oligotróficos em relação a clorofila-a

pH	OD	N-NH4	N-NO2	N-NO3	P-PO4	Ntotal	Ptotal	Clorofila.a
	mg/L	µg/L	µg/L	µg/L	µg/L	mg/L	mg/L	µg/L
9,1	6,3	69,7	343,7	367,9	410,7	1,4	0,8	8,3
9,3	4,8	87,0	347,7	335,0	431,4	1,2	0,7	8,6
8,0	3,6	97,7	83,9	649,8	480,0	3,1	1,0	9,3
8,7	4,4	87,5	96,1	447,7	474,1	3,2	1,8	9,5
8,9	4,0	14,6	54,8	1142,5	483,0	3,2	1,4	10,4
8,9	3,0	64,6	92,0	3197,8	452,0	2,1	1,2	10,8
8,7	0,9	21,7	55,5	2203,4	506,6	4,5	1,2	11,0
8,6	0,1	25,1	52,1	2567,2	503,6	4,9	1,1	11,1
8,7	4,6	16,8	48,1	2836,6	500,7	3,5	1,0	11,3
9,1	3,4	62,5	354,5	1020,5	435,8	1,7	1,1	12,0
8,7	0,6	19,8	46,7	1046,5	516,9	2,4	0,8	12,7
9,0	4,8	28,7	407,3	953,3	437,3	1,7	1,4	12,7
8,8	4,6	13,2	59,6	1991,7	488,9	3,1	1,1	12,7
8,6	3,9	70,6	332,1	1611,9	447,6	3,8	1,7	12,9
8,9	1,9	72,7	449,2	695,1	452,0	1,3	0,8	13,1
8,8	9,0	35,8	337,5	1352,2	381,2	3,6	0,8	13,1
8,9	9,4	21,2	4,1	1201,9	505,1	3,9	0,8	13,2
8,7	2,1	17,8	48,1	1274,1	575,9	4,1	0,9	13,4
8,4	6,6	53,8	272,6	1151,9	437,3	2,5	0,8	14,2
8,8	9,4	55,8	286,1	2190,0	437,3	5,1	0,8	14,2
8,7	8,2	35,3	161,7	1298,9	432,8	4,3	0,5	14,4
8,4	3,7	55,3	134,0	1314,0	418,1	3,7	1,0	14,5
8,5	5,1	46,6	361,2	1611,2	429,9	6,6	0,8	14,7
8,8	8,8	52,2	340,9	1251,1	416,6	1,9	1,0	14,8
8,6	3,6	54,8	10,9	1372,6	415,2	2,9	1,0	15,1
8,7	7,9	41,5	188,8	1976,1	421,1	5,8	1,0	15,8
8,7	4,2	26,6	100,8	899,1	593,6	4,6	0,6	16,3
8,6	1,4	57,0	167,8	826,9	620,1	4,6	0,9	16,4
8,6	5,3	23,9	105,6	610,6	543,4	3,3	0,7	16,4
8,5	3,0	51,7	34,6	974,6	516,9	2,9	1,0	16,6
8,5	2,2	57,5	25,1	978,0	370,9	2,7	0,8	16,6
8,2	6,4	28,2	272,2	1433,5	285,4	4,8	1,4	16,6
8,4	0,9	114,3	336,2	134,6	599,5	4,8	0,8	16,7
8,7	8,5	9,0	38,6	726,0	561,1	4,6	0,8	16,8
8,7	8,2	20,5	33,2	346,6	276,6	3,6	0,5	16,9
8,8	6,5	12,2	9,5	1562,5	592,1	3,5	0,7	16,9
8,4	1,2	78,2	308,5	495,2	627,5	3,5	0,8	16,9
8,3	7,0	121,4	6,2	1446,4	887,7	2,7	0,9	17,1
8,3	5,7	168,9	6,8	1295,7	956,1	3,2	1,0	17,5
9,5	6,2	28,3	257,7	1721,6	170,4	9,8	0,7	18,1

Continuação...

9,1	4,9	50,2	345,0	986,7	425,5	1,7	0,8	18,5
8,2	2,8	223,7	6,8	1367,8	861,0	3,2	1,0	18,9
8,4	5,8	88,5	6,8	1382,3	825,3	2,2	0,9	19,2
8,8	8,0	31,8	47,4	318,0	227,9	4,8	0,5	19,3
9,0	1,9	34,9	395,1	734,7	438,7	0,9	0,7	19,4
8,5	4,5	94,3	5,5	1106,6	941,2	4,3	0,9	19,6
8,8	5,7	49,2	6,2	272,7	185,1	4,1	0,4	20,6
8,3	3,6	165,2	5,5	1827,8	962,1	2,9	0,9	22,2
9,1	6,1	23,6	418,1	841,6	424,0	1,7	0,7	22,4
8,4	4,1	59,3	8,2	1643,5	899,6	2,9	0,9	22,4
8,3	5,0	107,9	6,8	1079,4	881,8	3,2	0,9	23,6
8,6	5,1	69,7	4,8	331,7	220,5	0,2	0,5	24,1
8,3	2,0	111,6	14,9	321,6	238,2	1,3	0,6	24,6
8,6	4,8	29,7	5,5	345,5	205,8	4,1	0,5	25,3
8,5	3,2	92,2	36,6	256,7	245,6	3,6	0,6	26,0
8,2	3,9	73,9	6,2	1331,0	902,6	2,2	1,0	26,6
8,5	2,8	37,9	4,1	318,0	248,5	3,9	0,6	27,6
8,3	1,3	142,3	1,7	262,7	183,7	3,2	0,5	28,3

Tabela 5 - Conjunto de dados eutróficos em relação a clorofila-a

pH	OD	N-NH4	N-NO2	N-NO3	P-PO4	Ntotal	Ptotal	Clorof.a
	mg/L	µg/L	µg/L	µg/L	µg/L	mg/L	mg/L	µg/L
7,8	4,1	79,3	253,0	1376,8	247,1	4,0	1,5	43,1
7,8	5,3	23,6	271,2	1390,3	334,1	5,1	1,5	45,0
8,1	4,3	28,7	291,2	1072,0	233,8	4,6	1,1	47,6
9,5	0,9	58,2	480,8	1365,4	208,7	9,3	0,9	47,9
9,3	4,6	20,0	345,6	1594,8	183,7	9,6	1,0	48,5
7,9	5,2	23,1	239,8	1593,8	171,9	4,9	1,5	48,6
8,2	6,5	12,4	166,4	1311,5	565,3	4,5	0,6	48,7
9,4	8,6	20,7	311,8	1107,7	177,8	6,7	0,8	49,5
8,2	2,8	7,7	296,6	1735,2	763,1	4,3	1,6	52,0
9,2	2,4	17,8	507,9	1426,2	198,4	6,5	0,7	55,4
8,0	5,4	14,9	284,7	1782,6	322,3	3,8	0,1	56,3
8,2	7,0	41,0	263,1	1650,3	322,3	3,9	1,5	57,2

Tabela 6- Conjunto de dados mesotróficos em relação a clorofila-a

pH	OD	N-NH4	N-NO2	N-NO3	P-PO4	Ntotal	Ptotal	Clorof.a
	mg/L	µg/L	µg/L	µg/L	µg/L	mg/L	mg/L	µg/L
8,3	5,9	6,2	281,0	1361,0	257,4	4,6	1,1	33,3
8,8	7,5	74,8	270,7	729,4	416,6	1,3	0,8	34,4
8,1	6,7	16,3	167,1	1815,7	569,8	4,2	0,6	37,4
8,1	2,9	16,4	295,6	1403,5	462,3	5,0	1,4	37,7

Tabela 8 - Conjunto de dados hipereutróficos em relação a clorofila-a

pH	OD	N-NH4	N-NO2	N-NO3	P-PO4	Ntotal	Ptotal	Clorofila
	mg/L	µg/L	µg/L	µg/L	µg/L	mg/L	mg/L	µg/L
8,1	6,5	9,0	138,7	1265,8	569,8	3,2	0,7	61,2
8,3	6,4	10,0	130,6	1403,3	553,4	3,9	0,7	64,3
8,2	2,8	22,7	34,6	1614,5	412,2	2,2	0,6	83,0
8,2	9,1	15,6	413,2	2663,8	304,6	2,5	0,3	84,8
8,1	6,6	159,4	217,2	1201,1	593,6	3,2	0,6	92,3
8,2	7,0	9,0	101,5	1315,5	554,9	2,7	0,7	95,9
8,6	9,1	36,1	7,1	552,9	309,0	3,9	1,0	98,0
8,5	7,8	12,0	12,9	2241,9	353,2	1,3	0,6	109,0
9,0	9,0	29,7	6,8	351,0	319,3	2,7	1,0	109,9
8,2	5,2	35,3	16,3	1315,5	407,8	4,7	0,5	112,0
8,3	6,4	20,5	23,7	1394,6	341,4	4,5	0,6	113,8
7,9	3,4	40,5	140,1	1537,9	429,9	10,1	0,7	114,9
8,5	5,5	31,9	3,8	1577,8	196,9	5,5	0,7	119,0
8,3	4,6	14,4	51,5	1842,8	356,2	10,6	0,3	123,7
8,1	4,9	22,2	96,8	1985,0	375,3	2,5	0,6	125,3
8,1	6,5	162,3	227,3	1378,5	577,3	3,4	0,6	126,4
8,7	5,0	33,9	77,1	977,7	375,3	9,5	0,9	127,6
8,8	7,7	27,1	9,5	769,2	320,8	4,4	0,7	143,2
8,0	4,2	29,0	10,9	1436,3	410,7	2,9	0,5	147,7
8,2	3,6	13,2	21,0	2709,8	366,5	9,0	0,6	149,6
8,3	5,2	10,5	21,0	2320,4	345,9	5,5	0,3	155,7
8,3	5,2	6,8	105,6	1297,2	563,9	3,7	0,7	157,1
9,5	7,3	31,7	12,5	1014,4	313,4	9,4	0,8	183,7
9,3	3,1	37,8	136,0	676,6	384,2	28,0	1,2	185,3
8,1	2,2	25,4	106,9	1080,6	416,6	6,1	0,6	186,2
8,4	3,9	27,5	0,7	481,7	279,5	6,6	0,4	208,4
8,1	4,6	91,4	17,7	1876,6	409,3	14,2	0,6	224,8
8,4	5,3	22,7	2,4	1430,4	183,7	4,3	1,0	225,7
8,5	4,4	24,6	1,7	1198,0	194,0	4,6	0,9	229,6
8,3	6,5	28,8	1,7	693,2	196,9	3,6	0,9	264,4
8,4	4,0	29,7	2,1	1105,1	254,4	8,1	0,6	293,7

ANEXO B – Metodologia de construção gráfica das interfaces

1 INTERFACES GRÁFICAS

1.1 GUIDE

As interfaces foram desenvolvidas através do GUIDE (Graphical User Interface Design Environment). O GUIDE é uma ferramenta do Matlab que permite a construção de interfaces (GUI) para facilitar a interação usuário-programa através de elementos como botões, listas suspensas, caixas de texto, gráficos entre outros. A linguagem de programação é a mesma utilizada em arquivos *.m*, no entanto, estes ficam “escondidos” dos usuários, de modo que o desenvolvedor é capaz de controlar que parâmetros podem ser modificados pelo usuário final. A seguir, será mostrado um exemplo de como foi desenvolvida uma das telas do índice.

1.2 Construção das interfaces

O primeiro passo para a construção de uma interface é iniciar o GUIDE através do comando *guide* na linha de comando do Matlab. A Figura 1 mostra a tela de *Quick Start* para criação de uma nova interface. Nesta tela, para escolher uma interface em branco, basta selecionar a opção *Blank GUI (Default)*. Ao escolher essa opção, a tela mostrada na Figura 2 é aberta.

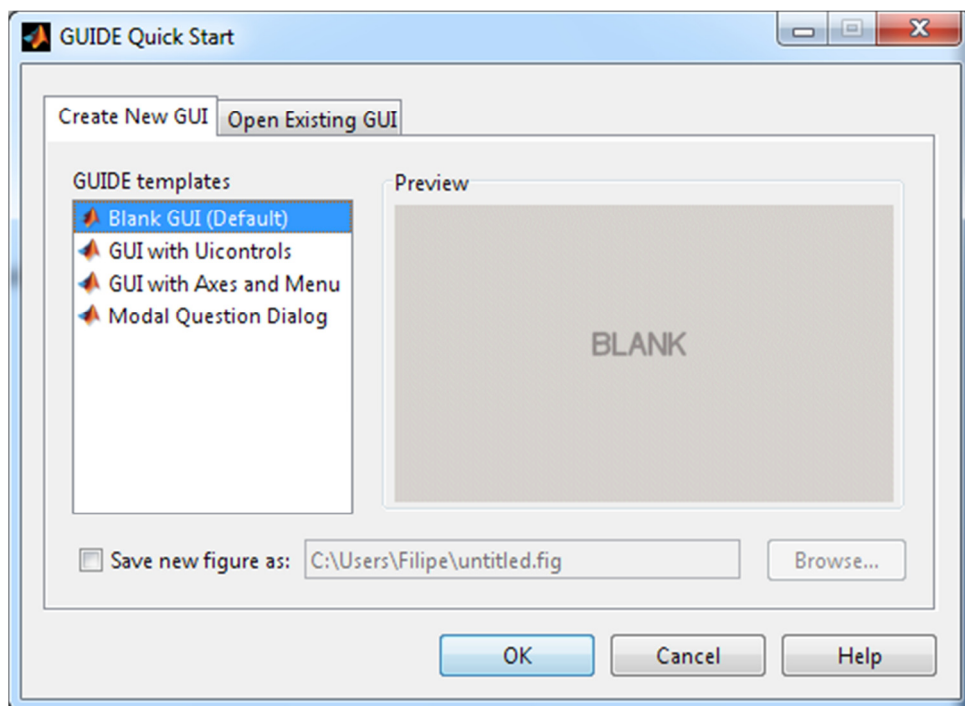


Figura 1. Tela para criação de uma nova interface.

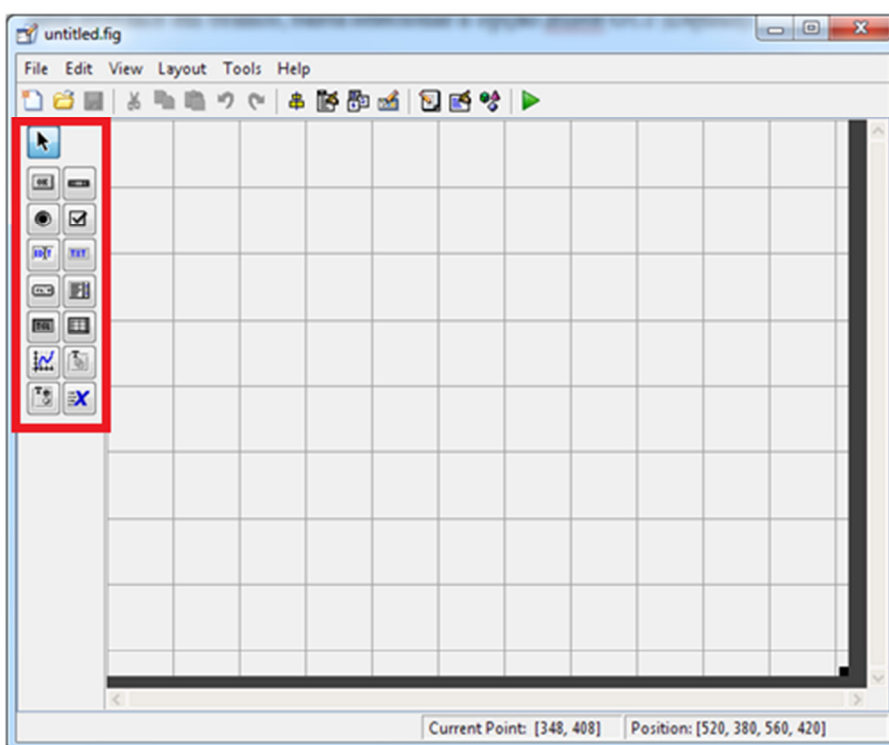


Figura 2. Uma interface em branco, com destaque para a barra de objetos.

Na Figura 2, os botões em destaque constituem a barra de objetos, onde é possível escolher os objetos que farão parte da interface, como botões, gráficos, caixas de texto, etc. Para adicionar um objeto, basta selecioná-lo, clicando sobre ele, e clicando em um espaço livre na interface. A Figura 3 mostra um botão sendo adicionado à interface.

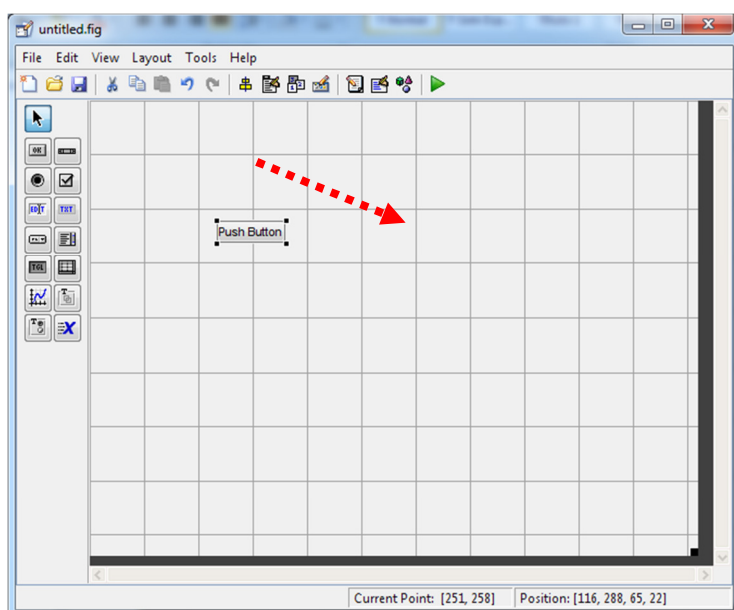


Figura 3. Botão sendo adicionado à interface.

Para associar ações aos objetos, existe um arquivo de extensão *.m* que guarda os respectivos comandos. O Matlab chama essas ações de *callbacks*. Para acessar o callback de um botão, por exemplo, basta clicar com o botão direito sobre o botão, acessar *View Callbacks* e em seguida clicar sobre *Callback*. A Figura 4 mostra como abrir o *callback* de um botão e a Figura 5 mostra o *callback* associado a esse botão.

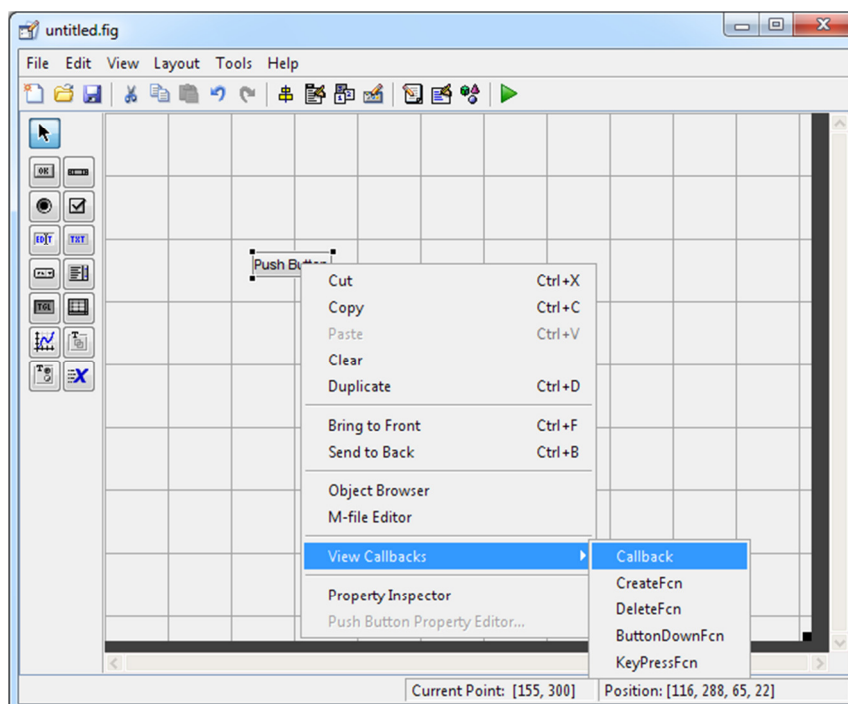


Figura 4. Passo a passo de como acessar o *callback* de um objeto.

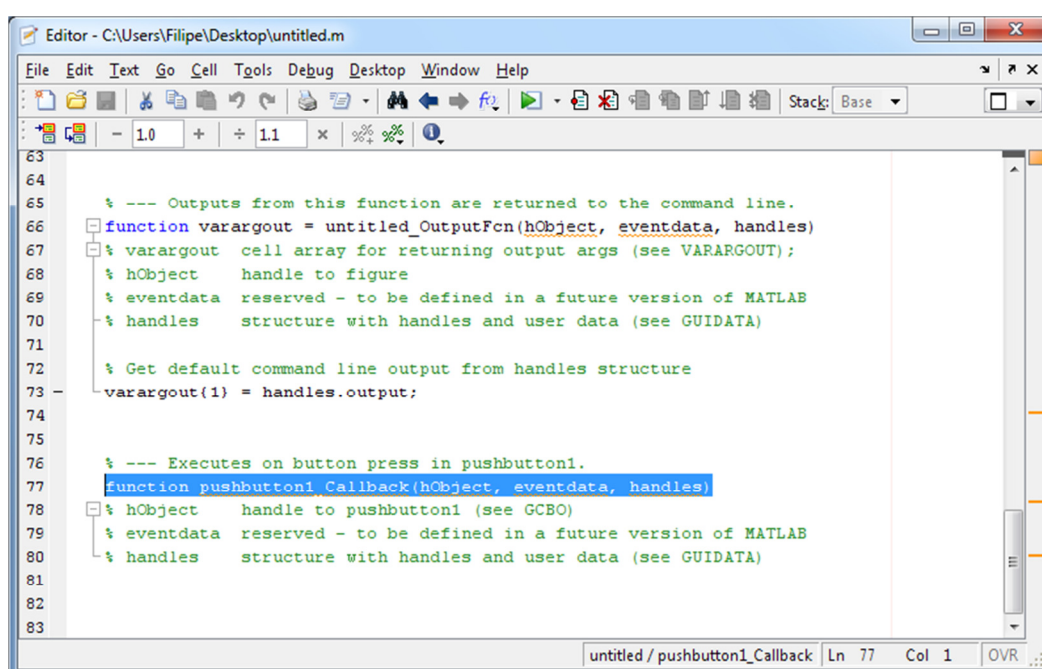



Figura 5. *Callback* associado a um botão.

Para fins de ilustração da metodologia de criação de interfaces, será mostrado como foi criada a tela principal para treinamento de redes neurais:

- 1) Inicialmente foi criada uma interface em branco no GUIDE, como mostrado na Figura 2.
- 2) Em seguida, foram adicionados dois painéis: um para o usuário escolher o tipo de rede neural e outro para conter as propriedades de cada tipo de rede neural. Os painéis foram adicionados utilizando o botão *Panel*  na barra de objetos. Para alterar o nome que é exibido no painel, basta selecioná-lo e no menu *View*, clicar em *Property Inspector*. Na tela que aparece, basta alterar a opção *Title* para o nome desejado. A Figura 6 mostra a interface com os painéis.

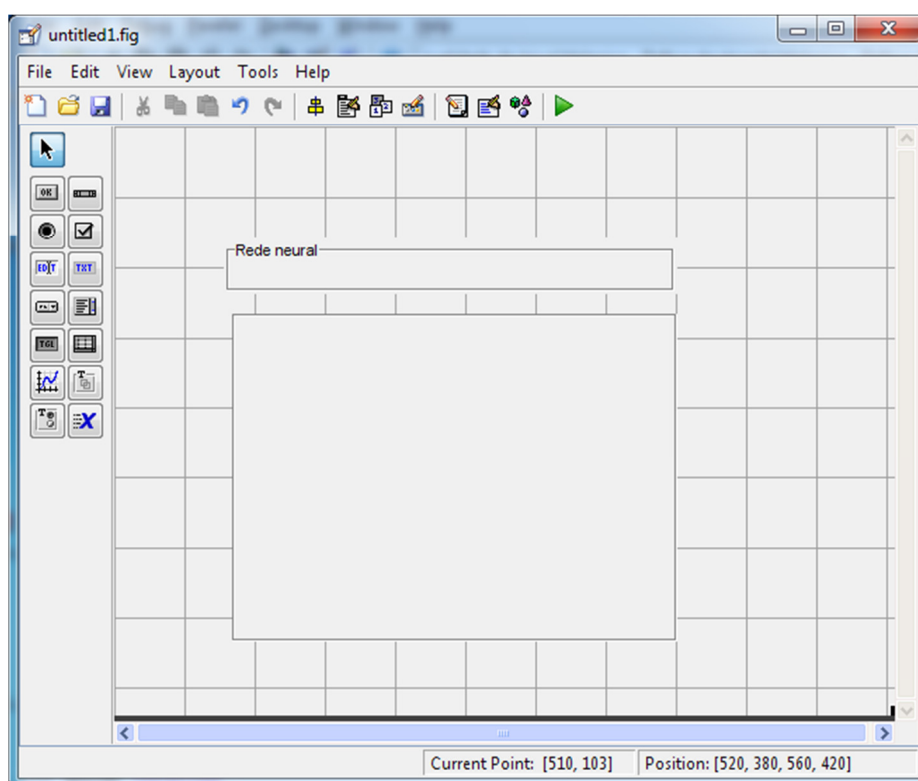




Figura 6. Interface de redes neurais com os painéis.

- 3) O próximo passo é inserir alguns botões e botões de rádio. Isso pode ser feito utilizando os objetos *Push Button*  e *Radio Button* . Para fazer a alteração dos nomes que são exibidos nesses objetos basta alterar a opção *String* no *Property Inspector*. A Figura 7 mostra a interface com os botões.

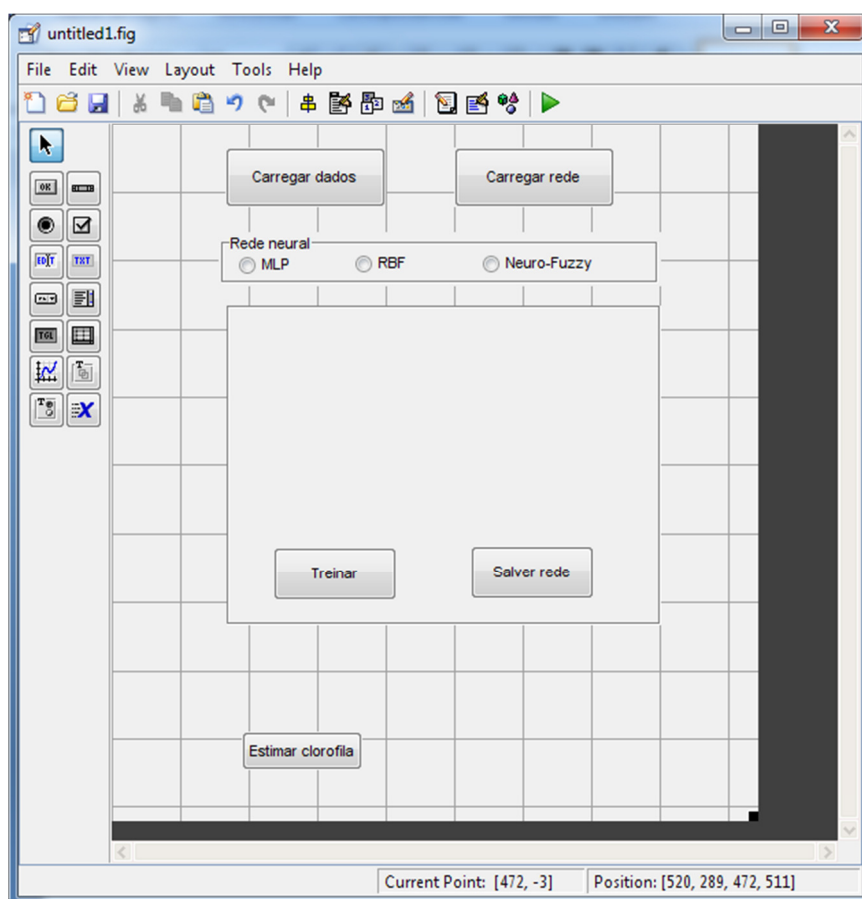





Figura 7. Interface com os botões.

4) Por fim, para introduzir os textos, foi utilizado o objeto *Static Text* , para as caixas de texto, *Edit Text*  e para os menus suspensos, *Pop-up Menu* . Para editar o texto do objeto *Static Text* e para manter a caixa de texto em branco, basta selecionar o objeto e na janela de *Property Inspector* alterar a opção *String*. Para o menu suspenso, para inserir as opções, basta alterar a opção *String* e dar quebra de linha para inserir as outras opções. A Figura 8 mostra a interface com todos os objetos necessários para o treinamento e a simulação de uma rede neural para inferência da clorofila.

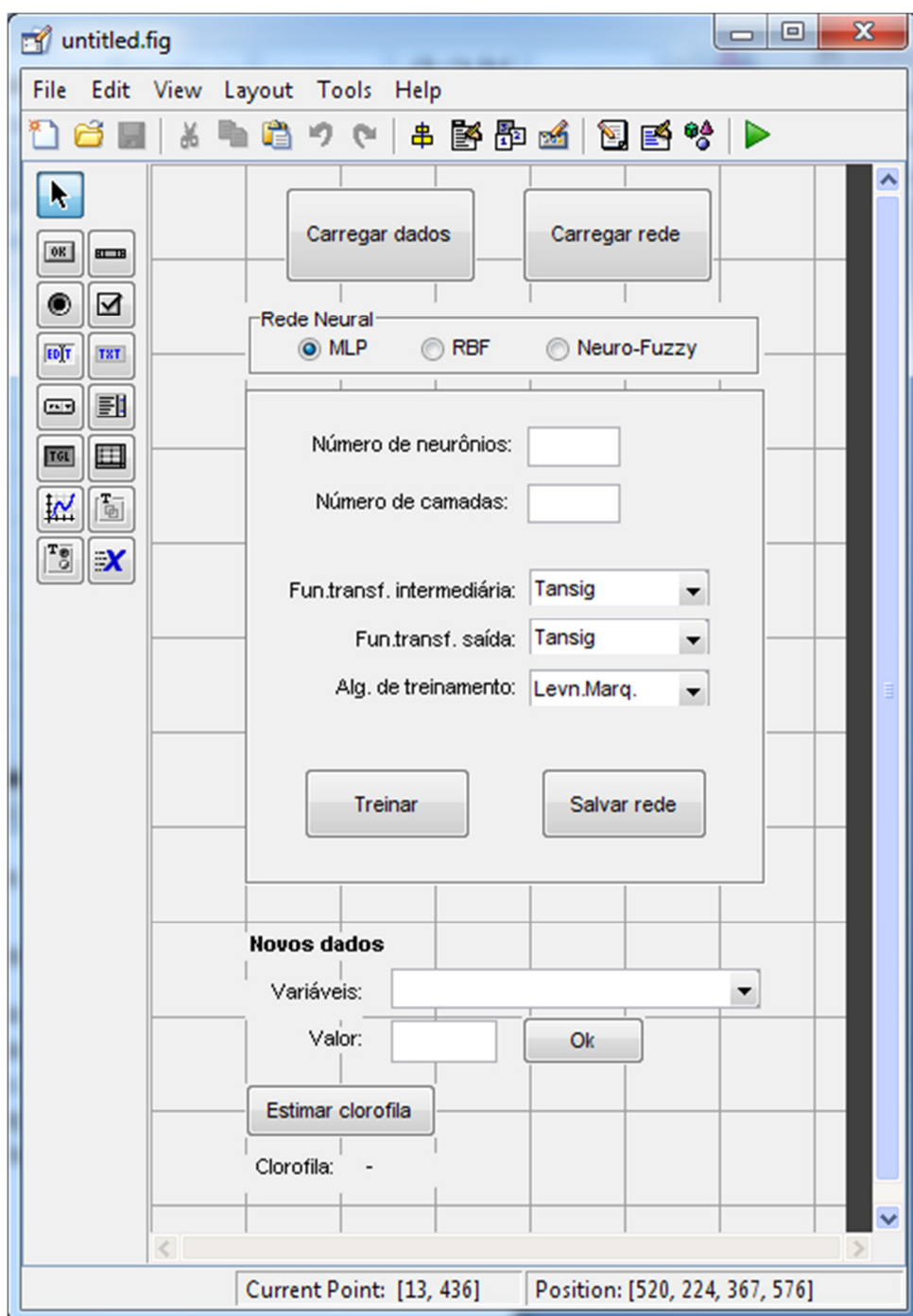


Figura 8. Interface da rede neural com todos os objetos necessários para o treinamento e simulação.

5) Para modificar o *callback* do botão para carregar os dados, por exemplo, bastou acessar o respectivo *callback* e adicionar os comandos necessários para carregar os dados a partir de uma planilha de extensão .xls (Figura 9).

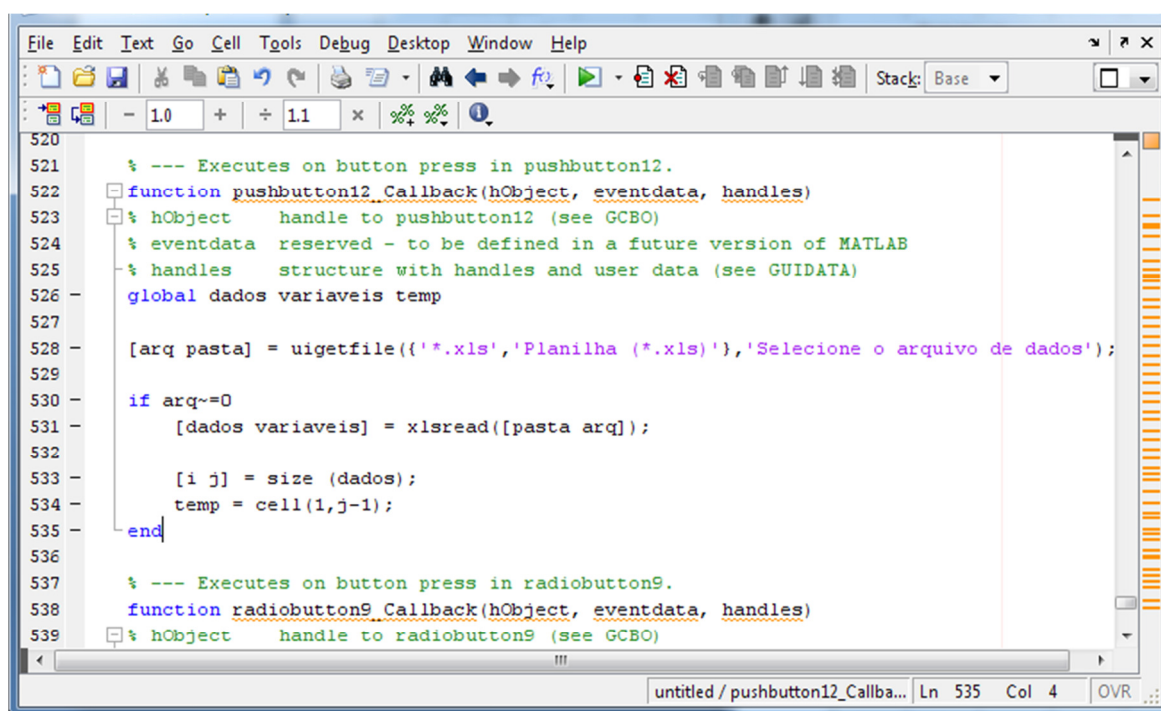


Figura 9. Callback do botão para carregar novos dados.

Os *callbacks* dos outros objetos foram construídos de maneira semelhante à mostrada no passo 5. Não é possível mostrar todos os passos para a construção desta interface uma vez que algumas estratégias foram usadas para fins de estética do programa, como sobreposição de objetos e utilização da propriedade de invisibilidade, e que não são escopo desta metodologia, uma vez que esta visa apenas os aspectos básicos utilizados na construção da interface e implementação dos comandos através de *callbacks*.

ANEXO C - Formatação de dados para a ferramenta de estimação de nível trófico

1) As variáveis devem ser dispostas em colunas, enquanto as amostras (dados experimentais), em linhas, conforme exemplificado na Figura 1.

pH	OD	N-NH4	N-NO2	N-NO3	P-PO4	Ntotal	Ptotal	saída- Clorof.a
8,30	6,36	10,01	130,59	1403,28	553,41	3,94	0,68	64,30
8,10	6,49	9,03	138,71	1265,75	569,82	3,22	0,74	61,15
8,30	5,20	6,84	105,57	1297,24	563,86	3,70	0,71	157,10
8,20	6,99	9,03	101,51	1315,51	554,90	2,66	0,67	95,89
8,10	6,55	159,37	217,16	1201,14	593,56	3,17	0,58	92,33
8,10	6,52	162,29	227,30	1378,50	577,35	3,39	0,59	126,36
8,15	9,06	15,61	413,21	2663,79	304,57	2,48	0,27	84,77
7,95	5,36	14,88	284,71	1782,65	322,27	3,85	0,10	56,30
8,10	4,62	91,38	17,65	1876,62	409,26	14,23	0,64	224,82
8,18	5,23	35,34	16,30	1315,46	407,78	4,71	0,52	112,02
8,19	2,80	22,67	34,56	1614,52	412,21	2,19	0,59	82,99
7,97	4,20	29,01	10,89	1436,26	410,73	2,91	0,50	147,68
8,08	2,16	25,35	106,92	1080,61	416,63	6,08	0,64	186,24
7,94	3,36	40,46	140,06	1537,86	429,90	10,12	0,68	114,95
8,31	6,38	20,48	23,74	1394,56	341,43	4,50	0,55	113,78
8,09	4,90	22,19	96,78	1985,00	375,35	2,55	0,56	125,30
8,45	7,85	11,95	12,92	2241,94	353,23	1,32	0,56	108,97
8,29	4,55	14,39	51,46	1842,81	356,18	10,63	0,34	123,72
8,33	5,18	10,49	21,03	2320,37	345,86	5,50	0,35	155,75
8,22	3,64	13,17	21,03	2709,80	366,50	9,04	0,57	149,64

Figura 1 – Exemplo de um arquivo de dados válidos

2) A primeira linha do arquivo deve conter os nomes das variáveis e sua grafia não interfere no desempenho da ferramenta.

3) A variável referente a saída, neste caso a concentração de clorofila-a, deve ser colocada na última coluna. O arquivo de dados deve sempre possuir essa variável (a variável de saída), caso contrário a rede neural não conseguirá ser treinada.

4) O arquivo deve ser salvo no formato de planilha .xls. O formato .xlsx não é suportado pela ferramenta.